Proceedings of the VIIth GSCP International Conference

Speech and Corpora

edited by

HELIANA MELLO
MASSIMO PETTORINO
TOMMASO RASO

FIRENZE
UNIVERSITY
PRESS

Proceedings e report

93

Proceedings of the VIIth GSCP International Conference

Speech and Corpora

edited by
Heliana Mello
Massimo Pettorino
Tommaso Raso

INDEX

PHONETIC STUDIES 221

Foreword

The 2012 GSCP International Conference was the 7th Conference of the Gruppo di Studio sulla Comunicazione Parlata (www.gscp.it), and for the first time it was held outside Italy. It took place at the Universidade Federal de Minas Gerais, in Belo Horizonte (Brazil) from February 29th to March 2nd. Its main theme was *Speech and Corpora*, and the conference was dedicated to the memory of Claire Blanche-Benveniste, reminded by Marie-Noëlle Roubaud and Frédéric Sabio.

The program can be seen at the address < http://www.letras.ufmg.br/gscp2012/>. It shows 4 plenary conferences, by Pier Marco Bertinetto, Philippe Martin, Plínio Barbosa and Douglas Biber; one round table, where the C-ORAL-BRASIL spontaneous speech corpus was presented (Tommaso Raso, Heliana Mello, Eckhard Bick, Emanuela Cresti and Massimo Moneglia); one workshop about Emotions, Attitudes and Illocutions, held by Klaus Scherer, Véronique Aubergé and João A. Moraes. Besides that, 62 oral presentations and 62 posters were presented.

The wide international origin of the authors is remarkable. They came from 21 countries in 4 continents. Brazil was the most present country, with 125 authors from 25 institutions, followed by Italy with 35 authors from 19 institutions, France (19 and 11), Germany (10 and 6), Portugal (8 and 4), USA (6 and 6) and 15 other countries with 32 authors from 24 institution.

These Proceedings present 89 papers. Their distribution reflects the main themes of the conference. First, the reminding of Claire Blanche-Benveniste, whose work was seminal in studying speech through corpora. Then five of the plenaries, which show how spontaneous speech studies have changed the way to look at linguistic phenomena, give a panorama of prosodic studies, focus on the relation between illocution and prosody and present, in two papers, the spontaneous speech corpus C-ORAL-BRASIL.

The other sections testify the importance that the conference gave to its main theme, with 12 papers dedicated to spoken corpora compilation and annotation, and with 5 papers dedicated to the connected field of speech technology and data bases. A second very important theme of the conference, present also in the workshop, is that of the relation between some pragmatic aspects of language and its relation with prosody. Therefore, one section is dedicated to illocutions and attitudes, and another section to information studies. A specific section is dedicated to speech pathologies, and four sections collect different works on phonetic studies, speech and linguistic analysis, speech and pragmatics and speech and sociolinguistics. It is worthwhile to underline that in all sections many papers are dedicated to the study of speech and second language studies.

Of course, the wide international origin of the participants led to contributions on several different languages: besides Brazilian and European Portuguese, Italian, French, English and German, we find contributions on Northern languages, Japanese, Amerindian languages, Vietnamese, Chinese, and many others. This online publication with Firenze University Press allows direct access to sound and video, in papers in which authors provided them.

The conference, together with this volume, bring therefore some important novelties: the first one was the internationalization of the GSCP association also for its conference seat; the second one was allowing the international scientific community to better know the important linguistic production in Brazil, and to Brazilian scholars and students to have easy access to many international scholars, giving birth to collaborations and new contacts between an important emerging country and the rest of the world in our scientific community; the third one was to give a new impulse to a very important field, that of the study of speech through corpora compilation.

In fact, empirical study in speech sciences cannot do without big resources organized for different scientific goals, statistically validated and technologically predisposed for quantitative studies. The importance of this methodology in spoken studies clearly emerges from the success of both the conference and its proceedings, and makes proud both the C-ORAL-BRASIL group (c-oral-brasil.org) and the GSCP association that organized the conference.

Heliana MELLO (UFMG-CNPq)
Massimo PETTORINO (Università di Napoli, L'Orientale)
Tommaso RASO (UFMG-CNPq-Fapemig)

Remembering Claire Blanche-Benveniste

Marie-Noëlle ROUBAUD, Frédéric SABIO

EA 4671 ADEF; Aix Marseille Université.; ENS de Lyon, IFE; 13248, Marseille, France & UMR 7309 CNRS;
Laboratoire Parole et Langage; Aix Marseille Université
mnroubaud@yahoo.fr, frederic.sabio@orange.fr

Abstract

This paper pays tribute to the French scholar Claire Blanche-Benveniste (1935-2010) whose contribution to linguistics is original and outstanding in many ways. In particular, she stands out as a pioneer in the field of corpus linguistics.

Keywords: Claire Blanche-Benveniste; spoken French; syntax; corpus.

Claire Blance-Benveniste

Claire Blanche-Benveniste (Lyon 1935- Aix-en-Provence 2010) acquired a thorough knowledge on medieval romance philology at the Sorbonne University in Paris, with such professors as Robert-Léon Wagner and Jean Boutière. She specialized in old Provençal and devoted the beginning of her career to female troubadours, more specifically the Countess of Die. After spending three years in Lebanon as a lecturer, she taught at the Universities of Lyon and the Sorbonne. In 1964, she contributed to the *Grammaire Larousse du français contemporain*, with Michel Arrivé, Jean-Claude Chevalier and Jean Peytard. In the same year, she was recruited by the French linguistics department of the University of Aix-en-Provence, headed by Jean Stéfanini.

She spent her entire career in Aix until 2000, when she became Professor Emeritus. Between 1994 and 2002, she worked as Director of studies at the prestigious Ecole Pratique des Hautes Études. In 2002, she directed the Paris Linguistics Society (*Société Linguistique de Paris*). She was elevated to be knight of the French « Légion d'honneur » in 2004, and was made Doctor Honoris Causa of the Katholieke Universiteit of Leuven in 2007.

With the passing of Claire Blanche-Benveniste, France lost a scholar whose influence extended far beyond her strict scientific domain which was the study of French language. Many messages of sympathy were received from many places around the word. All of them insisted on her outstanding academic achievements and the major role that she played throughout her career within the community of linguists. Among those numerous expressions of gratitude, here is the message sent by Morris Halle: "I always had the greatest respect and admiration for her both as a scholar and as an outstanding human being".

It would not be easy to give a comprehensive review of Claire's centers of interest, since her intellectual curiosity seemed to have no limits. She carried out research on the following fields:

- Syntax and flexional morphology of spoken and written French (among many studies: 1975, 1981, 1984 with J. Deulofeu *et al.*, 1990a, 1990 with M. Bilger *et al.*, 1999 with J.-P. Adam, 2000b, 2004, 2010);
- Orthography (1969 with A. Chervel, 2003a);
- Corpus design (1987 with C. Jeanjean, 2002 with C. Rouget & F. Sabio, 2005);
- Relationship between syntax and discourse (1979, 1990b);
- Language acquisition and children's linguistic productions (1982, 1998, 2001 with B. Pallaud, 2003b);
- Compared linguistics of Romance languages (2001, 2009);
- Simultaneous teaching of Romance languages (1997 with A. Mota *et al.*);
- The history of linguistics (2000a).

In all of those domains, she offered many precious contributions that where characterized by a flawless mastery of linguistic description and an exceptional clarity that she was able to demonstrate in both her publications and her oral presentations.

Unquestionably, Claire Blanche-Benveniste will particularly be remembered as a major specialist of the syntax of spoken French, a field in which she pioneered in the early seventies. She always insisted on the fact that spoken French was not to be conceived as a clearly distinct domain, but should simply be included into what she nicely called "le français tout court" ("French in itself"). In the book that she and C. Jeanjean wrote in 1987,

she made the following statement:

> "Quand on parcourt une documentation sur le français parlé depuis le début du vingtième siècle, on est frappé par la persistance de quelques grands mythes qui ont pour effet de 'séparer' ce qu'on appelle le 'français parlé' de l'ensemble de la langue ; on le voit retranché, mis à l'écart – pour le décrire comme pour l'encenser. Assimiler le parlé au populaire, c'est le retrancher du français légitime; y voir la source des innovations ou des conservatismes, c'est le retrancher dans le temps; opposer le parlé à l'écrit, c'est lui assigner une place bien à part (…). Toutes ces séparations sont faites, en général, sans la moindre étude sérieuse préalable. On sépare le français parlé du reste avant même de savoir en quoi il consiste, avant de l'avoir défini, comme s'il s'agissait là d'une évidence. Ces mythes séparateurs circulent dans ce qu'on appelle 'l'opinion commune', certes ; mais ils se glissent aussi dans les études de bien des spécialistes" (1987: 11).

> When considering the studies which have been devoted to spoken French since the beginning of the twentieth century, one is struck by the persistence of a few notions which all lead to a separation of what is commonly called "spoken French" from all of the other manifestations of language. Thus, spoken French always appears to be isolated and put aside, in order either to discredit or to praise it. Associating spoken French with popular speech means withdrawing it from legitimate language; considering it as the origin of linguistic innovation or conservatism means withdrawing it in time ; opposing spoken to written language means that it could be given a status of its own (…). All such distinctions are usually made without any careful prior study. Spoken French is separated from everything else before we even know what it consists of, before we even define it, as if it was obvious. Such separating myths certainly spread throughout what could be termed "common opinion", but they also creep into many researchers' studie" (1987: 11).

In the eighties, there was an urgent need for gathering data about spoken French, since reliable documentation was very scarce at that time in France; Claire Blanche-Benveniste, assisted by the others members of the *Groupe Aixois de Recherche en Syntaxe* (GARS), soon became fascinated about the various aspects of corpus design, which she considered as an extremely noble scientific task which needed to be achieved with extreme care, and which led her to develop rigorous transcription methods. Among the many difficulties raised by the transcription of oral documents, she often mentioned those relating to the listening process:

> "La difficulté à 'entendre' la langue parlée est plus grande qu'on ne pourrait le croire avant d'avoir essayé. Ce que nous entendons est un compromis entre ce que nous fournit la perception elle-même et ce que nous reconstruisons par l'interprétation" (1997a: 27).

> "'Hearing' spoken language is a more challenging activity than we could think before we experience it. What we do hear is a compromise between what is given by perception itself and what is reconstructed through interpretation" (1997a: 27).

Regarding transcription method, the GARS opted from the very beginning for a strictly orthographic presentation of the data:

> "La forme graphique des mots est celle des dictionnaires, y compris pour les majuscules sur les noms propres et les onomatopées […] Aucun trucage de l'orthographe n'est admis, même pas le procédé très répandu qui consiste à mettre une apostrophe pour signaler qu'une voyelle ou une consonne graphique, habituellement prononcée est absente" (1997a: 29-30).

> "The graphic transcription of words is that found in dictionaries, including for uppercase as the initial letter of proper names and for onomatopoeias […] No orthographic modification is allowed, not even the widespread method consisting of writing an apostrophe in order to indicate that a vowel or a consonant, which is usually pronounced, has been omitted" (1997a: 29-30).

Punctuation marks are banned from all transcriptions:

> "L'équipe du GARS a choisi de ne pas mettre la ponctuation, qui préjugerait trop vite de l'analyse à faire" (1997a: 34).

> "The GARS team chooses not to use punctuation marks, since they would prejudge the result of the analysis that has to be carried out" (1997a: 34).

Throughout her career, Claire Blanche-Benveniste defended the idea that corpus elaboration was a crucial aspect of linguistic research and she is fully recognized in France as a major proponent of what is now called "corpus linguistics". But although she proved outstandingly capable of defending the need for linguistic data, she remained convinced until the end of her life that this aspect of research has been largely neglected in France, to the extent that grammatical descriptions are unfortunately bound to remain fragmentary. In her latest

book which was published in 2010, she acknowledges that much effort is still needed in order to achieve a comprehensive description of spoken language:

> "Nous manquons encore d'instruments pour décrire la grammaire du français parlé dans toute son ampleur et dans toutes ses variétés. Il y faudrait de grandes quantités de données enregistrées et transcrites, c'est-à-dire de grands corpus de l'ordre de dix millions de mots, qui font défaut pour l'instant" (2010: 1).

> "We are still lacking instruments that would allow us to give a full and varied description of the grammar of spoken French. Large amounts of recorded and transcribed data would be necessary, that is large corpora of approximately 10 million words, which we do not have for the time being" (2010: 1).

Besides the need for corpora, she has always been willing to develop new methods of syntactic description: she was critical of both the framework of traditional grammar - which is not a valid method of linguistic analysis - and more formal theories which were essentially based on introspective data and showed very little interest about the description of authentic linguistic facts.

Therefore, much of her scientific activity became devoted to re-analyzing some major linguistic concepts in order to find out how adequate they where regarding the description of spoken productions. Most of these studies were published in the journal "Recherches sur le français parlé" (Studies on Spoken French) which she founded and supervised between 1979 and 2001.

One of the descriptive difficulties that had to be faced was the pervasive presence of those linguistic facts that would now be termed "disfluencies", like hesitations, fillers, repetitions, repairs and fresh starts (1983b, 1987, 1987 with C. Jeanjean, 1993). Claire attached great importance to all of these phenomena through which the "production process" was made visible:

> "L'oral spontané des conversations est un merveilleux observatoire du langage en train de se faire. Il livre, comme le dit Halliday, un processus de production, alors que nous étions habitués, par l'étude de l'écrit imprimé, à juger d'un produit fini. Il nous permet d'observer le producteur de langage en acte, de voir comment il lance un syntagme et le retouche, et comment il infirme ou confirme le discours qu'il est en train de faire" (1993, p. 16).

> "Spontaneous spoken conversations offer a wonderful observatory for language being created. They deliver, as Halliday says, a production process, whereas the study of written texts has taught us to evaluate a finished product. They

enable us to observe the language producer in action, to see the way he initiates a syntagm and modifies it, and the way he validates or invalidates his on-going discourse" (1993: 16).

Such characteristics explain why it would be unconceivable to undertake studies on spoken texts without considering their specific "production modes" ("modes de production"), about which Claire made numerous observations:

> "Loin d'être des obstacles qu'il faudrait supprimer pour accéder à l'analyse, les modes de production de la langue parlée sont de précieuses indications sur la structuration syntaxique" (1997a: 89).

> "Far from being obstacles that should be removed in order to conduct the analysis, the production modes of spoken language give valuable indications about syntactic structure" (1997a: 89).

In order to give a linguistic status to such phenomena which are so common in spontaneous spoken corpora, she undertook to reanalyze the notion of paradigm, with a desire to overcome the classical distinction between syntagmatic relations conceived as "in praesentia" relationships, and paradigmatic relations conceived as "in absentia" associations; she showed that in unplanned discourse, paradigamatic relations often take the form of "paradigmatic listings", which can be conveniently presented through what we call "syntactic grids" (1979 with B. Borel *et al.*, 1990b); let us take the following example:

> *il y avait des sacs d'olives + pas des sacs + des cartons + des cagettes d'olives*

> *there were bags of olives + not bags + boxes + crates of olives*

The speaker does not give the complement of the verb as one single Noun Phrase, but makes several successive trials in order to find the lexical version that suits him the best. Those four versions would be presented on different lines in order to show their paradigmatic relationship:

there were	*bags*	*of olives*
not	*bags*	
	boxes	
	crates	*of olives*

Another concept which demanded some clarification was the notion of "sentence" (2002). Claire Blanche-Benveniste did not consider such a unit as an adequate basis for the description of syntactical dependencies. She was convinced that the most useful units for grammatical analysis are those that can be

described on a clear morphological basis, such as Verb Phrases, Noun Phrases and so on. By contrast, she argued that sentence-units, which cannot be defined in a very precise way, should be avoided as a syntactic notion, especially when it comes to spoken language:

> "When describing language in our everyday lives we use units such as "word" or "sentence". We have learnt those terms along with the writing process and we apply them to spoken language, as for example in: "What was that word he used in his sentence?". There are some linguists who consider that these units may not be used scientifically to describe the spoken language because they are simply approximations. Furthermore they are characteristics of "practical knowledge" which may indeed accommodate the social rules of writing, but which is fundamentally different from "scientific know-how" (1997b: 21).

The difficulties involved by the concept of "subordination" are another theme which she has studied in a very original way.

> "Les séquences de verbes que l'on trouve dans les productions de français parlé posent à l'analyste des problèmes complexes, qu'il n'est pas facile de résoudre avec les notions courantes de subordination et coordination. Le concept de subordination se révèle trop pauvre ; il ne permet pas de rendre compte des divers degrés d'imbrications possibles entre deux verbes. Au lieu d'une seule relation de subordination, il en faudrait plusieurs, permettant de décrire une gamme de relations syntaxiques" (1983a: 71).

> "The sequences of verbs that we can find in spoken French pose complex problems of description, which are not easily solved by the use of such traditional concepts as subordination and coordination. The notion of subordination proves to be too poor; it fails to accurately reflect the diversity of intricacy relationship between two verbs. Instead of just one relation of subordination, we would need several different types in order to describe a wide variety of syntactic relations" (1983a: 71).

Some other major works focus on the analysis of elements without syntactic dependency, the description of dislocated sequences; the analysis of the way in which verbal constructions are organized according to a range of different "sentence types" (like cleft, pseudo-cleft, and so on); the distinction between micro- and macro-syntax.

We will end this very short evocation by pointing out that Claire's legacy to the community of linguists is outstanding: she had a truly rigorous and creative way to bring to light essential aspects of language and languages; and she always knew how to combine her exceptional erudition and a genuine curiosity for linguistic facts. She will be deeply missed.

References

Blanche-Benveniste, C., Chervel, A. (1969). *L'orthographe*, Paris: Maspero.

Blanche-Benveniste, C. (1975). *Recherches en vue d'une théorie de la grammaire française. Essai d'application à la syntaxe des pronoms*. Paris: Champion.

Blanche-Benveniste, C. (1979). Études de textes de français parlé: syntaxe et variation. In *Théories et pratiques de la sociolinguistique*. Rouen, pp. 305--314.

Blanche-Benveniste, C., Borel, B., Deulofeu, J., Durand J., Giacomi, A., Loufrani, C., Meziane, B. and Pazery, N. (1979). Des grilles pour le français parlé. In *Recherches sur le français parlé*, n° 2, pp. 163--206.

Blanche-Benveniste, C. (1981). La complémentation verbale: valence, rection et associé. In *Recherches sur le français parlé*, n° 3, pp. 57--98.

Blanche-Benveniste, C. (1982). La escritura del lenguaje dominguero. In E. Ferreiro & M. Gomez Palacio (Eds.), *Nuevas perspectivas sobre los proccessos de lectura y escritura*. Mexico, Editorial Siglo XXI, pp. 247--270.

Blanche-Benveniste, C. (1983a). Examen de la notion de subordination. In *Recherches sur le français parlé*, 4, pp. 71--115.

Blanche-Benveniste, C. (1983b). L'importance du « français parlé » pour la description du « français tout court' ». In *Recherches sur le français parlé*, n° 5, pp. 23--45.

Blanche-Benveniste, C., Deulofeu, J., Stéfanini, J. and Van den Eynde, K. (1984). *Pronom et syntaxe. L'approche pronominale et son application au français*. Paris: SELAF.

Blanche-Benveniste, C. (1987). Syntaxe, choix de lexique et lieux de bafouillage. In *DRLAV*, n° 36-37, pp. 123--157.

Blanche-Benveniste, C., Jeanjean, C. (1987). *Le français parlé. Transcription et édition*. Paris: Didier-Erudition.

Blanche-Benveniste, C. (1990a). Grammaire première et grammaire seconde: l'exemple de « en ». In *Recherches sur le Français Parlé*, n° 10, p. 51--73.

Blanche-Benveniste, C. (1990b). Un modèle d'analyse syntaxique « en grilles » pour les productions orales. In L. Tolchinsky (Ed.), *Anuario de Psicologia*, vol. 47, Barcelona, pp. 11--28.

Blanche-Benveniste, C., Bilger, M., Rouget, C., Van den Eynde, K. and Mertens, P. (1990). *Le français parlé: études grammaticales*. Paris: Editions du CNRS.

Blanche-Benveniste, C. (1993). Une description linguistique du français parlé. In *Le gré des langues*, 5, pp. 8--28.

Blanche-Benveniste, C. (1997a). *Approches de la langue parlée en français*. Gap/Paris: Ophrys.

Blanche-Benveniste, C. (1997b). The Unit in Written and Oral Language. In C. Pontecorvo (Ed.), *Writing*

Development. An interdisciplinary view, Amsterdam-Philadelphia: Benjamins, pp. 21--45.

Blanche-Benveniste, C, Mota, M.A., Uzcanga Vivar, I., Simone, R. and Valli, A. (1997). *Eurom 4: méthode d'enseignement de quatre langues romanes: Portugues, Espanol, Italiano, Français*. Firenze: Nuova Italia Editrice.

Blanche-Benveniste, C. (1998). Langue parlée, genres et parodies, *Repères*, 17, pp. 9--19.

Blanche-Benveniste, C., Adam, J.-P. (1999). La conjugaison des verbes: virtuelle, attestée, défective. In *Recherches sur le français parlé*, n° 15, pp. 87--112.

Blanche-Benveniste C. (2000a). La linguistique descriptive au XX° siècle. In Y. Michaux (Ed.), *Qu'est-ce que l'humain ? Université de tous les savoirs*, vol. 2, Paris: Odile Jacob, pp. 56--68.

Blanche-Benveniste, C. (2000b). Transcription de l'oral et morphologie. In M. Guille, R. Kiesler (Eds.), *Romania una et diversa*, Vol. 1, *Philologische Studien für Theodor Berchem zum 65. Geburstag*, Tübingen: Gunter Narr Verlag, pp. 61--74.

Blanche-Benveniste, C. (2001). Nouveaux apports de la grammaire contrastive des langues romanes. In I. Uzcanga Vivar, E. Llamas Pombo and J. M. Pérez Velasco (Eds.), *Presencia y renovación de la lingüística francesa*, Ediciones Universidad de Salamanca, pp. 41--54.

Blanche-Benveniste, C., Pallaud, B. (2001). Le recueil d'énoncés d'enfants: enregistrement et transcriptions. In *Recherches sur le Français Parlé*, n° 16, pp. 11--38

Blanche-Benveniste, C. (2002). Phrase et construction verbale, *Verbum* XXIV, 1-2, pp. 7--22.

Blanche-Benveniste, C., Rouget, C. and Sabio, F. (2002). *Choix de textes de français parlé: 36 extraits*. Paris: Champion.

Blanche-Benveniste, C. (2003a). L'orthographe. In Marina Yaguello (Ed.), *Le grand livre de la langue française*, Paris: Seuil, pp. 345--389.

Blanche-Benveniste, C. (2003b). L'oral des adultes parodié par les enfants, *L'École Valdôtaine*, n° 59, Cahier pédagogique, nouvelle série, Région autonome de la vallée d'Aoste, pp. 7--14.

Blanche-Benveniste, C. (2004). Le singulier et le pluriel en français parlé contemporain. In *Bulletin de la Société de linguistique de Paris*, T. XCIX, fasc. 1, pp. 129--154.

Blanche-Benveniste, C. (2005). Le corpus de français parlé du GARS. In E. Burr (Ed.), *Tradizione e Innovazione. Il parlato: teoria, corpora, linguistica dei corpora. Atti del VI Convegno SILFI*, Florence: Franco Cesrai Editore, pp. 57--76.

Blanche-Benveniste, C. (2009). Suggestions de recherches à mener pour entraîner la perception orale d'une langue romane à d'autres. In M.-C. Jamet (Ed.), *Orale e intercomprensione tra lingue romanze, Ricerche e implicazioni didattiche*, Universita Ca'Foscari Venezia: Le Bricole, pp. 19--32.

Blanche-Benveniste, C. (2010). *Le français. Usages de la langue parlée* (avec la collaboration de Philippe Martin pour l'étude de la prosodie). Leuven-Paris: Peeters.

PLENARIES

Speech and corpora: how spontaneous speech analysis changed our point of view on some linguistic facts

Philippe MARTIN

EA 3967, UFRL, Univ Paris Diderot Sorbonne Paris Cité
175, rue du Chevaleret. 75013 Paris, France
philippe.martin@linguist.jussieu.fr

Abstract

The elaboration of a grammar of spontaneous speech is of paramount importance. It allows focusing on pragmatic, information, syntactic and prosodic structures and leads to a better understanding as how they interact in actual speech. Indeed, many applications and industrial development in speech synthesis and recognition badly need coherent models to be integrated in their software, whereas today even successful systems rely mainly on word spotting if recognized speech does not simply results from oral written text reading. Other applications in oral language learning are also important by departing from traditions linguistic approaches based on written text. New tools are now becoming available to execute the main tasks involved in spontaneous speech studies: data speech recording, transcription, alignment, annotation. None of these tasks are trivial and require a sound expertise, but are essential for the future of linguistic studies. Current research in the domain of syntax-intonation interaction already revealed unexpected results for supposedly well-known prosodic items, such that sentence modality, congruence with syntax, stress clash, left and right dislocation, parenthesis, etc. These results could not have been discovered without careful analysis of actual spontaneous speech data, as the traditional available linguistic models, particularly in syntax, were, and still are, highly conditioned by the analysis of written text, a very specialized and limited mode of linguistic communication indeed.

Keywords: spontaneous speech; macrosyntax; intonation; transcription; alignment.

1.   Introduction

The last 60 years saw the advent of new and more and more sophisticated speech analysis tools which gave researchers the opportunity to test existing theoretical phonological models, especially those devoted to sentence intonation. Complex models elaborated from the linguist intuitions were tested against actual speech data, in well-defined production conditions first (laboratory speech), in various real life conditions later (spontaneous speech). Technology advances needed to perform satisfactory acoustic analysis were of paramount importance in these endeavours, and became gradually essential in the design of new corpora containing pertinent data in various discourse production conditions. At the same time, in prosodic studies, the quest for the correct and reliable measure of fundamental frequency became pivotal.

2.   Technological advances

Phoneticians were already using laboratory speech in the early XXth century. Rousselot (1901, 1908) for instance used a modified kymograph (Figure 1) to obtain rudimentary speech waveforms from which it was possible to derive values of laryngeal frequency in function of time. This was done by visual identification of the period or group of periods on the waveform. The duration of analyzed speech was of course quite limited and speakers had to be physically present to produce recordings.

Later the spectrograph appeared and it became possible to analyze speech segments of 2.4 s from speech recordings made elsewhere (Figure 2).
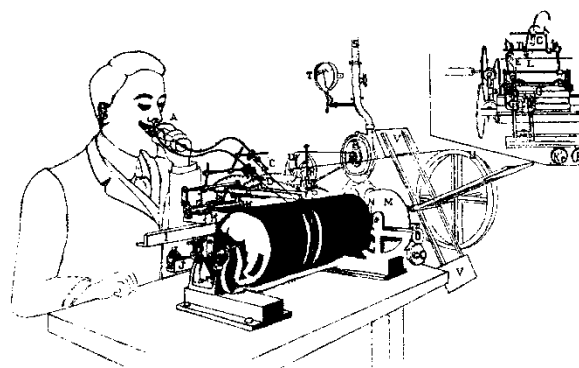
Fig. 1   Le kymographe de Rousselot (*Principes*, p. 1167)

Figure 1: Rousselot kymograph

Figure 2: Kay Elemetrics Sonagraph (the first model appeared in 1951)

Still, the visual identification and measure (from 10th harmonic for example, in order to achieve a

reasonable precision) was quite time consuming, not to mention that the spectrogram frequency scale was not always linear…

This rather painful evaluation of melodic curves lead to the development of specialized software programs such as "Pitch analyzers" (Signalyze, WinPitch, Praat,…).
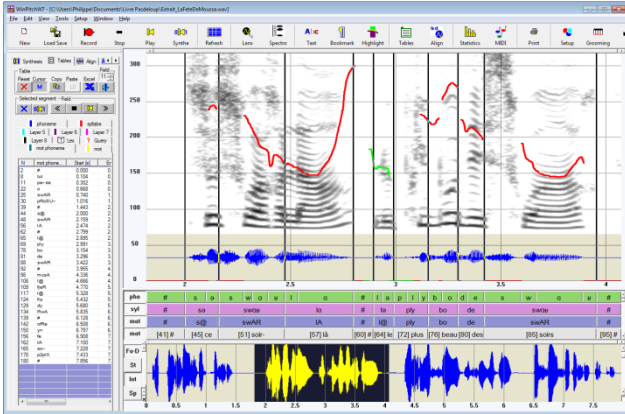
Figure 3: WinPitch display

More recently, elaboration of rather large spontaneous speech corpora lead to the development of specialized software programs such as WinPitch (2012) to transcribe, annotate and align recorded data.

3.   First results

Among the first changes of point of view pertaining to phonology, the use of a kymograph by Rousselot (1901, 1908) then by Grammont (1933) lead to a better understanding of stressed vowels duration. Later, the advent of the spectrograph made possible one of the first phonetic if not phonological, description of basic intonation patterns in French based on acoustical analysis (Figure 4) by Delattre (1966).

Figure 4: The 10 basic intonation patterns for French by Delattre (1966)

4.   Theoretical changes

Since two decades at least, the so-called Autosegmental-Metrical (AM) model has been dominant in intonation phonology. In this model, the prosodic structure organizes hierarchically prosodic events (PE) in three non-recursive levels: a first level assembles syllables σ, content words Wc (verbs, nouns adjectives and adverbs) and function words Wf (conjunctions, pronouns,…) into accentual phrases (AP); a second level groups AP into intonation phrases (IP) (Figure 5); finally a phonological utterance (PU) eventually groups sequences of IP.
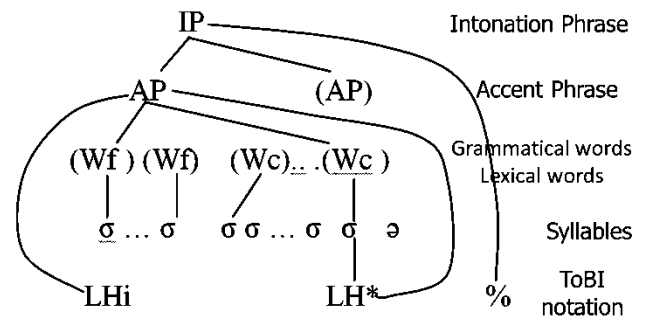
Figure 5: Autosegmental-Metrical prosodic structure

The prosodic events PE are aligned on accentual phrases specific syllables and are described as sequences of tones belonging to the ToBI notational system (tones and break indices). This system uses High (H) and Low (L) symbols to transcribe melodic targets as perceived or observed on fundamental frequency curves obtained from the speech signal acoustic analysis.

A revision of this model has been proposed recently to include an "intermediate phrase" (Figure 6).
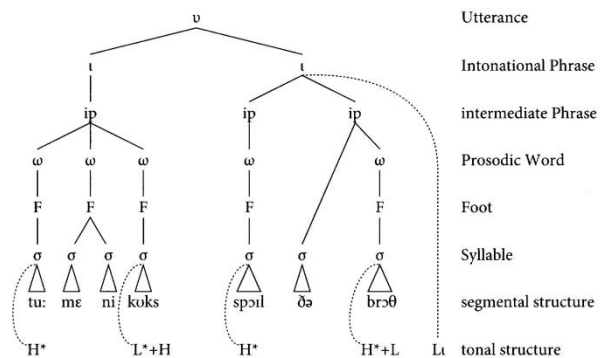
Figure 6: Modified Autosegmental-Metrical prosodic structure

An alternate approach has been proposed by Martin (1975, 1987) where the prosodic structure is a priori independent and associated to other structures organizing the sentence, syntactic, informational, etc. (Figure 7 below).

PY: Prosodic and syntactic structure

Syntactic Structure

Association between syntactic (upper tree) and prosodic structures (squared tree)
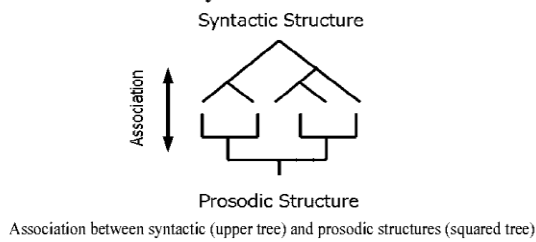
Figure 7: Independent prosodic structure associated to the syntactic structure

In this latter approach, the prosodic structure organizes hierarchically stress groups (i.e. prosodic words sequences of a maximum of 7+/- 2 syllables with only one lexical stress) normally formed with a content word (verb, noun, adjective or adverb) and one or more grammatical words (pronoun, conjunction,…) in dependency relation with content words. Furthermore, the prosodic structure is subject to the following constrains:

a. Stress clash: no consecutive stressed syllables;
b. Syntactic clash: no grouping of prosodic words whose corresponding text is dominated by distinct nodes in the syntactic structure;
c. Eurhythmicity: if more than one prosodic structure can be associated with a given syntactic structure, the most eurhythmic (i.e. with balanced number of syllables at each level) will be chosen by the speaker;

These models of the prosodic structure were able to describe various prosodic phenomena in French, such as:

1. Sentence Modality
2. Contrast of melodic slope
3. Congruence with syntax
4. Stress clash
5. Stress group
6. Left dislocation
7. Right dislocation
8. Parenthesis
9. Parallelism with syntax

Nevertheless, all these characteristics had to be reviewed when confronted to actual spontaneous speech data.

4.1   Modality

The last melodic contour, normally placed on the last stressed syllable in French, has been shown to be correlative with the declarative or interrogative modality of the sentence. But spontaneous speech data show many examples where the speaker uses ponctuants such as *hein*, *quoi*, *voilà*, etc. to signal the end of the sentence and at the same time a declarative modality (Figure 8). In such case, the last stressed syllable is placed on the ponctuant, which often carries a flat or even slightly rising melodic contour.
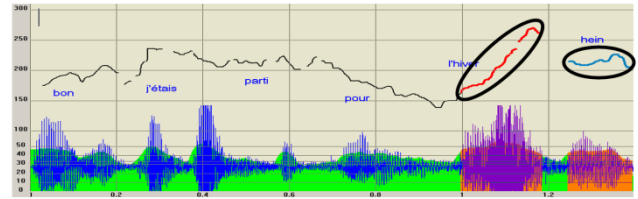
Figure 8: *bon j'étais parti pour l'hiver hein* sentence ended by a declarative ponctuant *hein* with a flat melodic contour

4.2   Contrast of melodic slope in French
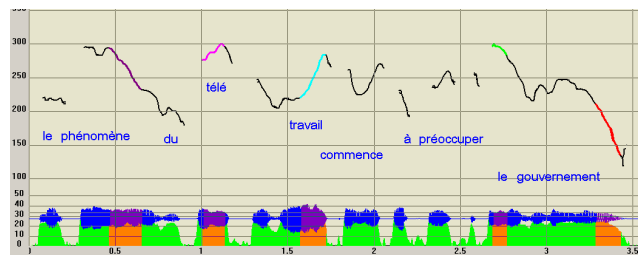
Figure 9: An example of contrasts of melodic slope indicating the prosodic structure for the read sentence *le phénomène du télétravail commence à préoccuper le gouvernement*

The prosodic structure in French is normally indicated by a contrast of melodic slope, correlative of a dependency to the right defining the hierarchical grouping of prosodic words. An example of read sentence *le phénomène du télétravail commence à préoccuper le gouvernement* is given Figure 9. The stressed syllable of *phéno**mèn**e* carries a falling contour indicating a dependency towards the rising melodic contour located on the stressed syllable of *télétra**vail*** to form the group [[*le phénomène*] [*du télétravail*]].

This larger group is itself integrated with the group [[*commence à préoccuper*][*le gouvernement*]] to constitute the complete sentence as indicated by the contrast between the rising contour on *phéno**mèn**e* and the falling contour ending the sentence and located on *gouverne**ment***.

Spontaneous speech data reveal other possible realizations of markers indicting the prosodic structure. A counter example is showed Figure10, where all melodic contours are falling, the contrasts indicating dependency to the right being implemented by differences in frequency height.

Examples of Figure 11 and Figure 12 demonstrate the neutralization process of melodic contours: when no further contrast has to be realized to indicate a lower level of prosodic words grouping in the structure, the contours take any shape as long as they are not to be confused with contours belonging to a higher level. Variants are thus possible, as schematized in Figure 13.
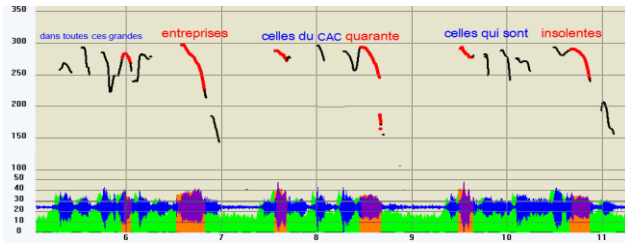
Figure 10: *dans toutes ces grandes entreprises celles du CAC quarante celles qui sont insolentes* pronounced by speaker SR with contrasts of fundamental frequency height using only falling melodic contours
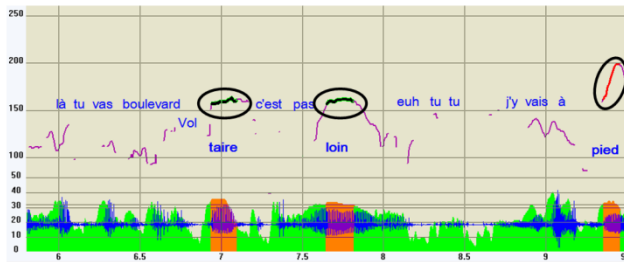
Figure 11: *tu vas boulevard Voltaire c'est pas looin euh tu tu j'y vais à pied* the contrast of melodic slope is neutralized and contours are realized flat, whereas the group ends with a rising contour.
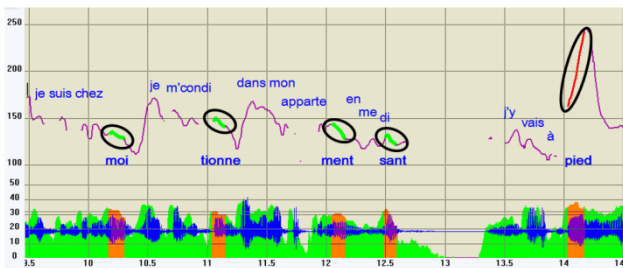
Figure 12: *je suis chez moi je me conditionne dans mon appartement en me disant j'y vais à pied* although the contrast of melodic slope is neutralized and contours are realized falling, whereas the group ends with a rising contour
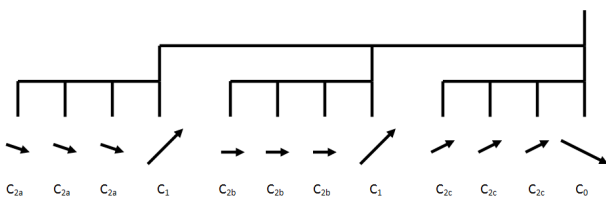
Figure 13: variants of melodic contours in a 2 level structure.

4.3   Congruence with syntax

The earlier assumed congruence between (macro)syntax and prosodic structure has been abandoned since some time. Figure 14 shows an example of non-congruence.
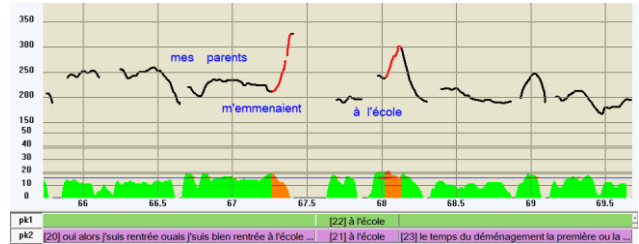
Figure 14: An example of non-congruence between macrosyntactic units …(*la première semaine*) (*mes parent m'emmenaient*] [*à l'école*) (*le temps du déménagement*]) (*la première ou la deuxième*…) CFPP2000 07-02. Syntactic phrasing is indicated by parentheses (). And prosodic phrasing by brackets []

In this example, the prosodic phrasing merges together the segment *à l'école* with *le temps du déménagement*, whereas *à l'école* is syntactically the complement of *m'emmenaient*.

4.4   Stress clash

Stress clash, already observed by Meigret (1550 !) should actually be revised according to the corresponding syntactic grouping of the prosodic words involved. When the corresponding words are grouped by the syntactic structure, a stress shift occurs. This is not the case when the words are grouped at different levels in the syntactic structure. Stress shift actually indicates the first case, as *liaison* in French in certain cases.

Figure 15: two cases of stress clash, inducing or not a stress shift according to the grouping of corresponding syntactic units

4.5   Stress group

As mentioned earlier, the prosodic word is defined as containing a content word (an open class word such as a verb, a noun, an adverb or an adjective) on which may depend one or more grammatical words (closed class words such as a pronoun, a conjunction, etc.). In French, it is easy to find counterexamples obtained by expansion, all stressed on the last syllable, as seen in the example below.

l'ar__moir__e
la petite ar__moir__e
la petite armoire __roug__e
la jolie petite armoire __roug__e

When the number of syllables exceeds a certain threshold (usually 7 +/- 2, depending on speech rate), a second stressed syllable must be realized, as in

la jolie petite ar__moir__e vert bou__teil__le

These simple examples show that speech rate is the important factor. When pronounced at a slow speech rate *la pe__tit__e ar__moir__e* would require two stressed syllables. The same process applies to rare long words in French, such as

an__ti__constitutionnelle__ment__ and
paraske__vi__dekatriapho__bi__e (fear of Friday 13*).*

which require (at least) two stressed syllables to be pronounced.

Conversely, a prosodic word may contain only one syllable as in *je te le demande __po__-__li__-__ment__* or s*i je te le demande __po__-__li__-__ment__, tu le feras ?* where the three syllables of *poliment* are stressed and separated by a short pause.

4.6   Left dislocation

In the literature (Mertens, 2008), the left dislocated segment is typically ended by an obligatory rising contour, as shown in Figure 16.
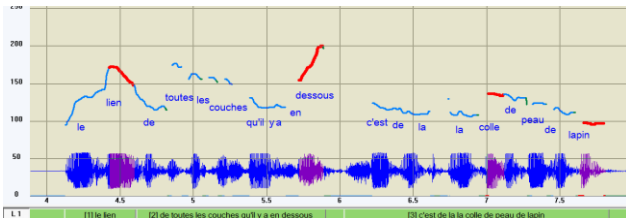
Figure 16: A prototype of left dislocation, the melodic contour ending the dislocated segment is rising *le lien de toutes les couches qu'il y a en dessous* before the Nucleus *c'est de la colle de peau de lapin* (ex.: Avanzi)

Figure 17: A counter example, where the prosodic structure is non-congruent to the left dislocation of *d'abord des passions* before the Nucleus *je m'en invente tous les jours* (ex.: Avanzi)

Nevertheless, spontaneous data contain many examples where the obligatory melodic rise is not found, as in Figure 17.

In this example, the prosodic structure merges the dislocated segment *d'abord des passions* with the main clause (nucleus) *je m'en invente tous les jours*, a phrasing non congruent with the left dislocation.

4.7   Right dislocation

A typical example of right dislocation (called in the macrosyntactic theory *postfix*) where the dislocated segment carries rather flat melodic contours on its stressed syllables is shown on Figure 18.

Figure 18: A typical example of a right dislocated segment *lui aurait di rostro* with flat and low melodic contours on stressed syllables (ex.: Avanzi)

An interrogative version of right dislocation is given in Fig 19. In this case, the final melodic contour of the dislocated segment (Postfix) is clearly rising.
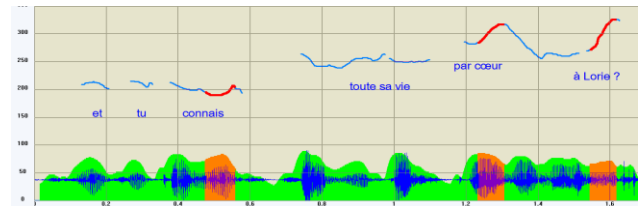
Figure 19: An example of interrogative right dislocation *et tu connais toute sa vie par cœur à Lorie ?*
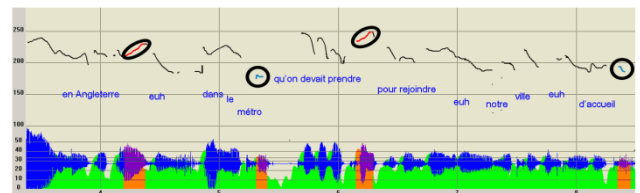
Figure 20: an example of "*complement rapporté"* where the complement of *le metro* is prosodically added to the nucleus. *Il y a eu un attentat en Angleterre euh dans le métro qu'on devait prendre pour rejoindre euh notre ville euh d'accueil*

It is thus possible to have a segment which follows the end of the sentence as indicated by a (declarative) conclusive contour.

Another but different example is shown in Figure 20, where the speaker initially intended to finish her sentence after *metro*, as indicated by the conclusive falling contour on the last syllable of *metro*. She then added the

complement *qu'on devait prendre pour rejoindre euh notre ville euh d'accueil* ended as well with a similar falling declarative contour.

4.8   Parenthesis

Traditionally, in the literature, parentheses are supposed to carry melodic contours of reduced variations, and a faster speech rate. Again, these characteristics are almost never found in spontaneous speech (Debaisieux et Martin, 2007). An example is shown Figure 21.
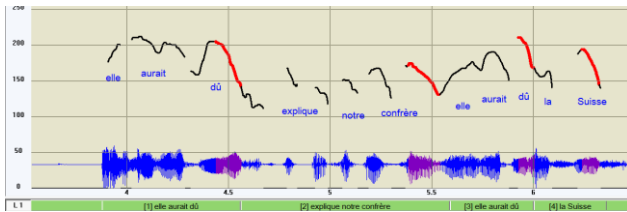
Figure 21: An example of parenthesis where melodic contour variations are not reduced *elle aurait dû explique notre confrère elle aurait dû la Suisse…*(ex.: Avanzi)

The parenthesis this example shows melodic variations and speech rate similar the the ones found in the main clause.

4.9   Parallelism with syntax

These various example lead to reconsider the parallelism with syntax assumed in most sentence intonation theories. Following C. Blanche-Benveniste and Martin (2011), the prosodic structuration operates *after* those effectuated by morphology and syntax. This appears clearly from the analysis of reprisals, when the speaker interrupts the flow of discourse in the middle of a stress group, and then starts over with a complete new stress group, never with an incomplete one.

5.   Constrains revisited

The observations presented above lead to reconsider to prosodic structure on the following points:

a. The prosodic word can contain one to a maximum of 7 +/- n syllables, depending on the speech rate, which actually determine in fine the maximum number of syllables that can form a prosodic word;

b. A prosodic word can contain more than one open class (lexical) word adjective, noun, adverb or verb. A grammatical word can be associated with one prosodic word (ex. **moi** in **moi** mon pa**pa** il est presi**dent**);

c. Stress clash induces the first stress involved in the clash to be shifted to the left or deleted only if the prosodic words involved are grouped together by syntax, i.e. if they are directly dominated by the same node in the syntactic structure;

d. The prosodic structure is more than often (at least in non-prepared speech) independent from

other structures organizing the sentence units (syntactic, informational, etc.);

e. In particular, one prosodic group can be associated with left dislocated syntactic segments together with the nucleus that follows;

f. The contrast of melodic slope in French (melodic rise with a dependency to the right towards melodic fall, melodic fall with a dependency to the right towards melodic rise, is not necessarily used by some speakers as other melodic features such as syllabic duration (ex. in whispered speech) or melodic frequency variation can ensure this function instead.

g. Furthermore, melodic contours which do not have to contrast with other melodic contours ending prosodic groups at a lower level in the structure (case of neutralization) can therefore present reduced frequency variations.

6.   Dynamic cognitive model

At this point, a sketchy revision of the concept of prosodic structure can be outlined, underlying the fact that the structure does not appear statically with all its melodic contours at once, but rather in sequence along the time axis, the contours being perceived and decoded one after the other by listeners.

Recent research (Gilbert & Boucher, 2007) suggests that these sequences of syllables are converted into higher linguistic units by one of three processes: a final stress syllable (in French), an identified rhythmic pattern or a direct pattern identification (i.e. the sequence is directly recognized as part of the lexicon).

In this process, acoustic features triggering the conversion of syllabic sequences in short term memory into higher rank linguistic units, be a final syllabic stress or a rhythmic pattern, are not identical along the sentence. On the contrary, a least for melodic contours, they are differentiated in order to allow the listeners to reconstitute the hierarchy intended by the speaker as a prosodic structure. In French, this process involves a dependency relation to the right, i.e. to the future prosodic events taking place along the time axis, and uses in priority features such as contrast of melodic slope, together with syllabic duration and melodic contour frequency span and height (Martin, 2009).

7.   Delta and Theta waves

These formal constrains governing the prosodic structure may find their justifications in recent neurophysiological investigations in speech processing. For instance, research in electro-encephalography suggests that the cortex Delta wave frequency range (1 to 4 Hz) governs stress groups size (maximum 7 +/- 2 syllables) as well as the eurhythmicity process, while Theta waves (frequency range 4 to 10 Hz) synchronize the perception of syllables by listeners.

EEG Theta and Delta waves are synchronized
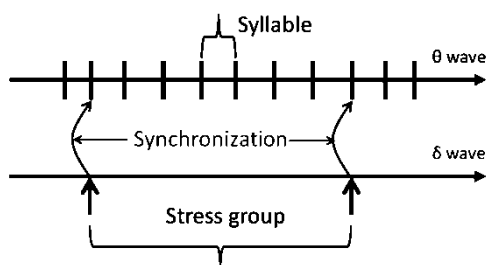
Figure 21: Process of synchronization between EEG Theta and Delta waves. Theta waves determine the minimum and maximum duration of syllables, whereas Delta waves synchronize the conversion and transfer of sequences of syllables into larger linguistic units (the stress groups or prosodic words)

The following cognitive interpretation of the prosodic structure rules can be proposed:

a. The 7 syllables rule reflects the memorization capacity of syllabic sequences;
b. The Stress clash rule would allow enough processing time for syllabic sequences conversion;
c. Eurhythmy corresponds to an optimization of the syllabic sequences conversion process;
d. The Syntactic clash rule prevents impossible syllabic sequences conversion;
e. Hesitations allow an interruption of the conversion process.

8.   Conclusion

An old vision of linguistic sees spontaneous speech data as full of "errors" compared to "correct" speech represented in written text. These views of correctness of language production lead today to phonological laboratory research for prosodic studies and the analysis of read speech only.

By contrast, spontaneous speech analysis shows how well established characteristics of the prosodic structure for instance had to be reviewed when confronted to actual data not found in laboratory speech.

Again, these "divergences" could have been (and have been) simply discarded as typical of every day's speech and did not really reflect the competence of the speakers. An alternate and more appropriate view would on the contrary lead to a revision of the model, in our case the prosodic structure constrains, allowing the theoretical views to evolve.

9.   References

Blanche-Benveniste, C. (2000). *Approches de la langue parlée en français.* Ophrys, Paris.

Blanche-Benveniste, C. & Martin, P. (2011) Structuration prosodique, dernière réorganisation avant énonciation. In Lefeuvre, E.M. (Ed.), *Unités syntaxiques et unités prosodiques*, Florence, Langue Française (170), pp. 127-142

Delattre, P. (1966). Les dix intonations de base du français. In *French Review* 40, pp. 1--14.

Debaisieux, J., Martin, P. (2007). Les parenthèses: analyse macrosyntaxique et prosodique sur corpus, *in* M.-J. Béguelin, M. Avanzi and G. Corminboeuf (Eds.), *La parataxe. vol. 2: Structures, marquages et exploitation discursive* Université de Neuchâtel, Suisse.

Gilbert, A., Boucher, V. (2007) What do listeners attend to in hearing prosodic structures? Investigating the human speech-parser using short-term recall. In *Proc. Interspeech* 2007, pp. 430--443

Grammont, M. (1933). *Traité de phonétique*, Delagrave, Paris.

Jun, S.-A. (1996) *The Phonetics and Phonology of Korean Prosody: intonational phonology and prosodic structure*, Garland Publishing Inc. New York.

Martin, P. (1975). Analyse phonologique de la phrase française, In *Linguistics*, 146, pp. 35--68.

Martin, P. (1987). Prosodic and Rhythmic Structures in French, In *Linguistics*, 25-5, pp. 925--949.

Martin, Philippe (2009). *Intonation du français.* Armand Colin, Paris.

Mertens, P. (2008). Syntaxe, prosodie et structure informationnelle: une approche prédictive pour l'analyse de l'intonation dans le discours. In *Travaux de Linguistique,* 56(1), pp. 87--124.

Meigret, L. (1550). *Le treté de grammere francoeze.* Réimpression chez Slatkine, Genève, 1972.

Rousselot, Pierre-Jean (1901-1908). *Principes de phonétique expérimentale*, Tomes 1 et 2, Didier, Paris.

WinPitch (2012). Available at: <www.winpitch.com>.

C-ORAL-BRASIL I: corpus de referência do Português Brasileiro falado informal. A general presentation

Tommaso RASO, Heliana MELLO

Universidade Federal de Minas Gerais

Av. Antônio Carlos, 6627, 31270-901 Belo Horizonte, Brazil

tommaso.raso@gmail.com, heliana.mello@gmail.com

Abstract

This paper presents the C-ORAL-BRASIL *corpus* (Raso & Mello, 2012), a spontaneous speech *corpus* of informal Brazilian Portuguese. The *corpus* is comparable in architecture and segmentation criteria with the four C-ORAL-ROM *corpora* (Cresti & Moneglia, 2005). C-ORAL-BRASIL presents 75% of private/familiar texts and 25% of public texts; for each context 1/3 of texts are monologues, 1/3 dialogues and 1/3 conversations. The *corpus* is text to speech aligned through WinPitch (Martin, 2005). Its main goal is to document the diaphasic variation with the widest range of different communicative situations. Segmentation of the speech flow is done through prosodic criteria in utterances and tone units. Utterances, defined as the smallest pragmatically interpretable unit, end with a prosodic break perceived as conclusive, while tone units end with a non- terminal break. Diastratic representation is also very well balanced. The diatopic represented is that of the Belo Horizonte metropolitan area. Transcriptions are made with criteria that aim to represent grammaticalization or lexicalization phenomena in speech, but attempting to maintain easy readability of the texts and consistency in transcribers' perception. Validation criteria lead to a Kappa of 0.86 for segmentation and a very low number of errors for transcriptions.

Keywords: Corpus; Brazilian Portuguese; Spontaneous Speech

1. Introduction

The C-ORAL-BRASIL (Raso & Mello 2012) [1] is a Brazilian Portuguese spontaneous speech *corpus*, especially representative of the mineiro diatopy, majorly from the metropolitan region of the state capital Belo Horizonte. The texts were recorded with sophisticated wireless equipment, in order to guaranty highly accurate acoustic quality, between 2006 and 2011. C-ORAL-BRASIL is structured in order to be comparable with the C-ORAL-ROM project corpora (Cresti & Moneglia, 2005) [2] for French, Italian, Spanish and European Portuguese. Here, we will list the central information about the corpus and the motivation for its architecture and sampling methodology, trying to show the advantages that they present for the study of spontaneous speech, mainly in a pragmatic perspective.

The corpus DVD contains:

i) the multimedia corpus, made up of the following archives for each text: audio (wav), transcription (rtf) and aligned file (xml) through WinPitch software (Martin, 2005), and txt file;

ii) the metadata: title, file name, participant abbreviation and their main sociolinguistic characteristics (gender, age, school level, occupation and role played in the interaction), recording date, place, context and topic, corpus branch, duration in time and number of words, acoustic quality, transcriber and revisers' names, and any commentary considered useful;

iii) the corpus tagged lexically and morphosyntactically (Bick, 2012 and in this volume) in full version (xml and txt) and in a simplified version (xml and txt);

iv) frequency lists, spreadsheets with interesting measurements and statistics about the informants;

v) a book, in pdf format in which audio examples are linked to the text, with the corpus description, a presentation of the theory behind it, the explanation for transcription and segmentation criteria and their validation, a discussion of the main speech measurements, and finally a description and discussion of the parser used for the lexical and morphosyntactic tagging.

The corpus transcription format follows CHAT (MacWhinney, 2000), implemented for prosodic annotation (Moneglia & Cresti, 1997); the corpus is segmented in utterances and tone units (Raso 2012b; Mello *et al.* 2012). The utterance is defined as the minimal unit with pragmatic autonomy. Its identification is marked by a prosodic break perceivable as terminal.

> Example 1 (*bfammn03*)[3]
> *ALO: *mas os filho também nũ são fácil também juntou os filho todo foram lá e trouxeram o corpo na força*
> [but the sons too they are not easy either they all meet (they) go there and bring the body by force]

The linguistic sequence in Example 1 can, in principle, be segmented in different ways. A simple

---

[3] All the example cited in this paper can be listened to in the C-ORAL-BRASIL DVD.

---

reading induces the interpretation of mas os filho também nũ são fácil também as an autonomous entity, since it is syntactically autonomous; the rest can be also interpreted as one or more entities. Nevertheless, by listening to the sequence, it is clear that there is just one autonomous entity, that is, one utterance, segmentable as follows:

*ALO: *mas os filho também nũ são fácil também / juntou os filho todo / foram lá e trouxeram o corpo na força //* 🔊

The double slash marks a terminal break, that is, the utterance frontier; the single slash marks a tone unit frontier. In fact, the first part of the sequence, which could seem autonomous in reading, it is not perceived as such through listening to the actual recording - 🔊.

Example 2 and 3 show that the same syntactic structure (in these cases a principal proposition followed by a relative clause) can be the locutionary content of one or more than one utterances, depending on the prosodic realization:

Example 2 (*bfamdl02*) 🔊
*BAL: *tá saindo de uma garrafinha que tem um bico muito pequeno //*
[It's coming out from a little bottle with a very small neck]

Example 3 (*bfamdl02*) 🔊
cê tá com um jarro d´água // que tem uma espessura assim //
[you have a water pottery that has a width like this]

Apparently in example 3 we have only one utterance. But listening to the sequence we realize that it perform two autonomous utterances - 🔊 and 🔊.

Example 3 shows that the same linguistic sequence can be interpreted by a reader as a negative assertion, while by listening to it, it is clear that it is an affirmative one preceded by another utterance that expresses refusal:

Example 4 (*bpubdl01*) 🔊
*PAU: *não // tá dando a altura daquele que a Isa marcou lá / né //*
[no // it has the height of that one that Isa marked there / isn´t it // also interpretable by the reader as: it doesn´t have the height of that one that Isa marked there / does it //]

Example 5 and Figure 1 show that the terminal break does not necessarily match with a pause. As the figure shows, there is no pause between the first and the second utterances, while utterance two and three are divided by a pause:

Example 5 (*bfamdl02*) 🔊
*BAL: *tá saindo de uma garrafinha que tem um bico

muito pequeno // então daquela coisa pequeninim nũ vai encher rápido // agora imagina cê pega um balde e joga dentro //
[It's coming out from a little bottle with a very small neck // so that little thing can't fill it quickly // now you imagine you fill it with a full bucket //]
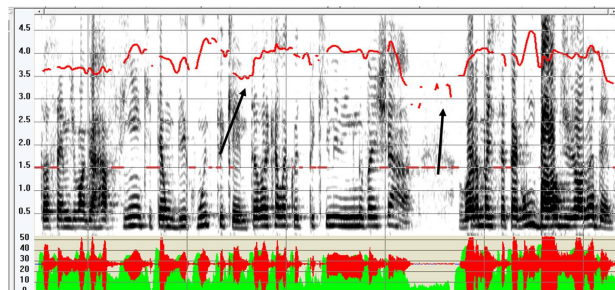
Figure 1: Example 5 in WinPitch software 🔊

The opposite is also true: a pause, even a long one, does not imply an utterance frontier, like example 6, where we have a pause of 1281 ms. inside the utterance:

Example 6 (*bpubdl11*) 🔊
*MAR: *o ensino tá [/1] tá assim / difícil / mas tá mais fácil / né hhh //*
[teaching is how can I say difficult but easier]

Therefore, only by listening to a verbal sequence it is possible to understand where a pragmatically interpretable reference unit ends. Hence, it is not possible to analyze speech without audio, nor is it possible to transcribe speech without marking the reference units that make it possible to segment it. That cannot be perceived by reading, nor be automatically measured through a pause (Moneglia, 2005).

These are the main reasons why the C-ORAL-BRASIL is sound-text aligned by utterance. Alignment is a crucial aspect in the study of speech. Without sound alignment, the text cannot be appropriately studied, since the audio source turns out to be unusable and unrecoverable for research. The real object of study, in this case, would be the transcription, which represents a special variety of writing, without the basic characteristics of speech, above all prosody. In our view, it is not possible to study speech without its acoustic information, that alone allows for the recognition of the main categories of speech, illocution being the basic one (Cresti, 2000b). In fact, example 4 shows that a pure syntactic and semantic analysis that does not result from an illocutive interpretation cannot account for the understanding of speech:

Example 7 (*bfamdl04*) 🔊
*KAT: *o quê //* [what //]
*SIL: *copos // copos de Urano / que tem aí //* [glasses // glasses from Urano / that are here //]
*KAT: *copos de quê //* [glasses from where //]
*SIL: *Urano //* [Urano //]

KAT: Urano // [Urano //]
SIL: é // Urano // Urano // [yeah // Urano // Urano]

It is only through the different illocutions that we can recover the different meanings of *Urano* in the different utterance. Its communicative function cannot be recovered just through its semantic and syntactic forms.

2.   The architecture

By spontaneous speech, we mean that speech is planned at the same moment it is performed, i.e. speech that does not perform a previously, totally or partially, planned text, like acted speech or even a previously planned discourse (Nencioni, 1983; Cresti, 2000a; Biber, 1988; Blanche-Benveniste *et al.*, 1990; Miller & Weinert, 1998; Givón, 1979; Moneglia, 2005, 2011). Spoken events that can be considered spontaneous show: i. a multimodal face-to-face interaction; ii. intersubjective reference to the deictic space; iii. mental programming at the same time as vocal performance; iv. contextually undetermined linguistic behavior, i.e. unforeseen behavior.

A long tradition of sociolinguistic studies (Berruto, 1987; Biber & Conrad, 2001; Biber *et al.*, 1998; Gadet, 1996a, 1996b, 1997, 2000, 2003; Halliday, 1989) focused on the value of sociological and contextual parameters to define speech qualities, and pointed to their variability. There are many types of spontaneous speech, and they vary according to the following parameters: a) the possible structural varieties of the communicative event (monologue, dialogue, conversation); b) the communicative channel; c) the sociological context, that is, the social domain of the event (family, private, public life); d) the programming conditions (partially or totally programmed *versus* non programmed speech); e) possible register and genre varieties; f) sociolinguistic factors (gender, age, school level, speaker's occupation); g) geographic origin; h) speech event task; i) topic.

Planning a spoken corpus is, therefore, a complex task that must ensure representativity of the principal variations explored by the different types of events in spontaneous speech (Berruto, 1987; Biber, 1988; De Mauro *et al.*, 1993; Gadet, 1996a, 1996b, 2003). The speech resources built so far, usually having technology needs as their objective (telephone information, health interactions, map tasking), were produced in controlled situations. This allows a very high acoustic quality, but represents restrict semantic domains, with highly foreseeable linguistic behavior. C-ORAL-BRASIL, like C-ORAL-ROM, collects data in natural context, which necessarily reduces acoustic quality and causes many more difficulties for recording. C-ORAL-BRASIL underwent a great effort to obtain the best acoustic quality for recordings in very different contexts, using sophisticated wireless equipment.

An important goal of this *corpus* is to achieve comparability with the C-ORAL-ROM *corpora*. Comparable corpora have been built for: written language, parallel corpora or corpora of the same specialized topic.

For spoken corpora the first case implies building up reading corpora, losing spontaneity. In speech, comparability can easily be reached only in strongly controlled situations. But if we assume that spontaneous speech is necessarily documented maximizing the textual variation, the consequence is that the more textual variation we have, the less comparability we obtain. Therefore, the comparability among the corpora of the C-ORAL projects results from the application of the same specific compilation parameters.

The C-ORAL-BRASIL *corpus* so far proposes only the informal part of a spontaneous speech corpus. The formal part is still to be completed. The informal corpus features 208,130 words, distributed in 139 texts of, on average, 1,500 words each. A few texts are bigger (up to 5,000 words) or smaller (only if they maintain textual autonomy). The 139 texts were divided in two contexts: private/familiar (159,364 words) and public (48,766 word**s**); for each context the texts were divided similarly among three interactional typologies: monologues, dialogues and conversations (dialogic texts with more than two main participants).

Texts are transcribed using the CHILDES-CLAN format (MacWhinney, 2000) implemented for prosodic annotation (Moneglia & Cresti, 1997). The prosodic annotation features the segmentation of the speech flow in utterances (double slash) and tone units (single slash)[4]; interrupted utterances (+) and retracting ([/n])[5] are also marked. Transcriptions follow traditional orthography, with significant exceptions due to the wish of capturing speech phenomena that can show processes of gramaticalizations and lexicalizations going on, so that they can be computed and statistically studied[6].

2.1   The pragmatic perspective and the diaphasic variation

A truly spontaneous speech *corpus* must portray in the best possible way situational variation. In fact, what conditions speech structuring the most is not speakers or topic variation. Especially under a pragmatic perspective, it is crucial to document the differences in verbal behavior depending on the different tasks speakers should perform in different situations. If on one hand the sociolinguistic tradition allows us to identify the main domains of formal speech, on the other, the possible situations in informal speech cannot be categorized. Therefore, while in formal speech it is possible to list a certain group of typical contexts, informal variation must be left open. The goal is, therefore, to document the widest range of situations, as no specific context can be considered, in principle, more typical than another. In order for this to be possible, considered that the cost (both economic and especially

---

[4] For the segmentation theoretical frame, see Cresti (2000); for the segmentation and validation methodology, see Mello *et al.* (2012), Raso and Mittmann (2009), Moneglia *et al.* (2010).
[5] The number means the quantity of retracted words.
[6] For the transcription criteria, see Mello and Raso (2009) and Mello *et al.*, (2012).

concerning time for recordings and transcriptions) of a spoken *corpus* is much higher than that for written ones, it is important to provide many different texts and reduce their size. The average of 1,500 words is sufficient for the interaction to be autonomous, since a text must show its syntactic and its pragmatic properties (Blanche-Benveniste *et al.*, 1990; Scarano, 2003), and at the same time allows the representation of a wide variety of situations.

Within the informal register, the partition in private/familiar versus public context documents the role that the participant plays, whether he acts as an individual, as for example in interactions with relatives or friends, or in a professional or institutional role, like, for example, in interactions between client and seller or student and professor or citizen and public officer, etc. Around 75% of the *corpus* represents the private/familiar context, since it normally occupies a larger space in human natural interactions.

Inside each context, there are three different typologies of interactions: i) a monologic typology, in which a speaker builds a spoken text, (almost) without any interaction; ii) a dialogic typology, in which two interlocutors interact; iii) a conversational typology, in which three or more speakers interact. The text characteristics are strongly conditioned by the interaction typology, especially in the opposition between monologic *versus* interactional[7]. It must be highlighted, however, that, differently from formal, informal register does not show, in principle, perfect monologic texts. Almost always there will be some kind of interaction. The criterion used to assign a text to this typology was the fact that the construction of a spoken text keeps developing even after the interlocutor's interventions which, in the majority of cases, are not considered by the speaker. The monologic typology is built by long turns and within them, by very articulated utterances with complex information structure and many tone units, stretching in a strongly processual way. The reference to the situational context is usually poor, while a great amount of cognitive contextualization is necessary. Depending on the textual typology of the monologic text, the more frequent illocutions change, but the illocutionary variation is poor. On the other hand, interactional typologies show short turns and small informationally patterned utterances; in these cases, the reference to the situational context is strong, making a high amount of verbal contextualization not necessary, while the illocutionary variation inside the same text is very high[8].

After this important distinction between monologic and interactive typologies, the most important factor of variation is dependent upon each typology. In monologic typology, speech structure depends mainly on textual genre: life tale, professional explication, argumentation, joke, recipe, story, etc. In dialogues and conversations, variation is basically due to the task speakers are performing: a chat between friends is much different from a couple's quarrel, or from an interaction between seller and client, or among the players in a football game, or between a personal trainer and the athlete, or between mother and crying child, or between two interactants performing a task together, etc. It is evident that in each activity the actions to be performed change completely, as well as the turn size, the amount of silence, etc.

These observations should be sufficient to understand how crucial the importance of real diaphasic variation is in a spontaneous speech *corpus*. Speech structuring variation cannot be documented through speaker's or topic variation. Different speakers perform the same action in basically the same way, and the change in topic in chats or interviews does not lead to structural variation, i.e. illocutionary and information structure variation (Cresti, 2000b).

2.2 Diastratic variation

Although diaphasic variation had been the main goal while building the *corpus* architecture, diastratic variation is very well documented. What is important methodologically is that, while diaphasic variation has no chance to be documented aiming only for diastratic variation, our methodology shows that diastratic variation is a natural consequence of the diaphasic one (Cresti & Moneglia, 2012).

C-ORAL-BRASIL features 362 speakers. For 68.23% of them gender, age, origin and school level are documented. The fact that more than 30% of the speakers are not fully documented is explained by the fact that they entered the recording context in an unforeseeable way. This strongly reinforces the point that the recording context was really uncontrolled. Moreover, they are responsible for only 1.91% of the *corpus* words. A cluster analysis is shown in table 1:

clusters	speakers
1 - 247 words	161
280 - 627 words	81
649 - 908 words	37
933 - 1016 words	16
1134 - 1400 words	26
1455 - 1663 words	17
1777 - 1994 words	7
2140 - 2455 words	10
2611 - 2901 words	2
3550 - 3738 words	2
4211 - 4327 words	2
6309 words	1
TOTAL	362

Table 1: Cluster grouping: number of words for speakers

---

[7] See Raso and Mittmann (2012).
[8] See Raso (2012b) and Mittmann and Raso (2012).

Table 1 shows that 44.5% of the speakers utter up to 247 words, accounting for just 3.92% of the *corpus*. Table 2 shows the cluster grouping these 247 speakers:

CLUSTERS	speakers
1 - 22 words	82
25 - 47 words	27
54 - 72 words	10
77 - 95 words	13
99 - 115 words	6
136 - 164 words	9
172 - 185 words	5
204 - 247 words	8
TOTAL	161

Table 2: Cluster grouping of speakers uttering up to 247 words

Table 2 shows that the great majority of the non-documented speakers (109) utters up to 47 words, and that more than half of them (82) utter up to 22 words. Table 1 shows also that the *corpus* features a small group of speakers (5) that utter more than 3,550 words each, representing 10.63% of the *corpus* words. These speakers appear in more than one recording in different situations and may be studied to see how the same speaker's structure varies in different contexts.

Gender balancing is perfect in terms of number of words: 50.36% of the speakers are female and 49.64% are male. In terms of speakers, 203 are female and 159 male (one informant utters just one word and his/her gender was not identified). Age balancing is also very good (measurements in words): 27.13% of the speakers belong to group A (from 18 to 25 years old); 30.28% to group B (from 26 to 40 years old); 31,01% to group C (41 to 60 years old); 8,05% to group D (more than 60 years old); 1.61% are underage and 1.91% are not documented for age. The *corpus* is very well balanced as far as speakers older than 18 years are concerned, considering that group D in Brazilian society is smaller than the others. As far as the number of speakers is concerned, 75 are in group A, 1 is registered in group A in one interaction and in group B later, 88 pertain to group B, 64 to group C, 15 to group D and 11 to group M.

Schooling is very well represented for mid and high schooling levels, the most relevant in the representation of the language synchronic standard use, but low schooling level is also sufficiently represented. Taking word numbers, 15.79% represent level 1 (no more than 7 years of school), 40.76% represent level 2 (up to college graduation if the degree is never used for their occupation), 40.66% represent level 3 (using college degree for their job or more than college graduation). As for number of speakers, 46 pertain to group 1, 101 to group 2, 104 to group 3, and one speaker is registered once in group 2 and once in group 3.

The last diastratic aspect is the speakers' occupation, which is an open category and cannot be treated like the previous ones. Looking at the metadata, the importance of occupations linked to the education field is clear. This happens for different reasons: because professors and students that worked in the *corpus* compilation appear in the recordings; because they looked for informants in their social environment (that of course is linked to the education system); because age group A is to a great extent formed by students. Nevertheless, in the group linked to education we find students and professors from different faculties, different level teachers, school directors and school clerks. But a significant part of the informants have occupations outside the education system: the *corpus* features many shop attendants and sellers, artists, public clerks, liberal professionals from very different fields (attorneys, doctors, psychologists, dentists, engineers, physiotherapists, etc.), housekeepers, technicians, brokers, craftsman, workers, masons, managers, farmers, and many other occupations.

2.3   Other aspect architecture

2.3.1.   Diatopy

As mentioned above, the diatopic variation of C-ORAL-BRASIL is essentially that of the *mineiro* variety of Brazilian Portuguese. A *corpus* of this size must concentrate in representing other variations inside one diatopy. The same happens with the C-ORAL-ROM Project *corpora*, which represent the regions of Madrid, Marseille, Florence and Lisbon (Cresti *et al.*, 2002). In all the *corpora,* speakers of other regions and countries are present, since a big metropolitan area implies a percentage of people from outside, but what is mandatory for each *corpus* is that more than 50% directly represents the chosen diatopy. For the C-ORAL-BRASIL this diatopy is the metropolitan area of Belo Horizonte, the state of Minas Gerais capital city. Table 3 shows the informants origin distribution:

Origin	Speakers
Belo Horizonte	138
Other cities in Minas Gerais state	89
Other Brazilian states	19
Other countries	2
Unknown	114
Total	**362**

Table 3: Speakers origin

Excluding the speakers without origin documentation which, as we have already seen, represent an irrelevant *corpus* percentage, 55.6% of the speakers are from Belo Horizonte and 35.9% from other cities in Minas Gerais state (many of them from cities inside the metropolitan area, like Contagem, Betim, Sete Lagoas, etc.). Therefore, 91.5% of the documented speakers represent the *mineiro* variety, and much more than 50%

that of the Belo Horizonte area.

2.3.2. Considerations about some texts

One of our main efforts was to reduce as much as possible, especially in dialogues, the incidence of chats and interviews, that is, situations in which the interlocutors do not perform any activity besides that of speaking. These are already the most well documented situations in oral *corpora*, the easiest to be recorded and also the less interesting if the goal is to document speech structuring. Among dialogues, only 8 in 48 texts can be considered chats or interviews. Among conversations, the incidence of chats is higher because this typology is more frequently characterized for lack of specific actions. For this typology, 17 texts in 42 can be considered chats, but no one of those occurred in public context.

Specific attention was given to recording in moving contexts, since static and dynamic actionality can be faced as two different actional macro-domains: 4 dialogues were recorded completely or partially between informants in a moving car while one of them was driving (*bfamdl03*; *bfamdl05*; *bfamdl08*; *bpubdl04*). But dynamic recordings are 19: among conversations, *bfamcv03* is a recording of friends playing snooker; *bfamcv05* is a recording of friends playing football (with a very high acoustic quality); *bfamcv10* is a recording of a group preparing lunch; *bpubcv01* is a recording of a visit to a blood donation centre; *bpubcv09* is a recording of a gym session; other conversations also feature dynamic parts. Among dialogues, besides those already mentioned, *bfamdl01* was recorded in a supermarket while two friends were shopping; *bfamdl04* is a recording of two maids cleaning the kitchen and other rooms; *bfamdl05* is the recording of a broker driving and showing different apartments to his sister; *bfamdl26* is the recording of mother and daughter cleaning the apartment; *bpubdl02* and *bpudl06* are recordings inside a store, while a client tries on shoes and dresses; *bpubdl03* is a recording of a gym lesson with a personal trainer; *bpubdl05* is a visit to a bee breeding; *bpubdl07* is a recording of two waiters preparing and serving pizza at a party; other dialogues also features dynamic parts.

A few texts are longer than the average. The decision to have these in the *corpus* was taken to document a longer textual development or due to the specific characteristics of the texts: *bfamdl09* and *bfamdl31* have around 3,000 words. The latter one is especially interesting, as it documents two parallel dialogues. In fact, two microphones were placed on two informants who were repairing windows at home and were expected to interact with one another. The distance between them caused their interaction to happen as expected only in some circumstances. Most of the time each of them interacts with two other unforeseen participants, and two different dialogues, without overlapping, went on. We thought it was interesting to document parallel dialogues, although this phenomenon happens also in parts of other recordings. Among monologues, *bfammn14* features more than 4.800 words: this monologue is by an

informant from Serra do Cipó[9], area whose linguistic variety is considered particularly interesting (another informant of the same area is documented in *bfammn29*); *bpubdl07* has more than 3.100 words: it is the recording of waiters preparing and serving pizza at the party. This recording documents a particularly interesting context, since the waiters move around and have a lot of small interactions with different interlocutors, giving rise to speech acts not easy to document but very common in real life, like greetings and thanking.

Especially among monologues, some texts are smaller than the average and have less than 1,000 words (1 among dialogues, 3 among conversations and 12 among monologues), and in a few cases only a few hundreds of words. They are all concluded textual entities. Three more recordings deserve some observations: *bfamcv06* is the recording of a birthday party; it features an entire text with overlapping voices, but it is clearly understandable; we thought it would be interesting to document this aspect of speech, although of course it appears, with less evidence, in other recordings. Recording *bfamdl12* documents the interaction between an infant and his mother: the infant cries and the mother talks to him trying to calm him down; although we have only one speaker, it is clear that this text documents an interaction and not a monologue, since the mother speaks in reaction to the infant's actions. Therefore, the text must be considered a dialogue, with different turns, reacting to different actions of the interactant. Something similar happens in *bpubdl03*, in which a personal trainer gives instructions to a client in order for him to perform gym tasks: the client is almost silent, but the trainer turns are interactive in regard to the client's movement.

2.3.3. An informationally tagged minicorpus

In order to study information structure and illocutions, two comparable *minicorpora* were selected for Brazilian Portuguese and Italian (Mittmann and Raso, 2012; Raso and Mittmann, 2012, and Mittmann *et al.* in this volume)[10]. The two *minicorpora* feature around 32,000 words and 5,500 utterances. The texts were chosen to represent the widest diaphasic variation, with a good acoustic quality without repetition of the same speaker. A complex manual system was used to tag the *minicorpora*. The tagging was double-validated: first by a rater agreement among the three taggers, and then through a comparison between the Italian and the Brazilian taggers. The criteria for the informational tagging are documented in many of the works by both the Italian and the Brazilian teams, and are based on the Language into Act Theory (Cresti, 2000a)[11]. The goal of the two teams is to tag the

---

[9] A region more or less 100 km from Belo Horizonte, south part of the Espinhaço Mountain range.

[10] A search on the tagged minicorpora is now possible in the IPIC Database (http://lablita.dit.unifi.it/ipic/), elaborated by Panunzi & Gregori (2012).

[11] For the Italian production see http://lablita.dit.unifi.it/; for the Brazilian production see www.c-oral-brasil.org. For the criteria of the informationally tagged *minicorpora*, see Mittmann and

minicorpora illocutionary and to make possible studies that can analyze speech crossing lexical-morphosyntactic, informational and illocutionary tagging, taking advantage of all the potential of the resource elaborated by Panunzi and Gregori (2012; see also Gregori and Panunzi in this volume) that allows for the automatic analysis of the three different levels.

3.   Methodological aspects

3.1   Acoustic quality

To make good acoustic quality recordings in natural context is a challenging task; nevertheless, it is crucial for spontaneous speech corpora. It is not enough to make recordings that allow transcription of just the main segmental aspects; a speech corpus that aims to document not only lexicon and morphosyntax, but also phonetics and pragmatics, must have a much higher acoustic quality and must be aligned. Of course, it is easy to get a good acoustic quality in lab recordings and controlled situations, but it is much more difficult to obtain it in natural contexts, specially with a truly diaphasic variation.

Properties	Label
Very high quality. Excellent microphone response. Almost the entire recording is appropriate for almost all phonetic study. Almost no overlapping. Almost no back noise. F_0 computation possible for (almost) the entire file.	A
High quality. Excellent microphone response. Most of the recording is appropriate for almost all phonetic study. Few overlappings. Almost no back noise. F_0 computation possible for (almost) the entire file.	AB
Medium quality. Good or medium microphone response. Many parts of the recording are appropriate for phonetic analysis. F_0 computation possible for most of the file. Not many overlappings and not much back noise.	B
Low quality. Medium microphone response. F_0 computation in at least 60% of the file. Even when F_0 computation is not trustable, the recording is clearly understandable.	BC
Low quality. Medium or low microphone response. F_0 computation possible in at least 60% of the file. A few parts may not be clearly understandable.	C

Table 4: Acoustic quality parameters

First, it is important to have a high quality wireless recording equipment[12], but is also crucial to plan very carefully the recording situations, to record much more

time for each session and many more texts than the number that will be transcribed, in order to choose those with better quality without affecting other variables. This project recorded three times the published texts, and each recording was, in average, four times longer than the published transcribed duration. Table 4 shows the characteristics of all the acoustic quality labels used for the *corpus*. Table 5 shows the acoustic quality of all the texts. All recordings are in wav format, usually in stereo.

60% of the recordings show high or very high quality. Only 23% show low quality. Naturally, acoustic quality tends to be lower in conversations, due to their nature   (more overlappings and a more difficult microphone   management). In principle, low quality should be accepted only when the recording is particularly interesting for diaphasic, diastratic or diatopic aspects and it is not possible for the recording to avoid quality problems, as, for example, the back noise in a supermarket.

	A	AB	B	BC	C	Total
bfamcv	8	11	4	6	5	34
bpubcv	1	2	---	1	5	9
bfamdl	7	14	6	5	3	35
bpubdl	5	2	2	1	1	11
bfammn	13	10	10	1	2	36
bpubmn	3	4	4	2	1	14
Total	40	43	25	14	18	139

Table 5[13]: Texts acoustic quality

The importance of the acoustic quality for a pragmatic analysis is crucial, which in its turn, can be better appreciated in the next section and considering the importance of prosody for illocutionary and informational studies (Raso, 2012b; Cresti, 2012; Mello & Raso, 2012a, 2012b; Moneglia, 2011; Mittmann & Rocha in this volume; Cresti in this volume).

3.2   Segmentations

Criteria of speech segmentation represent a very important and original aspect of the C-ORAL-ROM and C-ORAL-BRASIL projects[14]. How to segment speech is a very relevant and discussed task (Moneglia, 2005). In written texts, the nature of reference units being acknowledged to be higher than words is not controversial. It is possible to choose different units for analysis, but writing is always segmented in discrete objects pertaining to syntax (Abeillé, 2003). Recently, the question of how to segment speech started to be seriously discussed (Blanche-Benveniste, 1997; Biber *et al.*, 1999; Cresti,

Raso (2012).
[12] The equipment used for C-ORAL-BRASIL is described in the specifications, inside the DVD (Raso & Mello, 2012).

[13] The file names are build like in C-ORAL-ROM: b=Brazilian; fam=private/familiar;   pub=public;   cv=conversation; dl=dialogue; mn=monologue. A sequence number follows the abbreviation for the corpus branch.
[14] For C-ORAL-BRASIL, see Mello *et al.* (2012), Raso (2012b), Raso and Mittmann (2009) and Moneglia *et al.* (2010).

2000a, 2000b; Miller & Weinert, 1998; Quirk *et al.*, 1985). Even if the utterance is taken as the reference unit, its definition changes depending on the author (see the discussion in Scarano, 2003). Among different definitions, we can cite that by Biber *et al.* (1999) and Blanche-Benveniste (1997). Other definitions (like Voghera, 1992) normally imply a verbal nucleus. However, it is necessary to explain the fact, shown in many languages, that in spontaneous speech verbless autonomous units are around 30% (Cresti, 2001; Cresti, 2005a, 2005b; Raso & Mittmann, 2012). Therefore, definitions based on clause structure or predication cannot explain spontaneous speech.

In the C-ORAL-BRASIL, like in the C-ORAL-ROM, prosodic breaks are taken as the most relevant feature to determine utterance frontiers.[15] Tone units are those portions of speech separated by prosodic breaks and a general correspondence between tone units and information units has been recognized since Halliday (1976). Taking this point of view, it is possible to divide the speech flow into information units. In fact, the perceptual relevance of prosodic breaks is strong. However, this correlation is not sufficient to individualize utterances, since the correspondence between information unit and utterance is not bi-univocal. An information unit can be not coincident to an utterance, and can be just part of it. An utterance can be compounded by one information unit, but also by more than one, being, in this case, performed by more than one tone unit (Cresti, 2000a; Hart *et al.*, 1990). Even so, the correlation between prosodic break and utterance can be maintained, considering classic positions in linguistic studies (Crystal, 1975; Karcevsky, 1931) that identify the utterance with a terminal profile. On these bases, prosodic breaks that conclude utterances can be distinguished from those that are not conclusive (see audio examples in this paper). Consequently, utterances presenting a bi-univocal correspondence with tone units can be distinguished from those not presenting this feature.

In C-ORAL-BRASIL and C-ORAL-ROM the identification of terminal prosodic breaks was considered the heuristic method to determine utterance frontier. Each sequence concluded by a terminal break inside the speech flow is considered an utterance. This correspondence is based on the assumption that linguistic actions are necessarily correlated with prosody, that constitute the interface between illocutionary and locutionary acts. Performing illocutionary acts is therefore considered the main property that a linguistic event must present in order to be considered an utterance. The illocutionary force determines how the propositional content of the utterance must be interpreted. This explains why the utterance is defined as *the minimal linguistic unit that allows a pragmatic interpretation*.

Prosodic features allow a competent speaker to interpret linguistic activity. Competent speakers are very

skillful in detecting even subtle prosodic variations, if voluntarily produced (Hart *et al.*, 1990). This is what happens when a linguistically codified profile is performed to express an illocutionary force or an information unit (Raso, 2012b; Moneglia, 2011). Of course, segmentation can only identify the speech act frontiers, without labeling them. To identify the speech act conclusion and to label it are two different processes.

In the introduction we showed how speech is segmented in the corpus, and we briefly explained the main perceptual motivations for the segmentation. It is crucial to understand that speech and writing cannot be analyzed following the same criteria: prosody, absent in writing, is the main structural criterion of speech. Through prosody it is possible to identify the reference units of speech, to illocutionarily label these units, and segment the utterance in information units, identifying their specific functions (Cresti, 2000a; Moneglia, 2011; Raso, 2012b; Moneglia & Cresti, 2006). Example 8 tries to show how different the analysis in speech and writing can be:

Example 8
Não espera aqui em cima não
[do not wait here above no] or [no wait here above no]

In speech, the communicative value of this sequence depends on how we segment it:

a) *Não. Espera aqui em cima. Não.*
b) *Não espera. Aqui? Em cima? Não?*
c) *Não. Espera aqui! Em cima não!*
d) *Não, espera! Aqui em cima! Não!*
e) *Não. Espera! Aqui! Em cima! Não.*
f) *Não espera aqui em cima não!*
g) *Não espera aqui em cima? Não?*
h) *Não espera aqui? Em cima? Não!*
i) ...

These and other segmentation possibilities show how many speech acts we have in the sequence, and what type they are. Without prosody, we cannot make any of the following decisions:

1. Which are the frontiers in a sequence of speech, that allow us to individualize the different performed actions? Through the verbal sequence, the speaker is performing one, two or more actions? Which words pertain to each action?
2. What kind of speech act is being performed? A question, an order, a request, an assertion, an expression of surprise, etc.? How many information units compose the utterance? What are their specific functions (Cresti, 2012; Cresti & Firenzuoli, 1999, 2001)?

All those questions can receive an answer only through prosody. Note that even punctuation is far from

---

[15] For the relationship between prosodic breaks and utterance frontiers, see Simon (2004).

representing the possible illocutions and information structures. However, the main point is to underline that the reference unit in speech is the utterance, and that it corresponds to a speech act (Austin, 1962).

Of course the segmentation criteria need the segmentators to be trained and that the segmentations are statistically validated through inter-rater agreement[16]. We paid great attention to both, and the validation after the first revision, but before the last one, reached a Kappa (Fleiss, 1971) of 0.86, which means an excellent inter-rater agreement.

3.3   Transcriptions

An important implementation of C-ORAL-BRASIL was the choice of a specific set of transcription criteria for the segmental part. We wanted to capture a great quantity of phenomena that may be subject to grammaticalization and lexicalization, in order to study them with quantitative methodology and statistic criteria, also measuring their co-occurrence and the systemic relationship among them.

The criteria are based on the following parameters: i. the necessity to represent phenomena subject to grammaticalization and lexicalization (subject and negation cliticization, loss of verbal morphology, demonstrative reduction, articulated preposition contraction, loss of the verb *ser* in cleft constructions, government changes, aphaeresis, and many others); ii. the necessity to keep easy readability of transcriptions, excluding phenomena whose nature was exclusively phonetic, without evident grammatical effect; iii. the necessity to guaranty a coherent behavior of the transcribers, choosing clearly perceivable phenomena. An example of this last aspect is that of the cliticization of subject pronouns: while the distinction between tonic and clitic forms of the second and third person is relatively easy to perceive (*você(s)* versus *cê(s)* and *ele(s)* or *ela(s)* versus *e´, es, ea, eas*) the situation is different for the first person singular and plural; in this case we decide not to orthographically represent the opposition between tonic and clitic forms.

All the chosen phenomena are already known by linguists, but they were never documented through *corpora*. Only *corpus* based studies of spontaneous speech can truly document: a) how much these and other phenomena are really recurrent in spontaneous speech; b) to what extent they coexist and determine a deep change in the system; c) which are the most advanced phenomena that may trigger the others; d) what is their distribution based on sociolinguistic variations. If these phenomena were not marked in the transcription, it would not be possible to study them statistically. In fact, all the forms which differ from the orthographic tradition were implemented in the parser (Bick, 2012); this allows a large quantity of studies about on-going linguistic changes that would be impossible with manual techniques.

We want to emphasize the effort to choose, computationally implement and statistically validate the transcription criteria. They should be considered the most advanced stage for documentation of Brazilian Portuguese spontaneous speech at present.[17] We did two validations: the first one before the last revision and the second one after the last revision. The validation was divided in two sections: one section aimed to validate the transcription as a whole and another section concentrated only on the non-orthographic phenomena. The result was excellent: the errors in term of words in the transcription as a whole were 0.81%. The errors, considering as baseline only phenomena related with the non orthographic criteria, were only 0.43%. The phenomenon that presented more errors was that of articulated prepositions, with 3.28%. The baseline for the validation was 10% of the utterances of each text, chosen random.

4.   Acknowledgements

5.   References

Abeillé, A. (2003). *Treebanks Building and Using Parsed Corpora*. Dordrecht: Kluwer Academic.

Austin, L.J. (1962). *How to Do Things with Words*. Oxford: Oxford University Press.

Berruto, G. (1987). *Sociolinguistica dell'Italiano Contemporaneo*. Roma: La Nuova Italia Scientifica.

Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., Reppen, R. (1998). *Corpus Linguistics*: Investigating Language Structure and Use. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. Harlow-Essex: Pearson Education.

Biber, D., Conrad, S. (2001). Register variation: A corpus approach. In D. Schiffrin, D. Tannen and H. Hamilton (Eds.), *The Handbook of Discourse Analysis*. Oxford: Blackwell, pp. 175--196.

Bick, E. (2012). A anotação grammatical do C-ORAL-BRASIL. In T. Raso, H. Mello (Eds), *C-ORAL-BRASIL I: Corpus de Referência do Português Brasileiro Informal*. Belo Horizonte: Editora UFMG, 2012, pp. 223--254.

Blanche-Benveniste, C., Bilger, M., Rouget, Ch., van den Eyde, K. and Mertens, P. (1990). *Le Français Parlé*: Études Grammaticales. Paris: Éditions du C.N.R.S.

Blanche-Benveniste, C. (1997). *Approches de la Langue Parlée en Français*. Paris: Ophrys.

Cresti, E. (2000a). *Corpus di Italiano Parlato*. Firenze: Accademia della Crusca.

---

[16] For the training process and the validations, see Mello *et al.* (2012); Moneglia *et al.* (2010); Raso and Mittmann (2009).

[17] For a complete list and discussion of the transcription criteria, see Mello and Raso (2009); Mello *et al*., 2012; the specifications in the corpus DVD (Raso & Mello, 2012); the last two present also the transcription validation.

Cresti, E. (2000b). Critère illocutoire et articulation informative. In M. Bilger (Ed.), *Corpus, méthodologie et applications Linguistiques*. Paris: Champion, pp. 350--367.

Cresti, E. (2001). Per una nuova definizione di frase. In P. Bongrani, A. Dardi, M. Fanfani and R. Tesi (Eds.), *Studi di storia della lingua italiana offerti a Ghino Ghinassi*. Firenze: Le Lettere, pp. 511--550.

Cresti, E. (2005a). Enunciato e frase: teoria e verifiche empiriche. In M. Biffi, O. Calabrese and L. Salibra (Eds.), *Italia Linguistica: Discorsi di scritto e di parlato – nuovi studi di Linguistica Italiana per Giovanni Nencioni*. Siena: Protagon, pp. 249--260.

Cresti, E. (2005b). Notes on lexical strategy, structural strategies and surface clause indexes in the C-ORAL-ROM spoken corpora. In E. Cresti, M. Moneglia (Eds.), *C-ORAL-ROM: Integrated reference copora for spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins, pp. 209--256.

Cresti, E. (2012). The definition of Focus in the framework of the Language into Act Theory (LAcT). In H. Mello, A. Panunzi, A. and T. Raso (Eds.) *Pragmatics and Prosody. Illocution, modality, attitude, Information Patterning and Speech Annotation*. Firenze: FUP.

Cresti, E., Firenzuoli, V. (1999). Illocution et profils intonatifs de l'italien. In *Revue Française de Linguistique Appliquée*, IV, pp. 77--98.

Cresti, E., Firenzuoli, V. (2001). Illocution and intonational contours in Italian. In *Revue Française de Linguistique Appliquée* IV(2): pp. 77--98.

Cresti, E., Moneglia, M. (Eds.) (2005). *C-ORAL-ROM: Integrated Reference Corpora For Spoken Romance Languages*. Amsterdam: John Benjamins.

Cresti, E., Moneglia, M. (2012). In T. Raso, H. Mello (Eds.), *C-ORAL-BRASIL I: Corpus de Referência do Português Brasileiro Falado Informal*. Belo Horizonte: Editora UFMG, 2012, pp. 5--25.

Crystal, D. (1975). *The English Tone of Voice*. London: Edward Arnold.

De Mauro, T., Mancini, F., Vedovelli, M. and Voghera, M. (1993). *Lessico di Frequenza dell'Italiano Parlato*. Milano: ETAS.

Fleiss, J.L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, v. 76, pp. 378--382.

Gadet, F. (1996a). Niveaux de langue et variation intrinsèque. *Palympsestes*, 10: pp. 17--40.

Gadet, F. (1996b). Variabilité, variation, variété. In *Journal of French Language Studies*, 1: pp. 75--98.

Gadet, F. (Ed.) (1997). *Langue Française*, 115. Special Issue on *La Variation en Syntaxe.*

Gadet, F. (2000). Vers une sociolinguistique des locuteurs. In *Sociolinguistica*, 14: pp. 99--103.

Gadet, F. (2003). *La Variation Sociale en Français.* Paris: Ophrys.

Hallyday, M.A.K. (1976). *System and Function in Language*: *Selected Papers*. London: Oxford University Press.

Hallyday, M.A.K. (1989). *Spoken and Written Languages.* Oxford: Oxford University Press.

Hart, J., Collier, R. and Cohen, A. (1990). *A Perceptual Study on Intonation*: *An Experimental Approach to Speech Melody*. Cambridge: Cambridge University Press.

Karcevsky, S. (1931). Sur la phonologie de la phrase. In *Travaux du Cercle linguistique de Prague* IV, pp. 188--228.

MacWhinney, B.J. (2000). *The CHILDES Project*: *Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates, 2 v.

Martin, P. (2005). WinPitch Corpus: a text-to-speech analysis ans alignment tool. In E. Cresti, M. Moneglia (Eds.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins.

Mello, H. (2012). Os *corpora* orais e o C-ORAL-BRASIL. T. Raso, H. Mello (Eds), *C-ORAL-BRASIL I: Corpus de Referência do Português Brasileiro Informal*. Belo Horizonte: Editora UFMG, 2012, pp. 31--54.

Mello, H.; Raso, T. (2009). Para a transcriçao da fala espontânea: o caso do C-ORAL-BRASIL. *Revista Portuguesa de Humanidades*, v. 13, n. 1, p. 153-178.

Mello, H., Raso, T. (2012). Illocution, modality, attitude: different names for different categories. In H. Mello, A. Panunzi and T. Raso (Eds.), *Pragmatics and Prosody. Illocution, modality, attitude, Information Patterning and Speech Annotation*. Firenze: FUP.

Mello, H., Raso, T., Mittmann, M., Vale, H. and Côrtes, P. Transcrição e segmentação prosodica do *Corpus* C-ORAL-BRASIL: critérios de implementação e validação. In T. Raso, H. Mello (Eds), *C-ORAL-BRASIL I: Corpus de Referência do Português Brasileiro Falado Informal*. Belo Horizonte: Editora UFMG, 2012, pp. 125--176.

Miller, J., Weinert, R. (1998). *Spontaneous Spoken Language*. Oxford: Clarendon Press.

Mittmann, M.M., Raso, T. (2012). The C-ORAL-BRASIL informationally tagged minicorpus. In H. Mello, A. Panunzi, A. and T. Raso (Eds.), *Pragmatics and Prosody. Illocution, modality, attitude, Information Patterning and Speech Annotation*. Firenze: FUP.

Moneglia, M. (2005). The C-ORAL-ROM resource. In E. Cresti, M. Moneglia (Eds.), *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, pp. 1--70.

Moneglia, M. (2011). Spoken Corpora and Pragmatics. *Revista Brasileira de Linguística Aplicada*, Belo Horizonte, v. 11, n. 2, pp. 479--519.

Moneglia, M., Cresti, E. (1997). L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In U. Bortolini, E. Pizzuto (Eds.), *Il Progetto CHILDES Italia*. Pisa: Del Cerro, pp. 57--90.

Moneglia, M., Cresti, E. (2006). C-ORAL-ROM. Prosodic boundaries for spontaneous speech analysis. In Y. Kawaguchi, S. Zaima and T. Takagaki, (Eds.), *Spoken Language Corpus and Linguistics Informatics.* Amsterdam: John Benjamins.

Moneglia, M., Raso, T., Mittmann, M.M. and Mello, H. (2010). Challenging the perceptual relevance of prosodic breaks in multilingual spontaneous speech corpora: C-ORAL-BRASIL / C-ORAL-ROM. In Speech Prosody 2010, W1.09. *Satellite workshop on Prosodic Prominence: Perceptual, Automatic Identification*. Chicago.

Nencioni, G. (1983). *Di scritto e di parlato*: *discorsi linguistici.* Bologna: Zanichelli.

Panunzi, A., Gregori, L. (2012). An XML model for multi-layer representation of spoken language. In H. Mello, A. Panunzi and T. Raso (Eds.), *Pragmatics and Prosody. Illocution, modality, attitude, Information Patterning and Speech Annotation*. Firenze: FUP.

Raso, T. (2012a). O *Corpus* C-ORAL-BRASIL. In T. Raso, H. Mello (Eds), *C-ORAL-BRASIL I: Corpus de Referência do Português Brasileiro Falado Informal*. Belo Horizonte: Editora UFMG, 2012, pp. 55--90.

Raso, T. (2012b). O C-ORAL-BRASIL e a Teoria da Língua em Ato. In T. Raso, H. Mello (Eds.), *C-ORAL-BRASIL I: Corpus de Referência do Português Brasileiro Falado Informal*. Belo Horizonte: Editora UFMG, 2012, pp. 91--123.

Raso, T., Mello, H. (Eds.) (2012). *C-ORAL-BRASIL I: Corpus de Referência do Português Brasileiro Falado Informal*. Belo Horizonte: Editora UFMG.

Raso, T. Mittmann, M M. (2009). Validação estatística dos critérios de segmentação da fala espontânea no corpus C-ORAL-BRASIL. In *Revista de Estudos da Linguagem*, v. 17, n. 2, pp. 73--91.

Raso, T., Mittmann, M. (2012) As principais medidas da fala. In T. Raso, H. Mello (Eds.), *C-ORAL-BRASIL I: Corpus de Referência do Português Brasileiro Falado Informal*. Belo Horizonte: Editora UFMG, pp. 173--221.

Scarano, A. (Ed.) (2003). *Macro-syntaxe et Pragmatique. L'analyse Linguistique de l' Oral*. Roma: Bulzoni.

Simon, A.C. (2004). *La Structuration Prosodique du Discours en Français. Une Approche Multidimensionelle et Expérientelle.* Berne: Peter Lang.

Voghera, M. (1992). *Sintassi e Intonazione nell'Italiano Parlato.* Bologna: Il Mulino.

Grammatical annotation of the C-ORAL-BRASIL corpus

Eckhard BICK

Institute of Language and Communication, University of Southern Denmark
Odense, Denmark
eckhard.bick@mail.dk

Abstract

This paper describes the automatic annotation of the C-ORAL corpus with morphosyntactic tags. For this task, a modified version of the PALAVRAS Constraint Grammar parsers was used. In order to make the existing rule body work on speech data, a secondary, orthographically normalized data layer was introduced, maintaining the original transcription as well as prosodic and speech flow information while at the same time providing the parser with word forms in standard written Portuguese, using both pattern matches and a tailor-made lexicon. In addition to spelling variation and phonetic spellings, standardization was also applied to tokenization (non-standard contractions), overlaps and false starts. In order to accommodate for context-sensitive rules, syntactic "punctuation" was introduced based on prosodic break markers. The modified parser achieved correctness rates (F-scores) of 98.6% for part of speech, 95% for syntactic function and 99% for lemmatization. Experiments with unsegmented input showed that the use of prosodic breaks reduced syntactic error rates by two thirds, and PoS by half. However, the added effect of pause/break disambiguation affected only syntactic tags, not PoS tags, reflecting the two tag types' unequal dependence on long-distance contexts.

Keywords: Constraint Grammar; speech corpora; tagging; parsing; C-ORAL-Brasil.

1. Introduction

While linguistic interest in transcribed speech corpora has grown considerably in recent years, accessibility is often hampered by the lack of standardized markup and systematic searchability. Optimally, the necessary annotation should include not only phonetic issues, prosody, discourse structure etc, but also traditional morphosyntactic annotation. In this paper we will focus on how to integrate the latter with the former, and discuss the question whether and how a tagger-parser primarily designed for written language can be adapted to handle transcribed speech data. The work was carried out in the research context of the C-ORAL speech corpus project for Brazilian Portuguese (Raso & Mello, 2010, 2012), where morphosyntactic annotation was to be added automatically on top of an existing meta-annotation in the face of non-standard orthography and the absence of punctuation, preserving in-text speech flow markers etc.

Using automatic annotation, either on its own or as a pre-step for manual revision, is an obvious choice for a corpus of this size (~ 300.000 words). Thus, previous European C-ORAL sister projects employed statistical part of speech taggers for this task, such as the PiTagger system (Panunzi *et al*., 2004) for the Italian section, which had access to a lexicon-based analyzer, a standard lexicon (107.00 lemmas), a training corpus (50.000 words) and a special pre-dictionary covering about 2000 non-standard and dialectal forms. For the European Portuguese section, the Brill tagger (Brill, 1993) was used, trained on a written Portuguese corpus of 250.000 words. While no higher-level, syntactic annotation was attempted in the European C-ORAL, other speech corpus projects have opted for full treebank annotation, such as the Arabic treebank described by Maamouri *et al*. (2010), which combined manual selection of analyzer suggestion, followed by an automatic syntactic parsing stage.

2. Constraint Grammar parsing environment

For our own work we used the Palavras parser (Bick 2000) as a point of departure. Palavras is a Constraint Grammar (CG) parser that is mostly used for the annotation of written data, but has demonstrated great robustness in the face of genre variation - as, for instance, in the Linguateca project (linguateca.pt) and the CorpusEye corpora (corp.hum.sdu.dk). With lexical adaptation and various filter programs, the parser has also been used for non-standard language varieties, such as historical texts (Bick & Módolo, 2005). The Constraint Grammar paradigm (Karlsson, 1995) can be described as both a robust, modular disambiguation methodology for NLP, and a linguistic-descriptive convention, encoding linguistic analyses as token-based tags and function-mediated dependency structures. Both the method and the descriptive tradition offer a number of formal advantages for the annotation of non-standard language data such as speech. First, because CG systems have a modular architecture with a clear separation of lexica, analyzers and grammars (rule sets) for successive levels of analysis, it is relatively easy to add specialized lexica or morphological filters, as well as add specific grammar modules. Second, CG's token-based annotation, where even higher-level structural information is strictly token-based, allows a corpus project to maintain several layers of annotation in parallel (such as discourse markers as opposed to clause boundaries). Several speech annotation projects have made use of these advantages, such as Müürisep & Uibo (2006) for Estonian. In the Nordic Dialect Corpus (Bondi *et al*., 2009), CG output was used to train a DTT tagger (Schmid, 1994). In the European C-ORAL context, the Spanish section employed CG-inspired rules for part-of-speech disambiguation of morphological output from the GRAMPAL system (Moreno, 2003), and for the Palavras parser itself, Bick (1998) reports early experiments with a Constraint-Grammar-only solution in connection with the morphosyntactic annotation of the Brazilian NURC

corpus (Castilho, 1993).

Like other CG systems, PALAVRAS depends on a morphological analyzer to identify possible word form readings and uses thousands of context-sensitive rules to disambiguate ambiguous readings (so-called cohorts of reading lines). Higher-level information, such as syntactic and semantic tags, are iteratively mapped and disambiguated in consecutive modules.
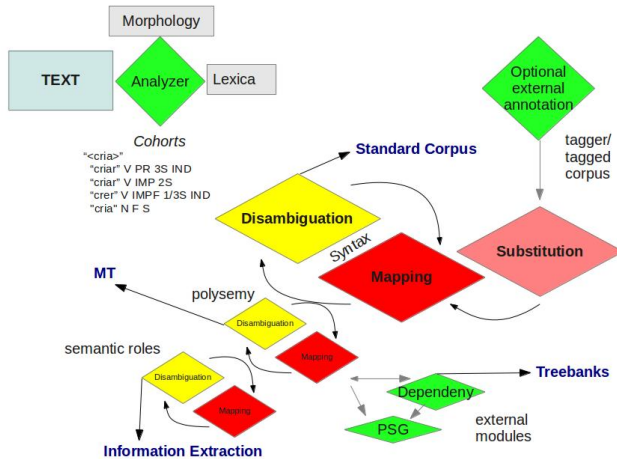
Figure 1: CG flow chart

The end result of this process is tokenized text with one token per line, followed by ordered tag fields:

Token   "Lemma"   <secondary tags>   POS MORPHOLOGY @SYNTAX §ROLE #n->m

where POS (part-of-speech) is followed by a class-dependent list of morphological features, such as number, gender and tense, and a syntactic function tag such as subject or object, and optionally a semantic role. Apart from these primary tag types, secondary tags may be added by lexicon lookup, providing contextual information for the parsing rules, e.g. valency class for verbs, or semantic prototype class for nouns. The #n->m field marks dependency relations from daughter (n) to mother (m), using running ID numbering.

3.   Project methodology

Given the rule-based and lexicon-dependent architecture of PALAVRAS, three challenges can be identified with regard to its application to oral data, affecting lexical recall on the one hand (3.2) and contextual disambiguation on the other (3.1 & 3.3). In many ways, the problems are similar to the ones encountered in the annotation of historical language data (Bick & Módolo 2005).

3.1   Text flow normalization

In order to maintain corpus meta information from other annotation layers, while still providing "running text" input to the PALAVRAS-analyzer, in-text markup for

turn-taking (e.g. LEO:), speaker overlap (e.g. <ô / mas>) and retractions (e.g. [/2]) was turned into <....> meta tags reminiscent of xml tags but without the projectivity restrictions of xml-trees (<LEO:>, <overlap-start>, <overlap-stop>, <retract:falando_em>)[1]. The annotation sample below exemplifies various types of meta tags, as well as lexical alterations (OALT) and general morphosyntactic mark-up[2]:

> ***LEO:*** *o Juninho <**foi**> //*
> ***GIL:*** *<ô / mas> / voltando à questão / **falando em [/2]** e também falando em povo mascarado / esse povo do Galáticos é muito palha / eu acho que es nũ deviam mais participar / e <**tal**> //*
>
> ***<LEO:>***
>   *o [o] <artd> DET M S @>N*
>   *Juninho [Juninho] <hum> <newlex> PROP M S @SUBJ>*
> ***<overlap-start>***
>   *foi [ser] <fmc> V PS 3S IND @FMV*
> ***<overlap-stop>***
>   *$;*
> ***<GIL:>***
> ***<overlap-start>***
>   *ô [ô] <newlex> IN @ADVL*
>   *$,*
>   *mas [mas] KC*
> ***<overlap-stop>***
>   *$,*
>   *voltando [voltar] V GER @IMV @#ICL-ADVL>*
>   *a [a] <sam-> PRP @<PIV*
>   *a [o] <-sam> <artd> DET F S @>N*
>   *questão [questão] <ac> N F S @P<*
>   *$,*
> ***<retract:falando_em>***
>   *e [e] KC*
>   *também [também] ADV @ADVL>*

---

1   It should be noted that non-inclusive bracketing overlaps of the type <a>   do occur in the corpus (crossings of overlap and retraction mark-up) and represent a general annotation problem, even for elaborate xml encoding schemes, since the latter do not envision non-projective (overlapping) tree structures, so the CG annotation chosen here can be said to be a fairly robust solution.

2   ***Tag abbreviations: POS:*** *V=verb, N=noun, PROP=name, ADJ=adjective, ADV=adverb, PERS=personal pronoun, DET=determiner, KS=subordinating conjunction, KC=coordinating conjunction, PRP=preposition, IN=interjection; **Morphology**: S=singular, P=plural, M=male, F=female, NOM=nominative, PR=present tense, IMPF=past tense, PS=preterite, IND=indicative, GER=gerund, INF=infinitive, PCP=participle, 1=1st person, 3=3rd person;* ***Syntax:*** *@SUBJ=subject, @ACC=direct object, @PIV=prepositional object, @SC=subject complement, @OC=object complement @ADVL=adverbial, @>N=prenominal, @N<=postnominal, @P<=argument of preposition,@FMV=finite main verb, @FS=finite subclause, @ICL=non-finite subclause; **Secondary tags:** <sam>=part of contraction, <artd>=definite article*

falando [falar] <vH> V GER @IMV @#ICL-<ADVL
em [em] PRP @<PIV
povo [povo] <HH> N M S @P<
mascarado [mascarar] <vH> V PCP M S @N<
$,
esse [esse] <dem> DET M S @>N
povo [povo] <HH> N M S @SUBJ>
de [de] <sam-> PRP @N<
o [o] <-sam> <artd> DET M S @>N
Galáticos [Galáticos] <org> <newlex> PROP M P @P<
é [ser] V PR 3S IND VFIN @FMV
muito [muito] <quant> ADV @<ADVL
palha [palha] <cm> N F S @<SC
$,
eu [eu] PERS M/F 1S NOM @SUBJ>
acho [achar] <vH> V PR 1S IND @FMV
que [que] KS @SUB @#FS-<ACC
*es **OALT** eles [eles] PERS M 3P NOM @SUBJ>*
*nũ **OALT** não [não] ADV @<ADVL*
deviam [dever] V IMPF 3P IND @FAUX
mais [mais] ADV @<ADVL
participar [participar] <vH> V INF @IMV @#ICL-AUX<
$,
e [e] KC
<overlap-start>
tal [tal] <diff> <KOMP> DET M/F S @<OC
<overlap-stop>
$;

The same procedure is used for so-called non-words, covering a few non-word surface strings without special markup ('hhh' and 'xxx'), as well as incomplete words (contractions), which are marked with an initial &-sign.

GIL: **hhh eu tenho **&dire***

<GIL:>
<nonword:hhh>
eu [eu] PERS M/F 1S NOM @SUBJ>
tenho [ter] <fmc> V PR 1S IND VFIN @FMV
<nonword:&dire>

Since PALAVRAS ignores <...> lines as corpus mark-up, it is left with what amounts to running, ordinary text, providing better syntactic matches for parsing rules.

3.2   Tokenization

Tokenization was also standardized, and largely performed as a preprocessing step. For instance, in order to match ordinary np and pp constraints, the parser needed to be fed two-word contractions as separate tokens. However, while all standard cases like *deles, naquele* etc. are already built-in, a number of frequent non-standard contractions *(*in order of frequency: *pa, pro, co, pros, prum, pos* etc*)* had to be treated separately. In some of these cases, readings ewere ambiguous, asking of CG-processing on top of preprocessing, as for *pra (para+a, para)*. 2% of utterances contained the Portuguese focus construction *é_que,* which was

transcribed as *que,* and therefore had to be disambuguated between a 2-token reading (focus particle) and a 1-token reading (conjunction or relative).

Post-tokenization (i.e. after morphological analysis) was used for the contractions that were less regular and/or more difficult to match with regular expressions. These cases were drawn from C-ORAL's normalisation lexicon, and their parts were word-form numbered and marked with OALT normalization tag (cp. chapter 3.3), in theory allowing any number of parts:

*pa despesa é bastante / **né** //*

pa OALT pra   [para] <sam-> PRP @ADVL>
a [a] <artd> <-sam> DET F S @>N
despesa [despesa] <mon> N F S @P<
é [ser] <vK> V PR 3S IND @FMV
bastante [bastante] <nh> ADJ M/F S @<SC
$,
<slash>
né OALT não   [não] ADV @ADVL>
né-2 OALT é   [ser] V PR 3S IND VFIN @FMV

3.3   Lexical and orthographic normalization

While maintaining the oral transcription forms as tokens, modified word forms were fed to the analyzer module where transcriptional orthography deviated from the written norm, and could not be recovered by the parser's own accentuation and affixation heuristics *(emitivi, ladim, estudemo)*. Thus, two new modules were added to the program chain, both with a manually maintained lexicon-file as input. The first program *(coral.inter)* handles specific or systematic standardizations and is run after preprocessing, before morphological analysis, while the second program *(postlex_pt)* is a regular morphological analyzer in its own right, with its own lexicon and inflexion rules, overriding PALAVRAS' heuristic analyses for unknown word forms, or adding additional readings to partially known forms, for contextual disambiguation. In both cases, both multi-word expressions and regular inflexional variation was covered on top of individual word forms. The two programs use lexica with 700 token normalizations and 2000 regular lexicon entries, respectively, both compiled by one of the C-ORAL authors (Heliana Mello) and then checked for consistency and compatibility to avoid unwanted interferences with PALAVRAS' existing core lexicon. .

An example for systematic normalization is the addition of first person plural -s for verbs *(comemoramo -> comemoramos, encontramo -> encontramos)*, which *coral.inter* accomplishes using string matches and a fullform lexicon that helps to avoid false s-additions to e.g. nouns like *balsamo, dinamo, esperramo.* l-r variation *(glandão - grandão)* was also covered but proved to be negligible in quantitative terms. Examples of lexicon-handled normalizations are abbreviations (a), word-initial a-drop and inflexional variation (c).

(a1) *emedebê -> MDB*
(b1) *inda -> ainda*
(b2) *roz -> arroz*
(c) *fazido -> feito*

While maintaining the original word form, standardized forms were added with an OALT:... prefix, and it is the standard form that annotation tags refer to:

meninim OALT menininho [menino] <DERS> N M S

The standardization lexicon also covers multi-word strings *(a'=aqui -> olha=aqui, c'=ocês -> com=vocês),* a fact that is also exploited at the tokenization/preprocessing stage. One advantage of multi-word normalization is that the individual parts provide disambiguation context for each other, allowing, for instance, the recognition of *a'* as *olha',* rather than the preposition or determiner reading, or the resolution of *n'* as *não* or *em* in *n'=era* and *n'=ocê,* respectively.

The second lexical add-on program is considerably more sophisticated than the normalization program, and allows both fullform and base form entries in its lexicon *(newlex_pt).* Regular inflexions of noun, adjective and verb forms will be recognized from the base form alone, but all irregular forms have to be entered separately. Like for the standardization lexicon, multi-word entries will also be visible to the preprocessor for tokenization (d1, b).

Due to the good general coverage of PALAVRAS, the lexicon contains few regular Portuguese nouns, but some inflected or complex noun forms (a2-3) proved useful to avoid the choice of a competing heuristic analysis, e.g. *caça-talentos* as plural- vs. singular-inflected. Also, the corpus contained a certain number of foreign words which are likely to be singular nouns, but may have endings that could trigger a heuristic (Portuguese) analysis as something else, e.g *remote* (c1). Even more important is it to list foreign non-noun words such as verbs (c3), adjectives (c4) or adverbs (c5), but these entries raise two problems that would have to be resolved if the lexicon were to be used in a more general setting (i.e. for other corpora): First, foreign words would need to be specified with *all* their readings, not only the one occurring in the corpus, e.g *shift* (c4) as both noun and verb. Second, also foreign entries would need full morphology, if they were to fully interact with their Portuguese context and CG-rules (e.g. agreement issues).

Two thirds of all entries were proper nouns (e1-3). Though these could be fairly safely recognized as such by PALAVRAS, their gender (and possibly number) is not easy to guess (e.g. *TIM* as feminine), and the addition of a semantic prototype reading (e.g. <hum>=human, <org>=organization, <Lciv>=town or state) provided valuable semantic context for CG rules, allowing, for instance, to unify the ±HUM feature on verbs and their subjects, allowing semantics-based disambiguation of word-class or syntactic function.

(a1) fazeção <activity> N F S
(a2) zenes N M P # termo de jogo
(a3) caça-talentos N M S
(a4) superbonitinha ADJ F S
(a5) superbem-arrumada ADJ F S
(b) mil-oitocentos-e=vovó=gostosa NUM M/F P
(c1) remote N M S # estrangeirismo
(c2) completed ADJ M/F S/P # estrangeirismo
(c3) save V # estrangeirismo
(c4) shift N M S # estrangeirismo
(c5) anche ADV # estrangeirismo
(d1) tu=tu X # onomatopéia
(d2) tuf X # onomatopéia
(e1) Titina <hum> PROP F S
(e2) TIM <org> PROP F S # operadora de telefonia
(e3) Timoftol <cm-rem> PROP M S
(f) agadê N M S # HD (harddisk)

3.4   Syntactic segmentation

A serious problem for the automatic analysis of transcribed speech is the lack of syntactic surface structure encoded as punctuation, which would normally be exploited to help segment clauses and phrases, and to provide the parser with syntactic windows for its rules, such as the uniqueness principle. In CG-terms, a comma is a member of the BARRIER set in many context rules, separating phrase-internal material from tokens belonging to another phrase. A breakdown of rule scopes in the Palavras grammar (Bick, 2000) shows that the share of so-called global rules (i.e. rules with context conditions spanning whole sentences) is substantial even for morphology (around 31%), and is very high for syntax, where most rules use unbounded contexts (> 80%). Without comma barriers and full stops such rules will act differently and produce more errors.

However, speech corpora usually provide other, prosodic means of segmentation. In some speech corpora, such as the NURC corpus version described in (Bick 1998), prosody is implicitly encoded by orthographic means such as vowel length *('u::m'),* stress *('esnoBAR')* and pauses *('eee').* This may further complicate normalization and also asks for the contextual disambiguation of pauses versus true syntactic breaks. In the C-ORAL corpus, on the other hand, prosodic segmentation was marked explicitly, at transcription time, using three different segmentation strengths:

1. major prosodic breaks (//), separating what functionally could be called utterances, equivalent to written language sentence separation;
2. discontinuation breaks (+) between utterances;
3. non-terminal prosodic breaks (/), separating what could be viewed as informational units.

Rather than making this information invisible to the parser by turning it into meta-tags (the strategy chosen for syntactic noise), we decided to replace the prosodic markers with standard punctuation, using a semicolon as the most obvious equivalent to the // terminal breaks (alternating with '...' for interruptions), and a comma for

the non-terminal breaks (/). Portuguese orthography does not use obligatory commas in all places where our transcription had a slash, but inspection of annotation results showed that the extra commas helped rather than hurt. Each comma candidate was assigned two potential readings, <break> and <pause>, and contextual CG rules were used to make the distinction and replace <pause> slashes with a meta tag rather than a comma, e.g.

(a) between a noun or a nominative pronoun to the left, and a finite verb to the right, a prosodic /-marker is treated as <pause> (subject - verb case)
(b) prosodic /-markers between a noun and another np are treated as <break> (appositions)

Of course, since this rule section had to be run *before* the parser's own rules (which it was supposed to help), linguistic context conditions had to be worded carefully and not too explicitly, taking into account the high morphological and PoS ambiguity of raw text input.

4.   Evaluation

We used the Constraint Grammar evaluation tool eval_cg to evaluate the modified parser on a randomly chosen transcription file (~ 2000 words), creating a gold-standard version by manual revision. In an ordinary CG setup, meta-markup and punctuation would align 100%, but in our case, matters were complicated by the pause/break disambiguation, where pause commas were removed in the gold file. On the one hand, this caused alignment problems for the evaluator, on the other hand, differences had to be identified and counted as recall errors. Other mismatches, caused by faulty splitting or non-splitting of ambiguous MWE's, were also counted as recall errors, e.g in the case of *"primeiro=que"* (conjunction vs. adjective/numeral + relative).

Overall, our system achieved correctness rates (F-scores) of 98.6% for part of speech, 95% for syntactic function and 99% for lemmatization:

	Recall	Precision	F-Score
Syntactic function	95.3	94.9	95
PoS (word class)	98.5	98.7	98.6
Morphology	98.4	98.6	98.5
Base form	98.6	99.4	99

Table 1: Performance

In order to judge the effectiveness of using prosodic break markers as punctuation, we also compared the standard run (with pause/break disambiguation) with a no-break run (/-marks ignored), a no-sentence run (both /, + and // ignored), and an all-break run (all /-marks turned into commas, *without* disambiguation). Since the gold file did have disambiguated commas, the evaluator was run in match-only mode, comparing tags only for matching tokens. Therefore, figures in the table below can only be compared with each other, and not with the original test run.

	no-sentence	no-break	all-break	pause / break
Syntactic function	86.2 (R: 86.5, P: 86.1)	90.7 (R: 91.0, P: 90.6)	93.7 (R: 93.3, P: 93.6)	**95.0** (R:95.3, P: 94.8)
PoS (Word class)	98,3	98,8	99,3	**99,4**
Morphology	98,1	98,6	**99**	98,7
Base form	99	99,1	**99,4**	**99,4**

Table 2: Influence of prosodic break markers

Clearly, exploiting prosodic break markers did improve performance at all levels. However, the effect was much more marked for syntax than for part of speech, lemmatization and morphology, reflecting the wider contextual scope of syntactic tags and the ensuing greater need for precise and correct segmentation. Interestingly, while syntactic performance can be further increased by pause/break disambiguation, this is not obvious for the more local tag categories. Thus, for inflexion tags (morphology), all-break performance was *higher* than for the pause/break run, and only for PoS a slight improvement was observed.

5.   References

Bick, E. (2000). *The Parsing System Palavras - Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*, Aarhus: Aarhus University Press

Bick, E. (1998). Tagging Speech Data - Constraint Grammar Analysis of Spoken Portuguese, in: *Proceedings of the 17th Scandinavian Conference of Linguistics*, Odense University

Bick, E. & Módolo, M. (2005). Letters and Editorials: A grammatically annotated corpus of 19th century Brazilian Portuguese. In C. Pusch, J. Kabatek and W. Raible (Eds.), *Romance Corpus Linguistics II: Corpora and Historical Linguistics (Proceedings of the 2nd Freiburg Workshop on Romance Corpus Linguistics, Sept. 2003)*. Tübingen: Gunther Narr Verlag, pp. 271--280.

Brill, E. (1992). A simple rule-based part of speech tagger. In *Proceedings of the workshop on Speech and Natural Language*. HLT '91, Morristown, NJ, USA: Association for Computational Linguistics, pp.112--116

Castilho, Ataliba de (Ed.) (1993). *Gramática do Português Falado*, vol.3, Campinas: Editora da Unicamp.

Johannessen, J. B., Priestley, J., Hagen, K., Åfarli, T. A.,. Vangsnes, Ø. A. (2009). The Nordic Dialect Corpus - an Advanced Research Tool. In K. Jokinen, E. Bick (Eds.): *Proceedings of the 17th Nordic Conference of Computational Linguistics* (NODALIDA 2009). NEALT Proceedings Series Volume 4

Karlsson, F., Voutilainen, A., Heikkilä, J., Anttila, A. (1995). *Constraint Grammar, A Language-Independent System for Parsing Unrestricted Text*. Berlin: Mouton

de Gruyter.

Maamouri, M. *et al*. (2010). From Speech to Trees: Applying Treebank Annotation to Arabic Broadcast News. In *Proceedings of LREC 2010*, Valletta, Malta, May 2010.

Moreno, A., Guirão, J.M. (2003). "Tagging a spontaneous speech corpus of Spanish". In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*. Borovets, Bulgaria, 2003. pp. 292--296.

Müürisep, K., Uibo, H. (2006). Shallow Parsing of Spoken Estonian Using Constraint Grammar. In P. J. Henriksen, P. R. Skadhauge (Eds.), *Proceedings of NODALIDA-2005 - special session on treebanking*. Copenhagen Studies in Language #33/2006

Panunzi, A., Picchi, E., Moneglia, M. (2004). Using PiTagger for Lemmatization and PoS Tagging of a Spontaneous Speech Corpus: C-Oral-Rom Italian. In M.T. Lino *et al*. (Eds.), *Proceedings of the 4th LREC Conference*, vol. 2, ELRA, Paris, pp. 563--566.

Raso, T., Mello, H. (2010). The C-ORAL BRASIL corpus. In M. Moneglia, A. Panunzi (Eds), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Universitá degli studi di Firenze, Biblioteca Digitale.

Raso, T., Mello, H. (2012). *C-ORAL-BRASIL I: corpus de referência da fala informal brasileira*. Belo Horizonte. Editora UFMG.

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing 1994*. pp. 44--49.

Panorama of experimental prosody research

Plínio A. BARBOSA

Instituto de Estudos da Linguagem, State University of Campinas, Brazil
Rua Sérgio Buarque de Holanda, 571 Zip code: 13083-859 - Campinas-SP, Brazil
E-mail: pabarbosa.unicampbr@gmail.com

Abstract

Starting by a definition, this paper presents a panorama of experimental prosody research. After briefly exposing the three main properties of an experimental work, testability, predictability and designability, as well as the selection of variables in experimental prosody research, the crucial concepts and methodological procedures involved in rhythm and intonation research are portrayed. These aspects are presented in functional terms, that is, they are explored as a means to reveal the functions of prosody in verbal communication. The concepts of prominence, acoustic salience, pitch accent, prosodic boundary, stress group, phrase stress, speech rate, phonetic syllable, pausing, tonal alignment and expressive speech are presented and illustrated with some examples from work on Brazilian Portuguese and Standard German. The procedures depicted in this work were the duration normalisation technique for rhythm research and the pitch accent and boundary tone annotation in intonation research. Some questions such as the terminological difference between "intonation" and "prosody", the stress- vs syllable-timing distinction, the difference between perceived and produced prosody are briefly presented and discussed.

Keywords: prosody; experimentation; rhythm; intonation; expressive speech.

1. Introduction

Experimental prosody can be defined as the area of research which applies the hypothetic-deductive method to prosodic studies via experimentation. This definition implies that experimentation in prosody research should preferably be developed in three steps of increasing complexity: observation, description and experimentation *stricto sensu.*

The observation of a prosodic fact is never naive, because formal instruction is necessary to see or to select what is relevant in terms of a variable under scrutiny (for a general reading about observation in science see Fleck, 1992, Beveridge, 1957: 102-105 and Bunge, 1998: 181-189). As an illustration, fundamental frequency (F_0) peaks can be of different heights but only some are relevant from the perceptual of from the linguistic points of view. Thus, a simple question such as "what is a linguistically meaningful F_0 peak?", needs a formal instruction to be appropriately answered.

Descriptive prosodic research is an important step of scientific discovery. It uses the formal devices of descriptive statistics or correlational methods to give measures of centrality, variation, amplitude and skewness in the former case or the correlation between two or more variables in the latter case. Several other measures can be used; we presented here the most common ones. The statistical descriptors reduce the degrees of freedom of the variables and give a first picture of the phenomena under scrutiny.

Experimentation is related to reproducibility, which is a key scientific component. That is why this step is so closely related to inferential statistics: "One of the first things which the beginner must grasp is that statistics need to be taken into account when the experiment is being planned, or else the results may not be worth treating statistically." (Beveridge, 1957: 19). Under certain conditions of control, a snapshot of a communicative instance (the corpus) is examined and the variables of interest are measured to infer, given the variation of the data, the behaviour of a population from which the data were obtained. Experimentation starts with a theory, which guides the observation of prosodic facts. The theory and the observed facts produce a set of hypotheses aiming at testing a model of prosody production or perception. To test this model, a set of hypothesis-derived measures extracted from the corpus are evaluated according to their validity as regards the hypotheses raised at the beginning of the experimental study. This last step allows the refinement or revision of the theory that motivated the study.

In section 2 some considerations and initial steps for carrying out an experimental work are given. In section 3, we present the functions of prosody. In sections 4 and 5 we respectively present the main conceptual and methodological in rhythm and intonation research. Section 6 gives some directions and key concepts of expressive speech research. The aim of this paper is not to present a review of prosodic research, but to give a panorama of experimental prosody research to stimulate the new comer to choose an area of research to work with.

2. Getting started in experimentation

2.1 Properties of an experimental work

In order to be scientific valid, a theory in experimental prosody research needs to satisfy three main properties: testability, predictability, and designability (for a similar view, see Xu, 2011).

Testability refers to the hypotheses raised by the experimenter. They should be well-formed, meaningful, and contain mechanisms to check whether they are true or false (Bunge, 1998: 309-315). The truth-conditions of an original hypothesis can be refined after experimentation, but the reformulated hypothesis should also be directly testable. Suppose that an experimenter posits the

hypothesis that stressed syllables are longer than unstressed syllables in Brazilian Portuguese (henceforth BP) based on previous experimental findings that suggested that syllable duration is the main correlate of stress in BP (Martini, 1991; Barbosa, 1996). This hypothesis is testable because we can design a corpus for comparing the duration of stressed syllables with that of unstressed syllables in similar conditions of production and then apply a two-sample statistical test to compute the probability of making a type-I error when rejecting the null hypothesis that both durations are the same. This does not mean that this kind of check is easy, considering the several elements that affect duration needing to be controlled. For instance, at the end of an utterance, an unstressed syllable with an identical phonemic composition of a stressed syllable (e.g., the second syllable of *papa* – pope) is longer than the latter because of the final lengthening phenomenon (Scott, 1980; Edwards *et al.*, 1991). This simple exception to the general finding would entail the refinement of the original hypothesis to: "non-pre-pausal unstressed syllables are shorter than stressed syllables". Additional exceptions can be discovered from subsequent experimental settings.

Predictability refers to the ability of predicting new outcomes under distinct experimental conditions. In order to do so, how to predict the values of the new outcomes must be explicit. This explicitness is associated with a model which can be conceived of as a set of rules or a set of equations. For instance, intonation models and rhythm models can generate F_0 and duration values for a particular utterance, which can be compared with the observed utterance for a certain number of speakers to assess the closeness between predicted and observed values, and, given the nature and/or extension of the errors obtained, evaluate the need for model refinement (some examples of either rhythm or intonation models can be found in Xu, 2011; Botinis *et al.*, 2001; Barbosa, 2006).

Designability refers to the possibility of conceiving an experimental protocol to test the hypotheses raised. The design of an experiment in prosody research is not easy. It includes the selection of variables for investigation, the choice of the statistical test to assess the hypotheses, the choice of the informants to record the corpus or of the subjects to listen to the set of stimuli of a perception test. The example of the stressed vs unstressed syllable mentioned before presents a high degree of designability. But it is not always like that. Suppose that a theoretical account of the relation between neuronal activity and speech perception states that a particular pathway is more activated when a subject listens to a C-to-V transition. Two ideal experimental designs could be: (1) to put electrodes directly in the areas along the pathway and to measure neuronal activity or (2) to make a lesion in some area in the pathway and study its consequences. It is unnecessary to explain the ethical problems involved in both designs. Researchers can cope with them by studying the aforementioned relationship in non-human mammals or by studying the consequences of

naturally-occurring lesions in human patients (see some studies reported by Scott & Wise, 2003).

2.2   The selection of the variables for study

There are three classes of variables in an experimental setting: independent, dependent and to-be-controlled. Independent variables are those manipulated by the experimenter and directly related to the hypotheses raised. It is important to know that they are not necessarily nominal or discrete.

Dependent variables are those which are measured and which are usually acoustic or articulatory correlates of the discrete or intervalar prosodic, independent variables.

To-be-controlled or nuisible variables are those that need to be controlled because their unpredicted (or unpredictable) influence can affect the dependent variables if we do not take enough care.

Let's examine three examples of these variables in prosodic research. First, the experiment about the duration of stressed vs unstressed syllables in BP presented above. In this case, the independent variable is STRESS, with two levels, "stressed" and "unstressed". The dependent variables are the acoustic duration of the syllables. The to-be-controlled variables are: phonetic context of the syllable, degree of prominence and boundary adjacent to the measured syllables, speech rate, and healthy state of the subject, among others. We cannot compare stressed vs unstressed syllables in words where the to-be-controlled variables differ because the non-chosen differences in these variables can also affect duration. In this case, it is not possible to infer the cause of the duration change. For instance, a stressed syllable in a word just after a previous focussed word can exhibit lesser duration than an unstressed syllable with a similar phonetic context in a word not in post-focal position. The ideal statistical test for comparing the mean duration of the syllables across the two levels of the STRESS variable is a t-test of independent variables, provided that the residue is normally distributed (otherwise the equivalent non-parametric test is Mann-Whitney. See Crawley, 2005 for a nice introduction and use of statistical tests).

As a second example, suppose you want to determine how many distinct boundary levels can be signalled by a relevant acoustic-prosodic parameter in BP. The independent variable is the height of the constituent immediately preceding the boundary in a hierarchy of linguistic domains. The dependent variable can be the duration of the syllable rhyme preceding the boundary (cf Barbosa, 2006). The to-be-controlled variables are all extraneous variables that affect the duration of the pre-boundary words but the boundary height in the hierarchy. The appropriate statistical test is clusterisation, which groups together, under certain conditions, the durations associated to the same statistical distributions. At the end of the process, the number of distinct boundary levels is the number of distinct statistical groups (see Whitman *et al.*, 1992 for research of boundary levels in American English, Barbosa, 1994 for Standard French

and Barbosa, 2006 for BP).

The final example concerns expressive speech. Suppose you want to predict the degree of arousal evaluated by a group of listeners from the acoustic-prosodic properties of an utterance. In this case, the independent variable is the set of acoustic-prosodic parameter values for the utterance. The dependent or predicted variables are the listeners' evaluation degrees, and the appropriate statistical test is multiple regression. The to-be-controlled variables are all the influencing factors that could explain the listeners' evaluation which are not based on what they hear from the acoustic information embedded in the speech signal, such as the lexicon, the habit of a listener in giving high grades, the health state of the listener that day, among others.

3. Functions of prosody

In terms of linguistic and paralinguistic uses, the following functions of prosody can be identified: (1) a discursive function such as to signal a turn in a dialogue, to signal that you are listening your interlocutor (backchannels such as "um-hum", "entendo" – (I) understand), to signal the modality of a sentence in a monologue, (2) a demarcative function aiming at signalling the edges of a prosodic constituent such as a phonological word or a stress group, (3) a prominence function aiming at signalling to the listeners the salience of a prosodic unit in relation to another one or in relation to the background units (see Barbosa, submitted, for examples of these functions and an introduction to prosodic research).

In terms of expressiveness, the following functions can be distinguished: attitudinal (attitude, personal stance) and affective (emotions such as sadness, joy and rage as well as other affects such as humour and traits of personality). Prosody can also convey indexical features such as gender, sex, dialectal and social origin, among others. Expressive and indexical features are found in every single utterance produced by a human subject because it's very hard to disguise aspects such as attitude, emotion and sex.

For an introduction to expressive speech research see the works by Fónagy (1986), Bolinger (1986) and Scherer (1984).

4. Conceptual and methodological aspects in rhythm research

More than a hundred definitions of rhythm can be given. Several of those proposed by Sauvanet (2000) highlight, in my sense, the two main components of rhythm, structuring and repetition: "Il y a rythme lorsqu'une structure évolue de manière périodique sur fond d'altération novatrice." (Wunenburger, 1992: 17) and "The essence of rhythm is the fusion of sameness and novelty; so that the whole never loses the essential unity of the pattern, while the parts exhibit the contrast arising from the novelty of their detail." (Whitehead, 1919: 198). Thus, there is periodicity and structuring in speech rhythm. Periodicity (sameness in Whitehead's terms) serves the production system because

it makes the control of the units produced easier. But an utterance with identical units would never signal anything to the listener. Then, it's necessary to build a structure to differ (novelty in Whitehead and Wunenburger) from the background. But what is repeated and what is modified to signal novelty? Essentially, syllables.

In BP, when a word is produced with acoustic salience, the acoustic parameters around the lexically stressed syllable are modified in relation to the background formed by the non-salient syllables. These acoustic parameters are F_0, duration, intensity, formant values, among others. In BP, salient syllables are often longer than non-salient ones. At strong syntactic boundaries or to signal a focussed item, these syllables are also higher in pitch (Barbosa, 2008). If the acoustically salient syllable is audible we say that the syllable and the word containing it are prominent, because these units catch the attention of the listener. Rhythm is the sensation caused by the succession of different degrees of syllabic prominence alternated with non-prominent syllables in the background (Barbosa, 1994).

Nowadays, rhythm research deals with the study of patterns of syllable-size duration along the utterances. In order to do so, it's necessary to separate segmental from prosodic information of syllable-sized durations. This is done by a technique of normalisation.

Duration normalisation allows to highlight with an accuracy of up to 80 % (Barbosa, 2010), the phonological words perceived as prominent or pre-boundary by the listeners. This is done by detecting normalised syllable-sized durations peaks in three steps. In the first step, the z-score of the phonetic syllable duration is computed. The phonetic syllable starts at the vowel onset of the realised phonological syllable and ends at the vowel onset of the next realised phonological syllable and is known in the literature as V-to-V unit (Barbosa, 2006). It has been used in rhythm research since a long time (cf Lehiste, 1970; Classe, 1939). By definition, the z-score, a common statistical measure, expresses the distance from the mean in units of standard-deviation. Then, if a z-score is -1.3 (it has no physical unit), this means that the duration is 1.3 standard-deviations distant from the mean, leftwards. Mean and standard-deviation can be obtained from a corpus containing all phones of a language, and it does not need to be from the same speaker, although it is recommended that the subject be from the same dialectal area (cf. Barbosa, 2006: 489 for values for these two descriptors in BP).

In the second step, a 5-point moving average technique is used to filter out additional sources of variation not related to perceived duration (for mathematical details see Barbosa, 2010 and Barbosa, 2006). The normalisation aims at minimising the effects of intrinsic duration and those of the number of segments of the V-to-V units.

The result of these two steps can be seen in Figure 1 for the utterance "Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara àquela maneira de viver.", uttered by a female speaker from São Paulo State. In the figure, five duration peaks around the

respective stressed syllables of five words can be seen: "entrado", "mosteiro", "ano", "adaptara", "viver". The three higher peaks are indicated, perceived by all listeners as pre-boundary or prominent. The peaks within the two other words are perceived as weakly prominent words. The peaks of normalised duration can be automatically detected by tracking the points where the derivative of the contour changes from positive to negative, which is the third step.
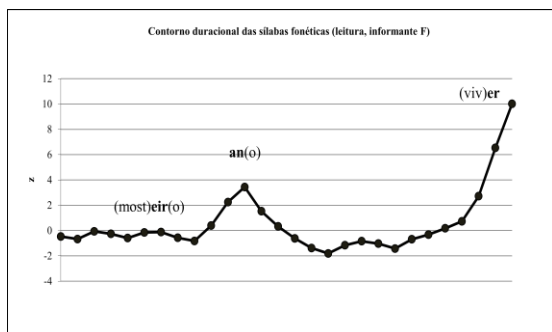
Figure 1: V-to-V normalised duration of the utterance "Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara àquela maneira de viver." by a female speaker

Each normalised V-to-V duration peak indicates an acoustic salience that, if perceived as a prominence, represents the position of a phrase stress. Because in BP the probability of a pre-boundary word be perceived as prominent is between 40 and 65 % (Barbosa, 2008) and because boundaries define the end of a domain, the association of normalised duration peaks to stress group boundaries is a convenient and appropriate decision for the need of automatic stress group delimitation. Despite the signalling of both prominence and prosodic boundary by longer durations, it is possible to distinguish the two functions when looking at the consequences of their implementation for the segments that make up the syllables. This difference can be found at least if the speaker highlights a word for signalling emphasis. In emphatic words, all segments of the lexically stressed syllable are lengthened, whereas, for words before a prosodic boundary, the stressed phonetic syllable is lengthened, that is, the vowel and the consonants following it (tautosyllabic or heterosyllabic). As an example in BP, let's choose the two sentences "Pedro vai casar, sabia?" and "Pedro vai CASAR, sabia?" The segments /a/ and /R/ are much more lengthened than /z/ in the word "casar" in the first sentence, whereas the segments /z/, /a/ and /R/ of the entire stressed syllable of the emphatic word in the second sentence are equally lengthened. This prosodic fact was experimentally demonstrated by Barbosa (2006: 309-317), and is found in several languages (see Tabain, 2003 for French and Byrd and Saltzman, 1998 for American English).

Another striking result of the normalisation technique is that the height of the duration peaks closely follows the degree of strength of the prosodic boundaries or

prominences: the strongest boundary is after the word "viver", followed by that after the word "ano". Without the application of this technique, the raw duration peak position and height do not correspond to valid prosodic functions as can be seen in Figure 2, where there are 12 peaks of duration. No listener perceives 12 prominent or pre-boundary words in this utterance. The normalisation procedure is basic in rhythm research and should be followed to reveal prosodic duration.
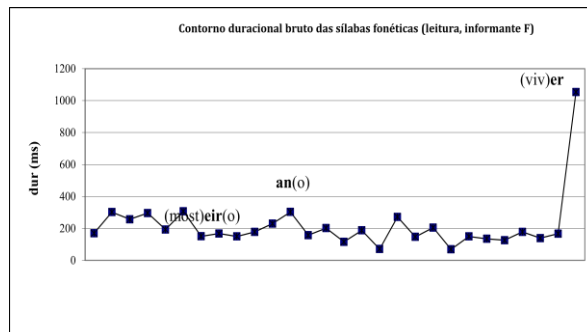
Figure 2: V-to-V raw duration contour of the same utterance of Figure 1

By analysing V-to-V normalised duration and other acoustic parameters such as vowel formant values, F_0 and spectral emphasis (Traunmüller & Eriksson, 2000), Arantes (2010) showed that duration and F_0 are entangled in the expression of secondary prominences in BP. Distinct from the prominence discussed so far, secondary prominences are realised in other positions than the stressed syllable. They signal the beginning of stress groups.

For applying the normalisation technique, the labelling of the phoneme-sized segments within each V-to-V interval is a necessary step that can be done manually or automatically (the *EasyAlign* tool developed by Goldman, 2011 delivers both phoneme-size boundaries and labels from an audio file. This tool was recently adapted to work on BP). As explained above, the normalised duration peaks can be used to define the right end of the stress groups in a right-headed language such as BP at this level. This allows both to count the number of phonetic syllables within the stress group as well as its duration automatically, which saves time and is useful for research on rhythm typology.

In fact, O'Dell and Nieminen (1999) showed that a tendency towards stress-timing (or syllable-timing) can be estimated from the ratio between the intersect and the slope of the linear regression line predicting stress group duration from the number of phonetic syllables within this group (see Barbosa *et al.*, 2009 for an application to evaluate the rhythmic differences between European and Brazilian Portuguese). Stress-timing concerns the alleged sensation that phrasally stressed syllables occur regularly in time, whereas syllable timing concerns the alleged sensation that syllables occur regularly in time. The literature on rhythm typology is very large, but some reviews on the theme can be found to get started (e.g.,

Barbosa, 2000, 2006; Bertinetto, 1989).

Speech rate is also a variable that needs to be taken into account in rhythm research. It can be defined either as the number of phonetic or as the number of phonological syllables per second. Both speech rate increase and decrease affect the syllable-sized durations throughout the utterances as shown by Barbosa (2006, 2007) for BP. That's why speech rate needs either to be controlled (in that case it is a to-be-controlled variable) for not influencing the results or it needs to be manipulated (in that case it is an independent variable) to study its effects on the corpus under study. Meireles and Barbosa (2008) have evaluated the possible contribution of speech rate increase for explaining the emergence of penultimate from antepenultimate lexical stress patterns in BP.

Pausing is another important component of the rhythmic structure of an utterance. A pause can be realised with a silent interval (silent pause) or with a lengthened V-to-V unit not followed by a silence (filled pause). Pause is a sensation of break caused by these two acoustic possibilities, among others. Pause can also signal a hesitation, when it is called a hesitative pause. Merlo (2012) has recently demonstrated that hesitation and hesitative pauses help maintain fluency during narrative and descriptive instances. Pause can also be a signal of a difficult in production, as in the case of dysarthria. Besides have shown that the longitudinal study of pausing reveals the benefit of therapy in dysarthric speech, the work by Vieira (2007) also revealed another striking aspect of pausing in pathological speech. Even though the number of silent pauses in dysarthric speech is higher than the number of silent pauses in the control group, the hierarchy of these pauses, revealed by the statistical distinction among their duration, signals the underlying linguistic structure also highlighted by the control group.

Production and perception mechanisms of rhythmic structure were recently studied in an integrative way by Barbosa & Silva (2012). They demonstrated that the rate and height of V-to-V normalised duration peaks, associated to speech rate explain up to 71 % of the variance of listeners' judgments about differences in manner of speaking of three BP subjects.

To sum up, in this section the roles of the prosodic functions of prominence and boundary to rhythm research were presented. To help revealing them in the production domain, the phonetic syllable was defined. The V-to-V normalised duration values throughout the utterance define the rhythmic structure associated to this utterance. This structure is characterised by a sequence of duration peaks of differing degrees which contributes to the perception of different degrees of prominence, secondary prominences and boundary strength. Pausing is an integral part of this rhythmic structure that can also be revealed by the same procedure. The alleged regular succession of phonetic syllables and phrasally stressed phonetic syllables was implicated in the definition of syllable- and stress-timing in rhythm research. Differences in the rate and degree of boundary and prominence of these variables explain differences in

perceived rhythm.

5.   Conceptual and methodological aspects in intonation research

The word "Manuel" in the example given in Figure 1 is perceived as prominent by the listeners even though there is no duration peak in the word. In fact, a rising F_0 contour within the word signals to the listeners the importance of this piece of information. The F_0 contour for the utterance can be seen in Figure 3, where the rising contour is represented by the symbol LH.
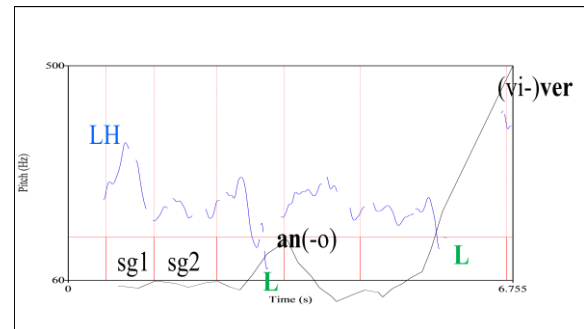
Figure 3: F_0 contour superposed to the V-to-V normalised duration contour of the same utterance of Figure 1."sg1" and "sg2" signal the first two stress groups. The first one ends at the syllable "tra" in the word "entrado". The second one ends at the syllable "tei" in the word "mosteiro"

This clearly tells that prosody perception in BP depends on at least two acoustic parameters: syllable duration and F_0 movement. In fact, it depends on all acoustic parameters that signal prosodic information, including intensity, voice quality and even vowel quality (e.g., the lower F1 value of the last /a/ of "papa", pope, also signals the penultimate stress pattern). It is the work of the experimenter to determine which parameters contribute more to perceived stress.

F_0 patterns also signal the prosodic functions of prominence and boundary. At strong syntactic boundaries it is common that both F_0 and duration signal the corresponding prosodic boundary (Barbosa, 2008). This can be seen in Figure 3, where the two main peaks of normalised duration, in "ano" and "viver", are accompanied by low levels of F_0, indicated with the L symbol.

Maybe because of the relevance of F_0 movements in signalling prominence and boundary in well-studied languages such as English, the term "intonation" is closely related to the term "prosody". That is why, before continuing it's necessary to say some words in this respect.

Hirst and Di Cristo (1998: 1-44) consider "prosody" as the general term including the lexical and post-lexical domains. For them, intonation is the study of the abstract relations in the post-lexical domain, independently of the acoustic parameter that signal these relations. In this sense

intonation embraces the study of pitch accent and boundary tone patterning, as well as the study of duration patterns throughout the utterances.

Another possible approach stems from studies on prosody perception and relies on the effects associated with the sensation of pitch, duration and loudness. For this approach, "prosody" is also the general term embracing the lexical and post-lexical domains, but "intonation", on the other hand, is restricted to the analysis of pitch variation throughout the utterances. Because the physical parameter that primarily controls the pitch sensation is F_0, the phonetic studies of intonation in this approach analyse the F_0 patterns throughout the utterances. It is this sense of intonation that we are using here. In this approach, "rhythm" is independent of "intonation" because it relies on the study of perceived syllable duration through the analysis of its main correlate, observed duration, as already depicted in the preceding section. Let's present some key concepts in intonation research.

Pitch accent is the intonation-related term for a prominence signalled by a F_0 movement, whereas the sensation of break is signalled by a boundary tone. Thus, pitch is not a synonym of F_0 peak or valley: it is a sensation that only can be evaluated by perception tests with real subjects. It cannot be measured in an objective way. Figure 3 illustrates an F_0 movement perceived as a pitch accent in BP. The movement has a rising shape (LH) and is followed by two low boundary tones (L). These two low tones in BP fulfil the function of signalling terminality, as we will see later in this section. The rising movement is defined with relation to alignment of the rising part of the contour with the stressed syllable, as can be seen in Figure 4, where the LH contour rising is entirely realised within the stressed syllable "lhões". Annotation of intonation-related prosodic functions is an important step to the study of intonation.
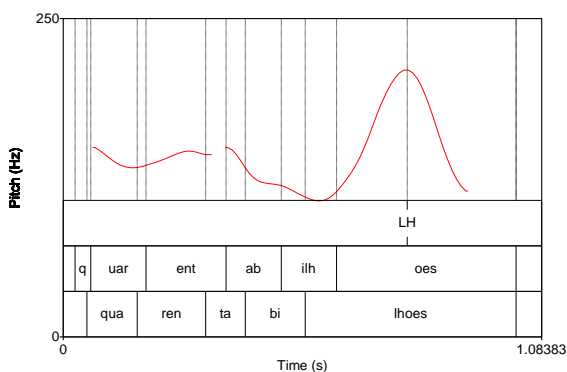
Figure 4: Illustration of the rising contour LH on the word "bilhões" from Lucente (2008)

Annotation systems such as ToBI, although largely adopted by researchers of American English (Silverman *et al*., 1992), German (Reyelt *et al*., 1996), and Spanish (Beckman *et al*., 2002), did not prove consistent across labellers (Wightman, 2002). By asking them to annotate pitch accent type by ear, the ToBI annotation procedure mixed up the roles of form and function in shaping

intonation patterning (Hirst, 2005). To avoid this, the best solution is to only ask the listeners to indicate whether a word is prominent or not, and whether a word precedes a prosodic break or not. After this phase, labels are assigned by examining the movement of the F_0 with relation to the stressed vowel (or stressed syllable). This was done by Lucente (2008) for studying focus in BP with the proposal of the DaTO system of intonation annotation. Recently, she extended the analysis to examining the relation between pitch accents and information status (Lucente, 2012). Examples of contour labels from the DaTO system can be seen in Figs. 4 to 9.
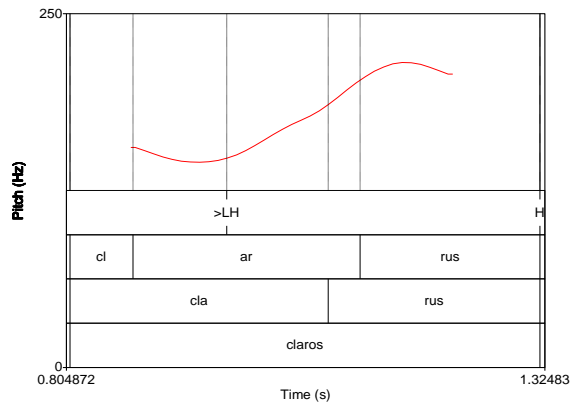
Figure 5: Illustration of the late rising contour >LH on the word "claros" from Lucente (2008)
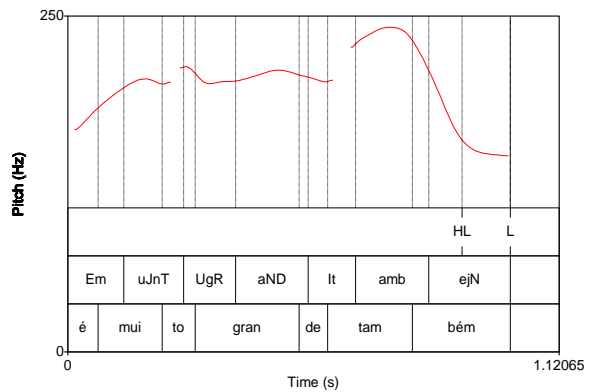
Figure 6: Illustration of the falling contour HL on the word "também" from Lucente (2008)

Figs. 4 and 5 illustrate the contrast between the rising and late rising contours. The F_0 peak occurs after the lexically stressed vowel in the latter case, whereas it occurs during the lexically stressed syllable in the former case. This contrast is similar to the one between the falling and late falling contours shown in Figs. 6 and 7. Observe in Figure 6 that in the HL contour, the low level of F_0 is attained during the lexically stressed syllable by a sharp fall from a higher position. This sharp fall is delayed in the late falling contour exhibited in Figure 7 where the lowest part of the F_0 contour levels out during the post-stressed syllable of the word "caras".

These differences are known as differences in tonal

alignment. Recent work on intonation has shown that tonal alignment with respect to the syllable is a crucial component of the intonation system of a language (see Xu, 2005).
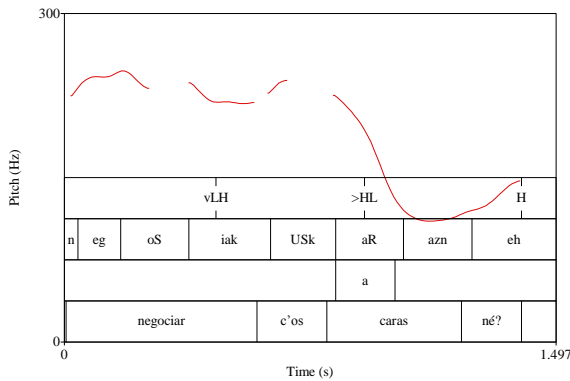
Figure 7: Illustration of the late falling contour >HL on the word "cara" from Lucente (2008)

The contours illustrated here are used by the speaker to signal prominence of the words onto which they are realised. Boundary tones are used to signal prosodic boundaries. Figure 8 shows the realisation of a low boundary tone (L) in spontaneous speech, also illustrated in Figure 3 in read speech. Low tones signal terminality in several Indo-European languages, although research about dialectal variability has shown that this picture is far from being simple (see Grabe, 2004 for prosodic variation in British English, where, in Newcastle English, almost 17 % of the declaratives are realised by a final high tone).
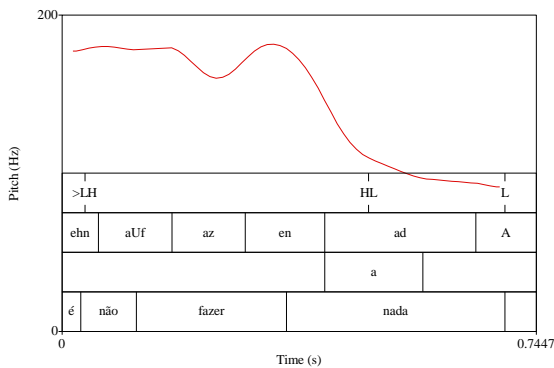
Figure 8: Illustration of the low tone contour L at the end of the word "nada" from Lucente (2008)

Figure 9 signals a high contour tone (H) in standard German storytelling. This high tone at the end of the utterance signals the listeners that there is more to come: that is, high tones in German signal non-terminality. In the same speaking style, non-terminal boundaries signalling the continuation of a story are realised by a rising-falling contour in BP, as shown in Figure 10 in two positions during the narration. Furthermore, Barbosa *et al.* (2011) showed that, in contrast with Standard German, during storytelling, BP –speaking subjects often maintain a high F_0 level between prominent words, as can be seen in

Figure 10 within the dotted ellipsis. In this excerpt, the speaker repeats part of the information she just gave, that the monk did not accustom with the routinely activities of the monastery. The stressed syllable of the word "acostumava" is extremely lengthened.
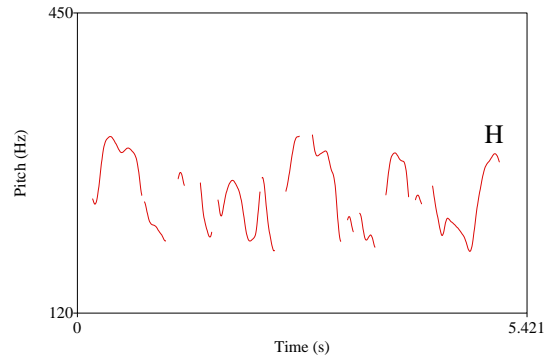
Figure 9: Illustration of the high tone H contour at the end of the word "gewöhnt" in a female speaker of read Standard German from Barbosa *et al.* (2011)
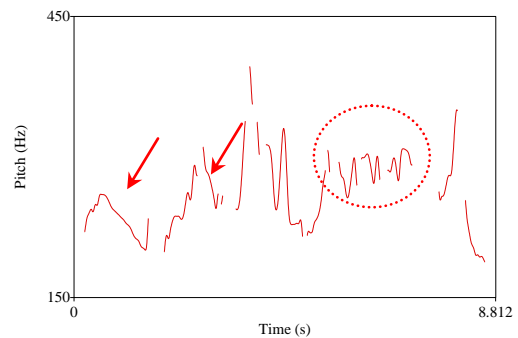
Figure 10: Continuative contours in the words "ele" and "que" (first two arrows from left to right) and high F_0 register (dotted elipsis) in the passage "ele não se acostumava com a rotina do [...]" during the narration of a BP female speaker

There is a close similarity between F_0 shapes for signaling yes-no questions and continuation of dialogue turns in BP. Both are signalled by rising falling contours whose difference relies on the alignment of the rising part of the contour. Figure 11 shows the rising-falling shapes in the same word "seguida" from the expression "em seguida" (in the following) realised by a male speaker from the State of São Paulo. It can be seen that the continuative contour rightwards is relatively low during the lexically stressed syllable with almost the entire rising realised during the post-stressed syllable /da/. On the other hand, the rising of the yes-no question contour leftwards resides in the stressed syllable /gi/. The difference in degree between the F_0 peaks in the two contrasting contours is related to the degree of emphasis the speaker put in the continuative contour. He could have realised the yes-no question with more emphasis, if necessary for communicative reasons. The crucial acoustic component for distinguishing yes-no questions

from continuative turns in BP is the delay of the rising part of the contour in the second case.
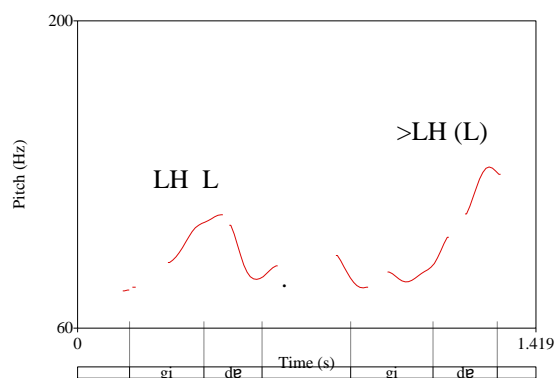
Figure 11: Contrast between yes-no question LH L (left) and continuative >LH L (right) contours in a male speaker of BP in the the word "seguida"

To sum up, intonation research needs an annotation system related to the classical prosodic functions of prominence, boundary marking and discourse event marking to be appropriately carried out. Recent research strongly suggests that annotation should relate F_0 contours to landmarks in the syllable. Defined functionally, pitch accents and intonational breaks can be adequately studied. Tonal alignment is a crucial element for distinguishing the contour types. Terminality and non-terminality are signalled by boundary tones which are different cross-linguistically. Intonational differences across languages can also be related to the way the F_0 curve between prominent intonational events is realised.

6.   Expressive speech research

The relevance of the vocal expression to signal affect was recognised at least as early as the XIXth century (Darwin, 1872 apud Scherer, 1986: 143). Scherer (1981) showed that naïve judges are more precise in assessing vocal than facial expression. The problem is to find out acoustic correlates for explaining this successful perceptual recognition. F_0 is certainly one of these parameters, at least as far as the study of high-arousal emotions are concerned (Scherer, 1986: 144; Frick, 1985: 418).

Emotion is only one of the possible affective states carried by the speech signal. Affect also includes mood, attitudes and interpersonal stances, preferences, and affective dispositions, as proposed by Scherer (1984). In comparison with the other affects, emotion is short in duration, it is more intense in terms of body responses, it triggers a simultaneous behaviour in other parts of the organism, and it is synchronous with the event that triggered the emotional behaviour. In everyday life, all affects are usually present in a single utterance. That is why the area of research dealing with affect in speech is called expressive speech research.

Several acoustic-prosodic parameters can be extracted from an utterance, which are relevant for expressive speech studies. The most used are F_0,

long-term average spectrum (LTAS), syllable-size duration and speech rate, as well as voice quality. Statistical descriptors such as mean, standard-deviation and skewness are used to evaluate the differences across different affects, such as the work on attitudes carried out by Moraes and colleagues in BP (Moraes, 2011; Moraes *et al.*, 2010; Rilliard *et al.*, 2012).

Another research approach in expressive speech studies is the evaluation of changes in expressiveness during sequences of utterances during conversations, as was done by Barbosa (2009) for BP. In this study, where circa 200 utterances extracted from a radio show were examined, an experiment designed to study the relation between perceived and produced expression was run out. The evaluation of the utterances was done by a set of judges and the prediction of the evaluation rates from a set of acoustic parameters. For predicting the rates, the set of utterances was split into two subsets, the training subset with 130 randomly chosen utterances from 12 subjects, and the test subset with the remaining 76 utterances. The training subset was evaluated by 12 judges, all of them undergraduate students of the first year in Linguistics. Four affect dimensions were evaluated by all judges in different days in two weeks. The dimensions were activation, involvement, valence, and dominance. The use of a dimensional approach in expressive speech research avoids the inter-subject variation in judgment if affective words are used due to idiosyncratic experience with each affect. Dimensional analysis has its limits: as it has been used, it could mask the dynamical aspects of affect change or rely only on the dimensions analysed to understand affect evaluation (see Scherer, 2000 for a criticism). These two drawbacks were avoided by using a Principal Component Analysis (PCA) to discover the main axes of variation in judgment when combining the dimensions chosen for analysis. All dimensions are evaluated within a 7-point differential semantic scale between two poles. Activation is a value between relaxed/calm and agitated/stimulated. Valence is a value between pleasant and unpleasant. Involvement is a value between involved and non-involved, and dominance is a value between under-control and submissive. This latter dimension was not reliably evaluated across judges and it was discarded.

Two factors in the PCA explained 97 % of the variance of the judgments, where factor 1 was related to arousal and explained 90 % of the judgments. To infer the judges' evaluation median rates for all utterances and dimensions, five classes of acoustic parameters were extracted: F_0, F_0first derivative (dF_0), intensity, spectral tilt (SpTt), and Long-Term Average Spectrum (LTAS). Up to four statistical descriptors were used for each class, producing twelve acoustic parameters: F_0 median, inter-quartile semi-amplitude, skewness, and 0.995 quantile; dF_0 mean, standard-deviation, and skewness; intensity skewness; spectral tilt mean, standard-deviation, and skewness; and LTAS standard-deviation. Spectral tilt is a correlate of vocal effort and was set to the difference of intensity in dB between the bands 0−1250 Hz and 1250−4000 Hz.

The spectral tilt descriptors and the dF_0 mean predict the new arousal dimension (factor 1) of the judgments' evaluations, with a correlation of 67 %. If these predicted-from-acoustics values are arranged chronologically in terms of the radio show participant, it is possible to detect changes in behaviour, as can be seen in Figure 12.
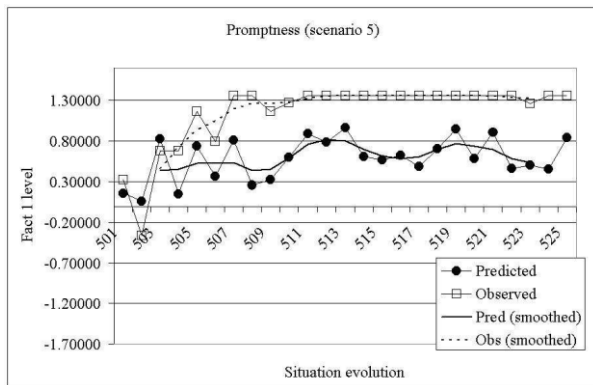
Figure 12: Predicted and observed values of arousal (promptness) in a scenario where the participant of a radio show is very irritated

From utterances 501 to 505 the participant talks to his daughter and from utterance 506 on, to the radio presenter. The observed contour shows, because evaluated by the judges, a saturation to a maximum level of arousal. This is not the case of the predicted-from-acoustics contour, which shows a trend to higher levels of arousal with some oscillations. The predicted levels are entirely based on acoustic parameters, contrary to the observed rates. These latter are also dependent on other influences, such as the semantic weight of the lexical items. In this situation, it is likely that the judges inferred the reasons for the participant's rage and decided to choose maximum levels of arousal, given the lexical items used by the participant. Nevertheless, the predicted values can be used to detect subtle changes in expression, such as the increasing of arousal from utterances 508 to 511.

The application to automatic detection of expressiveness is immediate.

7.   Acknowledgements

8.   References

Arantes, P. (2010). Integrando produção e percepção de proeminências Secundárias numa abordagem dinâmica do ritmo da fala. Doctoral Thesis. State University of Campinas.

Barbosa, P.A. (submitted). Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação.

Barbosa, P.A. (2009). Detecting changes in speech expressiveness in participants of a radio program. In *Proc. of Interspeech 2009 - Speech and Intelligence,* Brighton, UK. London: Causal Productions, pp. 2155 --2158.

Barbosa, P.A. (2008). Prominence- and Boundary-Related Acoustic Correlations in Brazilian Portuguese Read and Spontaneous Speech. In *Proc. Speech Prosody 2008*, Campinas, pp. 257--260.

Barbosa, P.A. (2007). From Syntax to Acoustic Duration: a Dynamical Model of Speech Rhythm Production. *Speech Communication*. 49, pp. 725--742.

Barbosa, P.A. (2006). *Incursões em torno do ritmo da Fala*. Campinas, Brazil: FAPESP/Pontes Editores.

Barbosa, P.A. (2000). "Syllable-timing in Brazilian Portuguese": uma crítica a Roy Major. In *D.E.L.T.A*, 16 (2), pp. 369--402.

Barbosa, P.A. (1996). At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration: emphasis on segmental duration generation. In *Cadernos de Estudos Lingüísticos*. 31, pp. 33--53.

Barbosa, P.A. (1994). Caractérisation et géenération automatique de la structuration rythmique du français. Thèse de Doctorat de l'INPG. Institut National Polytechnique de Grenoble, France.

Barbosa, P.A., Mixdorff, H. and Madureira, S. (2011). Applying the quantitative target approximation model (qTA) to German and Brazilian Portuguese. In *Proc. of Interspeech 2011*, Florence, Italy, pp. 2065--2068.

Barbosa, P.A., Viana, M.C., Trancoso, I. (2009). Cross-variety Rhythm Typology in Portuguese. In *Proc. of Interspeech 2009 - Speech and Intelligence.* Brighton, UK. London: Causal Productions, pp. 1011 --1014.

Barbosa, P.A., Silva, W. (2012). A New Methodology for Comparing Speech Rhythm Structure between Utterances: Beyond Typological Approaches. In H. Caseli *et al*. (Eds.), *PROPOR 2012, LNAI 7243*, Springer: Heidelberg, pp. 329--337.

Beckman, M. *et al*. (2002). Intonation across Spanish, in the Tones and Break Indices framework. *Probus*, 14, pp. 9--36.

Bertinetto, P. M. (1989). Reflections on the dichotomy "stress"- vs "syllable-timing". In *Revue de Phonétique Appliquée*, 91–92–93, pp. 99--130.

Beveridge, W.I.B. (1957). *The Art of Scientific Investigation.* New Jersey, USA: The Blackburn Press.

Bolinger, D. (1986). *Intonation and Its Parts: Melody in Spoken English*. Stanford: Stanford University Press.

Botinis, A., Granström, B. and Möbius, B. (2001). Developments and paradigms in intonation research. *Speech Communication*, 33, pp. 263--296.

Bunge, M. (1998). *Philosophy of Science. From Explanation to Justification.* New Brunswick, NJ: Transaction Publishers.

Byrd, D., Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*. 26, pp. 173--199.

Classe, A. (1939). *The Rhythm of English Prose*. Oxford: Blackwell.

Crawley, M.J. (2005). *Statistics. An Introduction using R.* Sussex, UK: John Wiley & Sons, Ltd.

Edwards, J., Beckman, M.E and Fletcher, J. (1991). The articulatory kinematics of final lengthening. In *J. Acoust. Soc. Am.* 89 (1), pp. 369--382.

Fleck, L. (1992 [1935]). Observation scientifique et perception en général. In J.-F. Braunstein (Ed.), *L'Histoire des sciences*. Paris: Librairie Philosophique J. Vrin, pp. 245--272.

Fónagy, I. (1986). Les langages de l'émotion, *Quaterni di semantica*. 7/2, pp. 305--318.

Frick, R.W. (1985). Communicating emotion: the role of prosodic features. *Psychological Bulletin.* 97 (3), 412--429.

Goldman, J.-P. (2011) EasyAlign: an automatic phonetic alignment tool under Praat. In *Proc. of the Interspeech 2001*. Florence, Italy, pp. 3233--3236.

Grabe, E. (2004). Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles, J. Peters (Eds.), *Regional Variation in Intonation.* Tübingen: Niemeyer, pp. 9--31.

Hirst, D. (2005). Form and Function in the Representation of Speech Prosody. *Speech Communication.* 46, pp. 334--347.

Hirst, D., Di Cristo, A. (1998). *Intonation systems: a survey of twenty languages*. Cambridge: Cambridge University Press.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, Massachussets: MIT Press.

Lucente, L. (2012). Aspectos Dinâmico-Funcionais do Foco e da Curva Entoacional do Português Brasileiro. Doctoral Thesis. State University of Campinas.

Lucente, L. (2008). Dato: Um Sistema de Notação Entoacional do Português Brasileiro baseado em Princípios Dinâmicos. Ênfase no Foco e na Fala Espontânea. Master's Thesis. State University of Campinas.

Massini, G. (1991). A Duração no estudo do acento e do ritmo em português. Master's Thesis. State University of Campinas.

Meireles, A., Barbosa, P.A. (2008). Lexical reorganization in Brazilian Portuguese: An articulatory study. *Speech Communication.* 50, pp. 916--924.

Merlo, S. (2012). Dinâmica temporal de pausas fluentes e hesitações na fala semi-espontânea. Doctoral Thesis. State University of Campinas.

Moraes, J.A. (2011). From a prosodic point of view: remarks on attitudinal meaning. In H. Mello *et al.* (Eds.), *Pragmatics and Prosody. Illocution, modality, attitude, information patterning and speech annotation.* Firenze: Firenze University Press, pp. 19--37.

Moraes, J. A. (1998). Intonation in Brazilian Portuguese. In D. Hirst, A. Di Cristo (Eds.), *Intonational Systems: a Survey of Twenty Languages*. Cambridge: MIT Press, pp. 179--194.

Moraes *et al.*, (2010). Multimodal perception and production of attitudinal meaning in Brazilian Portuguese In *Proc. of the Speech Prosody 2010 Conf.*

Reyelt, M *et al.* (1996). Prosodische Etikettierung des Deutschen mit ToBI. In D. Gibbon (Ed.), *Natural Language and Speech Technology, Results of the third KONVENS conference*, Berlin: Mouton de Gruyter, pp. 144--155.

Rilliard, A. *et al.* (2012). Prosodic analysis of Brazilian Portuguese attitudes. In *Proc. of the Speech Prosody Conf.* Shanghai.

Sauvanet, P. (2000). *Le Rythme et la raison. Rythmologiques.* Paris: Éditions Kimé.

Scherer, K.R. (2000). Psychological models of emotion. In J. C. Borod (Ed.), *The Neuropsychology of emotion*. New York: Oxford University Press, pp. 137--162.

Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin.* 99 (2), 143--165.

Scherer, K. R. (1984). On the nature and function of emotion: a component process approach. In K. R. Scherer, P. Ekman (Eds.), *Approaches to emotion*. Hillsdale, NJ: Lawrence Erlbaum, pp. 293--318.

Scherer, K. R. (1981). Speech and emotional states. In J. Darby (Ed.), *Speech evaluation in psychiatry*. New York: Grune & Stratton, pp. 189--220.

Scott D.R. (1980). Duration as a clue to the perception of a phrase boundary. *Journal of the Acoustical Society of America.* 71, pp. 996 --1007.

Scott, S. K.; Wise, R. J. S. (2003). PET and fMRI studies of the neural basis of speech perception. In *Speech Communication*, 41, pp. 23--34.

Silverman, K. *et al.* (1992). ToBI: a Standard for Labeling English Prosody. In *Proc. of the 2nd International Conference on Spoken Language Processing*. Banff, Canada, 2, pp. 867--870.

Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: articulatory results. *J. Acoust. Soc. Am.* 113 (5), pp. 2834--2849.

Traunmüller, H.; Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *J. Acoust. Soc. Am.* 107, pp. 3438--3451.

Vieira, J.M. (2007). Para um estudo da estruturação rítmica na fala disártrica. Doctoral Thesis. State University of Campinas.

Whitehead, A.N. (1919). *An Enquiry concerning the Principles of Natural Knowledge.* Cambridge, UK: Cambridge University Press.

Wightman, C. (2002). ToBI or Not ToBI? In *Proc. 1st Internat.Conf. on Speech Prosody*. Aix en Provence.

Wightman, C.W. *et al.* (1992). Segmental durations in the vicinity of prosodic phrase boundaries. In *J. Acoust. Soc. Am.*, 91, pp. 1707--1717.

Wunenburger,J.-J. (1992). *Les rythmes. Lectures et theories.* Paris: L'Harmattan.

Xu, Y. (2011). Speech Prosody: A Methodological Review. In *Journal of Speech Sciences*, 1(1), pp. 85--115.

Xu. Y. (2005). Speech melody as articulatorily implemented communicative functions. In *Speech Communication*. 46, pp. 220--251.

Illocution and intonation

João Antônio de MORAES

Universidade Federal do Rio de Janeiro (UFRJ), Brazil
Av. Horácio Macedo, 2151 – Cidade Universitária, Rio de Janeiro, RJ – 21941-917
jamoares3@gmail.com

Abstract

The question of how the expression of affect interferes with the nature of the illocution is addressed. Data on attitudinal intonation in Brazilian Portuguese are presented and the role of the presence of different types of "attitude" for establishing the value of an expression's illocutionary force meaning is discussed.

Keywords: intonation; illocution; speech acts; attitudes.

1. Taxonomies of illocutions

It is a well-known fact that taxonomies of speech acts usually comprise very large numbers (in fact many hundreds) of "individual" speech acts or illocutions (Austin, 1962; Vanderveken, 1998; Searle & Vanderveken, 2009). Austin, for instance, based on the number of performative verbs in English, reckoned nearly one thousand of different illocutions.

In the early days of Speech Act Theory in the 1960s and 70s, these inventories were usually based on the authors' introspective judgment and, occasionally, on the observation of written language, with imagined examples from which the context of the utterance was often erased; and they relied mainly on the presence of performative verbs or expressions.

Later, authors such as Searle & Vanderveken (1985/2009) spoke more explicitly about what were called "illocutionary force-indicating devices" (IFIDs). These are linguistic devices which indicate that the utterance is made with a certain illocutionary force. For instance, in Portuguese, as in English, the imperative mood indicates that the utterance is intended as a directive illocutionary act (an order, a request etc.); the words "I promise" are supposed to indicate that the utterance is intended as a promise; and so on. Besides performative verbs -- which must be in "performative conditions" (the verb in the first person singular of the present tense), which is rather rare in spontaneous language: in BP we hardly ever say "I order you to close the window" -- , other possible IFIDs in English include: the mood of the verb, the word order (which is less important in BP than in English or French), the presence of interrogative or exclamative morphemes, which characterize the traditional declarative, interrogative, imperative, exclamatory sentence types and intonation contours.

2. Oral language: the contribution of intonation

From the 1990s onwards, with the growing interest in the study of oral, spontaneous language, more and more emphasis has been placed on the importance of this latter element – the intonation contour or, in a broader sense, the prosody –by the adherents of the Teoria della Lingua in Atto, for instance (Firenzuoli, 2003; Cresti, 1998, 2000; Moneglia, 2011; Raso, 2012).

Indeed, many illocutions are typically "intonational" in that they display dedicated prosodic contours which, in the absence of other relevant factors, define the illocutionary force to be assigned to the utterance.

The big question then is: how many different pitch contours related to illocutions are there? In other words, how many are in fact different, and how many should be seen as variants of the same type? The answer to this question is not simple, for several reasons. I will focus on two aspects – both highly complex – (i) the relations between illocution and intonation and (ii) relations between illocution and affective states, particularly attitudes.

3. Illocution and intonation: the intonational homonymy issue

In addition to pitch contours proper, the practical task of classifying illocutions in spoken corpora underlines the importance of textual, contextual and situational elements in establishing the value of an expression's illocutionary force (Cutler, 1977, Cruttenden, 1986, Pierrehumbert & Hirschberg, 1990).

Three factors in particular play a crucial role in this respect:

i. text (locutionary) characteristics, like the proposition being in the in past or in future, grammatical person, special morphemes... (Searle's propositional content conditions);
ii. dialogical structure: the position the illocution occupies in a dialogical exchange (initiative *vs.* reactive),
iii. participation by other elements, such as voice quality and "visual" prosody (mainly facial gestures), characterizing intonation as a multimodal phenomenon.

These quite different factors participate – should I say conspire – and interact strongly in construction of the intonational meaning, or more generally the "communicative value" of an illocution, making it hard, if not impossible, to establish one-to-one equivalence between melodic contour and meaning.

As an illustration, Figure 1 shows a very common contour in BP. The double-rise contour with a higher melodic peak on the first stressed syllable, followed by a fall, forming a valley, and a second rise (less marked than the first) on the last stressed syllable, the latter peak aligning with the beginning of that syllable, causing a falling intra-syllabic configuration.
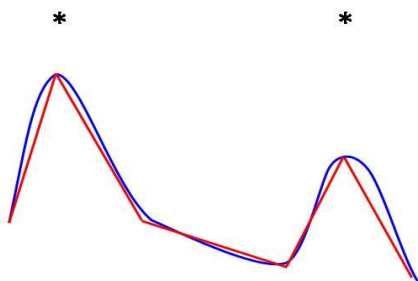
Figure 1: The double-rise contour in BP

That contour typically appears with imperative sentences, and has been associated with requests in BP (Moraes & Colamarco, 2007; Moraes, 2008; Bodolay, 2009; Queiroz, 2011). However – depending mainly on text/locutionary characteristics – it occurs with, and characterizes, other illocutions as well, such as yes-no questions (particularly rhetorical yes-no questions), exclamations, exhortations and even topic structures (which are not properly illocutions).

To illustrate this point, the same stylized double-rise contour was imposed on of 5 different sentences, as can be seen (and heard) in the next figures:

(i) Request:
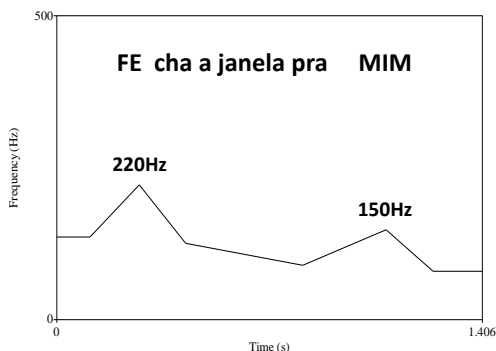
Figure 2: Stylized double-rise F0 contour of the utterance *Fecha a janela pra mim?* (Could you close the window for me?) ◀))

With this kind of sentence (imperative mood, second person singular, future action) the double-rise contour is typically interpreted as a request (Figure 2).

(ii) Yes-no question, often with rhetorical value:
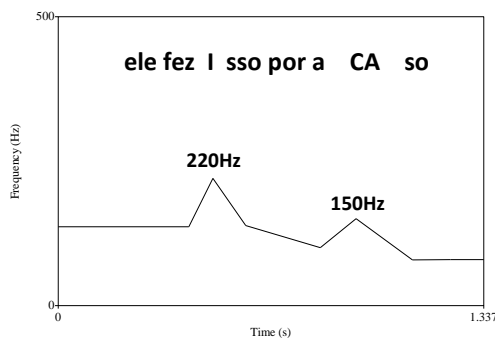
Figure 3: Stylized double-rise F0 contour of the rhetorical yes-no question *Ele fez isso por acaso?* (Did he do it, by any chance?) ◀))

This utterance (Figure 3) is preferentially interpreted as a rhetorical question, in which the speaker is assuming that the person referred to ("he") did not accomplish the action mentioned in the preceding context.

Interestingly, with the final rise contour as in Figure 4, the rhetorical sense is lost: we have then a "true" question, a request for information, with the meaning of "Did he do it by accident?" If both questions are answered in the negative, the truth value of these answers is different: in the first case, he did not do it and, in the second, he did (but not by chance), indicating that the negation's scope is distinct in each sentence.
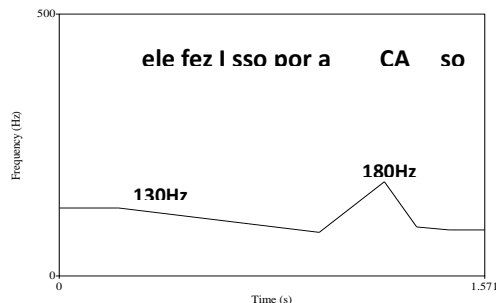
Figure 4: Stylized final-rise F0 contour of the real yes-no question *Ele fez isso por acaso?* (Did he do it by accident?) ◀))

(iii) Exhortation, invitation, encouragement:
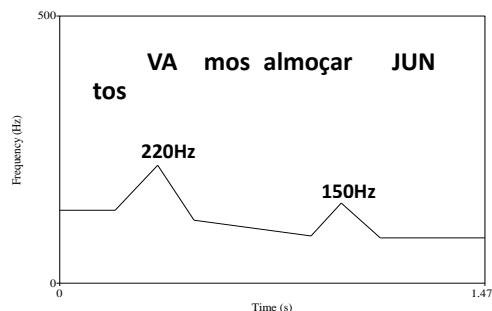
Figure 5: Stylized double-rise F0 contour of the exhortation *Vamos almoçar juntos?...* (Let's have lunch together?) ◀))

This utterance (Figure 5) *Vamos almoçar juntos* (Let's have lunch together?) (imperative mood, first person plural, future action) is understood as an exhortation, an invitation. Again, a final rise contour (Figure 6) causes, or at least favors, the real-question interpretation: "I don't remember, are we going to have lunch together?"
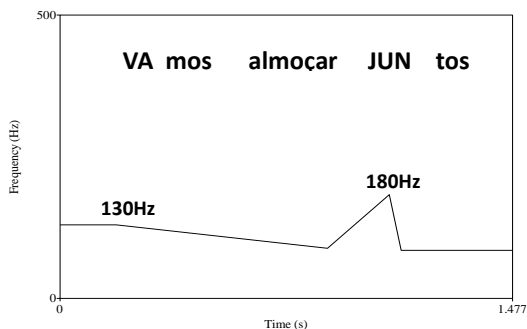
Figure 6: Stylized final-rise F0 contour of the real yes-no question *Vamos almoçar juntos?* (Are we going to have lunch together?) 🔊

(iv) Exclamation
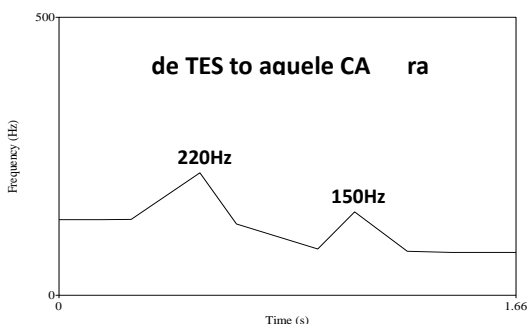
Figure 7: Stylized double-rise F0 contour of the 🔊 exclamation *Detesto aquele cara!..* (I hate that guy!...).
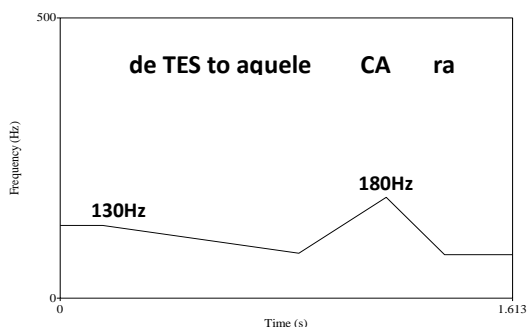
Figure 8: Stylized final-rise F0 contour of the yes-no 🔊 echo-question *Detesto aquele cara?* (Do I hate that guy?)

This utterance (Figure 7) (indicative mood, first person singular) often conveys the sense of an exclamation of personal, unexpected-information type (a sort of "revelation"). With a final rise contour (Figure 8), again the question sense appears, often a metalinguistic, echo-question:

Speaker 1 *Sei que você detesta aquele cara*. (I know

you hate that guy.) Speaker 2 *Detesto aquele cara?* (Do I hate that guy?)

(v) Finally, in the intonational phrase domain the same melodic contour appears also characterizing a Topic structure, as in Figure 9: *No Norte de Minas, [existia um... um ...[sujeito], meio aparentado com minha esposa...]* (In northern Minas, there was a ... [guy], some kind of relation to my wife.) In this spontaneous speech example from the C-Oral corpus (Raso and Mello 2012) the original melodic contour was slightly modified by F0 manipulation with Praat (it originally showed the higher peak in the final position; here I assume that, in Rio de Janeiro, these two contours are dialectal variants of the same topic pattern).
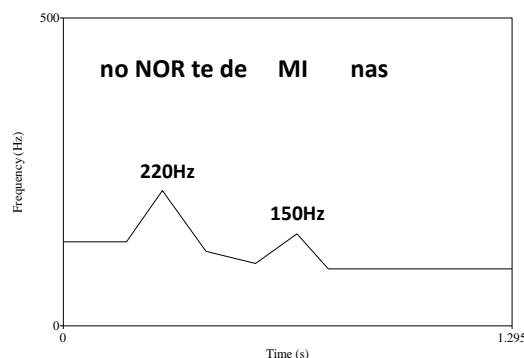
Figure 9: Stylized double-rise F0 contour in the Topic structure *No Norte de Minas, existia um... um ...[sujeito], meio aparentado com minha esposa.* (In northern Minas, there was a ... [guy], some kind of relation to my wife.) 🔊

Again, with the final rise contour, as in Figure 10, the sentence becomes a real question: *No norte de Minas?* (In northern Minas?)
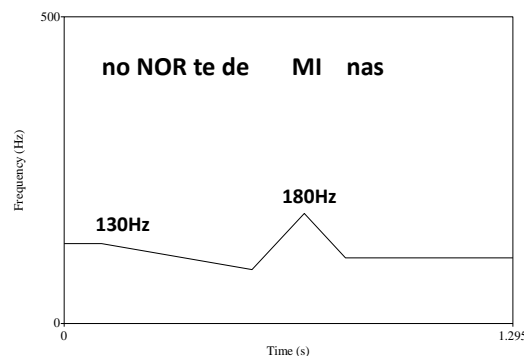
Figure 10: Stylized final-rise F0 contour of the real yes-no question *No norte de Minas?* (In northern Minas?) 🔊

This "context-dependence of intonational meanings", as the title of Ann Cutler's interesting study (1977) puts it, allows a massive reduction in the number of dedicated melodic patterns available and leads to a kind of widespread intonational homonymy phenomenon, to use an expression of Romportl's (1973). This, to some extent, explains the imbalance between the hundreds of illocutions described and the few dozen of melodic patterns assigned to them.

Conversely, the illocutionary act of posing a question (questioning) corresponds to different melodic contours according to the logical structure of the question, as shown in Figure 11.
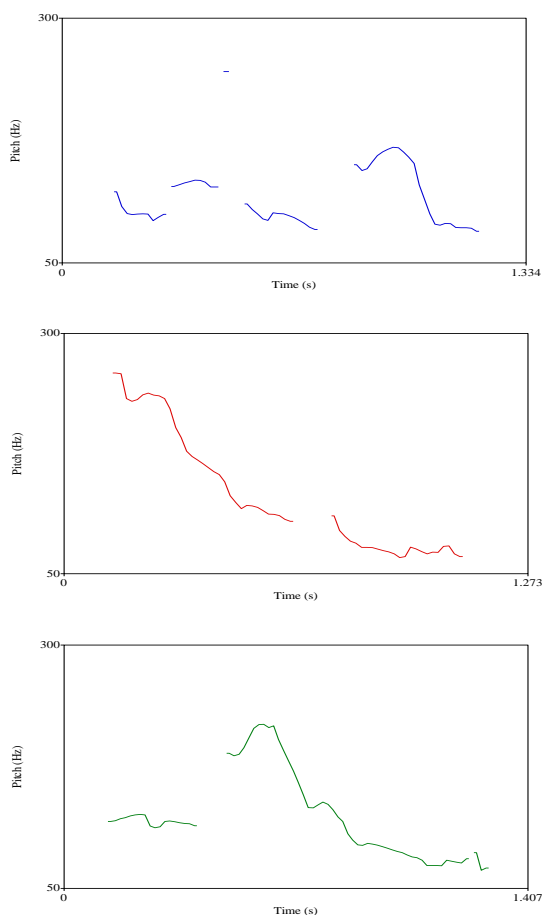
Figure 11: F0 contours of the yes-no question *Roberta dançava?* (Was Roberta dancing?) (top), the wh-question *Como ela dançava?* (How she danced?) (middle) and the alternative question *Ela dançava ou jogava?* (Was she dancing or playing?) (bottom)

4.   Types of affect

Ilocutions and affective states, especially attitude are closely related. The relevance of the speaker's attitudes and feelings in the composition of an illocutionary act is an especially important point. Indeed, it is hard, from a practical and theoretical point of view, to decide whether two melodic contours should be considered phonologically and pragmatically distinct or merely expressive variants of the same illocutionary act. As Couper-Kuhlen (1986) puts it: "The more basic problem may be that illocutionary contrasts shade into attitudinal contrasts and it is difficult to know where to draw the line."

Indeed, it seems – and some authors have proposed – that the illocutionary *versus* attitudinal contrasts, which overlap the grammatical versus expressive ones, behave rather like two categories arranged in a continuum rather than in discrete opposition. P. Léon (1993) for instance, proposes a continuum of five steps, going from raw

emotion to grammatical modality (Figure 12, in Appendix). This idea is also captured by the scheme proposed by Aubergé (2002), which distinguishes emotional from attitudinal and linguistic functions (Figure 13, in Appendix).

Scherer (2000, 2003) in turn have proposed a detailed design feature approach to distinguish five classes of affective states, in place of the traditional emotion versus attitude contrast:

- Emotions (e.g., angry, sad, joyful, fearful, ashamed, proud, elated, desperate),
- Moods (e.g., cheerful, gloomy, irritable, listless, depressed, buoyant),
- Interpersonal stances (e.g., distant, cold, warm, supportive, contemptuous),
- Preferences/attitudes (e.g., liking, loving, hating, valuing, desiring)
- Personality traits (or affect dispositions) (e.g., nervous, anxious, reckless, morose, hostile, jealous).

This typology is based on the behavior of seven parameters, rated in three degrees, H(igh), M(edium) and L(ow), as can be seen in the Table 1:

TIPES OF AFFECT

Design features	E	M	IS	PQ	AD
Intensity	H	M	M	M	L
Duration	L	M	M	H	H
Synchronization	H	L	L	L	L
Event focus	H	L	M	L	L
Appraisal elicitation	H	L	L	L	L
Rapidity of change	H	M	H	L	L
Behavior impact	H	L	M	M	M

Table 1: Features of Emotions (E), Moods (M), Interpersonal stances (IS) Preferences/attitudes (PA) and Affect dispositions (AD)

In this approach the traditional category of attitude (as opposed to emotions) is split into 4 new categories: moods, interpersonal stances, preferences/attitudes and affect dispositions. From a strictly prosodic perspective, the main concern is to establish to what extent these categories show different prosodic behaviors; that is, whether there are prosodic features that characterize these different categories of affective states, either because the

categories preferentially use different parameters, or use the same parameters in different ways.

5. Propositional attitudes and illocution

As regards the notion of "attitude" we have to refer to another of its senses, as used in the expression "propositional attitude", borrowed from the philosophical and contemporary logic tradition, since Bertrand Russell's work. This expression is used in intonational studies (Pierrehumbert & Hirschberg, 1990; Whichmann, 2000; Moraes *et al.* 2011, 2012) and in Speech Act theory as well, in contrast with social or interpersonal attitude (or "attitude" *tout court*). While interpersonal attitudes, or stances have to do with the speaker's behavior towards the hearer, a propositional attitude is a "psychological attitude towards a state of affairs" (Leech, 1983: 106), expressed by a proposition.

If we look at the theory of speech acts, we see that an illocutionary act (or rather a class of illocutionary acts) is often defined as the expression of a speaker's attitude toward the propositional content. So, as in Bach and Harnish (1979):

1. Constatives express the speaker's belief and the intention (desire) that the hearer form a like belief;
2. Directives expresses an attitude toward some prospective action, and the intention that the utterance be taken as a reason for the hearer's action;
3. Commissives express a speaker's intention and the belief that his utterance obligates him to do something; and
4. Acknowledgments ("expressive acts" for Searle, "behabitives" for Austin) express feelings regarding the hearer (or the speaker's intention that the utterance satisfy a social expectation to express certain feelings).

Searle regarded all illocutionary acts as condition-governed, and one of these conditions is the sincerity condition (or psychological state condition), which refers to the psychological state or attitude towards the proposition expressed by the speaker in performing an illocutionary act. Accordingly, the propositional attitude is one of the features or components that distinguishes the 5 classes of speech act.

In the performance of any illocutionary act with a propositional content, the speaker expresses some attitude, state, etc. to that propositional content (Searle, 1976). These attitudes or psychological states are: belief, desire, intention, and regret or pleasure, according to the type of act (respectively, Representative, Directive, Commissive, and Expressive). So the presence of these propositional attitudes seems crucial for distinguishing different types of speech acts, and prosodic features are possibly an important way to characterize illocutions.

6. Illocutionary/attitudinal (propositional) intonational contours

This hypothesis has been tested in an ongoing study of the production and perception of social (interpersonal) *versus* propositional attitudes in BP, conducted under the direction of Albert Rilliard as part of the PADE Project (Rilliard *et al.*, 2010). The purpose of this project is to examine attitudinal prosody cross-linguistically in languages such as French, Japanese, American English and Brazilian Portuguese, assessing the specific weight of visual and audio channels in its manifestations (Rilliard *et al.*, 2009; Moraes *et al.*, 2010).

It has been shown that, in BP, propositional and social attitudes in fact display differentiated prosodic behavior in both perception (Moraes *et al.*, 2010, 2011) and production (Moraes *et al.*, 2012). Sixteen attitudes were examined, six of which were social (arrogance, authority, seduction, contempt, irritation and politeness) and five propositional (doubt, obviousness, disbelief, irony and surprise), all expressed through the neutral declarative sentence "Roberta dançava" [Roberta was dancing/Roberta used to dance]. That same sentence, uttered as a yes-no question "Roberta dançava?" [Was Roberta dancing?/Did Roberta use to dance?] was spoken with the same six social attitudes and with four new propositional attitudes, namely, confirmation, incredulity, rhetoricity and surprise (Moraes *et al.*, 2011). Both studies also included the so-called "neutral" (respectively, assertive or interrogative) attitude.

Two Brazilian speakers were recorded and filmed while producing these sentences. The resulting audio and visual stimuli were submitted to an identification (forced choice) test with 30 subjects, who had to identify the speaker's attitude from the audio alone, from the image alone and, finally, from both information sources simultaneously.

The order in which the stimuli were presented was balanced: half the subjects judged video stimuli first and then audio stimuli (and finally both together), while the other half did things the other way round.

Subjects listened to/viewed the stimuli and gave their answers on a computer screen using a slider which, in addition to indicating the attitude chosen, also reported the relative intensity of the perceived attitude on a scale from 0 to 100.

The results for both modalities show not only that the propositional attitudes were in general significantly better recognized than social ones, but more specifically that the visual channel plays a much more important role than audio in recognition of social attitudes (Graphs 1 and 2 below).

Specifically for assertions, the audio channel for propositional attitudes returned a score of 61% correct answers (much higher than the 17% chance level), while for social attitudes it produced an average recognition of only 25% (near to the 14% chance level for this case). Although the contribution of the visual channel is very important in both, it is crucial in relation to social attitudes,

which are indeed visually dependent.
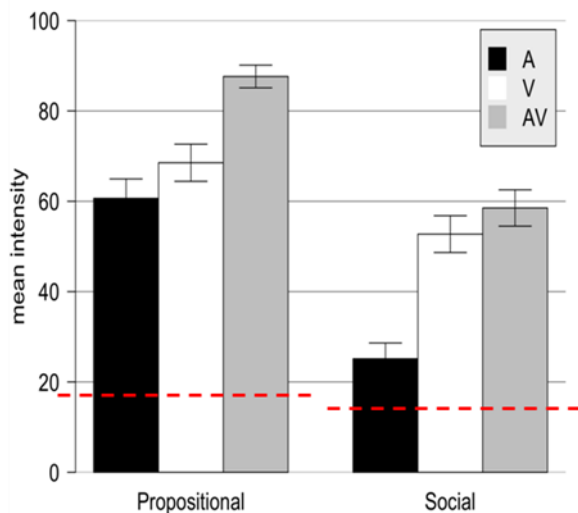
Graph 1: Assertive sentences: mean intensity of correct answers in each condition, for propositional and social attitudes, both speakers. A stands for audio condition, V for video and AV for both together

In interrogatives, almost the same results were obtained for audio stimuli: 60% for propositional and 28% for social attitudes, with the visual channel contributing less in relation to the propositional attitudes.
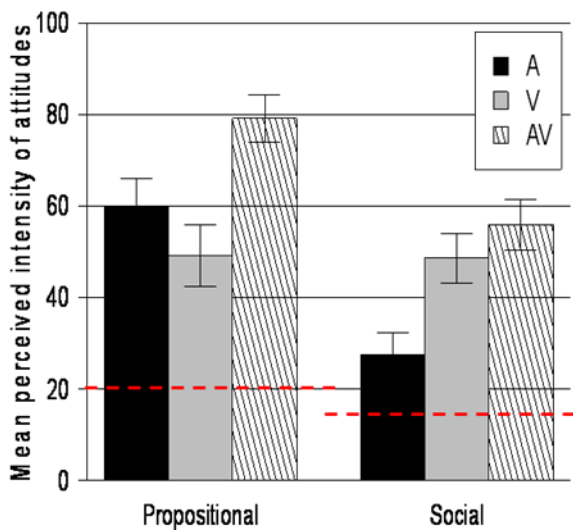
Graph 2: Interrogative sentences: mean intensity of correct answers in each condition, for propositional and social attitudes, both speakers. A stands for audio condition, V for video and AV for both together

Concerning production, the assertive sentence with neutral attitude can be characterized melodically by a moderate F0 fall in the final, nuclear position, specifically between the last pre-stressed and stressed syllables, which also assumes a falling internal configuration.

Looking at how social attitudes surface in melodic terms, one sees that they show rather subtle melodic distinctions (Figure 14), and that the neutral contour is basically preserved.
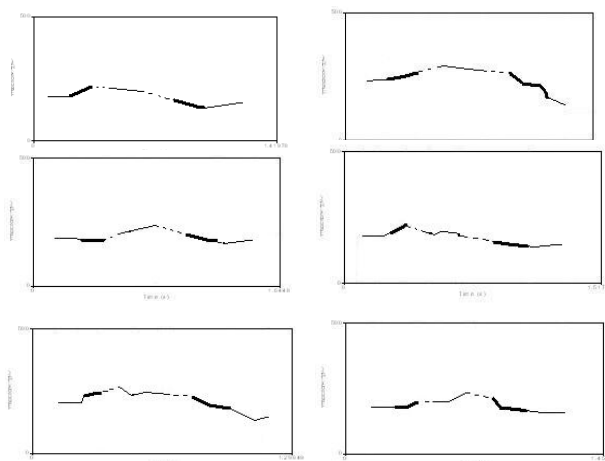
Figure 14: Stylized pitch contours of the assertive sentence 'Roberta dançava' [Roberta was dancing/ Roberta used to dance] uttered with six social attitudes, female speaker. The thicker line indicates the stressed vowels, the dotted line, voiceless consonants. From top to bottom: arrogance and authority; seduction and contempt; irritation and politeness
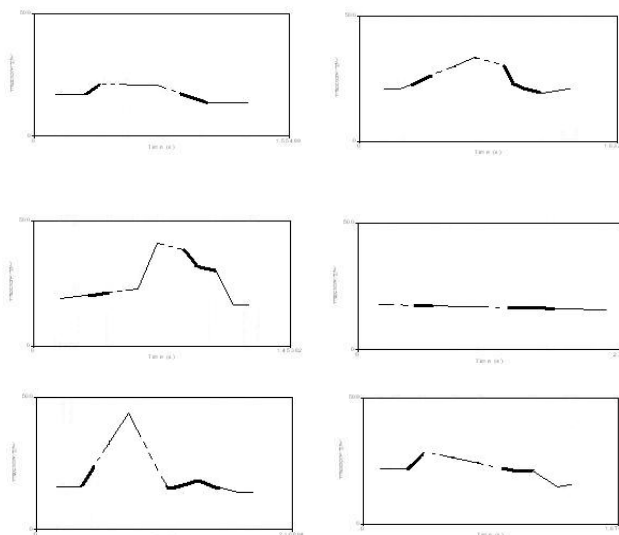
Figure 15: Stylized pitch contours of the assertive sentence 'Roberta dançava' [Roberta was dancing/ Roberta used to dance] uttered with neutral and five propositional attitudes, female speaker. The thicker line indicates the stressed vowels, the dotted line, voiceless consonants. From top to bottom: neutral and doubt; obviousness and disbelief, irony and surprise

On the other hand, most of the propositional attitudes examined here show important, punctual changes in the melodic contour (Figure 15), which modify its basic configuration (Moraes, 2011); that is why they are better perceived by the ear. These changes are located mainly in the nuclear position, more specifically the last stressed syllable, and/or in the contrast between this syllable and the preceding one. The tonal importance of the nuclear position has been confirmed by manipulating

the F0 at specific points in the melodic patterns of propositional attitudinal utterances, then validating by perception tests (Moraes, 2008).

Accordingly, in disbelief, both nuclear syllables are produced at a very low melodic level; in obviousness, the last stressed syllable is produced at quite a high level (for an assertive sentence); in irony the last stressed syllable assumes a typical, circumflex (rising-falling) shape; and doubt displays – among other things – a high last pre-stressed syllable. In addition in the duration level, irony, disbelief and doubt also display greater duration in general, especially a lengthening of the last stressed syllable. These major differences between the expression of social and propositional attitudes are observed among interrogatives as well.

The results of perceptive analysis (Moraes *et al*., 2010, 2011), acoustic analysis (Moraes *et al*., 2012) and even F0 manipulation experiments with resynthesis (Moraes, 2008) reinforce the idea that there are two independent prosodic systems: emotions + social attitudes *vs*. propositional attitudes, that in fact correspond to a large extend to different speech acts.

In the original scheme proposed by Aubergé (2002), the attitudinal functions are located halfway between the linguistic and non-linguistic functions. The proposal here is then to split the two categories of attitudes, putting social attitudes together with emotions, and propositional ones with speech acts.

Emotions and social attitudes do not conflict with speech acts or propositional attitudes: in fact they can be added to them without destroying the basic communicative value. Also, from a prosodic perspective, neither do they significantly disturb the basic melodic pattern - in fact, the pattern is largely preserved; to be more precise, it becomes a variant of the original (unmarked) pattern.

This means that the phonological representation of a particular illocutionary act spoken with different emotional or social-attitudinal values would be the same: there are no localized, punctual F0 changes, but global modifications in the overall pattern (register and tonal span), not to be represented in phonological form. With propositional attitudes and speech acts, the changes are local, discrete, leading to distinct phonological analyses.

Finally, regarding the participation of different "media" in the expression of affective meaning, our data reveal that the visual channel (facial stimuli) contributes more to the production and perception of social attitudes than the audio channel (prosody and voice quality).

7.   References

Aubergé, V. (2002). A gestalt morphology of prosody directed by functions: the example of a step by step model developed at ICP. In *Proceedings of Speech Prosody 2002*, Aix-en-Provence.

Austin (1962/1976). *How to Do Things With Words*, Oxford : Oxford Higher Education; 2nd edition.

Bach, K., Harnish, R. (1979). *Linguistic Communication and Speech Acts* M.I.T. Press.

Bodolay, A. (2009). Pragmática da entonação: a relação prosódia/contexto em atos diretivos no Português, Tese de Doutorado, Estudos Linguísticos, UFMG..

Couper-Kuhlen, E. (1986). *English Prosody*. Edward Arnold, London.

Cresti, E. (1998). Critère illocutoire et articulation informative. In *LABLITA*, preprint 2.

Cresti, E. (2000). Per una nuova classificazzione dell'illocuzione. In *LABLITA*, preprint 1.

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

Cutler, A. (1977). The context-dependence of "intonational meanings". In *13th Regional Meeting, Chicago Linguistic Society*, pp. 104--115.

Firenzuoli, V. (2003). Le forme intonative di valore illocutivo dell'italiano parlato. Tesi di Dottorato, Università degli Studi di Firenze.

Leech, G. (1983). *Principles of Pragmatics*. London: Longman.

Léon, P. (1993). De l'émotion aux conduites émotives et à la modalisation. In P. Léon (Ed.), *Précis de Phonostylstique*. Paris: Nathan.

Moneglia, M. (2011). Spoken corpora and pragmatics. In *RBLA*, Belo Horizonte, v.11 (2), pp. 479--519.

Moraes, J. (2008). The Pitch Accents in Brazilian Portuguese: analysis by synthesis, In P. Barbosa, S. Madureira and C. Reis (Eds.), *Proceedings of the Speech Prosody 2008*: *Fourth Conference on Speech Prosody*, pp. 389--397.

Moraes, J. (2011). From a prosodic point of view: remarks on attitudinal meaning, In H. Mello, A. Panunzi and T. Raso (Eds.), *Pragmatics and Prosody. Illocution, modality, attitude, information patterning and speech annotation*. Firenze: Firenze University Press, pp. 19--37.

Moraes, J., Colamarco, M. (2007). Você está pedindo ou perguntando? Uma análise entonacional de pedidos e perguntas na fala carioca no Português do Brasil. In *Revista de Estudos da Linguagem* vol 15 (2), 2007, pp. 113--126.

Moraes, J., Rilliard, A., Mota, B. and Shochi, T. (2010). Multimodal perception and production of attitudinal meaning in Brazilian Portuguese, *Proceedings Speech Prosody 2010*, 5 th International Conference, Chicago, 11 a 14 de maio de 2010.

Moraes, J., Rilliard, A., Erickson, D. and Shochi T. (2011). Perception of attitudinal meaning in interrogative sentences of Brazilian Portuguese. In *Proceedings of the 17 th ICPhS*, Hong Kong, China, 2011.

Moraes, J., Rilliard A., Erickson, D. and T. Shochi (2012). Acoustic analysis of a corpus of Brazilian Portuguese attitudes. In these Proceedings.

Pierrehumbert, J., Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, and M. Pollack (Eds.), *Intentions in Communication*. Cambridge, Mass.: The MIT Press, 1990, pp. 271--311.

Queiroz, H. (2011). Aspectos prosódicos das atitudes do locutor em atos de fala diretivos, Tese de Doutorado,

Estudos Linguísticos, UFMG..

Raso, T. (2012). O C-ORAL-BRASIL e a Teoria da Língua em Ato. In T. Raso, H. Mello (Eds.), *C-ORAL – Brasil I : Corpus de referência do português brasileiro falado informal*.

Raso, T., Mello, H. (Eds.) (2012). *C-ORAL – Brasil I : Corpus de referência do português brasileiro falado informal*. Belo Horizonte, Editora UFMG..

Rilliard, A. (2010). Projet Prosodie: Accents, Dialectes, Expressivité (PADE). Agence Nationale de la Recherche, France.

Rilliard, A., Shochi, T., Martin, J-C., Erickson, D. and Aubergé, V. (2009). Multimodal indices to Japanese and French prosodically expressed social affects. In *Language and Speech*, 52 (2/3), pp. 223--243.

Romportl, M. (1973). On the synonymy and homonymy of means of intonation. In M. Romportl (Ed.), *Studies in Phonetics*. Prague: Academia.

Scherer, K. (2000). Psychological models of emotion. In J. Borod (Ed.), *The Neuropsychology of Emotion*. Oxford: Oxford University Press.

Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. In *Speech Communication* 40, pp, 227--256.

Searle, J. (1976). A classification of illocutionary acts. *Language in Society* 5 (1), pp. 1--23.

Searle, J., Vanderveken, D. (1985/2009). *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

Vanderveken, D. (1998/2009). *Meaning and Speech Acts*: Volume 1, Principles of Language Use, Cambridge: Cambridge University Press.

Wichmann, A. (2000). The attitudinal effects of prosody, and how they relate to emotion, In R. Cowie, E. Douglas-Cowie and M. Schröder (Eds), *Proceedings, ISCA workshop on Speech and Emotion*, 2000 [SpeechEmotion-2000], pp. 143--148.

8. Appendix

raw emotion	emotion-feeling	intellectual emotion (attitude)	linguistic encoded emotion	grammatical modality
anger joy, fear sadness…	hate happiness anxiety longing…	admiration irony seduction …	emphatic stress…	assertion question, command…
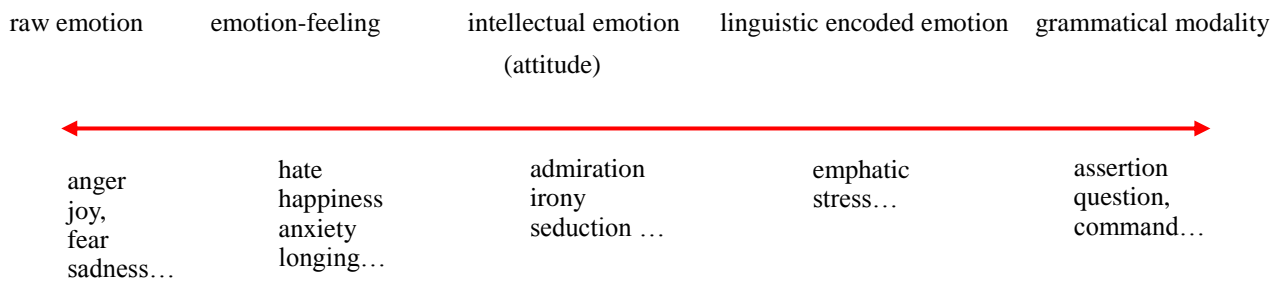
Figure 12: Scheme adapted from Léon (1993)

emotional functions	attitudinal functions	linguistic functions
- cortical	involuntary control voluntary control	+ cortical
states of the primary and secondary emotions of the speaker	*values of speaker's intentions*	*structures of enounciation*
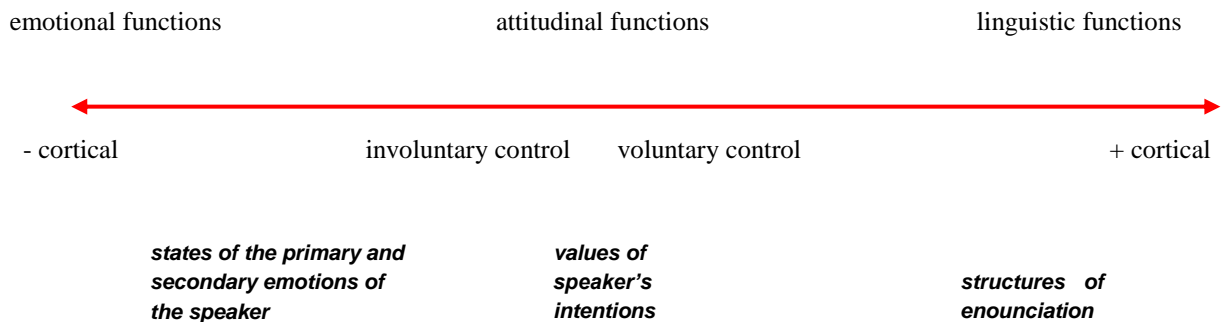
Figure 13: Scheme proposed by Aubergé (2002)

SPOKEN CORPORA COMPILATION AND ANNOTATION

Building a *corpus* for comparative analysis of language attrition

Lúcia de Almeida FERRARI

Universidade Federal de Minas Gerais (UFMG)

Av. Antônio Carlos, 6627 - Belo Horizonte/MG - Brasil

ferrari.lu@gmail.com

Abstract

The aim of this research is the study of first language attrition of Italian L1 in contact with Brazilian Portuguese. Language attrition is the gradual decline or the loss of a first or second language by an individual. This is a corpus-based study: a corpus of oral spontaneous speech was collected using eight different subjects. This corpus, composed of 21298 words, was compared with fourteen different texts from the Italian C-ORAL-ROM (Cresti & Moneglia, 2005). The results were then compared with those of previous studies by Raso and Vale (2007, 2009). The attrition of Italian L1 was confirmed, with a few differences that may deserve further and deeper analysis in future studies. The variation of the percentage of loss between the two researches seems to be mostly due to: 1) differences in typology of texts; 2) different diaphasic varieties; 3) different pragmatic contexts. The greater dissimilarities are noticed between the two reference corpora. Finally, data seems to confirm that attrition is a process that doesn't come to a halt after the first decade, but one that continues in time.

Keywords: attrition; corpus; Italian; clitics.

1. Introduction

This paper discusses the methodology employed to build a corpus for first language attrition study and the results obtained comparing it to previous researches.

The definition of L1 attrition is a "non-pathological decrease in proficiency in a language that has previously been acquired by an individual i.e. intragenerational loss" (Köpke & Schmid, 2004: 5). The process is due to two factors: the influence of L2 system and the lack of use of, and exposure to, the L1. In our case the study is about Italian L1 attrition in contact with Brazilian Portuguese.

Previous researches (Raso & Vale, 2007, 2009) on a group of clitics adopted the corpus methodology to investigate the degree of attrition of a group of Italians living in São Paulo for 20 to 30 years.

The aim of our research was to create a corpus with a greater diaphasic variety, in order to ensure the higher possible degree of spontaneousness.

The object of the study was the same group of clitics analysed by Raso and Vale, that is: *ci attualizzante*, *lessicalizzante* and *locativo*; *ne partitivo*, *argomentale* and *locativo* and the third person accusative clitics *lo*, *la*, *li*, *le l'*.

2. Corpus design and methods

Raso and Vale researches analysed a corpus extracted from a collection of interviews (Revista de Italianística, 1997), for a total of 18080 words, and compared it with an excerpt of the BADIP corpus (De Mauro *et al*., 1993) for a total of 18080 words.

To guarantee their complete acquisition of the language and some kind of meta-linguistic remark skills, the participants were all Italians, born and raised in Italy until the coming of age, with a high school degree obtained in Italy and, preferentially, a college degree.

In choosing the informants for our research we followed the same criteria; the required contact period with Brazilian Portuguese was of at least eight to ten years, as recommended by the attrition bibliography.

Eight different participants were selected: we were able to obtain various types of interactions, namely: a conversation between three people watching a soccer match on TV; five dialogues (one between a couple making dinner, one between two sisters, one about sports, one during a meal, and a discussion about doctors); and two monologues in which people spoke about their life experiences. Therefore, the resulting corpus reflects a higher degree of diaphasic variation than the one used by Raso and Vale. This is a key element for our study because it's correlated to a greater spontaneity of speech and can allow us to study the actual degree of attrition in real-world situations.

Our corpus has a total of 21298 words; as a reference corpus we selected fourteen different texts, the most similar to ours, from the Italian C-ORAL-ROM (Cresti & Moneglia, 2005), to a total of 21224 words. The choice of C-ORAL-ROM is due to it being a third generation corpus, highly spontaneous, transcribed in CHAT format (McWhinney, 1994), the same one we used in our corpus, and to the fact that all the digital recordings are available (as they are for our corpus).

The first step was to search our corpus and the Italian C-ORAL-ROM for excerpts containing the clitics we were studying and their collocations. Data were then normalized for comparison purpose. Every clitic was compared in normalized form and as a percentage.

The second step was to compare the results of the above described research with those of the studies by Raso and Vale. Again, all data had to be normalized. Several sets of data, as we will show, were extrapolated and compared, in order to point out similarities and differences between the results of both studies and to formulate hypotheses.

3. Data Collected

In the following section we will present the data we collected and the comparison made between our corpus and the reference one (C-ORAL-ROM), and between our findings and those of the Raso-Vale study. Each clitic will be examined separately and, at the end, we'll offer

our conclusions.

3.1   An overview

In this paper all data will be provided in their normalized form, to facilitate the comprehension of the comparison we made.

Our corpus, named in the below tables Raso-Ferrari corpus, presents a total of 191,09 occurrences of clitics every 10000 words, while the Italian C-ORAL-ROM presents 304,37.

CLITICS	Raso-Ferrari *Corpus*	Raso-Vale *Corpus*
TOTAL	191,09	179,18

Table 1: Normalized values (per 10000 words) in both attrition corpora studied

In percentage this means a 37,31% decrease compared to the reference corpus. Looking at the previous studies, the Raso-Vale corpus presents 179,18 occurrences, while the BADIP corpus presents 270,46 occurrences; in percentage, that's a 33,74% decrease. This difference is relatively small and our study seems to confirm the attrition of our test group.

Our data turn out to be much more interesting when the clitics are split, as seen in table 2: it's possible to observe considerable differences between the two studies. While in the Raso-Vale researches the number of *ci attualizzanti* increases by nearly 10%, our study shows a decrease of about 50%.

CLITICS	Raso-Ferrari *Corpus/Italian C-ORAL-ROM*	Raso-Vale *Corpus/BADIP*
Ci attualizzanti	-50,91	9,34
Ci lessicalizzanti	-54,72	-70,16
Ci locativo	-84,22	-38,47
lo, la, li, le, l'	-23,81	-45,39
Ne (total)	-25,86	-51,71
TOTAL	-37,21	-33,74

Table 2: Percentage variation between attrition studies

This is the most evident discrepancy, but there are others: in the case of the *ci lessicalizzanti* we can see a 70,16% decrease in the Raso-Vale corpus, greater than the 54,72% decrease registered in ours; the *ci locativo*, on the other side, shows a decrease of about 84% in our study, while in the Raso and Vale research is about 38%; third person accusative clitics show a decrease of nearly 24% in our corpus and about 45% in the Raso and Vale studies; finally, the total *ne* clitics show a decrease of about 26% in our studies and nearly 52% in the previous ones.

In an attempt to explain such remarkable differences, table 3 shows the normalized data of all corpora used in both researches to see how much weight each references corpora have in the total values.

CLITICS	Raso-Ferrari *Corpus*	Raso-Vale *Corpus*	Italian C-ORAL-ROM	BADIP
Ci attualizzanti	63,38	64,71	129,09	59,18
Ci lessicalizzanti	4,69	1,65	10,36	5,53
Ci locativo	1,4	13,27	8,95	21,57
lo, la, li, le, l'	111,27	91,81	146,06	168,14
Ne (total)	15,02	7,74	20,26	16,03
TOTAL	191,09	179,18	304,37	270,46

Table 3: Normalized values (per 10000 words) in all analysed corpora

As it can be easily seen, the two attrition corpora don't show such a huge difference as percentages could induce to believe. In fact, the values of the *ci attualizzanti* are mostly the same, while percentage data between the two studies suggested a considerable divergence. Also, third person accusative clitics don't show such a big difference in normalized values. The most significant differences are the *ci locativo* and total *ne* clitics, but, as we'll see, those discrepancies can be explained quite easily.

What is surprising is the strong difference perceptible between the two reference corpora and the two attrition corpora, and between the two reference corpora themselves. Table 3 shows clearly that the *ci attualizzanti* found in the Italian C-ORAL-ROM are more than twice than those found in the BADIP corpus: 129,09 versus 59,18, respectively. The other clitcs, excluding third person accusative clitics, also present two- or three-fold differences. We can assert, than, that the differences between the results of the studies can be due to the differences between reference corpora; but this isn't the only explanation, as we'll see analysing some particular cases.

3.2   The *ci locativo* clitic

As we can observe in table 4 below, a big difference could be seen in the use of the *ci locativo* both in the attrition corpora and in the reference corpora.

Data suggests that the Italian C-ORAL-ROM corpus has a much smaller number of occurrences of this clitic than the BADIP corpus. The same happens with the Raso-Ferrari corpus in comparison with the Raso-Vale.

CLITICS	Raso-Ferrari *Corpus*	Raso-Vale *Corpus*	Italian C-ORAL-ROM	BADIP
Ci locativo	1,4	13,27	8,95	21,57

Table 4: Normalized *ci locativo* (per 10000 words) in all analysed corpora

What can explain this behaviour? Our hypothesis is that both the Italian C-ORAL-ROM and the Raso-Ferrari corpora contain texts much more spontaneous than the other two corpora. The Raso-Vale corpus is mostly composed by interviews, where the speaker was asked about his migration and travels, so he would use this clitic much more than in a normal conversation. The same happens in BADIP, a corpus based on much less spontaneous types of interactions than C-ORAL-ROM.

So, the divergent data has their explanation in the different kind of texts and interactions that compose the corpora analysed.

3.3   The *ne* clitics

Another clitic that registers divergent values between corpora is *ne*. To better understand this behaviour it's necessary to split this clitic into its various functions and see the resulting figures as shown in table 5 below.

CLITICS	Raso-Ferrari *Corpus*	Raso-Vale *Corpus*	*BADIP*	*Italian C-ORAL-ROM*
Ne partitivo	8,92	3,31	8,84	13,19
Ne argomentale	6,1	4,42	5,53	7,06
Ne locativo	0	0	1,65	0
TOTAL *Ne*	15,02	7,74	16,03	20,26

Table 5: Normalized values (per 10000 words) in all corpora compared

We can see that the values of the Raso-Ferrari corpus and the Italian C-ORAL-ROM are higher than the other two corpora, both the attrition one and the reference one.

Again, what in our opinion may explain the divergent behaviour of these data is the different kind of texts that compose the corpora and the diaphasic variation in texts. The Italian C-ORAL-ROM is a much more modern corpus than BADIP and is representative of the actual spoken language in Italy. Proof of this is the higher number of *ne partitivi* in comparison with *ne argomentali*, less used in modern Italian, and the total absence of *ne locativi*, the latter being, as Russi (2008) supports, totally set aside nowadays. As data indicates, the Raso-Ferrari corpus also depicts this situation, whit a minor degree of attrition in relation to the Raso-Vale corpus. This last consideration induces us to think that, as our corpus is composed by interaction of people that have been living in Brazil for less longer than the ones who were interviewed for the Raso-Vale corpus, this may be a possible evidence of the fact that attrition continues in time and does not, as theorized by many scholars (for example Kopke and Schmidt, 2004), come to a halt after the first decade.

3.4   C*i* clitic in the verbs *esserci* and *averci*

The *ci* clitic can have various functions in Italian. We saw above that it can have a locative use but, as we'll explain, it can also be a particle lexicalizing a verb connected to it. In this paper we call *ci attualizzanti* the forms *esserci* and *averci*, where the grammaticalization is complete, and *ci lessicalizzanti* all the other forms, like *andarci* (going to a place) or *starci* (to agree to do something), independently of the degree of grammaticalization[1].

This distinction is important to understand our analysis. In the first place, as shown in table 3 above, in the Italian C-ORAL-ROM the number of *ci lessicalizzanti* is double respect to BADIP, the other reference corpus. This indicates, once again, the recency of the first corpus. In both attrition corpora the values decrease quite a lot, much more than in the Raso-Vale one, confirming our assumption that attrition increases over time.

If we observe *ci attualizzanti,* we can notice that both attrition corpora exhibit very similar values: the Raso-Ferrari corpus has 63,38 occurrences every 10000 words while the Raso-Vale has 64,71 occurrences. What is quite surprising is the strong difference between the two reference corpora. This time we expected a smaller number of occurrences in the Italian C-ORAL-ROM: again, the explanation lies in the broader diaphasic variation of the texts and in the spontaneity of them, as this form is a pretty comprehensive verb form.

We won't linger over the *ci lessicalizzanti* as the values are very small, to a point where it isn't possible to go further in our investigation.

On the other side, we will analyse in a little more depth both *esserci* and *averci*.

CLITICS	Raso-Ferrari *Corpus*	Raso-Vale *Corpus*	*BADIP*	*Italian C-ORAL-ROM*
esserci esistenziale	48,36	49,77	24,33	78,68
esserci presentativo	7,51	9,04	7,19	7,53
TOTAL esserci	55,87	59,18	31,52	86,22 / 66,9*
averci	7,51	5,53	27,65	42,87

Table 6: Normalized values (per 10000 words) of the *ci attualizzanti* in all corpora analysed (from Panunzi, 2010)

Again, we had to split the data, dividing the *esserci* form into *esistenziale*, when it can replace other verbal forms of existence; and *presentativo*, constituted by the form *esserci+SN+che pseudo-relativo*, a conformation that de-emphasize, from the cognitive point of view, the

---

1   A discussion about the funcions of *ci* with verbs can be found in Sabatini (1985, 1986) and Russi (2008).

structure of a totally new and rhematic phrase. Previous *esserci* and *averci* data were reviewed by Vale (2009) and we'll present them together with those by Panunzi (2010) for comparison purpose.

It's easy to notice that in both attrition corpora the *esserci esistenziale* values don't present significant differences. On the other way, the reference corpora exhibit a large difference in number of occurrences: 24,33 every 10000 words in the BADIP corpus and 78,68 in the Italian C-ORAL-ROM. In the case of the *esserci presentativo*, both the attrition corpora and the reference corpora don't show considerable differences, corroborating the fact that the informative function this form carries doesn't depend on textual variety.

We now have to explain the great difference in the *esserci esistenziale* between the reference corpora. In their studies Raso and Vale suggested, referring to the attrition corpus, that a high presence of this form would mean a lack of lexical variability. We can agree with this theory, but we can also assume that the main reason of a more than triple value of *esserci esistenziale* in the Italian C-ORAL-ROM in comparison to BADIP is due, once again, to the greater diaphasic variety of texts and, most of all, to their spontaneity. To support this hypothesis, table 6 presents the values of total *esserci* found in Panunzi (2010) who analysed the entire corpus of 300000 words. With a more general view it's possible to see that the differences between the two reference corpora continue to be quite noticeable, but smaller than the ones presented previously.

The case of *averci* is quite different. Raso and Vale suggested that future studies would show a smaller degree of attrition of this form, as it became widespread in Italy only after the migration of their informants. Instead of this, our research shows a quite similar level of attrition. As a possible explanation, we can propose that in this case the phenomenon may be due to the subjects of the research being mostly Italian teachers or translator, or individuals otherwise working in an Italian-speaking environment. As their professions require a high degree of proficiency, we can suppose that they tend to practice a higher level of self-control when speaking, especially when it comes to using a form that, while nowadays quite accepted in Italy, they perceive as incorrect or inaccurate.

3.5　The third person accusatives

We will now analyse the attrition of the third person accusatives *lo*, *la*, *li*, *le*, *l'*. As table 3 above shows, normalized data of all corpora don't seem to demonstrate a great difference of values between these clitics, and the degree of attrition seems quite low. But once again we have to split the data to obtain a more complete overview. In table 7 we can observe third person clitics divided by function and dislocation in the phrase.

In both attrition corpora, a first glance to the non-phoric dislocated constituents, informatively neutral, confirms the general opinion: the Raso-Ferrari corpus demonstrate an attrition process, albeit much lower than

the one shown by the Raso-Vale corpus. This could confirm our theory that attrition continues to grow even after the first decade of contact with the L2.

If we look at the phoric dislocated constituents we can see that the situation is much more complicated. In their researches Raso and Vale found that left anaphoric constituents have an increase in values comparing to BADIP, in contrast with the decrease of the total dislocated constituents and, to an even greater extent, of the right dislocated constituents.

CLITICS	Raso-Ferrari Corpus/Italian C-ORAL-ROM	Raso-Vale Corpus/BADIP	Raso-Ferrari Corpus/BADIP	Raso-Vale Corpus/Italian C-ORAL-ROM
Non-phoric dislocated constituents lo, la, li, le, l'	-18,17	-50,98	-32,22	-40,82
Left anaphoric dislocated constituents lo, la, li, le, l'	-21,01	26,02	-15,17	17,34
Right cataphoric constituents lo, la, li, le, l'	-65,85	-53,55	-63,63	-56,39
TOTAL lo, la li, le, l'	-23,81	-45,39	-33,82	-37,14

Table 7: Percentage variation between third person accusatives in a cross analysis of all corpora studied

Our research confirms the decrease of non-phoric dislocated constituents but exhibits a decrease in the left anaphoric constituents and a greater reduction of the right cataphoric constituents.

A cross-analysis of all corpora values can give us an answer about those incongruous results. First of all, it is quite evident that when both attrition corpora are compared with BADIP the results of left anaphoric constituents grow. Once more it seems that we have to investigate the kind of texts every corpus presents and the context of appearance of the object clitic.

In fact, the use of an anaphoric pronoun in Italian in thematized phrases is mandatory, in order to constitute the cognitive semantic bound of an illocution. If the semantic referent is clear to the listener, it's not necessary to constitute this cognitive semantic bound through a thematization, that requests the use of an anaphoric pronoun. To be clear, either in the Italian C-ORAL-ROM or in the Raso-Ferrari corpus, the texts are dialogical and very spontaneous: people know what are they talking about. The Raso-Vale and BADIP corpora, on the other hand, are more formal, with interviews or guided interactions, so people seems to be compelled to thematize the referents they are talking about, hence using the anaphoric pronouns much more.

In the case of the right anaphoric dislocations, it seems that the communicative situation effect plays a much smaller function, and a similar construction isn't

found in Brazilian Portuguese so, as it can be seen, the degree of attrition is higher.

4. Conclusion

This paper presented an L1 attrition corpus-based research. This study had the purpose to delve into this topic deeper than previous ones, building a new corpus with more up-to-date criteria.

As in previous investigations, attrition of Italian L1 in contact with Brazilian Portuguese is confirmed, with a few distinctions that we tried to explain.

The variation in percentage of loss between the two researches seems mostly be due to three reasons:

- differences in typology of texts;
- different diaphasic varieties;
- different pragmatic contexts.

The most relevant divergences can be noticed between the two reference corpora.

The facts above described can explain some seeming incongruous data, like the increase of the number of generic forms like *esserci* in the Italian C-ORAL-ROM or the absence of the *ne* locative clitic in our corpus.

Finally, smaller signs of attrition in our corpus in the case of third person accusative clitics can be a signal that the process doesn't come to a halt after the first decade but continues over time.

We are aware of the fact that the set of data we collected is still too small for a general overview of the L1 attrition discussion, so we hope that this subject and the questions that remain open could be answered by future studies.

5. References

Berretta, M. (1985). I pronomi clitici nell'italiano parlato. In G. Holtus, E. Radke (Eds.), *Gesprochenes Italienish in Geschichte und Gegenwart*, Tubingen: Narr, pp. 185--504.

Berretta, M. (1986). Per uno studio dell'apprendimento dell'italiano in contesto naturale: il caso dei pronomi personali atoni. In A.G. Ramat (Ed.), *L'apprendimento spontaneo di una seconda lingua*. Bologna: Il Mulino, pp. 329--352.

Cresti, E., Moneglia, M. (Eds.). (2005). *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Language*. Amsterdam-Philadelphia: John Benjamin.

De Mauro, T. *et al.* (1993). *Lessico di frequenza dell'italiano parlato.* Milano: EtasLibri.

Ferrari. L.A. (2010). A erosão linguística de italianos cultos em contato com o português brasileiro: asptectos do sistema pronominal. Dissertação (Mestrado) em Linguística – Faculdade de Letras da UFMG.. Belo Horizonte: UFMG..

Köpke, B., Schmid, M.S. (2004). *First Language Attrition. Interdisciplinary perspectives on methodological issues.* Amsterdam/Philadelphia: John Benjamin Publishing Company.

MacWhinney, B. (1994). *The CHILDES project: tools for analysing talk.* Hillsdale: Lawrence Erlbaum.

Panunzi A. (2010). *La variazione semantica del verbo essere nell'italiano parlato: uno studio su corpus.* Firenze: Firenze University Press.

Raso, T. (2009). Erosione dei clitici e strutture tematizzanti in italiani colti in contatto prolungato col portoghese brasiliano. In *Sintassi storica e sincronica dell´italiano. Subordinazione, coordinazione, giustapposizione. Atti del X Congresso della Società Internazionale di Linguistica e Filologia Italiana*, pp. 384--399.

Raso, T., Vale, H.P. (2009). A erosão lingüística em italianos cultos em contato prolongado com o português do Brasil: os clíticos e alguns efeitos na estrutura do enunciado. In *Revista de Italianística*, v. 16, pp. 1—22.

Revista de Italianística (1997). São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas da USP, n. 5, ano V. 291 p.

Russi, C. (2008). *Italian Clitics. An empirical Study*. Berlin New York: Mouton de Gruyter.

Vale, H.P. (2007). A erosão lingüística dos italianos cultos em contato prolongado com o português do Brasil: os clíticos. Belo Horizonte: Monografia apresentada no Curso de Graduação em Letras da Faculdade de Letras da UFMG..

Vale, H.P. (2009). A erosão dos clíticos verificada em um novo corpus: *esserci*, *averci* e *ci* lexicalizante (Oral presentation).

Annotating a corpus of spoken English: the Engineering Lecture Corpus (ELC)

Siân ALSOP, Hilary NESI

Coventry University
Priory Street, Coventry CV1 5FB, UK
alsops@uni.coventry.ac.uk, h.nesi@coventry.ac.uk

Abstract

This paper describes an approach to what we are calling the 'pragmatic' annotation of the Engineering Lecture Corpus (ELC). The ELC contains 70 English-medium engineering lectures from across the world, currently including Malaysia, New Zealand, the United Kingdom and Italy (www.coventry.ac.uk/elc). The lectures are in the form of videos, raw text transcripts and XML files encoded using traditional TEI methods, but also marked for a limited number of features which shed light on the specific nature of lecture discourse. These functions will be discussed in terms of: how the current working list was reached, markup and annotation processes, and possible uses of the complete corpus.

Keywords: lecture; engineering; annotation; corpora; pragmatic.

1.   Concept behind the corpus

Academic staff and students are increasingly moving from country to country to receive and deliver academic lectures. However, although English is often used as a lingua franca in higher education, and although lecture topics and syllabuses for disciplines such as engineering and medicine tend to be similar around the world, it is likely that different cultural norms and expectations will result in different lecture styles and structures in different local academic contexts. This suggests that staff and students may need to adjust the way they deliver and receive lectures in unfamiliar academic contexts, and that they may benefit from corpus linguistic insights when making these adjustments.

The corpus annotation of features other than syntax and part of speech is extremely time-consuming and encumbered by questions of subjectivity (Meyer, 2002; Leech, 2005; Smith, 2008). Some spoken corpora such as the London-Lund Corpus (LLC) (Garside *et al*., 1997) and the spoken component of the HKCSE business corpus (Warren, 2004; Cheng, 2004) have been manually encoded for prosodic features such as tone units, pitch and stress, but very few corpora have been annotated from a functional perspective, because of the labour intensive nature of such work, and because of the degree of interpretation it requires.

A number of small written corpora have been marked up in terms of generic moves and steps (see, for example, Durrant & Mathews-Aydinli, 2011), and classroom interaction in the *Singapore Corpus of Research in Education* (SCoRE) has been marked for pragmatic and pedagogical features (Peréz-Paredes and Alcaraz-Calero, 2009), but as far as academic lectures are concerned, progress with pragmatic mark-up has been very slow. Young (1994) identified a sort of generic move structure in academic lectures, consisting of various 'phases', each with a different communicative function, and Maynard and Leicher (2007) experimentally tagged a small subcorpus of 50 MICASE transcripts by identifying pragmatic features such as 'advice' and 'disagreement' in header metadata, but there does not seem to have been any prior attempt to mark up

an entire corpus of lectures to reflect their structure or purpose.

The largest British lecture corpus, the British Academic Spoken English (BASE) corpus (Nesi, 2001), is only encoded for part of speech, pausing, and contextual information. The BASE corpus annotation follows TEI (Text Encoding Initiative, www.tei-c.org) conventions so that it can be compared with other similarly encoded corpora, but TEI has not traditionally been used to signal the function of larger stretches of discourse, and appropriate coding strategies are still under development.

By annotating what we are calling 'pragmatic' features, we are able to identify and describe features that are typical of the discourse; in this case, engineering lectures. It will also allow us to compare the styles of English-medium engineering lecturers in different parts of the world, and explore what role English-medium instruction currently has in the discipline of engineering.

2.   The corpus

We have annotated six functions of the lecture within a cross-cultural corpus of 70 English-medium university level lectures across five areas of engineering (see Table 1). The ELC currently contains four subcorpora of lectures from: the United Kingdom (UK, four digit id. series: 1…), Malaysia (MS, id. series: 2…), New Zealand (NZ, id. series: 3…), and Italy (IT, id series: 4…).

		MS	NZ	UK	IT
area of engineering	civil	4		27	4
	mechanical	11	8	3	
	electrical		17		
	graphics		3		
	telecoms.				4
total lectures		15	28	30	7
total lecturers		9	4	5	4

Table 1: ELC holdings

3.   Categories annotated

The current set of six pragmatic features was arrived at through a three-stage process. The initial working list

was based on Nesi and Ahmed's (2009) set of 15 features (outlined in Table 2). For the first pass at annotation, the lead annotator, in collaboration with local experts, worked through samples from each of the four subcorpora, cycling between the original working list and the functions that actually occur in the corpus.

Using this data-driven approach to refine the pragmatic categories annotated resulted in the first adjustment to the working list. At this stage, it became clear that some of the functions identified in the original working list (or elements) needed to be expanded to include subcategories (or attributes), and some should be hierarchically demoted and subsumed under a more general umbrella category (see Table 2). Such changes included incorporating: 'review lecture content' and 'preview lecture content' as attributes of the umbrella element 'summary'; 'personal narratives' under 'storytelling', with the addition of the attribute 'professional narratives'; and the six independent types of humour that were originally identified were subsumed under a single unified element, which was expanded to include five more attributes and 'word play'. Two other

elements from the original working list ('reference to students' future profession' and 'greetings') and one partial element ('register' from 'register and wordplay') were not evident in sufficient quantity to justify their inclusion in the adjusted list when considered against the original criteria of identifying and describing typical engineering lecture discourse features.

The second pass at refining the clipboard was undertaken by a single researcher overviewing the entire corpus with the aim of ensuring consistency across all identified features. In this second adjustment, attributes of the 'summary' element were further expanded to identify reviews of previous and current lecture content, and previews of current and future lecture content. Attributes of the storytelling element were replaced; the distinction in genres of anecdote, exemplum, narrative and recount (cf. Plum, 1988; Martin, 2008; also see Alsop *et al*. *forthcoming*) were considered to be more useful than the former limited description of narrative type (as 'personal' or 'professional').

Nesi and Ahmed (2009)	1st adjustment		2nd adjustment	
	element	*attribute*	*element*	*attribute*
prayer	prayer		prayer	
housekeeping	housekeeping		housekeeping	
defining term	defining term		defining term	
review lecture content preview lecture content	summary	review lecture content preview lecture content	summary	review previous lecture content review current lecture content preview current lecture content preview future lecture content
personal narratives	storytelling	personal narrative professional narrative	storytelling	anecdote exemplum narrative recount
teasing self-recovery self-denigration black humour disparagement of out-group member mock threat	humour	bawdy humour black humour disparagement irony jokes mock threats playful humour teasing sarcasm self-denigration word play	humour	bawdy humour black humour disparagement irony jokes mock threats playful humour teasing sarcasm self-denigration word play
register and word play				
greetings				
reference to students' future profession				

Table 2: Refining the clipboard

The ELC is a growing corpus and we are constantly seeking new contributions from around the world. Because the pragmatic categories annotated are largely data-driven, we anticipate that further

adjustments to the working list of functions may be made as the corpus expands and a larger data set becomes available. An early example of the need to encode an unexpected category is that of 'prayer',

which only occurs in the Malaysian subcorpus. Given the highly technical content of large stretches of the language that currently constitute the ELC, we predict that further emphasis may need to be given to the way in which specialized vocabulary is conveyed. 'Defining', for example, could be subsumed under a new 'explaining' umbrella element, and further attributes (for example, 'categorising', 'equating', 'naming' and 'translating') added. Similarly, if storytelling emerges as a more prominent function as the corpus grows, it may be useful to revisit the original significance of describing 'personal' involvement and attribute another layer of annotation to the current categories by specifying whether the instance of storytelling is based on the lecturer's own experience or the experience of others.

4.   Examples of categories annotated

When identifying the boundaries of pragmatic categories, we have worked on the principle of including enough data so that the chunk of text annotated makes sense in isolation from its immediate context. Where boundaries were unclear, the widest scope was incorporated.

Some of the ELC categories are self-explanatory, such as 'prayer', or most usefully clarified by the subcategories attributed to them, such as 'humour' or 'story'. Some require further explanation. 'Housekeeping', in this context, refers to instances where lecturers talk about academic commitments and events external to the lecture. Also, 'defining' refers to the specific explanation of the meanings of technical terms in the ELC.

Given the inevitably somewhat subjective nature of the annotation process, we do not consider rigidly prescriptive definitions of the categories described to be either possible or desirable. Table 3, however, gives some examples from the current corpus.

Element	Attribute	Example of discourse
defining		so mathematically if we define the force the magnitude of the force as F and the angle that defines its direction to the horizontal is theta then simple trigonometry of triangles our horizontal component will be F cosine theta and our vertical component will be F sine theta simple enough (1001)
housekeeping		okay so there will be no class this Thursday and Friday because has been replaced here today (2010) how far have those certificates got well bring what's left down to the front and anybody else who wants their certificate come down to the front (1012)
humour	mock threat	I will open it up again for another two weeks except for the person whose phone's going off cause they're not gonna be able to sit down for about a month (1004)
	teasing	after a good lunch I'm sure you can answer what's the purpose of the horizontal curve (2009)
	irony	now today is a great day because we're going to allow the charge to move we're going to have current so don't get too excited (3005)
	self-denigration	if I would have to machine this I would pull my left hair out my few I have left (3019)
story	narrative	I hate to admit to this one but one site I was on we had cube failures and the reason was that when I'd been sending the cubes off I'd been having to break the ice on the top of the tank before I could get them out and um the tank had a heater in we just hadn't bothered to get the spark to wire it in and ah fairly obviously by the time the area manager appeared to ah come and have a look and see what had gone wrong it was all wired in and working fine and we said oh no no problem with that would we do a thing like that and ah but okay sort of nevertheless it caused endless hassle the fact that we'd had these cube failures if you keep them too cold they'll go down a low strength (1012)
summary	review previous lecture content	let's just review back what we did yesterday we talked about the refrigerator yeah we talked about the refrigerator and you were introduced to refrigerators and the heat pump (2017)
	preview previous lecture content	so what are we going to do today is we are going to wrap up chapter five the second law of thermodynamics yeah so today we should be able to determine finally the thermo efficiencies and the coefficient of performance for our ideal our reversible or our Carnot cycle (2019)
	review current lecture content	main three things that have come out of here though out of these tests is yield stress ultimate stress and modulus of elasticity (3026)
	preview future lecture content	in the next two lectures we're actually going to delve a little bit into material properties and then we're going to get back into the solid mechanics (3024)

Table 3: Examples of ELC pragmatic categories

5.   Markup and annotation processes

The ELC files have been created by merging two separately stored sets of information: the main body of raw transcribed lecture discourse and the header metadata. The spoken lecture content, varying in duration between 41-104 minutes, was videoed then transcribed as plain text by a local expert[1]. TEI compliant header information - such as title, recording equipment, main speaker information, etc. - was generated from a master spreadsheet and outputted in XML format to create a skeletal file, including empty 'body' tags. The transcribed plain body text was then merged into the body tags, and TEI compliant markup - including container elements to mark utterances according to speaker identifier, empty elements to mark pauses, gaps (for example, marking inaudible speech), and limited kinesic and vocal descriptions that were essential to context (for example, 'writes on board' and 'laughter') – was manually added.

We have distinguished this type of 'structural' *markup* from the *annotation* (c.f. Garretson, 2011) of pragmatic categories because the process by which the boundaries of the pragmatic categories are identified involves a subjective linguistic analysis. We think, therefore, that it should not be described in the same way as the identification of the objective structural components of a text, such as utterances.

In terms of the storage of these annotations, the boundaries of pragmatic categories were initially annotated inline alongside the structural markup. This posed a problem of validity for the XML metadata because the language of lectures often serves more than one function; a story, for example, may also be humorous, in full or in part, causing XML elements to overlap. Similarly, pragmatic categories can span various utterances – a lecturer delivering housekeeping information may be interrupted by a student asking a question, for example – which also results in malformed XML syntax. In addition to the methodological questions linked to storing annotation inline alongside markup, we did not consider using a system of workarounds to force the annotation into a well-formed state to be a desirable option.

Instead, we have decided to convert our current inline annotations into stand-off form and store them in separate XML files. The advantages of this system are that the subjective analysis is stored separately and multiple other layers of annotation can be done on the same text. In addition to the current pragmatic annotation, detailed kinesic or prosodic analyses could be

applied, for example. One consideration that may be seen as a disadvantage, particularly in a corpus of spoken language, is that the raw text must be static in order that the indices of the annotations in the stand-off files are correct. This means that the original transcripts must be completely accurate before stand-off files can be created, and the transcripts cannot be edited post-annotation.

We intend to use the Dexter suite of software (http://www.dextercoder.org/index.html) for further coding and analysis once the current inline annotation has been converted into stand-off form. To achieve the conversion, the current annotation (but not the TEI-compliant structural mark-up) will be stripped out and an XSLT stylesheet will be used to convert these 'pure' versions of the marked up texts into XML files that are readable by stand-off annotation software (in this case, DexML). Next in the conversion chain, a code file will be created by looping through the original text and, for each inline annotation found, locating the exact stretch of text, and then identifying the indices for that stretch of text and creating a code instance for it in the code file. The result will be one file containing the 'pure' text and a code file. The codes that used to be inline annotations will then be in the form of editable stand-off annotation.

6.   Possible uses of the corpus when complete

This data-driven process of pragmatic annotation will, we hope, eventually lead to the identification of linguistic features that typically realise the various purposes of lecture discourse. By encoding and then visualising these features we will be able to compare their location, duration and relative frequency in lectures delivered by local lecturers in different cultural contexts.

Looking at such data patterns allows one of two potential conclusions to be drawn. If significant consistency is identified in the way in which the annotated functions of language occur and are used across the subcorpora, we can conclude that key language functions are fundamental to the English-medium engineering lecture regardless of cultural context. We can then begin to build a model of the fundamental purposes of these lectures. If, on the other hand, significant variation in the uses of language functions is identified, we can begin to examine the role played by cultural difference in the delivery of the English-medium engineering lecture, regardless of consistency of language medium (English), discipline (engineering), and education level (undergraduate).

Our annotation system will be of interest to other corpus developers who intend to apply

---

[1] Further information on transcribing conventions can be found here: <www.coventry.ac.uk/elc>.

pragmatic mark-up, and our comparative findings will be of interest to EAP and ESP practitioners, staff developers, and all academics and students on the move.

7.   References

Alsop, S., Nesi, H., Moreton, E. (*forthcoming* 2013). The uses of storytelling in university Engineering lectures. In *ESP Across Cultures*, 10.

Cheng, W. (2004). //→ did you TOOK //↗ from the miniBAR// What is the practical relevance of a corpus-driven lanugage study to practitioners in Hong Kong's hotel industry? In U. Connor, T.A Upton (Eds.), *Discourse in the Professions*. Amsterdam: John Benjamins, pp. 141--166.

Durrant, P., Mathews-Aydınlı, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes*, 30(1), pp. 58--72.

Garside, R., Rayson, P. (1997). Higher-Level Annotation Tools. In R. Garside, G. Leech and A. McEnery (Eds.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London: Longman, pp. 179--193.

Garretson, G. (2011). Dexter: An introductory workshop. *BAAL Corpus Linguistics SIG*, Coventry University, December 9.

Leech, G. (2005). Adding Linguistic Annotation. In Wynne, M. (Ed.) *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, pp. 17--29.

Martin, J. R. (2008). Negotiating Values: Narrative and Exposition. *Bioethical Inquiry*, 5, pp. 41--55.

Maynard, C., Leicher, S. (2007). Pragmatic Annotation of an Academic Spoken Corpus for Pedagogical Purposes. In Fitzpatrick, E. (Ed.) *Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse*. Amsterdam: Rodophi, pp. 107--116.

Meyer, C. (2002). *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Nesi, H. (2001). A corpus based analysis of academic lectures across disciplines. In J. Cotterill and A. Ife (Eds.) *Language Across Boundaries*. London: Continuum Press, pp. 201--218.

Nesi, H., Ahmad, U. (2009). Pragmatic annotation in an international corpus of engineering lectures. *Conference of the American Association for Corpus Linguistics*, University of Alberta, October 10.

Peréz-Paredes, P., Alcaraz-Calero, J. M. (2009). Developing annotation solutions for online data driven learning. *ReCALL*, 21(1), pp. 55--75.

Plum, G.A. (1988). Text and contextual conditioning in spoken English: a genre-based approach. PhD Thesis. University of Sydney.

Smith, N., Hoffmann, S., Rayson, P. (2008). Corpus Tools and Methods, Today and Tomorrow: Incorporating Linguists' Manual Annotations. *Literary and Linguistic Computing*, 23(2), pp. 163--180.

Warren, M. (2004). //↓so what have YOU been WORKing on Recently //: Compiling a specialised corpus of spoken business English. In U. Connor and T. Upton (Eds.) *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins, pp. 115--140.

Young, L. (1994). University Lectures − Macro-Structure and Micro-Features. In J. Flowerdew (Ed.) *Academic Listening*. Cambridge: Cambridge University Press, pp. 159--176.

A multilingual speech corpus of North-Germanic languages

Janne Bondi JOHANNESSEN, Joel PRIESTLEY, Kristin HAGEN

The Text Laboratory, Department of Linguistics and Nordic Studies, University of Oslo

P.O. Box 1102 Blindern, 0317 Oslo, Norway

jannebj@iln.uio.no, joeljp@gmail.com, kristiha@iln.uio.no

Abstract

The Nordic Dialect Corpus project was initiated by the Scandinavian Dialect Syntax Network (ScanDiaSyn). In order to be able to study the North Germanic (i.e., Nordic) dialects, proper documentation of the dialect was needed. A corpus consisting of natural speech by dialect speakers was designed, in order to systematically map and study syntactic variations across the Scandinavian dialect continuum. The corpus was to be comprised of transcribed and tagged speech material linked to audio and video recordings. Further, it was decided that a user-friendly interface should be developed for the corpus, and that it should be available on-line. The corpus is now ready for use, and is described here.

Keywords: North Germanic languages; speech corpus; dialects; transcription; tagging; maps.

1. Introduction

The Nordic Dialect Corpus project was initiated by the Scandinavian Dialect Syntax Network (ScanDiaSyn). Documentation of the dialects was required, and it was decided that a corpus of natural, spontaneous speech was needed in order to systematically map and study syntactic variations across the Scandinavian dialect continuum. The corpus was to be comprised of transcribed and tagged speech material linked to audio and video recordings. Further, it was decided that a user-friendly interface should be developed for the corpus, and that it should be available on-line. The corpus is now ready for use and described in this paper.

The ScanDiaSyn network is a project umbrella where ten Scandinavian research groups collaborate.

Figure 1: The countries involved in the ScanDiaSyn project

The ten groups are spread across all of the five Nordic countries and one self-governed area. Three non-Nordic groups and a group working on Finnish dialect syntax liaise with the project through a NordForsk network. The core groups are from universities in Denmark, Faroe Islands, Finland, Iceland, Norway and Sweden.

Country	Informants	Places	Words
Denmark	81	15	242,885
Faroe Is.	20	5	62,411
Iceland	10	2	23,610
Norway	508	143	2,014,637
Sweden	126	39	307,861
Total	745	204	2,651,404

Table 1: The Nordic Dialect Corpus in numbers

The corpus is now installed in the Glossa corpus system for user-friendly search and results handling (Johannessen *et al.*, 2008; Johannessen, 2012).

There are a number of challenges that have had to be addressed, that we shall focus on in this paper:

data collection should be carried out in several different countries

- the recordings should be transcribed, with different transcription standards and types for the individual languages;
- the corpus, consisting of different languages should be tagged;
- different tags should refer to the same entities for uniform search possibilities;
- informant metadata (gender, age, sex etc.) should be used as filters for search;
- different geographical divisions should be specifiable (e.g. country, county, town);
- all text from all languages should be accessible in same search;
- transcriptions should be linked to audio and video;
- results should be available in a number of different ways, including different export formats;
- informant data should be plotted on map.

2.   Methodology for collecting speech

The corpus comprises recordings made in the five constituent countries of the Northern Germanic language area. From each country a number of sample points were selected specifically to capture dialectic variations.

There is some variation as to the combination of speakers in the corpus, given that the recordings were mostly done on national research funding and national research management.

In Norway, the Norwegian Dialect Syntax Project was funded by the Norwegian Research Council, a savings bank in North Norway and the University of Oslo. This ensured full funding of the recordings in a way that satisfied the criteria given by the researchers. From each point, four informants were identified, two men and two women, old and young. The informants were paired and asked to converse freely for approximately 30 minutes. Care was taken to create comfortable, informal surroundings, in order to encourage spontaneous, unaffected speech. Video equipment was set up, but the informants were left to themselves. Due to privacy legislation, a list of topics deemed off-limits was provided. This included subjects such as trade union and political party membership, as well as the naming of third parties, with the exception of public figures. Each informant also partook in a more formal interview, answering a standard set of questions. The Norwegian part also includes a number of old recordings from 1950–1980, provided by the Målførearkiv at the University of Oslo, and funded by the Norwegian Dictionary 2014 project.

The majority of the Swedish recordings (including Finland Swedish) were generously provided for use in the Nordic Dialect Corpus by the SWEDIA 2000 project. This project was originally aimed at collecting data for phonological research, but the data are mostly fully usable for our corpus, since this corpus, too, contains free speech. The Danish recordings were done by the Danish Syntax Project funded by the Danish Research Council, and contains six recordings from each place, but with no young people. The Faroese recordings were done on the ScanDiaSyn network budget (funded by the Nordic Research Council) and contains both young and old speakers. For Icelandic, the recordings have been less systematic, given a combination of funding and chronological synchronisation with the rest of the project. Some recordings have been generously provided by the University of Iceland, and some have been done by the network, using somewhat imperfect informants (linguists).

In spite of the diverse ways the recordings have been collected, the corpus is a unique source of spontaneous speech well suited for dialect research in syntax, but also for other linguistic disciplines.

3.   Transcription and tagging

All recordings have been transcribed with standard orthography. In addition, all the Norwegian recordings and some of the Swedish ones (those of the Övdalian dialect) have been transcribed in a more phonetic way, following (for Norwegian) the method described in Papazian and Helleland (2005) and (for Övdalian) the orthography standardised by the Övdalian language council Råðdjärum.

For each language, transcription software was used that inserts time codes directly into the transcribed text at suitable intervals, enabling the transcription to be presented with its corresponding audio and video. The transcriptions were partly done at a national level, and party in Oslo. Different software were used, but they were all adapted to the Transcriber format, which is the interchange format used in the project.

For the Norwegian and Swedish recordings that have also been phonetically transcribed, the process started with the phonetic transcription. These transcriptions were then translated to standard orthography using a program developed at the Text Laboratory, University of Oslo: an automatic dialect transliterator. The program takes as input a phonetic text and an optional dialect setting. Sets of text manually transliterated to orthography provide a good basis for training the program, enabling it to accurately guess the transliteration for further texts. The training process can be repeated, and the trained version can be used for similar dialects. Transcribing each recording twice therefore does not take as much as twice the time.

It is important that all words from the original phonetic transcription have an equivalent in the orthographic transcription. The two must be totally aligned for the results to be used in the corpus search system. Figures 3–5 below show how the phonetic transcription can be used in search and results presentation.

The languages are tagged individually with taggers for the respective languages. This means that each language has an individual tag-set decided by those who developed the taggers originally. The Danish transcriptions are lemmatised and POS tagged by a Danish Constraint Grammar Tagger developed for written Danish, see Bick (2003). The Faroese transcriptions first were tagged with a Constraint Grammar Tagger for written Faroese, see Trosterud (2009). Since spoken Faroese has a lot of words that are not approved in written standard Faroese, about half of the material was manually corrected after the Constraint Grammar tagging. Finally a TreeTagger was trained on the corrected material, and the rest of the transcriptions were tagged again. The Icelandic transcriptions were first tagged with a tagger for written Icelandic, see Loftsson (2008), and manually corrected afterwards. The orthographic version of the Norwegian corpus was lemmatised and POS tagged by a TreeTagger originally developed for Oslo speech. The Oslo speech tagger was trained on manually corrected output from the the written language Oslo-Bergen tagger, see Nøklestad and Søfteland (2008). The Oslo speech tagger was then further adapted to the dialect corpus. The Swedish subcorpus was tagged by a modified version of the TnT tagger developed by Kokkinakis (2003). The tagger was trained on the Swedish PAROLE corpus and manually

tagged orthographic Övdalian transcriptions. The tagger was applied to both the Swedish transcriptions and the orthographic versions of the Övdalian transcriptions.
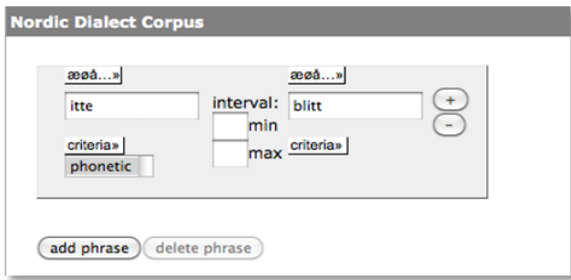
Figure 2: Searching for two words in sequence. The first is transcribed phonetically: *itte* for the orthographic word *ikke* 'not'
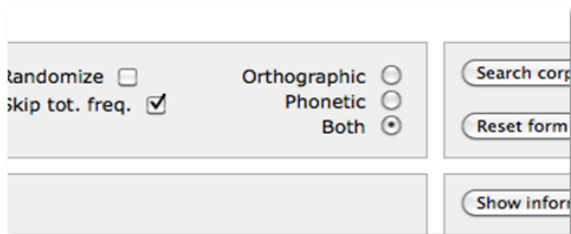
Figure 3: The *Both* button is ticked, in order to have both kinds of transcription presented in the search results
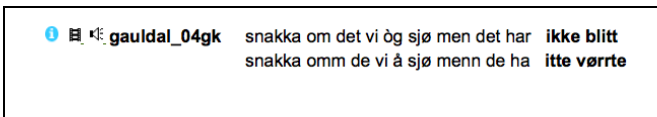
Figure 4: Part of the search result for the query in Figures 2 and 3
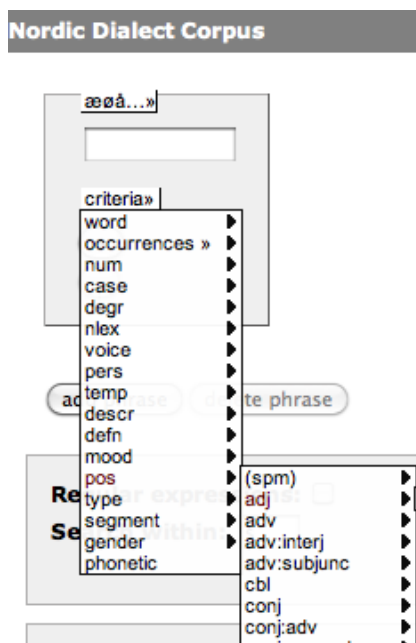
Figure 5: Querying for adjectives in the corpus

Each language subcorpus has its own tag-set, but the tags have been standardised in the search system, making it possible to search for the same category across all the corpora. The linguist can choose for example all adjectives to be shown, irrespecsive of language. This is illustrated in Figure 5.

4.   Metadata

The corpus has metadata relating to each informant and recording. There is information on the sex, age group, and place of origin; the latter being divided into country, region, area and place. Also, there is information on the year of recording, which is crucial for the Norwegian subcorpus, which contains both modern and old recordings, with 30–60 years between them. Finally, some recordings are distinguished according to genre: either interview or conversation.

The metadata can be used to create search filters for search in the corpus interface, as depicted in Figure 6.
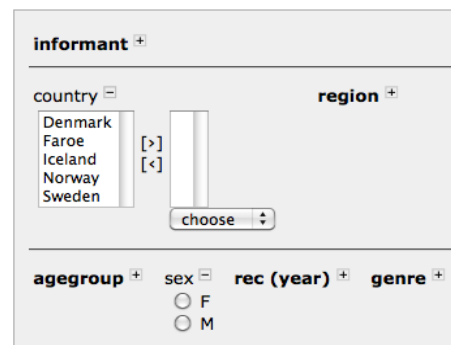
Figure 6: Metadata filter in corpus interface

The metadata is simply represented in a MySQL database, from which the corpus interface system Glossa picks the correct data according to the user's needs.
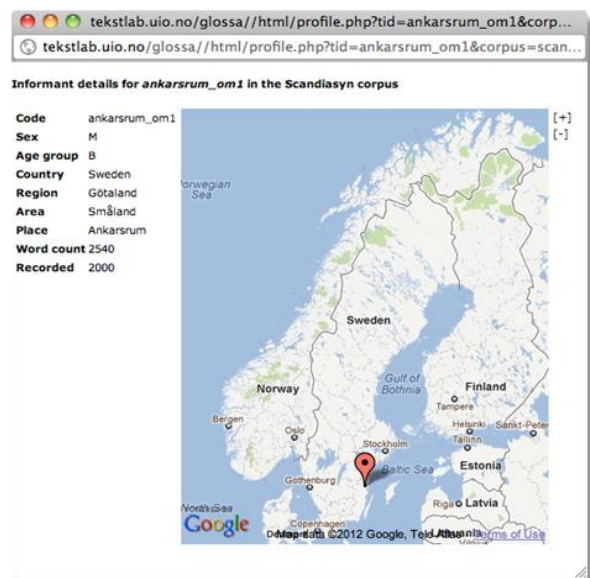
Figure 7: Metadata on each informant is available via a clickable button

Informant metadata can alternatively be found by clicking on the i-button (*i* for information) on the left of each concordance line in the results view, as in Figure 4, yielding the information displayed in Figure 7.

5.   Multilingual search

Users in the ScnDiaSyn network originally wanted the possibility for multilingual search. They imagined that if they wanted, say, all occurrences of the negation equivalent to 'not' in English, a full results list would appear for all languages. However, this would have required a full multilingual dictionary, which does not exist either in paper or digital format for the North Germanic languages.

Instead, we put a link on the search interface to a multilingual word-list (*Tvärslå*) compiled by several previous language technology projects, including ScanLex in which the first author of the present paper was also in charge. This way the user can look up the equivalents of particular words in the other languages. The multilingual list is far from comprehensive, and also contains wrong language equivalents, since it is partly developed using automatic methods.

The search system Glossa allows for disjunctive searches, making it possible for several strings to be looked up at the same time. This is illustrated in Figure 8, for the orthographic versions of 'not' for Faroese, *ikki*, Swedish, *inte*, Danish and Norwegian, *ikke*, and Icelandic, *ekki*.

Figure 8: Disjunctive search for the word for 'not' in several languages

6.   Links to audio and video

The user can click on the film or sound symbol to get the desired multimedia display. Figure 9 depicts the display.
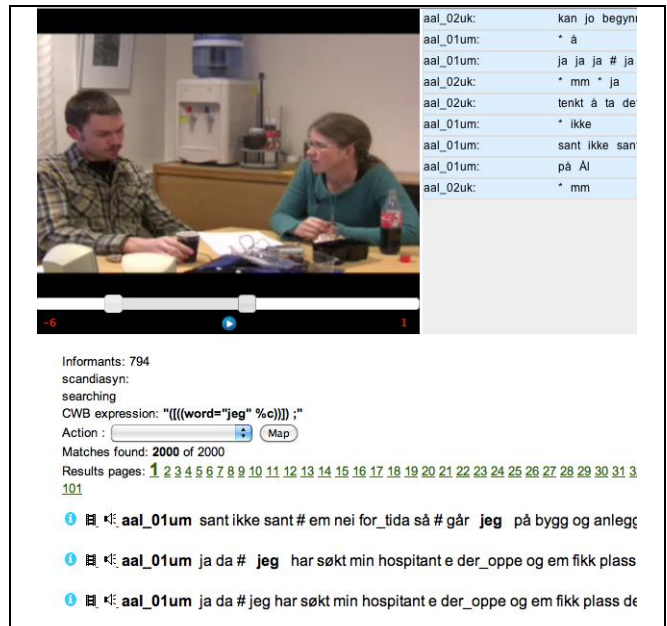
Figure 9: Results with selected video presentation

The transcriptions have time codes, implemented as XML tags, at regular intervals, inserted at the time of transcription. This way there is a direct link between text and audio and video files, to be used by the corpus search system. These files are made available in Flash or Quicktime (according to the user's choice).

7.   Results presented on maps

For a corpus aimed at dialect research, getting results on a map view is very useful. The place of origin for each informant is located by GIS coordinates and the Google Maps API is used. Since every item in the corpus is connected to an informant, it means that for each word, string, piece of word or syntactic construction, there is a geographical location.

We have incorporated two ways of displaying results via maps. One way is that all hits are simply marked on the map. Figure 10 shows a search that asks for all hits where in a subordinate clause the negation *ikke* or *inte* (Norwegian, Danish, Swedish) precedes the subject. The geographical distribution is shown in Figure 11 below.
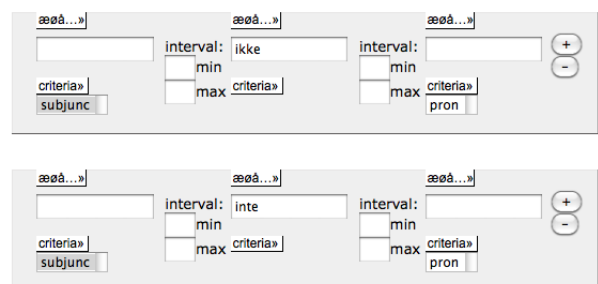
Figure 10: A search for subjunction+negation+pronoun

Figure 11: Results for the search for subjunction+negation+pronoun in Figure 10
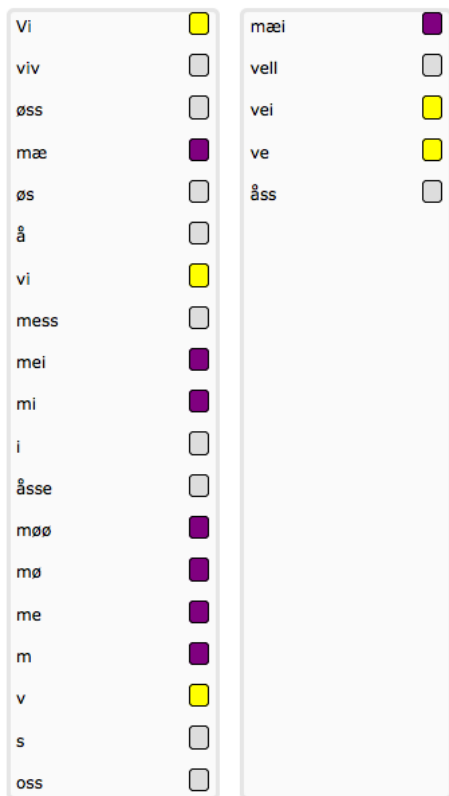
Figure 12: Chart for colouring in the phonetic variants of the pronoun *vi* 'we' in Norwegian

It has been debated in the literature whether this word order is allowed (see Johannessen & Garbacz, 2011). The red dots on the map in Figure 11 show where the hits are. Even if there are more recording places in Norway than in Sweden and Denmark, cf. Table 1, we see immediately that there are many more places where this

construction is found in Norway than in especially Sweden. Since stress patterns also interfere with the generalisations, it is necessary of the user to listen to selected results, but the first picture given by the map is a very useful start.

The other way to use maps is only possible for those search results that belong to a set of two transcriptions. All the phonetic varieties are presented on a chart with the option of colouring each according to any classification one might be interested in.

In Figure 12 a chart can be seen of all the phonetic versions of the word *vi* 'we' in Norwegian. We have chosen to colour those variants that are pronounced with an initial bilabial /m/ sound with a deep violet colour, while the initial /v/ sounds are coloured yellow. The result is shown in Figure 13.

It should be quite clear from the map example that the opportunity of using a corpus combined with maps is an excellent way of finding isoglosses. The geographical limits for a phenomenon are readily apparent on the map. It should be noted also that dialect maps are not a new thing. However, in the past, researchers rarely had the chance to cover many places, so the present corpus may contain data that has never been known before. Secondly, the old maps were rarely the result of spontaneous speech, but rather of words and lists given by the researcher to the informants. The present solution, with a corpus of spontaneous speech as a direct basis for maps, gives good opportunities for both a comprehensive and a correct view of the geographical language variation.

Figure 13: Map of two phonetic variants of the pronoun *vi* 'we' in Norwegian: /m/ variants are coloured violet, while /v/ variants are coloured yellow

8.   Conclusion

We have presented the Nordic Dialect Corpus. We have shown how challenges posed by researchers in this project initiated by linguists have been met. The corpus contains recorded speech from five different languages. provides access to audio and video, as well as transcriptions – many of which are both phonetic and orthographic. All transcriptions are tagged. Everything is accessible in the Glossa search system, with monolingual or multilingual search options, specified linguistically with additional possible metadata. There are different options for results handling that we have not focused on here. However, we have shown how the map options work, and how this way of combining a corpus with a map solution provides advanced possibilities for identifying and representing isoglosses in a simple way.

9.   Acknowledgements

10.   References

Bick, E. (2003). PaNoLa - The Danish Connection, In H. Holmboe (Ed.) *Nordic Language Technology, Årbog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004 (Yearbook 2002).* Copenhagen: Museum Tusculanum, pp. 75--88.

Glossa. Available at:
<www.hf.uio.no/tekstlab/glossa.html>.

Kokkinakis, S.J. (2003). En studie över påverkande faktorer i ordklasstaggning. Baserad på taggning av svensk text med EPOS. Göteborg University.

Johannessen, J.B. (2012). The Corpus Search and Results Handling System Glossa – a Description. To appear in *Chung-hwa Buddhist Journal*.

Johannessen, J.B., Garbacz, P. (2011). Fältarbete med Nordic Dialect Corpus. In *Acta Academiae Regiae Gustavi Adolphi 116,* pp. 169--176.

Johannessen, J.B., Nygaard, L., Priestley, J. and Nøklestad, A. (2008). Glossa: a Multilingual, Multimodal, Configurable User Interface. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08).* Paris: European Language Resources Association (ELRA).

Loftsson, H. (2008). Tagging Icelandic text: A linguistic rule-based approach. In *Nordic Journal of Linguistics 31.1.*

Nordic Dialect Corpus. Available at:
<http://tekstlab.uio.no/glossa/html/?corpus=scandiasyn>.

Nøklestad, A., Søfteland, Å. (2007). Tagging a Norwegian Speech Corpus. *NODALIDA 2007 Conference Proceedings. NEALT Proceedings Series.*

Papazian, E., Helleland, B. (2005). *Norsk talemål.* Kristiansand: Høyskoleforlaget.

Text Laboratory. Available at:
<http://www.hf.uio.no/tekstlab/>.

Trosterud, T. (2009). A constraint grammar for Faroese. In *NODALIDA 2007 Conference Proceedings. NEALT Proceedings Series.*

Tvärslå. Available at:
<http://ordbok.nada.kth.se:8070/ordbog_module/ordbog>.

Formation and annotation of North AMPER project's *corpus*

Regina CRUZ[1,2]**, Ilma SANTO, Camila BRITO**[1,2,6]**, Rosinele LEMOS**[4]**, Isabel DOS REMÉDIOS**[1,5]**, João FREITAS**[5]**, Elizeth GUIMARÃES**[5]**, Lurdes MOUTINHO**[3]

[1]UFPA; [2]CNPq; [3]Universidade de Aveiro; [4]SEDUC; [5]Master student; [6]Scholarship PIBIC

Av. Augusto Correa, s/n – Campus do Guamá – Belém (PA) – 66075-900

regina@ufpa.br, ilmasanto@hotmail.com, camila.brito@ilc.ufpa.br, rosinele_lemos@terra.com.br, isabelrochar@yahoo.com.br, freitas606@hotmail.com, elizeth@ufpa.br, lmoutinho@ua.pt

Abstract

The study of the linguistic diversity on Pará state has as its aim to understand the main factors which cause the linguistic diversity in this region and the importance of these factors on the verbal manifestations of the people who speak the language in use: the Amazon's Portuguese variety. On this paper we present how the formed corpora for the study of prosodic features of Brazilian Portuguese (PB) linguistic varieties spoken in Amazon are being organized, processed and annotated. The Prosodic Multimedia Atlas of Northern Brazil aims to verify the prosodic variations of Amazon PB to provide a sociolinguistic configuration of prosodic level of Pará state. So far the formed corpora are from the following cities: Belém, Bragança, Baião and Cametá. There are also three corpora in progress: Abaetetuba, Belém islands and Marajó island, all of which were formed according to the guidelines of AMPER project, following strictly its methodology, from the selection of informants to the protocol of the data collection.

Keywords: AMPER project; prosodic variations; Amazon; Brazilian Portuguese.

1.   Introdução

This paper aims mainly to present how the formed corpora are being organized, processed and annotated for the study of the prosodic characteristics of the linguistic varieties of Brazilian Portuguese (PB) spoken in Amazon. This study is closely linked to the AMPER[1] project, whose aim is to supply the acoustic and prosodic characterization of the Romance languages, as well as an *online* multimedia atlas (Contini *et al*., 2002: 227-230; Moutinho *et al*., 2001: 245-252). In relation to the Portuguese system, eleven institutions participate for the description of its three main varieties: European Portuguese, insular European Portuguese and Brazilian Portuguese (PB).

UFPA has already been participating of this Project since 2007, responsible for the Multimedia Prosodic Atlas of Northern Brazil. Currently, four atlas are in progress: a) Belém (Brito, in progress; Guimarães, in progress); b) Abaetetuba (Remédios, in progress); c) Marajó (Freitas, in progress), Baião (Lemos, in progress). This project has one atlas already finished. It belongs to Cametá (Santo, 2011).

2.   The AMPER-North project

Since its entry in the AMPER project, UFPA's team has already formed *corpora* of spoken Portuguese in the following places: a) Belém (Santos Jr., 2008; Cruz *et al*., 2008; Cruz & Brito, 2011); b) Bragança (Castilho, 2009); c) Baião (Lemos, in progress); and d) Cametá (Santo, 2011; Santo & Cruz, 2011). The formation of these corpora was made according to the AMPER Project guidelines, following its methodology, since the selection of informants until the protocol of the data collection. A detailed description of these methodological procedures is shown on item 3.

As this project has as an aim to form a *Prosodic Multimedia Atlas of Northern PB*, other three corpora are still in prediction of formation: a) of Abaetetuba's city (Remédios, in progress); of Belém's isles (Guimarães, in progress; Brito, in progress) of Marajó's isle (Freitas, in progress). We have also a prediction to the formation of *corpora* from the cities of Mocajuba, Óbidos, Santarém and Breves.

On the map below there is the localization of all the inquest points that are covered for this project in Para State nowadays.
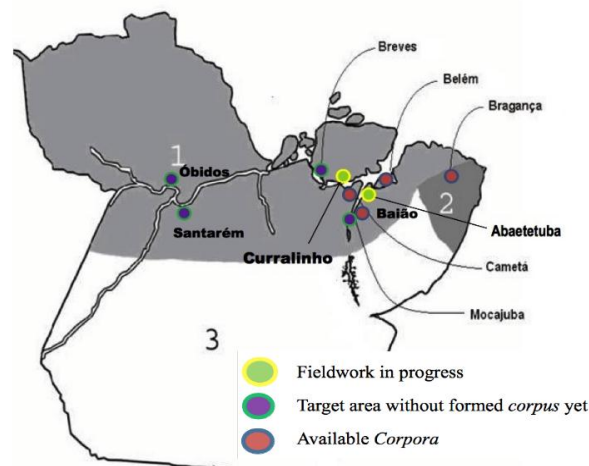
Figure 1: Map 01 – The localities attained for The AMPER-North Project. It was adapted from Cruz (2012: 205)

Cassique (2006 *apud* Cruz, 2012) presents a new dialectal division of Pará State from Silva Neto (1957) that has been considered by UFPA's researchers linked to AMPER-POR Project, so it has been used as base to the choice of the target localities of its project.

According to this dialectal division, the selected localities for this project's investigation belong to the

---

[1] http://pfonetica.web.ua.pt/AMPER-POR.htm

regional PB of Pará State (cf. Zone 1 of map 1). Bragança is the only one that belongs to another dialect called bragantino (cf. Zone 2 of map 1).

The PB spoken in Pará State is considered by Silva Neto (1957) how being of *canua cheia de cucus de pupa a prua*. It has the main dialectal mark that is a rising on the back vowels on the stressed syllable (Rodrigues, 2005).

For this reason, the **Prosodic Multimedia Atlas of Northern PB** will register exactly the prosodic variations of the PB spoken in Pará State and supply a sociolinguistic configuration on the prosodic level of this variety of PB.

On the first version of this project it was possible to move forward on the formation of corpora. Unfortunately, it was not possible to explore them yet, except for these two corpora that have been explored: Cametá city (Santo & Cruz, 2011; Santo, 2011) and Belém city (Cruz & Brito, 2011).

3. Methodological procedures that are adopted on the formation of the corpora

On this study they were adopted all the methodological procedures determined by the general coordination of the AMPER Project.

As one of the goals of the AMPER Project comprises a contrastive analysis of the studied dialects, the *corpus* was recorded for the varieties of Brazilian Portuguese. It is made up of six replicates of sixty-six sentences of the corpus-based AMPER project for the Portuguese language. Each constituent of the phrases have a corresponding image, since it is not allowed any contact of the speakers with the written sentences. Therefore, during the fieldwork, there is the visual representation of the sentences which means that slides are shown to informants as a way of graphic stimuli for the production of 396 sentences to be generated. The set of sentences that form the corpus of the project AMPER follows previously established phonetic and syntactic criteria.

Since the vowels have the most relevant information regarding the prosodic curve and taking into account the characteristics of the stress structure of the Portuguese, there have been chosen words that represent the different stress structures (oxytone[2], paroxytone[3] and proparoxytone[4]) in various positions on the sentence[5].

The sentences were syntactically set up, so as for the present Subject – Verb – Complement (SVC). In relation to the intonation, they were designed to accommodate the neuter modes, affirmatives, declaratives and global interrogatives. Therefore, the sentences utilized on the recordings are of the type of SVC and its extensions to include prepositional phrases. As for the syntactic structure, all sentences have only: 1) three characters (Renato, pássaro and bisavô); 2) three adjectival phrases (nadador, bêbado and pateta); 3) three prepositional phrases of place ('de Mônaco', 'de Veneza' and 'de Salvador'); 4) a single verb (gostar).

During the collection of the data, six repetitions of a set of phrases are asked to each speaker of a set of phrases in the *corpus* (in random order), being selected for acoustic analysis the top three repetitions, in order to establish the meaning of three acoustic parameters: duration, fundamental frequency (F0) and intensity.

As it was determined by the general project, for the selection of informants were taken into consideration the following criteria: 1) age (above 30 years old); 2) schoolar level (elementary school, high-school and college); and 3) residence time in the town (only local indigenous people). Based on these criteria, six informants were selected, three males and three females, who participated on the data collection. It is, therefore, a stratified sample. Each informant has received a code which contains information on his/her profile. On the Table 1 below, we can see the codification adopted by AMPER-North project.

Portuguese Dialect	Brazilian Region	City		School Level		Sex
		Code	description	Code	description	
B (Brazil)	E (North)	0	Belém	1	The Lowest school level	female
		1	Cotijuba – Belém's island	2		male
		2	Abaetetuba	3	high-school	female
		3	Bragança	4		male
		4	Curralinho	5	College	female
		5	Cametá – urban region	6		male
	F (North)	4	Mocajuba – rural region			
		5	Mocajuba – urban region			
		6	Cametá – rural region			
		7	Outeiro – Belém's island			
		8	Mosqueiro – Belém island			
		9	Baião			

Table 1: Codification of the speakers adopted by AMPER-North Project

In total six sound files were obtained by investigating the localities. The rate of sample of the sign is 44.100 Hz, 16 bits, mono. All the data collection was made in the informant's own house.

4. Characterization of the corpora

Therefore the AMPER-North Project's *corpus* is composed of **198** sentences, in total **1.188** sentences by informant, which contains samples of the linguistic

---

[2] The oxytone words used are: 'o bisavô', 'de Salvador', 'nadador'.
[3] The paroxytone words used are: 'o Renato', 'de Veneza', 'pateta'.
[4] The proparoxytone word used are: 'o pássaro', 'de Mônaco', 'bêbado'.
[5] The syntactic positions considered on the assembly of the corpus AMPER sentences are noun phrases and prepositional phrase.

varieties that are spoken in Belém, Cametá, Baião and Bragança. Below Table 2 contains the size in hours of recording of each formed corpus.

City	Code	Duration corpus size	Source
Belém	BE0	3h 36 min. 25 sec.	Brito (in progress)
Baião	BF9	6h 33 min. 15 sec.	Lemos (in progress)
Bragança	BE3	2h 21 min. 21 sec.	Castilho (2009)
Cametá	BE5	3h 16 min. 39 sec.	Santo (2011)

Table 2: Total size of formed *corpora* of the AMPER - North Project in hours of recording

The Project in itself organizes the formed *corpora*, but the availability *on line* of the *corpus* is responsibility of the   general coordination of AMPER-POR Project.

This project already supplied the list above of the varieties of Belém (BE0) and Cametá (BE5) to the general coordination and, therefore, the data on this list are already in the site of AMPER-POR Project[6].

5.   Tendencies of spoken Portuguese in the northern of Brazil: preliminary analysis

Until the present day, the obtained results refer to the physical parameters - intensity, duration and F0 – in relation to the kind of Portuguese stress and to the syntatic aspects controled by AMPER Project on the construction of its *corpus*. At the moment, two Brazilian Portuguese varieties spoken in Amazon were analised: Cametá and Belém.

The preliminary analysis made with the data (Santo & Cruz, 2011; Cruz & Brito, 2011) indicated that, in general, the F0, duration and intensity measures complement one another to establish a distinction between statement and interrogative in Brazilian Portuguese spoken in Cametá (PA).
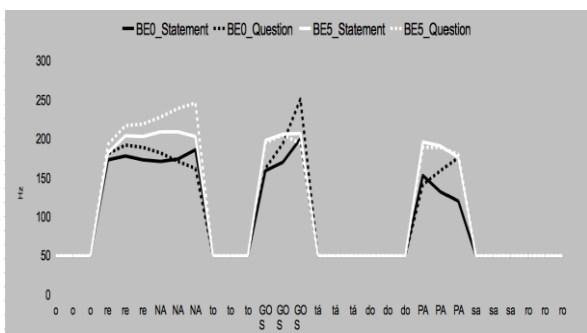
Figure 2: Comparison between the F0 variation meaning of the sentence twp – O **Renato** gosta do **pássaro** - on both the modalities – declarative (full line) and interrogative (dashed line), spoken by a female speaker with low educational level from Belém - BE0

(black) – and another speaker of the same social profile from Cametá dialect – BE5 (white)

We can equally to state that the important variations of the three controlled acoustic parameters, that establishes the difference between the two modalities, occur preferentially on the stressed syllable of the nuclear element of the phrases and/or on the last stressed syllable of the statement.

It is important to consider the meaning of F0 variations. Note that the more important variations occur just on the stressed syllable of the statement. We have showed above, on the two pairs of phrases – Figures 2 and 3 -, that the nucleus of the sintagma occupies firstly the position of subject of the phrase and after occupies the last position of the verbal complement to verify that the stressed syllable has the movement of variation of F0 which is more important on the sentence.
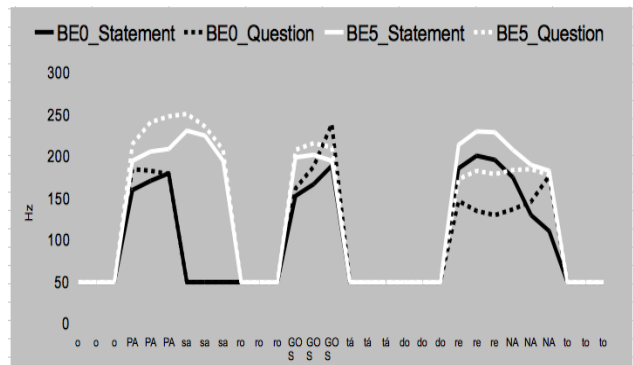
Figure 3: Comparison between the F0 variation meaning of the sentence pwt - O **pássaro** gosta do **Renato** - on both the modalities – declarative (full line) and interrogative (dashed line), spoken by a female speaker with low educational level from Belém - BE0 (black) – and another speaker of the same social profile from Cametá dialect – BE5 (white)
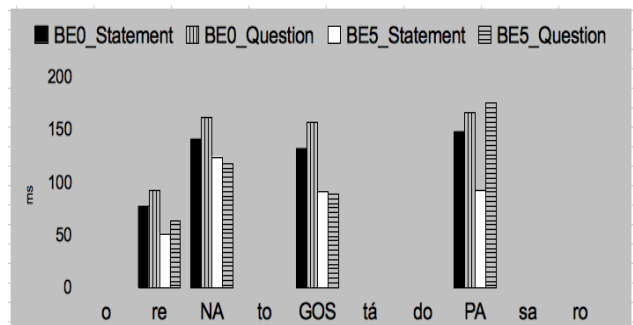
Figure 4: Comparison between the meaning of the duration (ms) on the sentence twp - O **Renato** gosta do **pássaro** – on both the modalities – declarative and interrogative - spoken by a female speaker with low educational level from Belém – BE0 and another speaker of the same social profile from Cametá dialect – BE5

---

[6] http://pfonetica.web.ua.pt/AMPER-POR.htm

The parameter of duration (ms) seems to act like a complement of the variations of F0 on the distinction of the two modalities that were analyzed, just as it is possible to observe on figures 4 e 5.

While the parameters of F0 and duration seems to complement themselves on the caracterization of both modalities declarative and interrogative on the varieties of the North of Brazil, the intensity seems not to be a significant physical parameter on the distinction of the two modalities in question, like we can note on the graphic of the Figures 6 and 7 below.
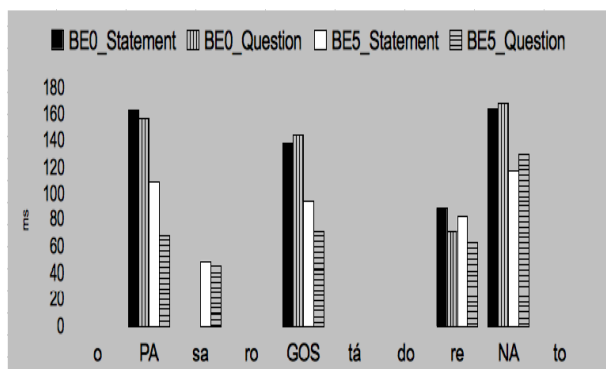
Figure 5: Comparison between the ms variation meaning of sentence pwt - O **pássaro** gosta do **Renato** - on both the modalities – declarative and interrogative - spoken by a female speaker with low educational level from Belém – BE0 - and another speaker of the same social profile from the Cametá dialect – BE5
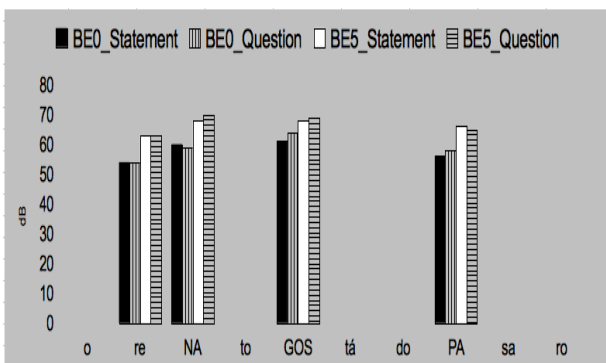
Figure 6: Comparison between the meaning of dB on the sentence twp - O **Renato** gosta do **pássaro** – on both the modalities – declarative and interrogative - spoken from a female speaker with low education level from Belém – BE0 - and another speaker of the same social profile of the Cametá dialect – BE5

Therefore the data have demonstrated that the measures of F0 are responsible for the principal difference between the two analyzed modalities – declaratives and interrogatives – it establishes an alteration on the movement of curve of F0 just on the stressed syllables of the nucleus of the final sintagmas of each sentence.

It is important to point out once again that the last stressed syllable of the phrase is the one that registeres the more important movement of the distinction between both modalities. For this reason, it has been our base hypothesis to be verified on the corpora of the
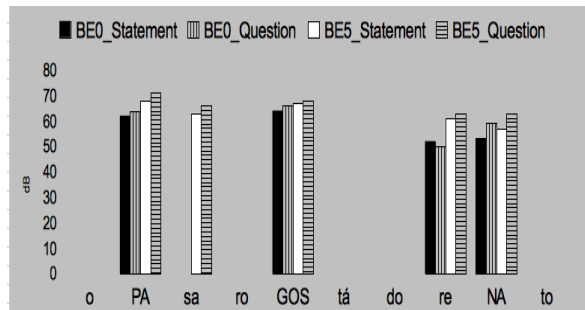
project here outline.

Figure 7: Comparison between the meaning of dB on the sentence pwt - O **pássaro** gosta do **Renato** - on both the modalities – declarative and interrogative - spoken from a female speaker with low education level from Belém – BE0 - and another speaker of the same profile of the Cametá dialect – BE5

6. Conclusion

The previous version of this project, whose period of execution includes since March 2009 until February 2012, has composed a corpora for the following Atlas: a) Belém – BE0 – (Cruz & Brito, 2011); b) Bragança – BE3 – (Castilho, 2009); c) Cametá – BE5 – (Santo & Cruz, 2011; Santo, 2011); and d) Baião – BF9 – (Lemos, in progress). Currently there is a planned fieldwork to the formation of other three *corpora*: f) Abaetetuba (Remédios, in progress); g) Belém islands (Guimarães, in progress); and h) Marajó island (Freitas, in progress). The Project has the exploration and the acoustic analysis of the corpora from Belém (Cruz & Brito, 2011) and from Cametá (Santo & Cruz, 2011; Santo, 2011).

7. References

Brito, C. (in progress). *Atlas prosódico multimédia do Português do Norte do Brasil – AMPER-POR: variedade lingüística da zona rural de Belém (PA)*, (Plan for Undergraduate Research) Belém: UFPA/ILC/FALE.

Castilho, F. (2009). Formação de Corpora para o Atlas Dialetal Prosódico Multimídia do Norte do Brasil: Variedade Lingüística de Bragança (PA). (Graduation Final Monograph). Bragança: UFPA/Campus de Bragança/Faculdade de Letras.

Contini, M. *et al*. (2002). Un Projet d'Atlas Multimédia Prosodique de l'Espace Roman. In B. Bel, I. Marlien (Eds.), *Proceedings of the 1st International Conference on Speech Prosody*. Aix-en-Provence: Laboratoire Parole et Langage, pp. 227--230.

Cruz, R. (2012). Alteamento vocálico das médias pretônicas no português falado na Amazônia

Paraense. In S.H. Lee (Ed.). *Vogais além de Belo Horizonte*. Belo Horizonte (MG): Faculdade de Letras da UFMG, pp. 194--220.

Cruz, R., Brito, C. (2011). Prosodic Multimedia Atlas of Belem city (Brazil): an overview. In *Atas do V congresso de Fonetica Experimental*. Caceres (Spain), October from 25th to 28th.

Freitas, J. (in progress). Atlas Prosódico Multimédia do Município da ilha do Marajó (PA). Master Dissertation. Curso de Mestrado em Letras, Universidade Federal do Pará, Belém (PA).

Guimarães, E. (in progress). Atlas Prosódico Multimédia da Belém Insular (PA). Master Dissertation. Curso de Mestrado em Letras, Universidade Federal do Pará, Belém (PA).

Lemos, R. (in progress). Atlas Prosódico Multimédia do Município de Baião (PA). Master Dissertation. Curso de Mestrado em Letras, Universidade Federal do Pará, Belém (PA).

Moutinho, L. *et al*. (2001). Contribuição para o estudo da variação prosódica do Português Europeu. In F. Sánchez Miret (Ed.), *Actas do XXIII CILFR* (Salamanca, Espagna, 22-28 Set. 2001). Vol. 1. Tübingen: Niemeyer, pp. 245--252.

Remédios, I. (in progress). Atlas Prosódico Multimédia do Município de Abaetetuba (PA). Master Dissertation. Curso de Mestrado em Letras, Universidade Federal do Pará, Belém (PA).

Rodrigues, D. (2005). Da zona urbana à rural/entre a tônica e a pretônica: alteamento /o/ > [u] no português falado no município de Cametá/Ne paraense - uma abordagem variacionista. Master Dissertation. Curso de Mestrado em Letras, Universidade Federal do Pará, Belém (PA).

Santo, I. (2011). Atlas Prosódico Multimédia do Município de Cametá (PA). Master Dissertation. Universidade Federal do Pará, Belém (PA).

Santo, I., Cruz, R. (2011). *Atlas Prosódico Multimédia do Município de Cametá (PA): uma visão geral*. In *Caderno de Resumos III Colóquio de Prosódia*. Belo Horizonte (MG), Jun.

Silva Neto, S. (1957). *Introdução ao Estudo da Língua Portuguesa no Brasil*. 4th ed. Rio de Janeiro: Presença.

Desafios da formação de *corpus* nas zonas de migração do Norte do Brasil

Regina CRUZ[1,2], Edson GOMES[1,3], Jany Eric FERREIRA[1,3], Soelis MENDES[1], Emanuel FONTEL[1]

[1]UFPA; [2]CNPq; [3]Aluno de Mestrado

Av. Augusto Correa, s/n – Campus do Guamá – Belém (PA) – 66075-900

regina@ufpa.br, edsongomes@ufpa.br, eric_qf@yahoo.com.br, soelis@ufpa.br, emanuelfontel@hotmail.com

Abstract

This work aims show how they formed sociolinguistic corpora for study of Brazilian Portuguese spoken in migration areas of Northern Brazil. We approach mainly difficulties happened during fieldwork of two UFPA's teams: i) Vozes da Amazônia project's team linked to PROBRAVO and b) ALIPA project's team linked to ALIB. Both projects aim to identify and map Amazon dialects. We show here the whole methodology of these projects: aims, nature of study, research context, speaker selection, collection of data, composition of corpora and researchers report of their experience. Between some difficulties we found: i) to meet speakers whose profile is right for each project; ii) non-availability of speakers in order to collaborate of collection of data; iii) the rejection of the people to face the recorder; iv) the fact the interviewer is not of location (Mendes, in progress). In the other hand, we noted that when the researcher lives in target city and the project uses the same methodological criteria of Bortoni-Ricardo (1985) for formation of corpus, this frame changes and researcher obtains a strong collaboration of the speakers (Ferreira, in progress).

Keywords: sociolinguistic corpora; Amazon Brazilian Portuguese; interdialectal contact.

1. Introdução

O presente trabalho tem como objetivo principal demonstrar como se afigura árduo o processo de formação de *corpora* sociolinguísticos em zonas de grande fluxo de migração na Amazônia paraense. Dar-se-á principal enfoque às dificuldades enfrentadas no trabalho de campo realizado pela equipe do Projeto Vozes da Amazônia sediado na UFPA que, por sua vez, é vinculado diretamente ao Diretório de Pesquisa Nacional PROBRAVO[1]. Três regiões foram selecionadas para uma nova fase de investigação desse projeto: Marabá (Mendes, em andamento), Aurora do Pará (Ferreira, em andamento) e Breves. Duas outras localidades estão previstas: Breu Branco e Parauapebas. No âmbito do ALiPA[2], outro projeto de formação de corpora também sediado na UFPA, identificam-se as mesmas dificuldades da equipe do **Vozes da Amazônia**. Por essa razão, um cotejo entre os dois projetos é aqui estabelecido. Para tanto, esposam-se considerações em torno dos objetivos e da metodologia empreendida em cada programa de investigação, além de breves relatos que ressaltam a atuação de cada pesquisador na respectiva comunidade linguística em que atua.

2. Projetos sociolinguísticos do Norte do Brasil

Em um Estado com dimensões continentais como é o caso do Pará, era de se esperar que houvesse forte variação no falar da população, principalmente pelo fato de essa constituição populacional ter-se dado por diferentes processos de ocupação territorial. Nesta seção apresentaremos os projetos **Vozes da Amazônia** e **ALiPA,** os quais têm o propósito de identificar e mapear os dialetos paraenses. Os projetos investigam os contatos interdialetais, resultado de processo de migração para a Amazônia paraense. No tocante a esse aspecto, a pesquisa em 6 (seis) pontos da mesorregião sudeste do Pará, empreendida por Gomes (em andamento), no projeto ALiPA, é o trabalho que trata mais especificamente da influência de outras regiões do Brasil nesta referida mesorregião.

2.1 Projeto Vozes da Amazônia

A versão atual do projeto **Vozes da Amazônia** prioriza uma investigação da identidade sociodiscursiva do amazônida nas regiões onde se atesta contato interdialetal decorrente de fluxo migratório intenso, motivado por projetos econômicos na região Amazônica, o que inclui o tratamento de aspectos culturais, sociais, históricos, político e ideológicos. Mapear a situação sociolinguística diagnosticada por Cruz (2012) relativamente à Amazônia paraense é o objetivo central do Vozes, em outras palavras o projeto busca identificar a influência de fatores extralinguísticos e identitários na configuração dos dialetos da Amazônia paraense, considerando o cenário sociohistórico da região e o fluxo migratório ali registrado. O projeto encontra-se vinculado a dois *campi* da UFPA - o de Belém e o de Marabá - e conta com a infraestrutura destes para a execução de suas atividades. A equipe atual**,** responsável pela condução das investigações, é composta por 2 (dois) alunos de Mestrado, 2 (dois) bolsistas de Iniciação Científica e 3 (três) pesquisadores titulados, todos com vínculo direto com a UFPA, além da coordenadora geral.

2.2 Projeto ALiPA

O **ALiPA** é um projeto de pesquisa ligado ao laboratório de linguagem da UFPA. Esse projeto tem por objetivo a construção do Atlas Geo-Sociolinguístico do Pará. Neste sentido desenvolve estudos cuja finalidade é identificar, analisar e mapear a variação linguística do português falado no Estado do Pará, integrando a dimensão social, que permitirá melhor compreensão dos mecanismos internos envolvidos na variação, especificamente, fonética, morfossintática e semântico-lexical. O projeto

---

[1] relin.letras.ufmg.br/probravo.

[2] http://www.ufpa.br/alipa.

utiliza metodologia do ALIB[3]. Para sua a execução, foram selecionados no Estado 50 (cinquenta) pontos de inquérito. Desses cinquenta pontos, mais de quarenta já foram coletados, restando alguns pontos na mesorregião sudeste, dentre esses enquadram-se os seis pontos de inquérito de (Gomes, em andamento).

3. Regiões mapeadas

Como o objetivo de ambos os projetos – **Vozes da Amazônia** e **ALiPA** – é compor um panorama histórico, antropológico e social do Pará, assim como identificar fatores sociais favorecedores da variação dialetal do português da Amazônia paraense, falado nas regiões de forte migração interna, faz-se necessário relacionar aspectos de variação inter e intradialetal. Por essa razão, à medida que se caracteriza sociolinguisticamente o português falado em Marabá, Aurora do Pará, Tucurui, Curionópolis, por exemplo, obtém-se o panorama geral das zonas de migração do Estado. Na nova fase do projeto **Vozes** no Estado do Pará**,** os municípios de Breves, Aurora do Pará e Marabá, em destaque verde no mapa (1), foram selecionados para a realização da pesquisa, tanto nas suas zonas rurais quanto urbanas. No caso do Projeto **ALiPA**, há um número maior de regiões contempladas, entretanto, o presente trabalho trata particularmente das localidades de Tucuruí, Itupiranga, São João do Araguaia, Curionópolis, Santana do Araguaia e São Felix do Xingu, indicadas de azul, no referido mapa.

Mapa 1: Indicações das localidades pesquisadas

4. Procedimentos metodológicos adotados por projetos

Os projetos da UFPA aqui descritos, apesar de terem como ponto em comum o tipo de região investigada, no caso as zonas de forte fluxo migratório no Estado, a metodologia adotada por ambos na formação de seus *corpora* é bem diferente como veremos nesta secção.

4.1 Como trabalha o Vozes da Amazônia?

O **Vozes da Amazônia** parte do conceito de Redes sociais como um conjunto de ligações que se estabelecem entre indivíduos. Segundo Bortoni-Ricardo (1985), nesse tipo de estudo o foco da investigação está na caracterização das relações entre os indivíduos, através da qual se pode explicar seus comportamentos, inclusive comportamentos linguísticos. Outro conceito importante é o de grupo de referência, que serve de alavanca à construção da identidade do indivíduo, o qual tenta modelar seu discurso segundo o daqueles que atende às suas expectativas psicossociais e com os quais busca identificação. A figura 1, abaixo, ilustra as relações que podem explicar o comportamento linguístico, em conformidade com o que propõe a referida autora.
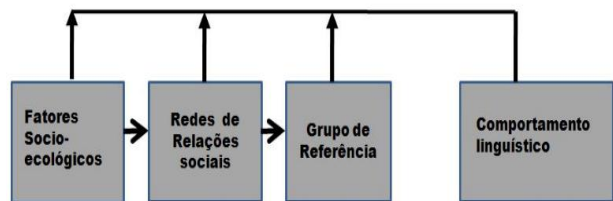
Figura 1: Relação estabelecida entre as partes componentes do modelo utilizado por Bortoni-Ricardo (1985)

A composição do *corpus* ocorre a partir de dois grupos de informantes: ancoragem e controle. O grupo de ancoragem possui 24 informantes (12 de cada sexo) e o de controle, 12 informantes (6 de cada sexo) que devem necessariamente ter algum vínculo de parentesco com membros do grupo de ancoragem, como filhos, netos ou sobrinhos. Todos os informantes são distribuídos em três faixas etárias: a) de 15 a 26 anos; b) de 30 a 46 anos e; c) acima de 50 anos. A coleta de dados é realizada por meio de narrativas de experiência pessoal. O trabalho de Mendes (em andamento) atesta que esse tipo de procedimento metodológico tem sido eficaz nos dois grupos de informantes, para os quais se pergunta sobre a origem de cada um e sobre a percepção que cada informante detém da cidade antes de ele lá ter-se instalado etc. Além desses aspectos, firma-se atenção a todas as orientações da técnica apresentada por Tarallo (1988). Registre-se, ainda, que os dados estão sendo coletados por meio de gravadores digitais. Uma vez o trabalho de campo concluído, o tratamento dos dados seguirá todas as etapas previstas em um estudo sociolinguístico, a saber: (i) transcrição dos dados nos

---

[3]

http://twiki.ufba.br/twiki/bin/view/Alib/MetodologiaGeral

moldes da análise da conversação (Castilho, 2003); (ii) triagem dos grupos de força (Câmara Jr., 1969); (iii) transcrição fonética dos vocábulos que contenham marcas dialetais alvo, utilizando-se o alfabeto SAMPA; (iv) codificação dos dados e; (v) tratamento quantitativo VARBRUL.

4.2   Como trabalha o ALiPA?

O projeto ALiPA contempla o número de 50 (cinquenta) localidades, distribuídas por seis microrregiões do Pará, levando-se em consideração a extensão de cada região, os aspectos demográficos, culturais, históricos e a natureza do processo de povoamento da área. Para compor o *corpus*, foram selecionados 4 (quatro) informantes por localidades: dois do sexo feminino e dois do masculino, distribuídos nas faixas etárias de 18 a 30 anos e 40 a 70 anos. Eles devem ser filhos da localidade pesquisada, assim como os pais; devem ter, no máximo, a 4ª série do fundamental e exercer profissões que evitem mobilidade. A coleta de dados é realizada com uso dos questionários fonético-fonológico, morfossintático e semântico-lexical. No estudo de Gomes (em andamento), está sendo aplicado apenas o questionário semântico-lexical. Os dados estão sendo coletados em equipamentos sonoros, como gravador digital, gravados em CD e em outros equipamentos de informática para posterior tratamento. Vencida essa etapa, os dados serão transcritos grafematicamente e transferidos para as cartas lexicais. Também vem sendo usado o recurso fotográfico como meio de registrar através de imagens o homem e o espaço em que habita.

5.   Caracterização dos *corpora* formados

Até o presente momento, o *corpus* formado conta com amostra de 14 (quatorze) informantes (sendo oito do grupo de ancoragem e seis do grupo de controle) da variedade linguística de Marabá (Mendes, em andamento), 18 (dezoito) informantes do grupo de ancoragem e 8 (oito) do grupo de controle da variedade linguística de Aurora do Pará, localidade que está sendo pesquisada por Ferreira (em andamento). No total, há 36 (trinta e seis) informantes por variedade. Ambos trabalhos fazem parte do Vozes da Amazônia. Para a pesquisa de Ferreira (em andamento), na fase de trabalho de campo, foram estabelecidas visitas prévias aos informantes, para um primeiro contato, o que permitiu preliminarmente a criação de um certo vínculo de intimidade com os informantes. Esse grau de intimidade favoreceu para criar um clima, o mais descontraído possível, entre os participantes da pesquisa, o que é essencial para uma boa coleta de dados. O fato de contribuírem para o trabalho de alguém conhecido deixava os informantes bastante alegres e descontraídos, o que amenizava o estranhamento causado pelo aparato técnico, presente no momento da entrevista, que ocorreram na própria casa dos informantes. Há de se destacar, contudo, que, se por um lado o grau de intimidade que se estabeleceu entre entrevistador e informantes, representou, naquele momento, facilidade,

a falta de informantes e de instrumentos técnicos suficientes para o grupo de pesquisa têm sido entraves para a composição dos *corpora*. Ferreira (em andamento) afirma que há enorme dificuldade para se identificar informante tanto do sexo masculino quanto do feminino na faixa etária de 26 a 46 anos. A maioria dos migrantes possui mais de 50 anos e instalou-se no município nas décadas de 60 e 70. Os de faixa etária mais jovem apresentam-se em menor quantidade, o que dificulta o trabalho para encontrá-los. Pesquisas dessa natureza, no qual se tem critérios para a seleção de informantes, nem sempre são fáceis de serem executados posto que o andamento da pesquisa depende do total de informantes necessários a sua realização. Isso tem ocasionado o atraso de todo o trabalho de campo de Ferreira (em andamento). O fato de não se ter uma quantidade suficiente de gravadores digitais para coletar os dados também constitui outro entrave. Os poucos investimentos injetados em trabalhos dessa natureza afetam diretamente em sua realização. Gomes (em andamento) até o presente momento conta com um *corpus* contendo amostra de 20 (vinte) informantes, sendo 10 (dez) homens e 10 (dez) mulheres, de um total de 24 (vinte e quatro) informantes. 10 (dez) informantes são da faixa etária de 18 a 30 anos e 10 (dez) são da faixa etária de 40 a 70 anos. De cada localidade - Santana do Araguaia, São Félix do Xingu, Tucuruí, Curionópolis e São João do Araguaia - foram entrevistados 4 (quatro) informantes, faltando apenas 4 (quatro) informantes de Itupiranga. Para coletar os dados, Gomes (em andamento) teve que se deslocar em dois momentos: em julho de 2011, para Santana do Araguaia, São Félix do Xingu e Tucuruí; em fevereiro de 2012, para Curionópolis e São João do Araguaia. A coleta de dados ocorreu a partir de entrevistas realizadas, na maioria das vezes, nas casas dos informantes, o que não foi muito bom, devido às interferências de curiosos que, em alguns momentos, respondiam às perguntas. Outras, contudo, foram realizadas fora da casa, às margens do Rio Xingu, por exemplo, o que facilitou o trabalho, mas houve momentos difíceis, em que foi preciso fazer a entrevista sob o sol ou embaixo de chuvisco, por falta de local apropriado e para não perder o informante. Pelo fato de as entrevistas terem sido realizadas em localidades distantes, no sudeste do Pará, foi preciso montar um planejamento de deslocamento. Mesmo os informantes sendo pessoas desconhecidas do entrevistador, houve sucesso no trabalho, porque, em todas as cinco localidades, foram encontradas pessoas dispostas a ajudar na coleta de dados. O fato de alguém ter ido de longe e em condições desfavoráveis para campo, sensibilizava os informantes, os deixava satisfeitos e os tornava mais propensos a contribuir, dando informações dos seus respectivos lugares, muitas vezes esquecidos. A maior dificuldade por que se passou foi conseguir pessoas que se encaixassem nas exigências do projeto ALIPA. As 20 (vinte) gravações foram realizadas em gravador digital Olympus Linear PCM Recorder LS-10. A próxima etapa constitui-se em

trabalhar os dados para verificação da variação que ocorre dentro da mesorregião objeto da pesquisa, e desta em relação às outras mesorregiões do Estado do Pará, para se obter um retrato, o mais fiel possível, do falar paraense. A seguir figura (5.1) apresenta a síntese do total de informantes por pesquisa, num total de 60.
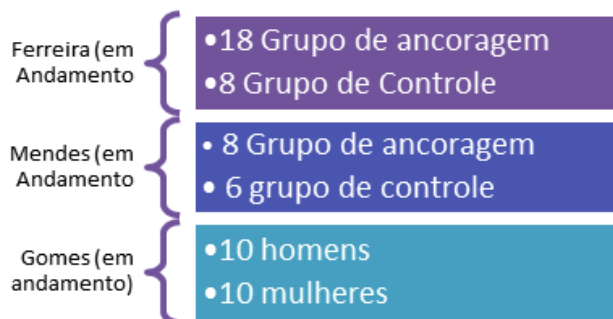
Figura 5.1: Síntese do Total de Informantes por pesquisa

6.   Dificuldades impostas pela realidade amazônica das zonas de migração

Tanto Mendes (em andamento) quanto Fagundes (em andamento), que também faz parte da equipe do Vozes da Amazônia, estão tendo grande dificuldade na obtenção dos dados necessários. Uma das dificuldades encontradas para realização da pesquisa está em encontrar pessoas que se encaixem no perfil do projeto e a não disponibilidade dos falantes localizados para participar da pesquisa. Mesmo se tendo a preocupação de deixar o informante o mais a vontade possível com a presença da equipe e com a do gravador, a recusa da parte de algumas pessoas dá-se, na maioria das vezes, sem um motivo aparente. Invariavelmente, os que se recusam fazem-no simplesmente com a afirmação de que não aceitam participar e, diante disso, não são feitas mais investidas, pois é necessário que o informante sinta-se à vontade. Outras vezes, o falante, em função da presença do gravador, sente-se inibido e se recusa a participar. Além disso, há também a incidência daqueles que marcam a entrevista e não comparecem ao encontro. Diante dessas situações, muitas vezes, tenta-se marcar nova coleta, no entanto, isso não garante a presença do informante nessa nova oportunidade. A situação é mais grave, ainda, quando o entrevistador, além de não ser morador nativo da localidade, utiliza critérios sociolinguísticos para a formação de amostras mais adequados para estudos variacionistas clássicos, como, por exemplo, o critério de selecionar apenas informantes nativos da comunidade pesquisada ou que tenham ido morar para lá ainda criança, como é o caso de Gomes (em andamento), para quem os aspectos históricos e sociais da localidade investigada não são importantes, tendo em vista os objetivos da pesquisa que empreende atualmente. O fato de a pesquisa de Gomes (em andamento) estar localizada na mesorregião Sudeste do Pará, dificulta a coleta dos dados, porque alguns critérios adotados pelo projeto ALiPA, como a exigência de informantes nascidos na localidade, vão de encontro ao histórico da região, onde a população é constituída, em sua maioria, por migrantes de outras localidades do país. Enquanto nas outras mesorregiões a população rural é constituída, em sua maioria, por habitantes nascidos na localidade e portadores de característica "cabocla", na mesorregião Sudeste (Sul do Pará) verifica-se exatamente o contrário: nas zonas rurais muitos habitantes moram nos Projetos de Assentamento (PAs), o que faz com que, muitas vezes, seja mais fácil encontrar um morador nascido na cidade que, embora desenvolva suas atividades na zona rural, nasceu em zona de cidade, ou seja, zona urbana, tanto que o deslocamento da população dessa região se dá mais para os Estados do Centro-Oeste, como Tocantins, Goiás, Brasília, do que para Belém, capital do Pará. Surpreendentemente, quanto à abordagem dos informantes, Gomes (em andamento) sentiu pouca dificuldade, pois quase sempre as pessoas estavam dispostas a colaborar na pesquisa. Sua maior dificuldade foi, sem dúvida, identificar informantes com o perfil exigido. Alguns funcionários da EMATER foram peças-chave para a localização de informantes, principalmente em Santana do Araguaia e São Félix do Xingu. Outro fato surpreendente foi o engajamento de outras pessoas que acabaram colaborando diretamente, ao terem se sensibilizado com o descolamento do pesquisador e com os objetivos de seu trabalho de pesquisa. Tais fatos acabaram afigurando-se pontos positivos na implementação dos trabalhos e na colaboração dos moradores locais. Por outro lado, constatou-se que, quando o entrevistador é um morador da localidade alvo e se utiliza dos parâmetros propostos por Bortoni-Ricardo (1985) para a formação do *corpus*, este quadro de dificuldades não é verificado e o pesquisador consegue obter forte colaboração dos informantes, é o que relata Ferreira (em andamento). De qualquer forma, a experiência de estar realizando este tipo de pesquisa tem sido bastante rica, não só pelo fato de permitir perceber a grande recorrência do fenômeno analisado, no caso de Ferreira (em andamento), as vogais médias pretônicas, sem o que a pesquisa restaria inviável, mas também, e sobretudo, pelas possibilidades advindas do contato estabelecido com as pessoas nativas da região e pela interação com elas estabelecida, tanto na condição de amigo, de conhecido, quanto na de pesquisador que observa e analisa os ricos aspectos que envolve o fenômeno da variação linguística, evidenciada na fala natural. Nesse sentido, constata-se, ainda mais, a pertinência da ideia comum de que a interação independe de formalismos linguísticos. Fortalece a compreensão dos modos como o trabalho de campo imprime melhor reflexão ao trabalho de coleta e análise, além de permitir a verificação dos aspectos que, nessas condições, dão ensejo a ocorrência de fala natural, tão ansiada pelos pesquisadores da sociolinguística. É preciso ressaltar que Mendes (em andamento), em sua pesquisa na comunidade linguística de Marabá, embora com dificuldades na seleção dos informantes e, em especial, na realização das entrevistas, constatou que

uma entrevistada ficou tão envolvida emocionalmente com a condução de sua narrativa, que, quando o tempo estipulado para todas as entrevistas chegou ao fim, ou seja, 15 minutos, a informante pediu para continuar contando casos de sua sofrida vinda para Marabá. Outro ponto de destaque, que certamente poderá auxiliar na condução de pesquisas em desenvolvimento, refere-se a uma participação no Congresso Internacional de Linguística Histórica, em homenagem ao Prof. Dr. Ataliba Castilho, realizado na USP, no mês de fevereiro de 2012, quando Mendes, ao relatar à Profa. Dra. Odete Pereira da Silva Menon, (UFPR) a dificuldade vivenciada na realização das entrevistas em Marabá, foi orientada a contar com a autoridade de pastores evangélicos no contato com informantes. De fato, essa estratégia tem contribuído para selecionar novos informantes, permitindo, desse modo, vislumbrar, logo em breve, a consecução das entrevistas com os 18 (dezoito) falantes restantes.

7. Conclusão

Nos dias atuais, as mudanças ocorrem muito rapidamente e, em consequência, as transformações na língua também. Essa realidade tem impulsionado estudos linguísticos que visam a recuperar e/ou registrar os falares de diversas comunidades linguísticas, porquanto, assim procedendo, registram-se não só fenômenos linguísticos observados, mas também as memórias linguística e discursiva da comunidade da região estudada. Ao identificar e mapear dialetos nas regiões de migração do norte do Brasil, os projetos ALiPA e VOZES da AMAZÔNIA cumprem seu papel social, pelos motivos acima expostos. No entanto, essa tarefa nem sempre é tão simples e possível de ser concretizada, em virtude das dificuldades e desafios impostos pela pesquisa de campo. Desse modo, as dificuldades que envolvem a composição dos *corporas*, em regiões de migração, no norte do Brasil, apresentam-se tanto no projeto **Vozes da Amazônia** quanto no projeto ALiPA. Dentre as que merecem destaque, elencamos a dificuldade de encontrar pessoas que se encaixem no perfil dos projetos e a não disponibilidade dos falantes locais para participar da pesquisa; a recusa das pessoas perante o gravador e, por vezes, a escassez de equipamentos e ferramentas adequadas. O fato de o entrevistador não ser da localidade também dificulta a coleta de dados, haja vista que esse aspecto produz estranheza e desconfiança por parte dos informantes. Por outro lado, constatou-se que quando o entrevistador é um morador da localidade-alvo e o projeto utiliza os critérios metodológicos de Bortoni-Ricardo (1985) para a formação do *corpus,* este quadro de estranheza e de desconfiança não se apresenta, por conseguinte, o pesquisador consegue obter uma forte colaboração dos informantes. E mais, o grau de conhecimento entre entrevistador e informante favorece a coleta de dados, na medida em que possibilita a incidência de falas muito próximas ao natural, mitigando, desse modo, um dos paradoxos do observador. Foi possível apontar também

estratégias que facilitam ou, pelo menos, diminuem os transtornos de muitos pesquisadores. Uma delas acena para o entrosamento que deve haver entre entrevistador e líderes comunitários ou religiosos na busca de informantes, fato que pode favorecer o contato entre os envolvidos na pesquisa de campo. Esperamos que as problemáticas aqui apresentadas não sirvam para desmotivar aqueles que se interessam ou intencionam trilhar as veredas da pesquisa linguística. Do contrário, esperamos que as considerações aqui esposadas sirvam como demonstrativo de que as dificuldades, sejam de cunho metodológico, sejam de outro caráter, não devem se sobrepor a imperiosa e nobre tarefa do pesquisador de descrever o funcionamento da língua em todos os seus matizes e possibilidades.

8. Referências

Bortoni-Ricardo, S.M. (1985). *The urbanization of rural dialect speakers: a sociolinguistic study in Brazil*. Cambridge: Cambridge University Press, 265p.

Câmara Jr., J.M. (1969). *Estrutura da Língua Portuguesa*. Petrópolis: Vozes.

Cardoso, S. (2010). *Geolinguística: tradição e modernidade*. São Paulo: Parábola.

Castilho, A. (2003). *A língua falada no ensino do português*, 5ª. Edição. São Paulo: Contexto.

Cruz, R. (2012). Alteamento vocálico das médias pretônicas no português falado na Amazônia Paraense. In S.H. Lee (Ed.), *Vogais além de Belo Horizonte*. Belo Horizonte (MG): Faculdade de Letras da UFMG, pp. 194--220.

Fagundes, G. (em andamento). Alteamento das Vogais Médias Pretônicas no Português da Amazônia Paraense: a influência do dialeto dos migrantes no português falado em Breves (PA). Dissertação (Mestrado em Letras), Universidade Federal do Pará, Belém (PA).

Ferreira, J.E. (em andamento). Variação do /s/ no falar aurorense. Em andamento. Dissertação (Mestrado em Letras), Universidade Federal do Pará, Belém (PA).

Gomes, E. (em andamento). Variação Lexical em Seis Municípios da Mesorregião Sudeste do Estado do Pará. Em andamento. Dissertação (Mestrado em Letras), Universidade Federal do Para, Belém (PA).

IBGE. Censo 2010. Disponível em: <www.censo2010.ibge.gov.br>. Acesso em: 07 de mar. de 2012.

Mendes, S. (em andamento). Vozes da Amazônia: a realização das vogais médias pretônicas na comunidade linguística de Marabá. Marabá: UFPA/Campus do Sul e Sudeste do Para. (Projeto de Pesquisa).

Tarallo, F. (1988). *A pesquisa sociolinguística*. São Paulo: Ática. (Série Princípios).

Compiling a Multilingual Spoken Corpus

Ekaterina LAPSHINOVA-KOLTUNSKI, Kerstin KUNZ, Marilisa AMOIA

Saarland University
Universität Campus 66123 Saarbrücken Germany
e.lapshinova, k.kunz, m.amoia@mx.uni-saarland.de

Abstract

The present paper describes the compilation of the spoken part of an English-German corpus, which has been created for the investigation of cohesion. The corpus is one of the few existing resources supporting contrastive studies of cohesion and, to our knowledge, the only one permitting a contrastive analysis of spoken registers in the two languages. In addition, our corpus data offer further research potentials for contrastive linguistics and translation studies as well as for numerous NLP research areas.

Keywords: corpus compilation, spoken corpus, multilingual corpus, corpus annotation, cohesion.

1.   Introduction

The present paper describes the compilation of the spoken part of an English-German corpus, which has been created for the investigation of cohesion. The corpus is one of the few existing resources supporting contrastive studies of cohesion and, to our knowledge, the only one permitting a contrastive analysis of spoken registers in the two languages. In addition, our corpus data offer further research potentials for contrastive linguistics and translation studies as well as for numerous NLP research areas.

1.1   Aims

The main objective of the present paper is to compile the spoken part of a multilingual corpus to investigate cohesion in German and English. Our long-term linguistic research interest is in the analysis of cohesive resources provided by both language systems and their instantiations in texts. More precisely, we are concerned with the exploration of contrasts in form, frequency and function of cohesive devices and meaning relations established to other textual elements. We aim to analyse these phenomena across and between languages, registers and modes.

1.2   Motivation

Comprehensive accounts of cohesion are only existent from a largely systemic and monolingual perspective, see e.g. (Halliday & Hasan, 1976; Brown & Yule, 1983; Schubert, 2008 and Esser, 2009) for English, and (De Beaugrande & Dressler, 1981; Vater, 2005; Brinker, 2005) for German. Empirical analyses (both monolingual and contrastive) in the area of cohesion mainly deal with indiviual cohesive devices, cf. (Bosch *et al*., 2007) or (Gundel *et al.,* 2004). Empirical analyses of cohesion in spoken discourse exist for German, e.g. (Ahrenholz, 2007) and English, e.g. (Gundel *et al.*, 2004 and 2005; Eckert & Strube 2001). These however, are limited to the investigation of individual phenomena, and mostly examine personal pronouns or demonstratives. To our knowledge, there is only one contrastive empirical analysis by (Schreiber, 1999) comparing English and German. It includes a relatively broad range of cohesive

phenomena, however it uses excerpts of French and German corpora to illustrate particular phenomena rather than presenting a contrastive interpretation of findings from a statistical analysis.

These studies seem to suggest that particular cohesive devices exhibit a tendency to occur either in registers of spoken language only or with a much higher frequency than in written discourse, see e.g. (Schreiber, 1992; Ahrenholz, 2007). Our preliminary extractions from registers of written language [1] underpin these observations. For instance, they show that occurrences of the German demonstrative pronouns *der, die, das* and particular constructions of substitution are rarely traced in typical registers of written language and appear with a much higher frequency in those written registers that approximate spoken language, such as fiction or political speeches[2]. In addition, dialogic sequences of our fiction subcorpus point to instantiations of cohesive ellipsis which seem to be restricted to spoken discourse. These first findings call for a corpus which allows to integrate differences between written and spoken registers so as to establish a comprehensive model of cohesion in English and German. To our knowledge, there are no corpus resources to support our research goal. The existing ones are either monolingual, e.g. ICE, cf. (Greenbaum, 1996) for English or "Deutsch heute", cf. (Brinckmann, 2008) German, or compiled for special purposes, e.g. SCOTS corpus, cf. (Anderson, 2007) or Verbmobil, (Hinrichs *et al.*, 2000). Some of them also contain non-native data, e.g. ICLE described in (Granger, 2008) and LINDSEI, cf. (Gilquin *et al.,* 2010).

2.   Theoretical Background

There are substantial gaps in the area of text-based contrastive modeling for the two languages under analysis, especially text-based empirical accounts of mechanisms of textuality are absent. System-based text/discourse grammars commonly deal with specific questions of textuality only. While the literature in English mainly

---

[1] cf. (Kunz *et al*., 2009; Klein, 2007 and Birster, 2007).

[2] The extractions were done on the CroCo corpus, cf. (Neumann, 2005)

focuses on linguistic resources for establishing textuality, e.g. (Halliday & Hasan, 1976; Brown & Yule, 1983; de Beaugrande & Dressler, 1981), the German literature frequently takes as its starting point general pragmatic, cognitive and semantic principles of coherence, which are reflected in linguistic phenomena, cf. (Linke *et al.*, 2001; Brinker, 2005; Vater, 2001). These methodological differences lead to noticeable differences in the range of phenomena considered. In general, monolingual text-/discourse treatments inform us about the coherence-building systems of a language and are structured by type and/or function of the system (e.g. (co-) reference, conjunctive relation, lexical/semantic relations, etc). We define cohesive resources (devices) as a set of lexico-grammatical items that function as resources allowing to transcend the boundaries of the clause. For our classification of general categories, we follow the one by (Halliday & Hasan, 1976), according to which cohesion includes five categories: *reference, substitution, ellipsis, conjunctive relations, lexical cohesion*.

3.   Corpus Compilation

3.1   Data Collection

Our multilingual spoken corpus contains two registers: interview and academic speech. These registers are added to the eight registers of written language of the already existing corpus, cf. (Kunz & Lapshinova, 2011): popular-scientific texts, tourism leaflets, political essays, corporal communication, instruction manuals and websites, prepared speeches, fictional texts. Especially the latter two registers are considered to lie at the borderline between written and spoken discourse. In order to create the German-English spoken corpus, we extract parts of already existing speech corpora and collect our own data, cf. table 1.

subcorpora	German (GO)	English (EO)
INTERVIEW	BACKBONE-DE	ELISA BACKBONE-EN
ACADEMIC	GECCo spoken collection	MICASE

Table 1: Sources for the GECCo spoken part

For English, we use the data of the MICASE corpus, the English part of the BACKBONE corpus and the ELISA corpus. The Michigan Corpus of Academic Spoken English (MICASE) is a collection of nearly 1.8 million words of transcribed speech – almost 200 hours of recordings) from the University of Michigan and includes lectures, classroom discussions, lab sections, seminars, and advising sessions, cf. (Simpson *et al.,* 2002). The BACKBONE pedagogic corpus contains corpora of video-recorded spoken interviews with native speakers of various European languages, cf. (Kohn**,** 2011). The ELISA corpus contains interviews with native speakers of English talking about their professional career (e.g. in

tourism, politics, the media or environmental education), cf. (Braun**,** 2006). The data from the corpora were extracted according to criteria such as nationality of speaker, type of speech event, degree of speaker interaction. For German, we use the German part of the BACKBONE corpus, which contains interviews with German native speakers (including variants of German). This subset is comparable to the interviews in ELISA and the English part of the BACKBONE corpus. In addition, we compile our own corpus of spoken academic discourse consisting of lectures from all departments of the Saarland University. The lectures were recorder by VISU (Virtual University of Saarland) and have been transcribed by our team according to the transcription guidelines described below.

3.2   Problems in Spoken Data Compilation

In the process of data collection for the German part of spoken academic discourse, we have encountered a number of practical problems. For instance, we initially planned to include recordings of seminars for the analysis of dialogues. However, the seminars in Germany turn out to be less interactive and dialogic than assumed and hence, do not correspond to their English counterparts. Moreover, the collected student presentations constitute prepared speech and thus lack the authentic character of spontaneous speech. Therefore, our German academic corpus currently consists of lecture recordings which are comparable in their speech conditions to the English lectures.

Besides that, we had to apply manual transcription methods which is very labour- and time-consuming. Yet, the recorded data was found to contain too much noise to permit an automatic transcription (speech recognition). Moreover, manual transcription requires the formulation of transparent transcription guidelines. Since the English data was transcribed according to differing guidelines we elaborate a consistent scheme for tags in both languages to annotate extra-linguistic information (example (1)), linguistic variants (example (2)) and repairs and repeats (example (3)).

1) LAUGHTER:
   \<laugh>text\</laugh>
   CONTEXTUAL EVENTS:
   \<writing_on_board>
   \<writing_on_board> text \</writing_on_board>
   \<door>
   \<break type="gasp">
2) EO-INTERVIEW
   \<turn speaker="Lauren"> \<alternative text="yep"> Yes \</alternative>, absolutely. \<alternative text="yeah">Yes\</alternative>, I \<break/> \<alternative text="yeah"> yes \</alternative>, absolutely \</turn>
   GO-INTERVIEW
   \<turn speaker="Stefan"> \<alternative text="We'mer"> Wenn wir \</alternative> die Netze erreicht \<alternative text="hand"> haben \</alternative>, \<alternative text="weret"> werden \</alternative> die Netze \<alternative

text="gehobe'"> gehoben </alternative>, es sind Stellnetze.</turn>

3) REPEAT:
so it's <repeat text="an awful"> an awful</repeat> lot of different cultures, different religions, different countries that people are from, which is great.
REPAIR:
So <repair text="they're"> <break/> they do </repair> struggle to settle in and <repair text="it's just for"> <break/> you know, it's our place</repair> really.

In order to guarantee comparability in frequency and function of cohesive devices between the written and spoken registers we had to restrict each register to 10-14 texts with around 34 thousand tokens each. The existing registers of written language contain both comparable and parallel texts of English and German. However, for the spoken registers, only comparable texts are available, cf. table 1. One possible solution for obtaining aligned texts would be to create interpretations for the existing originals. Interpreted texts, however, are produced under very specific conditions and are affected by various constraints such as time pressure, limited short-term memory capacity, linearity and others, see e.g. (Gumul, 2010) and (Pöchhacker, 2001). They are not considered as reflecting spontaneous speech on the one hand and differ considerably from translations, on the other hand. We thus consider to integrate transcriptions of films and their synchronizations in our corpus, although these are subject to other limitations described, for example, by (Herbst, 1994) and (Döhring, 2006).

4. Corpus Annotation

The spoken registers of the multilingual corpus are annotated on the same level as its written part:

1) **token level:** words, lemmas, parts-of-speech;
2) **chunk level**: sentences, syntactic and semantic chunks and their grammatical functions;
3) **cohesion level**: cohesive devices and cohesive chains;
4) **text level**: registers;
5) **extra-linguistic level**: meta information.

The automatic annotations of parts-of-speech, chunks and their grammatical functions are obtained with the help of the Stanford Parser, cf. (Marneffe *et al.*, 2006). Cohesive devices, such as conjunctive relations, personal and demonstrative reference, substitution, ellipsis and lexical cohesion, are semi-automatically annotated with a tool based on the YAC recursive chunker, cf. (Kermes, 2003) which utilises the CWB Perl-Modules developed within the framework of YAC, cf. (Kermes & Evert, 2001) and (Kermes & Evert, 2002). We also apply the MMAX tool, cf. (Müller & Strube, 2006) for the manual correction of these annotations. Disambiguation of cohesive devices is based on the analyses described in

(Kunz & Steiner, *in progress*) and (Kunz, 2010).

We also aim at annotating reference and lexical chains in our corpus. For this, we apply one of the existing systems for coreference resolution, the Stanford Coreference Resolution System described by (Lee *et al.*, 2011). Our preliminary evaluation tests, see (Amoia *et al.*, 2012), show that the system does not perform with the desired accuracy. Therefore, we also plan to manually improve annotations for this category of cohesion.

The corpus metadata include not only the information on speaker, such as age, sex (female, male, unisex, undefined), profession (translator, teacher, professor, student, etc.) and role (interviewer, interviewee, lecturer, etc.), but also the information on register analysis: field (experiential domain and goal orientation – argumentation, exposition, instruction, narration, description and persuasion), tenor (number of speakers, agentive role – monologic or dialogic, social role – equal, up or down, social hierarchy – expert to expert, expert to layperson, layperson to expert, layperson to layperson, social distance – formal or not) and mode (language role – ancillary or constitutive, channel – graphic, phonic or electronic, and medium – written, written to be spoken, spoken).

5. Corpus Querying

The corpus can be queried with the Corpus Query Processor (CQP, (Evert, 2005)), which allows us to detect candidates for cohesive devices by means of regular expressions, offering several functionalities for extraction (e.g., context expansion) and sorting purposes (e.g., counting, grouping of results). CQP allows two types of attributes: positional (e.g. for part-of-speech and morphological features) and structural (e.g. for chunks, registers or extra-linguistic information). These attributes are employed for CQP-based queries which include string, parts-of-speech, chunk, register and further constraints, cf. table 2.

Query elements	meaning
[word="and" & .cohesive_device="conj" & .text_register="INTERVIEW" & .tenor_numberOfSpeakers="2" & .speaker_ager="31-50" & .tenor_socialRole="equal"]	word *and* which is cohesive conjunction in interviews only with 2 speakers only aged between 31-50 in equal social role

Table 2: Example of a CQP query

The present CQP query delivers a list of concordances, as shown in example for the cohesive conjunction *and* (4).

4) 8: My name's Norma Holt **and** I actually come from the Wirral Peninsula which is on the west coast of Liverpool, which is Lancashire...

29: which is Lancashire, **and** we have Cheshire on one side and north Wales on the other.
188: the nice seaside is, if you like, all the big houses are, **and** it's more countryside, more of the farming...
296: However, over the years certainly it has changed **and** now it's very much a Liverpool accent ...
304: ... now it's very much a Liverpool accent **and**, you know, which I'm not saying I disapprove of it ...
325: I think it's a lazy speech **and** you need to actually think about what you're saying.
348: My nephew sometimes'll speak to me in the Liverpool accent **and** I'll say, please speak to me in English ".

Moreover, the sorting, counting and grouping functionality of CQP allows us to extract frequency information, as shown in table 3 (for English only as the German ACADEMIC part is still under construction). The obtained frequencies of cohesive phenomena can then be evaluated in terms of their distribution across registers, languages and modes.

For instance, table 3 displays the frequencies per million words of all cohesive occurrences of the form *one* in its function as nominal substitute. What the table nicely illustrates is that some registers show more commonalities in their distribution of cohesive *one* than others, and most notably that there is a considerable difference in frequency between the spoken and the written registers of our subcorpus. In addition, the two registers FICTION and speech are closer to the spoken registers then others. This may be due to the fact that FICTION contains text passages imitating spoken dialog and that SPEECH was written to be spoken. Thus, ACADEMIC seems to be at one end of the spoken written continuum of our corpus and SHARE at the other end (at least as far as cohesive one is concerned) with FICTION and SPEECH taking a somewhat middle position.

	register	Cohesive *one* per 1M
spoken	INTERVIEW	949,84
	ACADEMIC	2769,33
written	FICTION	378,42
	SPEECH	199,65
	ESSAY	85,72
	SHARE	83,74
	INSTR	110,60
	TOU	124,02
	POPSCI	142,26
	WEB	166,18

Table 3: Frequencies delivered by CQP

6.   Conclusion and Future Work

We have compiled a spoken corpus for English and German that is enhanced with annotations on several linguistic and extra-linguistic levels. Our corpus architecture not only allows a text-based contrastive analysis of cohesion in German and English but also permits a comparison of various spoken and written registers. Therefore, the findings based on our resources will not only complement the existing research gaps in cohesion but also enrich contrastive grammars with a systematic account of discourse phenomena in written vs. spoken mode. Moreover, both the developed resources as well as our findings on cohesion will provide valuable insights for language teaching and translator training and will open up new research options for various fields.

In the future, we aim at expanding corpus with further registers, e.g. internet forums, TV talk shows and reports. Besides that, we will develop further procedures to automatically annotate cohesive devices and relations. We also plan to enhance our spoken corpus with translations. The corpus will be available for querying online within the CLARIN-D initiative.

7.   Acknowledgements

8. References

Ahrenholz, B. (2007). Verweise mit Demonstrativa im gesprochenen Deutsch. Grammatik, Zweitspracherwerb und Deutsch als Fremdsprache. Berlin u. New York: de Gruyter.

Amoia, M., Kunz, K. and Lapshinova-Koltunski, E. (2012). Coreference in Spoken vs. Written Texts: a Corpus-based Analysis. In *Proceedings of LREC-2012*. Istanbul, Turkey.

Anderson, W. (2007). The SCOTS Corpus: a resource for language contact study. In P.S. Ureland, A. Lodge and S. Pugh (Eds), Language Contact and Minority Languages in Europe. Studies in Eurolinguistics, Vol. 5. Berlin: Logos Verlag

De Beaugrande, R.A., Dressler, W.U. (1981). Einführung in die Textlinguistik. Tübingen: Niemeyer.

Birster, L. (2007). Kohäsionsmittel im Englischen und Deutschen – ein Vergleich anhand ausgewählter Phänomene. *Diploma thesis*. Universität des Saarlandes, Fachrichtung 4.6. Angewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen.

Brinker, K. (2005). Linguistische Textanalyse: Eine Einführung in Grundbegriffe und Methoden. Berlin: Schmidt.

Braun, S. (2006). ELISA – a pedagogically enriched corpus for language learning purposes. In S. Braun, K. Kohn and J. Mukherjee (Eds.) Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods. Frankfurt am Main: Peter Lang, pp. 25--47.

Brinckmann, C., Kleiner, S., Knöbl, R. and Berend, N.

(2008). German Today: an areally extensive corpus of spoken Standard German. In *Proceedings 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakesch, Marokko.

Brown, G., Yule, G. (1983). Discourse Analysis. Cambridge: Cambridge University Press.

Döhring, S. (2006). Kulturspezifika im Film: Probleme ihrer Translation. Berlin: Frank & Timme.

Eckert, M., Strube, M. (2001). Dialogue Acts, Synchronising Units and Anaphora Resolution. In Journal of Semantics 17 (1), pp. 51--89.

Esser, J. (2009). Introduction to English Text-linguistics. Frankfurt a.M. u.a.: Peter Lang.

Evert, S. (2005). The CQP Query Language Tutorial. IMS, Universität Stuttgart.

Evert, S., Kermes, H. (2002). YAC – A Recursive Chunker for Unrestricted German Text. In M.G. Rodriguez, C.P. Araujo (Eds.), *Proceedings of LREC-2002*, Las Palmas, Spain, pp. 1805--1812.

Evert, S., Kermes, H. (2001). Exploiting large corpora: A circular process of partial syntactic analysis, corpus query and extraction of lexicographic information. In P. Rayson, A. Wilson, T. McEnery, A. Hardie, and S. Khoja (Eds.), *Proceedings of the Corpus Linguistics 2001 Conference*, Lancaster, England, pp. 332--340.

Gumul, E., Lyda, A. Disambiguating Grammatical Metaphor in Simultaneous Interpreting. In J. Maliszewski (Ed.) *Discourse and terminology in Specialist Translation and Interpreting*. Frankfurt am Main: Peter Lang, pp. 87--100.

Granger S. (2008). Learner Corpora. In Lüdeling, A., M. Kytö (Eds), *Handbook on Corpus Linguistics*. Mouton de Gruyter.

Gilquin, G., De Cock, S. and Granger, S (2010). Louvain International Database of Spoken English Interlanguage. (CD-ROM+ handbook). Presses universitaires de Louvain, Louvain-la-Neuve.

Greenbaum, S. (1996). Oxford English Grammar. Oxford: Clarendon Press.

Gundel, J.K., Hedberg, N. and Zacharski, R. (2005). Pronouns without NP Antecedents: How do we know when a pronoun is referential. In A. Branco, T. McEnery, and R. Mitkov (Eds.), *Anaphora Processing: Linguistic, Cognitive and Computational Modelling*, John Benjamins, pp. 351--364.

Gundel, J.K., Hedberg, N. and Zacharski, R. (2004). Demonstrative Pronouns in Natural Discourse. In *Proceedings of DAARC-2004 (the Fifth Discourse Anaphora and Anaphora Resolution Colloquium)*, Sao Miguel, Portugal.

Halliday, M.A.K., Hasan, R. (1976). *Cohesion in English*. London, New York: Longman.

Herbst, T. (1994). *Linguistische Aspekte der Synchronisation von Fernsehserien*. Phonetik, Textlinguistik, Übersetzungstheorie. Tübingen: Niemeyer.

Hinrichs, E.W. , Bartels, J., Kawata, Y., Kordoni, V. and Telljohann, H. (2000). The Tübingen treebanks for spoken German, English, and Japanese. In W. Wahlster

(Ed.), *Verbmobil: Foundations of Speech-to-Speech Translation, Artificial Intelligence*, Springer-Verlag, Berlin, Heidelberg, New York, Barcelona, Hong Kong, London, Milan, Paris, Singapore, Tokio*, pp. 550--575.

Klein, Y. (2007). Übersetzungsspezifische Eigenschaften – eine korpusbasierte Studie am Beispiel der Kohäsion. *Diploma thesis*. Universität des Saarlandes, Fachrichtung 4.6. Angewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen.

Kohn, K. (2001). Final report of the LLP BACKBONE project. Public part. University of Tübingen, Germany.

Kunz, K., Lapshinova-Koltunski, E. (2011). Tools to Analyse German-English Contrasts in Cohesion. In H. Hedeland, T. Schmidt, and K. Worner (Eds.). Multilingual Resources and Multilingual Applications. *Proceedings of the Conference of the German Society for Computational Linguistics and Language technology (GSCL)-2011*.

Kunz, K., Steiner, E. (*forthcoming*). *Towards a comparison of cohesion in English and German – contrasts and contact. Functional Linguistics*. London: Equinox Publishing Ltd.

Kunz, K., Maksymski, K. and Steiner, E. (2009). Suggestions for a corpuslinguistic analysis of cohesion. Deliverable No. 3 of the GECo Project. Available at: <http://fr46.uni-saarland.de/uploads/media/GECo_AP 3.pdf>.

Kunz, K. (2010). *Variation in English and German Nominal Coreference. A Study of Political Essays.* Frankfurt am Main: Peter Lang.

Lee, H., Peirsman, Y., Chang, A., Chambers, N., Surdeanu, M. and Jurafsky, D. (2011). Stanford's Multi-Pass Sieve Coreference Resolution System at the CoNLL-2011 Shared Task. In *Proceedings of the CoNLL-2011 Shared Task*, 2011.

Linke, A., Nussbaumer, M. and Portmann, P.R. (2001). *Studienbuch Linguistik*. 4 edition. Tübingen: Niemeyer.

Marneffe, M.C., MacCartney, B. and Manning, C.D. (2006). Generating Typed Dependency Parses from Phrase Structure Parses. In *Proceedings of LREC-2006*.

Müller, C., Strube, M. (2006). Multi-Level Annotation of Linguistic Data with MMAX2. In S. Braun, K. Kohn, J. Mukherjee (Eds.), *Corpus Technology and Language Pedagogy*. New Resources, New Tools, New Methods. Frankfurt: Peter Lang, (English Corpus Linguistics, Vol.3), pp. 197--214.

Pöchhacker, F. (2001). *Dolmetschen. Konzeptuelle Grundlagen und deskriptive Untersuchungen*. Tübingen: Stauffenburg.

Neumann, S. (2005). *Corpus Design. Deliverable No.1 of the CroCo Project*. Available at: <http://fr46.uni-saarland.de/croco/corpus_design.pdf>.

Schreiber, M. (1992). *Textgrammatik – Gesprochene Sprache – Sprachvergleich. Proformen im gesprochenen Französischen und Deutschen*. Frankfurt a.M.: Peter Lang.

Schubert, C. (2008). *Englische Textlinguistik*. Eine Einführung. Berlin: Schmidt.

Simpson, R.C., Briggs, S.L., Ovens, J. and Swales, J.M.

(2002). *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.

Vater, H. (2001). *Einführung in die Textlinguistik*. 3 edition. München: Fink.

Vater, H. (2005). *Referenz-Linguistik*. München: Fink.

LINDSEI-BR: an Oral English Interlanguage Corpus

Heliana MELLO , Luciana ÁVILA, Tufi Neder NETO, Barbara ORFANÒ

Universidade Federal de Minas Gerais, Faculdade de Letras – UFMG

Av. Antônio Carlos, 6627 – Pampulha – Belo Horizonte, MG 31270-901 Brazil

heliana.mello@gmail.com

Abstract

Corpus Linguistics has been more than instrumental in the study of interlanguage. It has made it possible for researchers not only to have access to large quantities of varied interlanguage samples but also process these data both for individual language features as well as for a host of other elements, sucha as interlanguage feature comparison. Presently there are many interlanguage corpora available to researchers and teachers, both written and oral, and this has afforded a spurt in interesting findings as far as the manyfold processes involved in language acquisition are concerned. In this paper, we will present a new English interlanguage corpus under compilation in Brazil, the LINDSEI-BR. It is associated with a larger project - the COBAI; the Brazilian Oral Corpus of Learner English is a repository of spoken interlanguage data that aims to gather varied subcorpora of Brazilian learner English with the main purpose of providing data for the study of interlanguage features within the frame of second language acquisition research. The larger project was launched in 2011 and so far it is concerned with the compilation of the LINDSEI-BR, a component of the Louvain-based project Louvain International Database of Spoken English Interlanguage.

Keywords: interlanguage; leaner oral corpus; Brazilian Portuguese; English.

1. Introduction

Corpus Linguistics has been more than instrumental in the study of interlanguage. It has made it possible for researchers not only to have access to large quantities of varied interlanguage samples but also process these data both for individual language features as well as for a host of other elements, such as interlanguage features at a given acquisition stage, comparative error analysis, among others. Presently there are many interlanguage corpora available to researchers and teachers, both written and oral, and this has afforded a spurt of interesting findings as far as the manyfold processes involved in language acquisition are concerned. In this paper, we will present a new English interlanguage corpus under compilation in Brazil. It is associated with a larger project - the COBAI. The Brazilian Oral Corpus of Learner English (COBAI) is a repository of spoken interlanguage data that aims at gathering varied subcorpora of Brazilian learner English with the main purpose of providing data for the study of interlanguage features within the frame of second language acquisition research. The project was launched in 2011 and so far it is concerned with the compilation of the LINDSEI-Brazil, a component of the r the LINDSEI international project, which will be presented in this paper.

The Louvain International Database of Spoken English Interlanguage (LINDSEI) project is an international initiative coordinated at the Centre for English Corpus Linguistics, at the Université Catholique de Louvain (cf. Gilquin, De Cock & Granger, 2010). The LINDSEI project encompasses seventeen different interlanguage subcorpora, compiled with the same parameters and transcribed following the same guidelines. The LINDSEI project is the oral counterpart for the ICLE – International Corpus of Learner English, compiled by the same team of researchers under the direction of Sylviane Granger (cf. Granger, 2003; Granger *et al*., 2009).

The LINDSEI-BR is being compiled following the international project guidelines. At present we have achieved our recording goal of fifty recordings and their transcription is underway. The recording informants were university, high intermediate to advanced level students of English as a second language. The recordings covered three different tasks: a narrative about a chosen set topic by the informant, free discussion with the interviewer and the description of a pictured scene. Each recording is on average twenty minutes long and features quasi spontaneous speech patterns. For each recording there is an accompanying learner profile that covers the learner's language history and other elements that might have contributed to her/his process of language acquisition, besides having information about the interviewer and the actual interview itself. The transcription guidelines include a code for each recording, speakers' turns, and the marking of several speech features, such as: overlapping, pauses, backchannelling, contractions, truncation, among others.

2. LINDSEI-BR participant profiles

Following the LINDSEI guidelines, all participants recorded are third, fourth year students of English. The participants are recruited by the researcher and are aware that they are contributing their speech to the compilation of a corpus. All participants have to fill in willingly a learner profile in which information about their acquisition history is reported through the number of years of study, context of English learning, etc. In order for a recording session to be incorporated into the corpus, participant permission is necessary.

The participants in the LINDSEI-BR study in a major federal university in Brazil and have chosen English as their major. Many are already ESL teachers, although this per se does not mean a level stage of acquisition among informants, as there might be sever

different proficiency levels even among ESL teachers.

3.   The recordings

Recording sessions took place in the first semester of 2011, and were carried at the Laboratory for Empirical and Experimental Language Studies (LEEL) at UFMG. The interviewer is a Brazilian with high proficiency level in English as a foreign language. It was not possible for the LINDSEI-BR team to arrange for a native speaker of English to carry the recordings. This is a shortcoming of the project since conversations might evolve differently between native and non-native speakers versus two non-native speakers of English with the same mother tongue background.

The recordings were carried with the following equipment: Recorder Marantz PMD660 Professional Solid State Recorder, unidirectional wireless microphones Sennheiser ME 4 clip-on (cardioid), cable Sennheiser CL100 (conectors XLR and mini jack 1/8"), receivers Sennheiser EM 100 G2 A, Sennheiser EK 100 G2 A, Sennheiser EK 100 G3and transmiter Sennheiser SK 100 G2 A and Sennheiser SK 100 G3.

Recording files are wav format and in general have good acoustic quality. Some sessions have some background noise but this does not prevent understandability.

4.   Some remarks about the transcriptions

Transcriptions are being carried at present by undergraduate research assistants. The transcribed files are revised by the project coordinator. No intertranscriber validation process has been carried so far but this is one of the goals the project to-do list.

Transcriptions follow the guidelines made available by the Leuven LINDSEI team and encompass the following aspects: a header, <h nt="BR" nr="BRXXX">, which indicates that participant number XXX is a native speaker of Brazilian Portuguese; turns are marked for interviewer <A> and interviewee <B>, each turn end carries the corresponding end tag, either </A> or </B>; overlapping is annotated at its onset with the tag <overlap /> in the undergoing turn and also at the beginning of the overlapper's turn, however its end is not annotated; the British orthographic convention is followed. There are several specific guidelines that cover empty pauses, filled pauses and backchannelling,   unclear passages, anonymisation, truncated words, contracted forms, non-standard forms, dates and numbers, some phonetic features, among others. An example of a transcription is given below:

Example 1:

<B> <overlap /> and: I'm going talk about a movie . that I saw ... <overlap /> (erm) . the: Inception .. with Leonardo DiCaprio . and I . I thought that . is a: very . good movie .. very interesting .. and: <overlap /> .. I don't   know it's so .. (eh) . first of all the the: . photograph= of the movie . is amazing . the: . special effects . that they use . is very nice . (erm) and it was

the first movie that I saw with my boyfriend= <laughs> <overlap /> and . we stayed for . three hours . in the cinema . and we: (eh) tired and the movie . (eh) (er) . how can I say </B>

As can be seen above, there are some specific markings such as: some end of words are followed by colons, which indicate last syllable lengthening (eg. and:); there are fillers such as (erm); non-verbal sounds are annotated (eg. <laughs>); truncated words are marked with = (eg. photograph=); silence is annotated through dots (eg. .., meaning 1-3 seconds).

The transcriptions do not contemplate pronunciation interlanguage features. In order for phonetic-based studies to be carried using this material, further annotation must be added.

5.   Future directions

LINDSEI-BR is still on the making; therefore, much remains to be done in order for it to be ready to be offered to researchers. However, plans are for the transcription process to be concluded within the year 2012. Additionally, some analysis has already been carried using data provided by this corpus, especially focusing on phonetic-phonological aspects of interlanguage speech (Medina, 2012).   Future plans upon transcription completion include the addition of interlanguage feature annotation in order to facilitate researchers' use of the corpus.

6.   References

Gilquin, G., De Cock, S. and Granger, S. (2010). *The Louvain International Database of Spoken English Interlanguage*. Handbook and CD-ROM. Presses universitaires de Louvain, Louvain-la-Neuve.

Granger, Sylviane. (2003). The International Corpus of Learner English: a new resource for foreign language learning and teaching and second language acquisition research. In *TESOL Quarterly*, 37(3), pp. 538--546.

Granger, S., Dagneaux, E., Meunier, F. and Paquot, M. (2009). *The International Corpus of Learner English* (Version 2). Handbook and CD-ROM.   Presses universitaires de Louvain, Louvain-la-Neuve.

Medina, V.H. (2012). L1 Brazilian Portuguese phonetic interference on L2 English. B.A. Thesis. Universidade Federal de Minas Gerais.

Bi-national bi-modal bi-lingual corpora of child language

Ronice Müller de QUADROS, Diane LILLO-MARTIN, Deborah CHEN PICHLER

Universidade Federal de Santa Catarina; University of Connecticut; Gallaudet University

ronice.quadros@ufsc.br, lillo.martin@uconn.edu, deborah.pichlerr@gallaudet.edu

Abstract

This paper discusses projects involving the building of corpora of sign language acquisition data. We developed a methodology to collect, to transcribe and to store data from different contexts of acquisition. The corpora include deaf children, from deaf parents; deaf children, from hearing parents; hearing children, from deaf parents (Codas) and deaf children with cochlear implants. There are two sign languages involved: Brazilian Sign Language and American Sign Language and two spoken languages, in the bilingual bimodal cases, that are, Brazilian Portuguese and American English. The complexity of building these corpora includes development of patterns of transcription and the organization of the same metadata system. In this process, we are developing manuals, database and software to make the data available and comparable across the languages. One example of software that we present in this paper concerns Sign ID, that is, it is software to indicate identities for each sign that is part of the database. The Sign ID software helps us make the annotations more consistent across transcribers. This kind of work is making it possible to compare data from these languages.

Keywords: sign language; corpora; and language acquisition.

1. Introduction

In order to address numerous linguistic research questions, we have been building several corpora of sign language acquisition data. Until recently, our focus had been on sign language only with deaf children, from deaf parents, acquiring sign language as native language. In this case, we built corpora of longitudinal data collected over a long period of time: these corpora included spontaneous data, with interaction of the child from 1-4 years old and an adult (usually the Deaf mother or a Deaf experimenter),. On the Brazilian side, there is also data from deaf children with hearing parents. In this context, a Deaf experimenter interacts with the child in sessions alternating with the hearing mother. All the analyses done so far indicate that in the specific context of deaf children with deaf parents, the sign language acquisition is parallel to spoken language acquisition (see Lillo-Martin, 1999 and Newport & Meier, 1985 for reviews of some of this). However, there are also findings showing that certain aspects of language acquisition in this context show modality effects (e.g. Meier & Newport, 1990; Marentette & Mayberry, 2000; Meier, 2006). On the other hand, in the context in which the deaf child has limited contact with sign language, there is a lot of variability in the language development reported by different researchers, but it seems that even in these contexts in which *input* is not conventional, because the child has parents learning sign language and restricted or no access to sign language, the child develops his/her signing skills better than his/her parents, showing that the child is able to make better use of the mental language system (e.g. Singleton & Newport, 2004; Goldin-Meadow, 2003; Goldin-Meadow & Mylander, 1984, 1990, 1998; Quadros & Cruz, 2011).

Now we are expanding our work to include bimodal bilingual children acquiring both a sign language and a spoken language, building comparable corpora across two sign/spoken language pairs: Brazilian Sign Language and Brazilian Portuguese on the one hand, and American Sign Language and American English on the other. We are again collecting longitudinal data with babies from 1 to 4 years old, and adding experimental data with children from 4 to 7 years old.

We use different sets of researchers (deaf and hearing) to emphasize appropriate target language use, assuming the child's interlocutor sensitivity (Petitto *et al.*, 2001), but we also recognize that code-blending is simply a part of the language system being acquired.

We reorganized the form of the database used with the longitudinal data and we built a new database for the experimental studies. The experimental studies include a set of 24 tests, evaluating different language aspects, such as, morphology, phonology, syntax, discourse and pragmatics. The goal of the tests is to provide a comprehensive profile of each bilingual child's developing competency in Libras (Brazilian Sign Language) and Brazilian Portuguese, or ASL (American Sign Language) and American English.

The data in sign and in speech adds considerable complexity to the already challenging prospect of corpus building. In this presentation, we explore some of the issues we have faced already and those we expect to face, in the context of our linguistic goals.

Recent research on childhood bilingualism has indicated that although children have two separate developing grammatical systems from very early on, there are instances of cross-linguistic influence, where grammatical structures from one language seem to exert a temporary influence on the child's grammar of the other language (e.g. Hulk & Müller, 2000). An important question is to identify the loci of such influences based on linguistic criteria. In order for us to address such issues, we are developing corpora from individual children acquiring both a sign language and a spoken language. Many of the same data collection issues arise as those for projects investigating only sign language (see Baker & Woll, 2005 for some best practices in this domain). However, in our current project, it turns out that there are specific things for which additional practices are needed; for instance, we frequently observe code-blended

language (the use of signs and speech produced simultaneously) as well as unimodal productions (Bogaerde & Baker, 2005, 2009; Emmorey *et al.*, 2008). Language- or modality-specific properties as well as universals are found to be very interesting in these contexts. In this paper, we will present the organization of the sign language acquisition corpora developed on both sides of the project: Brazil and the United States of America.

2.   Metadata

The metadata of the children is organized through documents that are shared with researchers involved in the different steps of the investigation: data collection involving filming, transcribers, people that organize the data for specific purposes and people that analyse the findings. The main topics of the documents are the following:

LONGITUDINAL
- Protocol of the child (nickname of the child, for example, EDU)
- Number of the section (from 000 up to the number of the sections collected, for example, EDU001, EDU002, EDU003)
- Date of the filming
- Age of the child (years;months.days)
- Target language
- Duration of the session
- Adults involved in the session
- Other participants involved in the session
- Comments
- Transcribers
- Checker/reviser of the transcription
- Organizer of the data for each purpose (for example, for WH analysis, for Modality analysis, etc.)

EXPERIMENTAL
- Name of the test
- Nickname of the child
- Condition (Coda, Deaf, CI, Coda adult)
- Date
- Age
- Language
- Duration
- Comments
- Transcriber
- Reviser

The whole database is organized in a computer server. See Figure 1 for an illustrative sample of this organization. There are two main folders: the original archive ("*acervo*") and the production. The first one has the original videos. The second one has the compressed videos for manipulation by the people that access the videos, as well as transcription and analysis files.

The production folder includes the experimental data and longitudinal data in separate sections. First we

discuss the longitudinal data. The basic organization is to list the children in separate folders. Each child's folder will include the folders for each session containing the video and the transcript files (the basic one and the ones with the specific organization for specific purposes). The transcription is done using ELAN software producing eaf files with separate tiers of annotation capturing different types of information (see also below).

For the experimental studies, the basic organization is to have the folders with the places and years in which the fairs happened. Within each place, the folders are separated by test. These folders are further divided into two sets of data by child: one for those whose data is without restriction *("sem restrição")*, and another for restricted data ("*com restrição*"). The restrictions are related to the kind of access people have to the videos. Some of the parents do not want students to have access to the videos of their child or for the researchers to use frames of the videos in conferences, for example. Within these two folders based on restriction, the children, then, are listed with the video and the eaf or the form of the test scanned with the results, depending on each test.

In the case of the experimental studies, the database is organized as well as using FileMakerPro (Figure 2 in Appendix). This database includes all four languages. Then, it facilitates the comparison among the experimental results over the four languages.

Figure 1: Example of the organization of the database

3.   Designing annotation patterns

Following video collection, we invest considerable energy in the production of transcripts, to be used in conjunction with the videos for linguistic analyses. Following our earlier sign-only research, we use ELAN for time-locked videos with transcription (http://www.latmpi.eu/tools/elan/).

For bilingual research, we designed a different template so that both languages are parent tiers, to optimize the study of (sequential or simultaneous) bimodal productions. See Chen Pichler *et al.* (2010) for a detailed description of our ELAN tier structure and transcription conventions (cf. Figure 3 and Figure 4, in Appendix).

The general principles that guide the annotation of the data are to create a machine-readable record of language samples, not necessarily sufficient for the reader to reproduce in exactly the same way, but so that the records can be searched to find all occurrences of given phenomena of interest (in the way described by Johnston, 2001, Johnston & Schembri, 2007; Miller, 2001; Pizzuto & Pietrandrea, 2001). In addition to having a basic annotation of the utterance in each language, we use multiple annotation parses focusing on different phenomena. This documentation of the data is the foundation for our analysis decisions.

Where it is possible, we follow the CHILDES conventions established for child language data (MacWhinney, 2000) in transcribing both speech and sign (though we do not use BTS) http://childes.psy.cmu.edu/manuals/chat.pdf. When the CHILDES conventions conflict with our sign-specific goals, we create new conventions to be followed for transcribing both sign and speech. It is important to keep the sign and speech transcriptions comparable.

4.   Sign IDs

Finally, we see a number of important implications and extensions of the system we are developing. For example, we are creating a specific identification for each sign to be used in our transcripts (in the same spirit of Johnson, in preparation, for Australian Sign Language), what we call "Sign ID". Because there is no commonly accepted writing system for sign languages, sign researchers generally rely on a system of glossing; however, traditional transcription does not assign a consistent gloss for each sign, but different glosses depending on context and other aspects of the signed utterance. This means that it is very difficult for researchers to identify the locations of interest in a transcript using a search function to discover all occurrences of a particular sign. Analysis must proceed at a much slower pace of hand searching transcripts one utterance at a time. In order to facilitate and expand the analysis of data collected in the parent project, we developed a sign ID lexicon containing the vocabulary items used most frequently by the children we are studying. Sign IDs are word labels chosen to represent each sign root systematically, so that every use of the sign

has the same label, despite contextual or morphological differences which affect how the sign is interpreted. By using sign IDs in our transcripts, we are able to conduct our analyses more efficiently, using a wider range of data. The sign ID lexicon addresses the problem of transcript searchability and greatly facilitates the analysis of data collected for sign language corpora. This helps to standardize annotations and it can be more freely accessed by other researchers.

On the Brazilian side, we have been developing the sign IDs database by feeding it with the signs over which transcribers had doubts regarding transcription. We have periodic meetings to discuss these signs, then we christen each and add it to the ID list (www.idsinais.libras.ufsc.br) (see Figure 5 in Appendix for the Sign ID screen). The search system has filters based on sign language parameters (132 handshapes divided in 13 groups and 8 locations). An example with a group of handshapes chosen as a parameter to search for a specific sign is given in Figure 6 and the results of this search are shown in Figure 7, in Appendix.

The sign ID specifications include identification of the sign, Portuguese translation, English translation, written sign, handshape groups, handshapes, location and sign video. The searching may be done through handshapes, locations, handshape groups, location groups, the sign ID or the first letter of the sign ID.

On the American side, the development of an ID gloss database has taken into consideration the needs of different research groups across the country, each of which uses a different system for writing signs. The database was set up so that different local groups can enter their own information about each sign, and each group can also view the information entered by the others. This approach will facilitate the comparison of transcriptions used across different groups, and may eventually lead to greater convergence in the glossing systems used.

5.   Conclusion

One of our major goals has been cross-site comparability, that is, establishing the same criteria, approach to data collection, ELAN template, and general transcription principles to be used across our three universities. The metadata and data are shared through the use of a common server, as well as online services including Google docs and Dropbox. The analyses of the results are being conducted through regular meetings and we are on the right track to answer our research questions (e.g., Lillo-Martin *et al.*, 2010; Chen Pichler *et al.*, 2010; Quadros *et al.*, in press).

We have not yet resolved the following linguistic issues, but we hope that our project will contribute to their discussion in the field as a whole. Does bimodal bilingualism lead to cross-language influence different from that found in mono-modal bilingualism (e.g., due to code-blending, or use of non-manuals)? When bimodal bilinguals code-blend, are they choosing grammatical structures which are permitted in both languages for maximum accommodation? What kinds of syntactic

representations can account for code-blends? These are the types of research questions our project can address through the use of the corpora we are now building.

Our template and corpus-building decisions can be applicable to the development of adult only bimodal bilingual corpora. In addition, many similar issues arise in the study of co-speech gesture, and researchers in this area may take advantage of aspects of our procedures. And, we hope that our collaboration across continents may contribute to and promote cross-linguistic research on sign languages as well.

6.   Acknowledgements

7.   References

Baker, A., Woll, B. (Eds.) (2009). Sign language acquisition. Amsterdam: John Benjamins.

Bogaerde, B. van den, Baker, A.E. (2005). Code-mixing in mother-child interaction in deaf families. In *Sign language & linguistics*, 8(1-2), pp. 151--174.

Bogaerde, B. van den, Baker, A.E. (2009). Bimodal language acquisition in Kodas (kids of deaf adults). In M. Bishop, S.L. Hicks (Eds.), *Hearing, mother-father Deaf: Hearing people in Deaf families*, Washington, DC: Gallaudet University Press, pp. 99--131.

Chen Pichler, D., Hochgesang, J., Lillo-Martin, D. and Quadros, R. M. (2010). Conventions for sign and speech transcription of child bimodal bilingual corpora in ELAN. In *Language, Interaction and Acquisition, 1*, pp. 11--40.

Chen Pichler, D., Quadros, R.M. and Lillo-Martin, D. (2010). Effects of Bimodal Production on Multi-Cyclicity in Early ASL and LSB. In J. Chandlee, K. Franich, K. Iserman, and L. Keil (Eds.), *A Supplement to the Proceedings of the 34th Boston University Conference on Language Development*. Available at: <www.bu.edu/linguistics/BUCLD/supp34.html>.

Emmorey, K., Borinstein, H.B., Thompson, R. & Golan, T.H. (2008). Bimodal bilingualism. In *Bilingualism: Language and cognition*. 11(1), pp. 43--61.

Goldin-Meadow, S. (2003). *The resilience of language: what gesture creation in deaf children can tell us about how all children learn language*. New York: Psychology Press.

Goldin-Meadow, S., Mylander, C. (1984). Gestural communication in deaf children: The effects and noneffects of parental input on early language development. Monographs of the Society for Research in Child Development, 49 (3–4, Serial No. 207).

Goldin-Meadow, S., Mylander, C. (1990). Beyond the input given: The childs role in the acquisition of language. In *Language*, 66, pp. 323--355.

Goldin-Meadow, S., Mylander, C. (1998). Spontaneous sign systems created by deaf children in twocultures. In *Nature*, 391, pp. 279--281.

Hulk, A., Müller, N. (2000) Bilingual first language acquisition at the interface between syntax and pragmatics. In *Bilingualism: Language and Cognition* 3 (3), 2000, Cambridge University Press, pp. 227--244.

Johnston, T. (in preparation). *From archive to corpus: Transcription and annotation in the creation of signed language corpora, manuscript.* Department of Linguistics, Macquarie University, Australia.

Lillo-Martin, D. (1999). Modality effects and modularity in language acquisition: The acquisition of American Sign Language. In T. Bhatia & W.C. Ritchie (Eds.), *Handbook of Language Acquisition,* San Diego: Academic Press, pp. 531--567.

Lillo-Martin, D., Quadros, R.M., Koulidobrova, H. and Chen Pichler, D. (2010). Bimodal Bilingual Cross-Language Influence In Unexpected Domains. In J. Costa, A. Castro, M. Lobo and F. Pratas (Eds.), *Language Acquisition and Development: Proceedings of GALA 2009*, Newcastle upon Tyne: Cambridge Scholars Press, pp. 264--275.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Marentette, P., Mayberry, R. (2000). Principles for an emerging phonological system: A case study of acquisition of American Sign Language. In C.D. Chamberlain, J.P. Morford and R. Mayberry (Eds.), *Language Acquisition by Eye*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 51--69.

Meier, R. (2006). The form of early signs: Explaining signing children's articulatory development. In B. Schick, M. Marschark and P.E. Spencer (Eds.), *Advances in Sign Language Development by Deaf Children*, New York: Oxford University Press, pp. 202--230.

Meier, R.P., Newport, E.L. (1990). Out of the hands of babes: On a possible sign advantage in language acquisition. In *Language*, 66, pp. 1--23.

Newport, E.L., Meier, R.P. (1985). The acquisition of American Sign Language. In D.I. Slobin (Ed.), *The Cross-Linguistic Study of Language Acquisition, Volume 1*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 881--938.

Petitto, L.A., Katerelos, M., Levi, B., Gauna, K., Tetrault, K. and Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. In *Journal of child language.* 28(2), pp. 453--496.

Quadros, R.M., Lillo-Martin, D. and Chen Pichler, D. (in press). Early effects of bilingualism on WH-question structures: Insight from sign-speech bilingualism. In *Proceedings of GALA 2011*. Newcastle upon Tyne:

Cambridge Scholars Press.

Singleton, J.L., Newport, E., (2004) E. When learners surpass their models: The acquisition of American

Sign Language from inconsistent input. In *Cognitive Psychology* 49. pp. 370--407.

8. Appendix

Figure 2: FileMakerPro

Figure 3: ELAN in the context of Bibibi Project with the basic tiers for the child

Figure 4: ELAN in the context of Bibibi Project with the specific tiers for modality analysis

Figure 5: ID screen for Libras
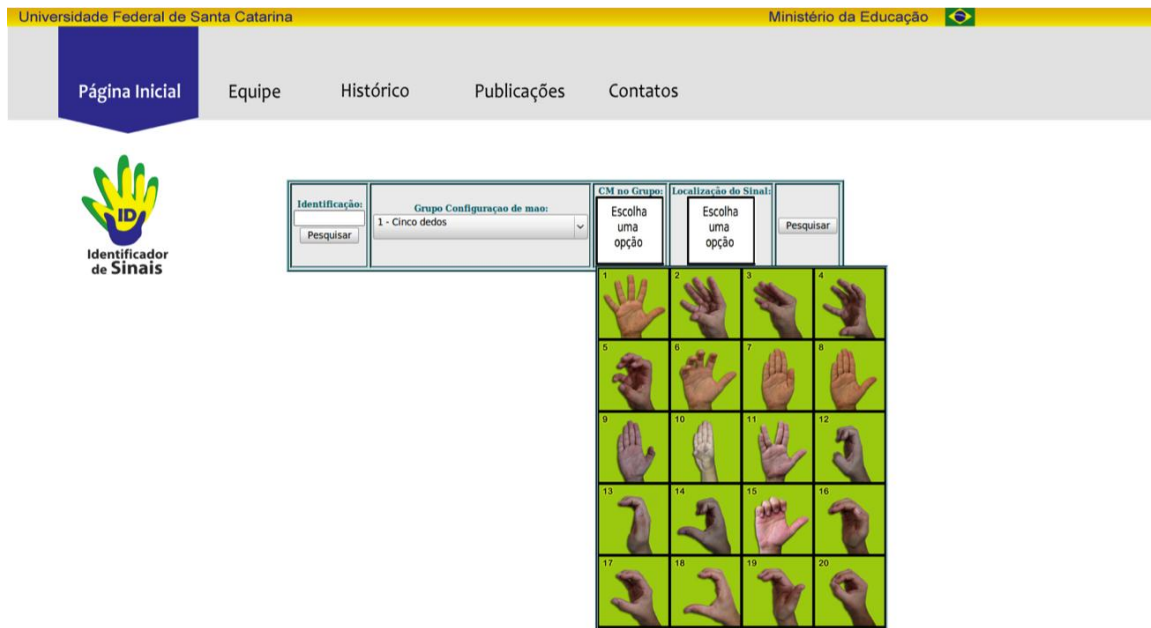
Figure 6: ID searching system: Handshape selection
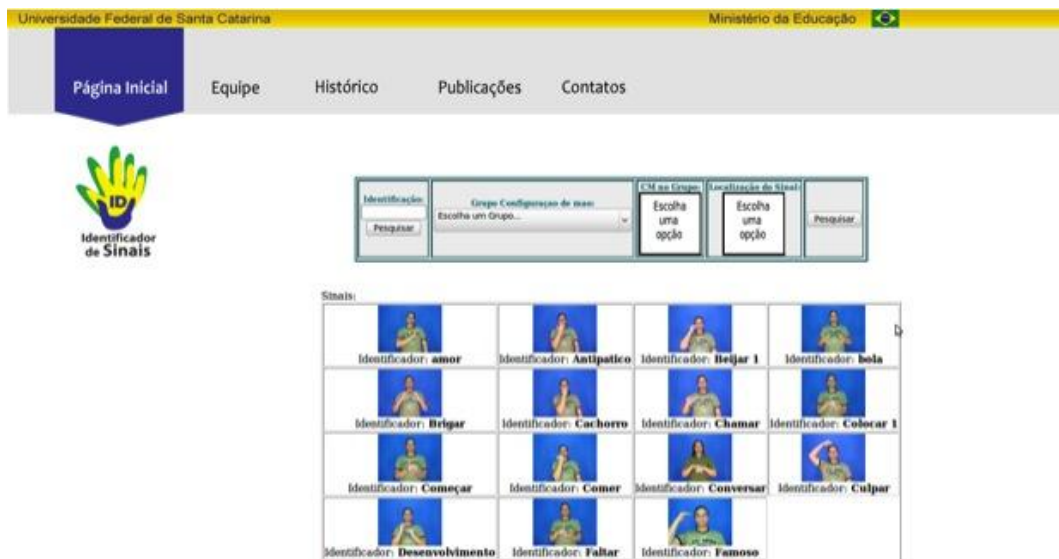
Figure 7: ID result of a search

C-Or-DiAL (Corpus Oral Didáctico Anotado Lingüísticamente) y la enseñanza del español

Carlota Nicolás MARTÍNEZ

Università degli Studi di Firenze

Via Santa Reparata, 93 – 50122 Firenze

carlota.nicolas@unifi.it

Abstract

En este artículo se describen las características, el proceso de elaboración y la utilidad didáctica de C-Or-DiAL. Este corpus está formado por 118.756 palabras que proceden de alrededor de diez horas de grabación de los siguientes géneros discursivos: conversación y diálogo (29%), entrevista informal (51%), conversaciones con tema preestablecido (13%) y clases o conferencias (7%). El texto etiquetado de la transcripción está precedido de una cabecera con informaciones generales sobre la elocución (participantes, situación, tema y palabras claves) y otras más específicas con indicaciones y propuestas para la enseñanza de la lengua (nivel del alumnado con el que usar la sesión, lista de palabras poco usadas, aspectos lingüísticos y funciones comunicativas que se pueden aprender en esa elocución). Este rico corpus está a disposición de quien quiera renovar el modo de enseñar y de aprender la lengua oral espontánea.

Keywords: Corpus orales del español; enseñanza de la lengua; géneros discursivos; base de datos.

1. Características y elaboración de C-Or-DiAL

1.1 Qué es

Es un corpus de la lengua oral espontánea recogida en grabaciones y transcrita ortográficamente, etiquetada prosódicamente y con las funciones comunicativas anotadas.

Es un corpus que además de ser un recurso para la investigación puede ser utilizado como material para la enseñanza de la lengua española. Para facilitar este uso se ofrecen en la cabecera de cada texto indicaciones y propuestas específicas para la enseñanza (nivel del alumnado con quien usar el texto, lista de posibles palabras desconocidas, observaciones lingüísticas).

1.2 Quién lo ha hecho

Proyectado y estructurado con la ayuda de Massimo Moneglia y Alessandro Panunzi. Creación de la base de datos: Lorenzo Gregorio. Fundamentos teóricos: Emanuela Cresti. Transcripciones: alumnos de Lengua Española de los cursos 2005-2012 de la Università di Firenze, corregidas por Carlota Nicolás. Colaboración para la anotación de las funciones: Martina Viliani. Grabaciones, fragmentación de las grabaciones en sesiones, reelaboración de los criterios de transcripción y revisiones globales: Carlota Nicolás.

1.3 Cuándo se ha hecho

La primera grabación se hizo en el 2004.

El corpus se introdujo en la base de datos en noviembre del 2012. El libro sobre C-Or-DiAL se publica en julio de 2012.

Las transcripciones son revisadas y corregidas periódicamente.

1.4 Cuánto material contiene

Es un corpus de dimensión media: 118.756 palabras transcritas que proceden de alrededor de diez horas de audio.

Son 240 sesiones compuestas por la transcripción de los correspondientes 240 audios; estos ha sido extraídos de las 72 horas de grabaciones hechas en los últimos 9 años.

En la Tabla 2 (Appendix) se muestra el número de palabras de cada uno de los géneros que forman la estructura de C-Or-DiAL.

1.5 Dónde se ha hecho

Las grabaciones de C-Or-DiAL se han hecho en Madrid con el apoyo técnico del Laboratorio de Lingüística Computacional de la Universidad Autónoma de Madrid.

Las transcripciones y toda la elaboración del corpus se ha hecho en la Università di Firenze.

1.6 Dónde consultar el corpus C-Or-DiAL

1.6.1. C-Or-DiAL base de datos

Las sesiones C-Or-DiAL se pueden extraer de la base de datos que lo aloja en LABLITA (Laboratorio de Linguistica Italiana) Università di Firenze (lablita.dit.unifi.it/app/C-Or-DiAL/index.php).

En la Tabla 3 (Appendix) se muestra la página *Acceso a la sesiones de C-Or-DiAL* desde la que abrir cada texto y cada audio del corpus, y donde consultar las informaciones sobre cada sesión: *Título y tema*, *Tipología de los textos*, *Número de hablantes*, *Situación*, *Número de palabras*, *Minutos*, *Uso didáctico* (lablita.dit.unifi.it/app/C-Or-DiAL/corpus.php). También se accede al corpus desde la *Búsqueda avanzada* (lablita.dit.unifi.it/app/C-Or-DiAL/search.php) que utiliza listas cerradas con informaciones sobre: *Tipología de* texto, *Palabras clave*, *Nivel de uso didáctico* y *Funciones comunicativas*.

1.6.2. Libro de C-Or-DiAL

Se ha editado en el 2012 el libro *C-Or-DiAL (Corpus Oral Didáctico Anotado Lingüísticamente)* publicación

de LICEUS EDICIONES en dos formatos: en papel, acompañado de un cd, y en formato electrónico. Esta publicación contiene el corpus con todas sus sesiones y, además, una detallada descripción sobre la elaboración, las características y los posibles usos didácticos de C-Or-DiAL.

2.   Estructura y contenidos de C-Or-DiAL

2.1   Qué estructura tiene

2.1.1.   Macroestructura

C-Or-DiAL contiene 240 sesiones compuestas por las transcripciones y los audios correspondientes.

Estas sesiones tienen diferentes tamaños y géneros discursivos.

La distribución de C-Or-DiAL en géneros discursivos se ve en la Tabla 1 (Appendix), en la que se evidencia, mediante cuatro parámetros de clasificación, el rasgo de espontaneidad en la lengua que predomina en este corpus.

Los tamaños de las sesiones son:

- 89 audios de hasta 2 minutos (01:58:37 horas);
- 67 audios de 2 a 3 minutos (02:46:48 horas);
- 57 audios de 3 a 4 minutos (03:09:47 horas);
- 27 audios de 4 a 8 minutos (02:18:15 horas).

2.1.2.   Microestructura

Cada texto de las sesiones está compuesto por la cabecera y el texto transcrito.

La cabecera tiene datos y metados:
- Informaciones sobre las características básicas de la sesión (número de minutos y de palabras, grabación de la que procede el fragmento, nombres de los archivos y de los transcriptores y revisores);
- Informaciones del contenido del texto (tema, informaciones sobre los participantes, situación en la que transcurre la elocución);
- Indicaciones y propuestas específicas para la enseñanza (nivel del alumnado con el que usar la sesión, lista de palabras poco usadas y de interés para ser estudiadas, aspectos lingüísticos y funciones comunicativas que se pueden aprender en esa elocución).

2.2   A quién se ha grabado

Los participantes de C-Or-DiAL, todos ellos anónimos, son más de 50. Denominados con tres letras mayúsculas que mantienen en todas sus intervenciones. La cultura de los participantes es media-alta (universitarios en general). Son personas de mediana edad (entre 30 y 60 años), solo ocho de estas personas tienen menos de 10 y más de 70 años.

El 99% de las personas es de Madrid.

2.3   Cómo se recoge el habla

El 30% de las grabaciones se han hecho sin que las personas supieran que se les estaba grabando; en estos casos se ha pedido el permiso para utilizarla al acabar la grabación. El 7% son grabaciones hechas en salas de conferencias y en aulas; el 63% restante se ha hecho pidiendo permiso a los participantes antes de iniciar la grabación, en estos casos la relación de amistad o la situación familiar hacía que la grabadora no fuera un impedimento para que se hablara con gran naturalidad.

2.4   Cuándo y dónde se recoge el habla

Las grabaciones se han hecho en distintos momentos del día.

Los lugares de las grabaciones han sido los normales de la vida cotidiana: casas particulares, cafés o bares. Han sido grabadas en lugares de trabajo las conferencias, las clases, las específicas sesiones de trabajo, y cinco entrevistas de las 20 realizadas.

2.5   Cómo se transforma el habla de la grabación en el texto de la transcripción

El primer paso para crear las sesiones ha sido fragmentar las grabaciones originales de larga duración en audios de pequeño tamaño (ver 2.1.1.). En cada audio se habla al menos de un tema claro, este ha sido el criterio de fragmentación.

Estos audios se fueron entregando a los alumnos de la Università di Firenze de los cursos de Lengua Española del 2005 al 2012 que tuvieron la obligación de hacer con cada uno su transcripción como parte del programa del curso.

Para la transcripción se usaron reglas que derivan de las usadas en C-ORAL-ROM.

Las correcciones y control final de todas las transcripciones es responsabilidad de Carlota Nicolás.

3.   Utilización de C-Or-DiAL

3.1   Qué uso se puede hacer de C-Or-DiAL

C-Or-DiAL está diseñado como corpus para la investigación y para la enseñanza de la lengua oral.

Hasta ahora C-Or-DiAL se ha utilizado como valioso contenedor de muestras reales de la lengua oral espontánea, en el que analizar sus características. En estos años de trabajo con los alumnos de Lengua Español se ha constatado que hacer transcripciones es un modo muy válido para el análisis de la lengua oral.

La transcripción se ha revelado como un método de aprendizaje de gran impacto, pues despierta en el alumno actitudes hacia el aprendizaje poco desarrolladas al trabajar con otros métodos más usuales. La labor del transcriptor no solo es una práctica de minuciosidad y de concentración muy pedagógica, sino que aporta este patrimonio:

- la atención obligada para entender un audio habitúa a escuchar con especial atención;
- el traslado del audio a la escritura (aunque solo sea una trascripción ortográfica sin seguir las normas de puntuación)   enseña a diferenciar

estas dos modalidades;

- marcar los rasgos prosódicos que dependen de la percepción del transcriptor los hace reconocer conscientemente;
- la colocación en la transcripción de las etiquetas obliga a hacer un análisis solo posible si se han aprendido algunas características fundamentales de la lengua oral que son representadas por estas etiquetas.

3.2   Con quién y para quién utilizar C-Or-DiAL

C-Or-DiAL puede ser utilizado en la enseñanza de la lengua española con alumnos de todos los niveles, con ayuda del profesor o, sin ella, realizando su estudio en autonomía.

3.4   Cuándo y dónde utilizar C-Or-DiAL

En cualquier momento del proceso de aprendizaje de la lengua española se pueden incluir, para su estudio, las sesiones de C-Or-DiAL. Un profesor de lengua sabrá adaptar cada sesión al nivel del alumno. C-Or-DiAL es requerido para que el alumno tenga contacto con el español real espontáneo que es el español que necesita comprender y con el que se debe expresar.

Para trabajar con C-Or-DiAL es necesario el uso de un laboratorio informático para que el alumno pueda acceder a las transcripciones y a los audios individualmente y pueda con su propio ritmo trabajar con este material.

4.   Relación entre el desarrollo de las habilidades personales del estudiante y la práctica de las destrezas lingüísticas

Al trabajar con C-Or-DiAL se activa la concentración, la percepción auditiva y la necesidad de segmentar lo escuchado para poder llegar a la comprensión oral.

La comprensión oral de los textos del C-Or-DiAL lleva a ejercitar el análisis, la deducción, la inducción y la síntesis.

A partir de los textos de C-Or-DiAL se pueden hacer ejercicios de imitación y recreación lo que conlleva la práctica de la expresión oral, la interacción oral y la expresión escrita.

5.   Tres propuestas de actividades para distintos tipos de aprendientes

5.1   Contacto inicial con una sesión de C-Or-DiAL

Actividades preferidas por aprendientes de estilo pragmático y de estilo activo:

- Audición;
- Reconocer variantes prosódicas;
- Coger notas de lo que se oye;
- Buscar las palabras clave;
- Escribir el tema;
- Separar y reconocer palabras, locuciones y

colocaciones;

- Reconocer funciones comunicativas;
- Observar aspectos gramaticales.

5.2   A partir de la sesión de C-Or-DiAL

Actividades preferidas por aprendientes de estilo pragmático:

- Dramatización a partir del texto
- Cambiar entonaciones del texto y observar los efectos;
- Escribir lo dicho en el texto con la estructura de una obra dramática;
- Hacer un guión cinematográfico añadiendo movimientos y situación ;
- Escribir el resumen de lo sucedido;
- Inventar lo anterior o lo posterior dicho o sucedido entorno al texto.

5.3   Utilización de los recursos de C-Or-DiAL

Actividades preferidas por aprendientes de estilo teórico y de estilo reflexivo:

- Aprender particularidades prosódicas;
- Subdividir los enunciados;
- Analizar las peculiaridades discursivas;
- Reconocer diferencias entre géneros discursivos;
- Controlar las funciones comunicativas relevantes en el texto;
- Observar la estructura temática;
- Conocer la estructura dialógica;
- Aprender palabras, locuciones y colocaciones nuevas.

6.   Conclusiones

El mejor modo de concluir esta descripción de C-Or-DiAL y de su uso es presentar una sesión en la que se observan algunas de sus cualidades. Es una conversación entre amigas que no sabían que eran grabadas. Se puede observar en ella su espontaneidad, un modo ejemplar de estructurar la narración, una cierta riqueza de vocabulario, además de otros muchos detalles que se pueden encontrar, y que serán especialmente apreciados por los profesores que buscan materiales reales y ricos para sus estudiantes.

7.   Appendix

7.1   Transcripción

@Archivos:
conv_03_UNA_CHIQUITA_JAPONESA.txt,
conv_03_UNA_CHIQUITA_JAPONESA.wav 🔊))
@Título: una chiquita japonesa
@Participantes: CAR, Carlota (mujer, C, 3, profesora, Madrid, vive en Italia desde hace más de 20 años)
PIZ, Pizca (mujer, C, 3, archivadora, Madrid)
ANG, Ángeles (mujer, C, 3, traductora, Madrid, vive en Bélgica desde hace 25 años)

ISA, Isabela (mujer, C, 3, arquitecto, Madrid)
MAI, Maite (mujer, C, 3, editora, Madrid)
VIR, Virginia (mujer, C, 3, gestora, Madrid)
@Relación entre los participantes: compañeras de colegio desde los 6 años hasta los 17, se ven en raras ocasiones
@Situación: en el salón de casa de PIZ a media tarde
@Tema: el sorprendente modo de viajar de una jovencita japonesa que ha sido huésped en casa de MAI en verano
@Palabras clave: juventud
@Uso didáctico: A2
@Nivel para la comprensión del texto: B1
@Palabras nuevas: japonés, autobús, maletón, bromear, marcharse, violador, pelos de punta, dar tumbos, ámbito, agarrar, hala
@Funciones comunicativas: 1.7 narrar, contar, describir, referir y relatar, 2.2 dar una opinión, valorar, 6.16 introducir palabras de otros y citar
@Observaciones lingüísticas: enunciados complejos; incisos; organización del discurso; enunciación ininterrumpida; citas
@Duración y número de palabras: 00:01:21 - 295
@Transcriptores y revisores: Carlota Nicolás Martínez
@Grabación original: 03_AMIGAS.wav, 2004, Madrid, 01:44:29
*MAI: 1.7 y este verano tuvimos en casa a una japonesa / una cría de veintiún años / vino a casa [///] había estado tres meses en Sevilla / estudiando español // nos aparece / fuimos a buscarla a la estación de [/] de autobuses / te aparece una japonesita así jovencísima con un maletón / con su ordenador portátil / con el que &mm se comunicaba con su familia claro tal // dices pero esta chica / aquí / en España / primero a Sevilla / luego se viene a [/] a Madrid / a casa de un amigo / 6.16 que le decía a Ramón / pues porque somos gente decente / pero es que puedes aterrizar < en casa de un >

6.16 ...
*PIZ: < yyy claro > //
*ANG: < en cualquier &sit > //
*MAI:/ se conocían de un foro en el que hay españoles que estudian japonés / y japoneses que estudian español / un foro de Internet /
*PIZ: ya //
*MAI:/ ¿ sabes ? y dices / y de pronto cogen la maleta / y se < colocan en el otro lado del mundo > /
*PIZ: < del Japón > ...
*MAI:/ a casa de uno que se llama Ramón
*TTT: yyy
*MAI:/ y que lo has conocido ... y yo / luego le bromeaba / porque después de casa / se marchó a Barcelona a casa de otro del foro / 6.16 yo le decía de broma / &eh ¿ ha llegado ya a casa del violador del Ensanche 6.16 1.7 ?
*CAR: yyy
*MAI:/ 2.2 porque es que dices / es que a mí me pone los pelos de punta / ¿no? 2.2 //
*ANG: sí //
*MAI:/ 1.7 los padres de esta chica se quedan tan contentos / en Japón //
*ANG: 2.2 bueno no / tan contentos no / es que veintiún años ya / si no están contentos ... < va a ser peor > 2.2 //
*TTT: yyy
*PIZ: < les va a dar igual ¿no? > //
*MAI:/ < y la niña / dando tumbos > ... había estado / en otros viajes en Australia / en el norte de Marruecos / en París / en no sé qué / dices / realmente es que para estos el mundo es todo // o sea su [/] su ámbito es todo // agarran la maleta se suben en un avión y < ¡hala! / por todas partes > 1.7 //
*ANG: < se largan > //
*XYZ: xxx

7.2 Tablas

Géneros discursivos y porcentaje de tiempo en C-Or-DiAL	Parámetros de clasificación de la espontaneidad de los textos			
	Lazos familiares o de intimidad	Lugar familiar (casa, café, bar, jardín)	Papel determinado de los hablantes	Tema u objetivo preestablecido
conversaciones 24% conv	100%	100%	-	2%
diálogos 5% dial	100%	100%	-	-
entrevistas 54% entr	99%	99%	100%	-
charlas 5% char	100%	80%	-	100%
fin predeterminado 2% finp	100%	100%	-	100%
trabajo 6% trab	85%	-	100%	-
clases 5% aula	-	-	100%	100%
conferencias 2% sala	-	-	100%	100%

Tabla 1: Clasificación de espontaneidad de los géneros discursivos de C-Or-DiAL
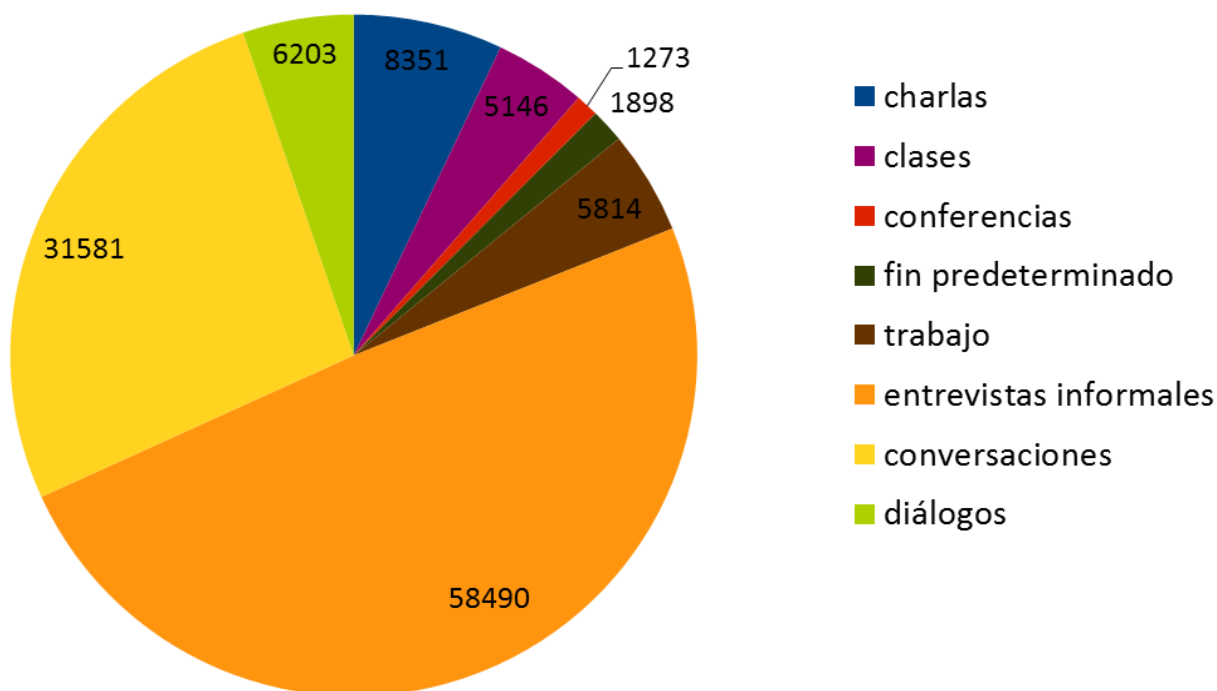
Tabla 2: Proporciones de palabras por género

Tabla 3: Sitio de C-Or-DiAL en LABLITA. Página de acceso directo a los archivos

Extension of the LECTRA corpus: classroom LECture TRAnscriptions in European Portuguese

Thomas PELLEGRINI[1], Helena MONIZ[1,2], Fernando BATISTA[1,3], Isabel TRANCOSO[1,4], Ramon ASTUDILLO[1]

[1]Spoken Language Systems Lab, INESC-ID, Lisbon, Portugal; [2]Faculdade de Letras, Universidade de Lisboa, Lisbon, Portugal; [3]ISCTE-IUL - Instituto Universitário de Lisboa; [4]Instituto Superior Técnico, Lisbon, Portugal
thomas.pellegrini@inesc-id.pt, helena.moniz@inesc-id.pt

Abstract

This paper presents the recent extension of the LECTRA corpus, a speech corpus of university lectures in European Portuguese that will be partially available. Eleven additional hours of various lectures were transcribed, following the previous multilayer annotations, and now comprising about 32 hours. This material can be used not only for the production of multimedia lecture contents for e-learning applications, enabling hearing impaired students to have access to recorded lectures, but also for linguistic and speech processing studies. Lectures present challenges for automatic speech recognition (ASR) engines due to their idiosyncratic nature as spontaneous speech and their specific jargon. The paper presents recent ASR experiments that have clearly shown performance improvements on this domain. Together with the manual transcripts, a set of upgraded and enriched force-aligned transcripts was also produced. Such transcripts constitute an important advantage for corpora analysis, and for studying several speech tasks.

Keywords: lecture domain speech corpus, ASR, speech transcripts, speech alignment, structural metadata, European Portuguese.

1. Introduction

This paper aims at a description of the corpus collected within the national project LECTRA and its recent extension. The LECTRA project aimed at transcribing lectures, which can be used not only for the production of multimedia lecture contents for e-learning applications, but also for enabling hearing-impaired students to have access to recorded lectures. The corpus has been already described in (Trancoso *et al*., 2008). We describe the recent extension of the manual annotations and the subsequent automatic speech recognition and alignment experiments to illustrate the performance improvements compared to the results reported in 2008. The extension was done in the framework of the METANET4U European project that aims at supporting language technology for European languages and multilingualism. One of the main goals of the project is that languages resources are made available online. Thus, the LECTRA corpus will be available through the central META-SHARE platform and through our local node: *http://metanet4u.l2f.inesc-id.pt/.*

Lecture transcription can be very challenging, mainly due to the fact that we are dealing with a very specific domain and with spontaneous speech. This topic has been the target of much bigger research projects such as the Japanese project described in Furui *et al.* (2001), the European project CHIL (Lamel *et al.*, 2005), and the American iCampus Spoken Lecture Processing project (Glass, 2007). It is also the goal of the Liberated Learning Consortium [1], which fosters the application of speech recognition technology for enhancing accessibility for students with disabilities in the university classroom. In some of these projects, the concept of lecture is different. Many of our classroom lectures are 60-minute long, and quite informal, contrasting with the 20-minute seminars used in (Lamel *et al.*, 2005), where a more formal speech can often be found.

After a short description of the corpus itself and the annotation schema in Sections 2 and 3 respectively, ASR experiments are reported in Section 4. Section 5 describes the creation of a dataset that merges manual and automatic annotations and that provides prosodic information. Section 6 presents the conclusions and the future work.

2. Corpus description

The corpus includes seven 1-semester courses: Production of Multimedia Contents (PMC), Economic Theory I (ETI), Linear Algebra (LA), Introduction to Informatics and Communication Techniques (IICT), Object Oriented Programming (OOP), Accounting (CONT), Graphical Interfaces (GI). All lectures were taught at Technical University of Lisbon (IST), recorded in the presence of students, except IICT, recorded in another university and in a quiet office environment, targeting an Internet audience. A lapel microphone was used almost everywhere, since it has obvious advantages in terms of non-intrusiveness, but the high frequency of head turning causes audible intensity fluctuations. The use of the head-mounted microphone in the last 11 PMC lectures clearly improved this problem. However, this microphone was used with an automatic gain control, causing saturation in some of the recordings, due to the increase of the recording sound level during the students' questions, in the segments after them. Most classes are 60-90 minutes long (with the exception of IICT courses which are given in 30 minutes). A total of 74h were recorded, of which 10h were multilayer annotated in 2008 (Trancoso *et al.*, 2008). Recently additional 11 hours were orthographically transcribed. Table 1 below shows the number of lectures per course and the audio duration that was annotated, where V1 corresponds to the 2008 version of the corpus, *Added* is the quantity of added data, and V2 corresponds to the extended actual version.

---

[1] www.liberatedlearning.com

	# Lectures			Duration		
	V1	**Added**	**V2**	**V1**	**Added**	**V2**
LA	5	+3	8	2h25	2h30	4h55
GI	3	+1	4	2h50	0h51	3h41
CONT	6	+1	7	4h40	1h02	5h42
ETI	3		3	3h11		3h11
IICT	4		4	1h37		1h37
OOP	5	+1	6	4h00	2h22	6h22
PMC	2	+5	7	2h00	4h09	6h09
Total	28	+11	39	20h43	+10h54	31h37

Table 1: Number of lectures and durations per course

For future experiments, the corpus was divided into 3 different sets: Train (78%), Development (11%), and Test (11%). Each one of the sets includes a portion of each one of the courses. The corpus separation follows a temporal criterion, where the first classes of each course were included in the training data, and the final classes were included in the development and test sets. Figure 1 shows the portion of each course included in each one of the sets.
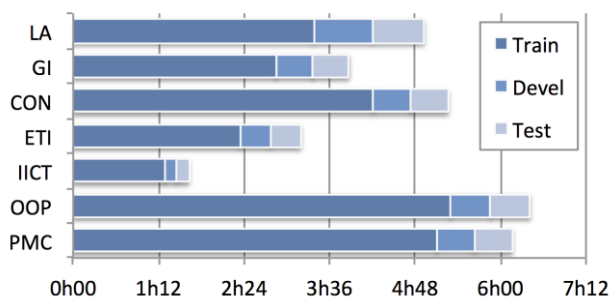
Figure 1: Corpus distribution

3. Corpus annotation

The orthographic manual transcriptions were done using Transcriber[2] and Wavesurfer[3] tools. Automatic transcripts are used as a basis that the transcribers corrected. At this stage, speech is segmented into chunks delimited by silent pauses, already containing audio segmentation related to speaker and gender identification and background conditions. Previously, the annotation schema comprised multilayers of orthographic, morpho-syntactic, structural metadata (Liu *et al.*, 2006; Ostendorf *et al.*, 2008), *i.e.*, disfluencies and punctuation marks, and paralinguistic information as well (laughs, coughs, etc.). The multilayer annotation aimed at providing a suitable sample for further linguistic and speech processing analysis in the lectures domain. The extension reported in this work respects the previous schema, however does not comprise the morpho-syntactic information tier, since automatic classifications of part-of-speech (POS) tags and of syntactic parsing is automatically performed, initially by

---

Marv (Ribeiro *et al.*, 2003) and more recently by *Falaposta* (Batista *et al.*, 2012). Thus, the extension of the annotation comprises the full orthographic transcription, enriched with punctuation and disfluency marks and a set of diacritics fully reported in Trancoso *et al.* (2008). Segmentation marks were also inserted for regions in the audio file that were not further analyzed (background noise, signal saturation).

Three annotators (with the same linguistics background) transcribed the extended data. However, two courses could not benefit from the extension for different reasons: the IICT, since no more lectures were recorded, and the ETI due to the fact that the teacher did not accept to make his recordings publicly available.

Due to the idiosyncratic nature of lectures as spontaneous and prepared non-scripted speech, the annotators reported in the five sessions of the guidelines instructions two main difficulties: in punctuating the speech and in classifying the disfluencies. The punctuation complexities are mainly associated with the fact that speech units do not always correspond to sentences, as established in the written sense. They may be quite flexible, elliptic, restructured, and even incomplete (Blaauw, 1995). Therefore, to punctuate speech units is not always an easy task. For a more complete view on this, we used the summary of grammatical and ungrammatical locations of punctuation marks for European Portuguese described in Duarte (2000). The latter is related to the different courses and the difficulty in discriminating the specific types of disfluencies (if it is a substitution, for instance), since the background of the annotators is on linguistics. To sum up, the guidelines given to our annotators were: the schema described in Trancoso *et al.* (2008) and the punctuation summary described in Duarte (2000).

The general difficulty of measuring the inter-transcriber agreement is due to the fact that two annotators can produce token sequences of different lengths. This is equivalent to measuring the speech recognition performance, where the length of the recognized word sequence is usually different from the reference. For that reason, the inter-transcriber agreement was calculated for pairs of annotators, considering the most experienced[4] as reference. The standard F1-measure and Slot Error Rate (SER) (Makhoul *et al.*, 1999) metrics were used, where each slot corresponds to a word, a punctuation mark or a diacritic:

$$F1 - measure = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \ , \ SER = \frac{\text{errors}}{\text{ref\_tokens}}$$

where *ref_tokens* is the number of words, punctuation marks and diacritics used in the reference orthographic tier, and *errors* comprise the number of inserted, deleted or substituted tokens.

The inter-transcriber agreement of the three annotators is based on a selected sample of 10 minutes of

---

speech from one speaker involving more than 2000 tokens. The selection of the sample has to do with the reported difficulties of the annotators, in annotating disfluencies (e.g., complex sequences of disfluencies) and also punctuation marks. Table 2 reports the inter-transcriber agreement results for each pair of annotators. The table shows the number of (Cor)rect slots, (Ins)ertions, (Del)etions, (Sub)stitutions, (F1)-measure, and slot accuracy (SAcc), which corresponds to 1-SER. There is an almost perfect agreement between A1 and the remaining annotators, and a substantial agreement between the pair A2-A3. These results may well be the outcome of a thorough process of annotation in several different steps and with intermediate evaluations during the 5 guidelines instruction sections. Moreover, several other annotators for other corpora already tested the guidelines here in use.

Annotator	Cor	Ins	Del	Sub	F1	SER	SAcc
A1-A2	1714	67	79	224	0.852	0.184	**0.816**
A1-A3	1632	38	34	351	0.808	0.210	**0.790**
A2-A3	1480	81	97	444	0.735	0.308	**0.692**

Table 2: Evaluation of the inter-transcriber agreement

4. ASR experiments

Transcribing lectures is particularly difficult since lectures are very domain-specific and speech is spontaneous. Except the IICT lectures where no students were present, students demonstrate a relatively high interactivity in the other lectures. Nevertheless, since only a lapel microphone was used to record the close-talk speech of the lecturers, the audio gain of the student interventions is very low. The presence of background noise, such as babble noise, footsteps, blackboard writing noise, etc. may difficult the speech processing, in particular the Speech / Non-speech detection that feeds the recognizer with audio segments labelled as speech. Typical WER reported in the recent literature is between 40-45% (Glass *et al.*, 2007).

4.1 Overview of our ASR system

Our automatic speech recognition engine named Audimus (Neto *et al.*, 2008; Meinedo *et al.*, 2008) is a hybrid automatic speech recognizer that combines the temporal modeling capabilities of Hidden Markov Models (HMMs) with the pattern discriminative classification capabilities of Multi-Layer Perceptrons (MLPs). The MLPs perform a phoneme classification by estimating the posterior probabilities of the different phonemes for a given input speech frame (and its context). These posterior probabilities are associated to the single state of context independent phoneme HMMs.

The most recent ASR system used in this work is exactly the ASR system for EP described in (Meinedo *et al.*, 2010). The acoustic models were initially trained with 46 hours of manually annotated broadcast news (BN) data collected from the public Portuguese TV, and in a second time with 1000 hours of data from news shows of several EP TV channels automatically transcribed and selected

according to a confidence measure threshold (non-supervised training). The EP MLPs are formed by 2 hidden layers with 2000 units each and have 500 softmax output units that correspond to 38 three state monophones of the EP language plus a single-state non-speech model (silence) and 385 phone transition units which were chosen to cover a very significant part of all the transition units present in the training data. Details on phone transition modeling with hybrid ANN/HMM can be found in (Abad & Neto, 2008).

The Language Model (LM) is a statistical 4-gram model that was estimated from the interpolation of several specific LMs: in particular a backoff 4-gram LM, trained on a 700M word corpus of newspaper texts, collected from the Web from 1991 to 2005, and a backoff 3-gram LM estimated on a 531k word corpus of broadcast news transcripts. The final language model is a 4-gram LM, with Kneser-Ney modified smoothing, 100k words (or 1-gram), 7.5M 2-gram, 14M 3-gram and 7.9M 4-gram. The multiple-pronunciation EP lexicon includes about 114k entries.

These models, both AMs and the LM, were specifically trained to transcribe BN data. The Word Error Rate (WER) of our current ASR system is under 20% for BN speech in average: 18.4% for instance, obtained in one of our BN evaluation test sets (RTP07), composed by six one hour long news shows from 2007 (Meinedo *et al.*, 2010).

4.2 ASR results

A test subset was selected from the corpus in 2008, by choosing one single lecture per course. In (Trancoso *et al.*, 2008), preliminary ASR results were reported on this test set, showing the difficulty to transcribe lectures. Very high word error rates (WER), 61.0% in mean, were achieved for a subset of various lectures chosen as a test set. It has a vocabulary of around 57k words. Applying this recognize without any type of domain adaptation, obviously yielded very bad results

Table 3 illustrates the performance of the old and the recent systems without and with adaptation of the LM for the recent system. Our recent system, which was described in the previous section, achieved a WER of 45.7% on the same test subset, hence showed a 25.0% relative reduction. The lexicon was almost twice the size of the one of the previous system. Further improvements were achieved with a 44.0% WER. This performance was obtained by interpolating our generic broadcast news 4-gram LM with a 3-gram LM trained on the training lecture subset. 100-best hypotheses were generated per each sentence and rescored with this LM and a RNN (implementation of the Brno University (Mikolov *et al.*, 2011)). This RNN was trained only on the lecture train subset.

An analysis of the ASR errors showed that most of the misrecognitions concerned small function words, such as definite articles and prepositions, the backchannel word "OK" also appeared to be very often misrecognized. Then, words specific to each jargon of the courses also were error-prone. For instance variable names in the Linear

Algebra lecture, such as "alfa", "beta", "vector" were often substituted. In the PMC lecture, words such as "MPEG", "codecs", "metadados" (metadata), "URL" were subject to frequent errors.

ASR system	LM adapt?	OOV (%)	WER (%)
2008	no	_	61.0
2011	no	2.8	45.7
2011	yes	1.7	44.0

Table 3: Comparison of the ASR results reported in 2008 and obtained with our most recent system. OOV stands for out-of-vocabulary words

5.   Enriched annotations

The ASR system is able not only to produce automatic transcripts from the speech signal, but also to produce automatic force-aligned transcripts, adjusting the manual transcripts to the speech signal. Apart from the existing manual annotations of the corpus, automatic force-aligned transcripts have been produced for the extended version of the corpus, and will be available in our META-SHARE node. These force-aligned transcripts were updated with relevant information coming from the manual annotations, and finally enriched with additional prosodic information (Batista *et al.*, 2012). The remainder of this Section provides more details about this process.

5.1   Automatic alignment

Force-aligned transcripts depend on a manual annotation and therefore do not contain recognition errors. A number of speech tasks, such as the punctuation recovery, may use information, such as pause durations, which most of the times is not available in the manual transcripts. On the other hand, manual transcripts provide reduced or error-free transcripts of the signal. For that reason, force-aligned transcripts, which combine the ASR information with manual transcripts, provide unique information, suitable for a vast number of tasks.

An important advantage of using force-aligned transcripts is that they can be treated in the exact same way as the automatic transcripts, but without recognition errors, requiring the same exact procedures and tools. However, the alignment process is not always performed correctly due to a number of reasons, in particular when the signal contains low energy levels. For that reason, the ASR parameters can be adjusted to accommodate the manual transcript into the signal. Our current force-alignment achieves 3.8% alignment word errors in the training, 3.1% in the development, and 4.5% in the evaluation sets.

5.2   Merging manual and automatic annotations

Starting with the previously described force-aligned transcripts, we have produced a self-contained dataset that provides not only the information given by the ASR system, but also important parts of the manual transcripts. For example, the manual orthographic transcripts include punctuation marks and capitalization information, but that is not the case of force-aligned transcripts, which only includes information, such as: word time intervals, and confidence scores. The required manual annotations are transferred by means of alignments between the manual and automatic transcripts.

Apart from transferring information from the manual transcripts, the data was also automatically annotated with part-of-speech information. The part-of-speech tagger input corresponds to the text extracted from the ASR transcript, after being improved with the reference capitalization. Currently, the Portuguese data is being annotated using *Falaposta*, a CRF-based tagger robust to certain recognition errors, given that a recognition error may not affect all its input features. It accounts for 29 part-of-speech (POS) tags and achieves 95.6% accuracy.

The resulting file, structured using the XML format, corresponds to the ASR output, extended with: time intervals to be ignored in scoring, focus conditions, speaker information for each region, punctuation marks, capitalisation, disfluency marks, and POS information.

5.3   Adding prosodic data

The previously described extended XML file is further improved with phone and syllable information, and other relevant information that can be computed from the speech signal (*e.g.*, pitch and energy). The data provided by the ASR system allows us to calculate the phone information. Marking the syllable boundaries as well as the syllable stress are achieved by means of a lexicon containing all the pronunciations of each word together with syllable information, since these tasks are currently absent in the recognizer. A set of syllabification rules was designed and applied to the lexicon, which account fairly well for the canonical pronunciation of native words, but they still need improvement for words of foreign origin. Pitch (f_0) and energy (E) are two important sources of prosodic information, currently not available in the ASR output, and directly extracted from the speech signal. Algorithms for automatic extraction of the pitch track have, however, some problems, *e.g.*, octave jumps; irregular values for regions with low pitch values; disturbances in areas with micro-prosodic effects; influences from background noisy conditions; *inter alia*. We have removed all the pitch values calculated for unvoiced regions in order to avoid constant micro-prosodic effects. This is performed in a phone-based analysis by detecting all the unvoiced phones. We also had a calculation cost to eliminate octave-jumps. As to the influences from noisy conditions, the recognizer has an Audio Pre-processing or Audio Segmentation module, which classifies the input speech according to different focus conditions (*e.g.*, noisy, clean), making it possible to isolate those speech segments with unreliable pitch values.

After extracting and calculating the above information, all data was merged into a single data source. The existing XML data has been upgraded in order to accommodate the additional prosodic information.

6.   Conclusions

This paper described our lecture corpus in European Portuguese, and its recent extension. The problems it raises for automatic speech recognition systems were illustrated. The fact that a significant percentage of the recognition errors occurs for function words led to us believe that the current performance, although far from ideal, may be good enough for information retrieval purposes, enabling keyword search and question answering in the lecture browser application. ASR performance is still poor but as stated in Glass *et al.* (2007), "accurate precision and recall of audio segments containing important keywords or phrases can be achieved even for highly-errorful audio transcriptions (*i.e.*, word error rates of 30% to 50%)". Together with the manual transcripts, a set of upgraded and enriched force-aligned transcripts were produced and made available. Such transcripts constitute an important advantage for corpora analysis, and for studying a number of speech tasks. Currently, the LECTRA corpus is being used to study and perform punctuation and capitalization tasks, and spontaneous speech phenomena. We believe that producing a surface rich transcription is essential to make the recognition output intelligible for hearing impaired students. Six courses of the corpus will be soon available to the research community via the META-SHARE platform.

7.   Acknowledgements

8.   References

Abad, A., Neto, J. (2008). Incorporating Acoustical Modelling of Phone Transitions in a Hybrid ANN/HMM Speech Recognizer. In *Proc. Interspeech*, pp. 2394-2397, Brisbane.

Batista, F., Moniz, H., Trancoso, I., Mamede, N. and Mata, A.I. (2012). Unified Data Representation for Prosodically-based Speech Processing. In *JOSS – Journal of Speech Sciences* (submitted).

Batista, F., Moniz, H., Trancoso, I., and Mamede, N.J. (2012). Bilingual experiments on automatic recovery of capitalization and punctuation of automatic speech transcripts. In *IEEE Transactions on Audio, Speech and Language Processing, Special Issue on New Frontiers in Rich Transcription*, 20(2), pp. 474--485.

Blaauw, E. (1995). On the Perceptual Classification of Spontaneous and Read Speech. PhD Diss, Research Institute for Language and Speech, Utrecht.

Duarte, I. (1995). *Língua Portuguesa, Instrumentos de Análise*. Universidade Aberta.

Furui, S., Iwano, K., Hori, C., Shinozaki, T., Saito, Y. and Tamura, S. (2001). Ubiquitous speech processing. In *Proc. ICASSP*, Salt Lake City.

Glass, J., Hazen, T., Cyphers, S., Malioutov, I., Huynh, D. and Barzilay, R. (2007). Recent Progress in the MIT Spoken Lecture Processing Project. In *Proc. Interspeech*, Antwerp.

Lamel, L., Adda, G., Bilinski, E. and Gauvain, J. (2005). Transcribing lectures and seminars. In *Proc. Interspeech*, Lisbon.

Liu, Y., Shriberg, E., Stolcke, A., Hillard, D., Ostendorf, M., and Harper, M. (2006). Enriching speech recognition with automatic detection of sentence boundaries and disfluencies. In *Transactions on Audio, Speech and Language Processing*, 14(5), pp. 1526--1540.

Makhoul, J., Kubala, F., Schwartz, R. and Weischedel, R. (1999). Performance measures for information extraction. In *Proc. of the DARPA BN Workshop*.

Meinedo, H., Viveiros, M. and Neto, J. (2008). Evaluation of a live broadcast news subtitling system for Portuguese. In *Proc. of Interspeech*, Brisbane, Australia.

Meinedo, H., Abad, A., Pellegrini, T., Neto, J. and Trancoso, I. (2010). The L2F Broadcast News Speech Recognition System. In *Proc. Fala*, Vigo, pp. 93--96.

Mikolov, T., Deoras, A., Kombrink, S., Burget, L. and Cernocky, J.H. (2011). Empirical evaluation and combination of advanced language modeling techniques. In *Proc. Interspeech*, Florence pp. 605--608,.

Moniz, H., Batista, F., Trancoso, I. and Mata, A.I. (2012), Prosodic context-based analysis of disfluencies. In *Proc. Interspeech*, Portland, Oregon.

Neto, J., Meinedo, H., Viveiros, M., Cassaca, R., Martins, C. and Caseiro, D. (2008). Broadcast News Subtitling System in Portuguese. In *Proc. ICASSP*, Las Vegas.

Ostendorf, M., *et al*. (2008). Speech segmentation and spoken document processing. *IEEE Signal Processing Magazine*, 25(3):59--69.

Ribeiro, R., Oliveira, L.C. and Trancoso, I. (2003). Morphossyntactic information for TTS systems: comparing strategies for European Portuguese. In *Proc. Propor*, Springer-Verlag, LNAI, Faro, pp. 143--150.

Trancoso, I., Neto, J., Meinedo, H. and Amaral, R. (2003). Evaluation of an alert system for selective dissemination of broadcast news. In *Proc. Eurospeech*, Geneva.

Trancoso, I., Nunes, R., Neves, L., Viana, M.C., Moniz, H., Caseiro, D. and Mata, A.I. (2006). Recognition of Classroom Lectures in European Portuguese. In *Proc. Interspeech*, Pittsburgh.

Trancoso, I., Martins, R., Moniz, H., Mata, A.I. and Viana, M.C. (2008). The LECTRA Corpus - Classroom Lecture Transcriptions in European Portuguese. In *Proc. LREC*, Marrakech.

A constituição de um *corpus* de italiano falado para o estudo de pedidos e pedidos de desculpas: considerações sobre a validade interna e externa dos dados

Elisabetta SANTORO

Universidade de São Paulo

Av. Prof. Luciano Gualberto, 403 Cidade Universitária

CEP 01060-970 – São Paulo – SP – Brasil

esantoro@usp.br

Resumo

O texto aqui apresentado pretende discutir questões ligadas à constituição de um *corpus* de italiano falado, coletado a partir de gravações em áudio e vídeo, para um estudo que se insere no âmbito da pragmática linguística e procura investigar dois atos de fala específicos, a saber, pedidos e pedidos de desculpas. Propõe-se, em especial, uma reflexão sobre a validade externa e a validade interna dos dados, sendo que, a partir desses conceitos, será possível pensar nas características das pesquisas realizadas com dados coletados a partir de diferentes metodologias, além de se poder imaginar uma "hierarquia" de metodologias, da mais livre à mais controlada. Se, por um lado, metodologias muito abertas permitem uma elevada validade externa dos dados, mas não são muito adequadas para o estudo de fenômenos específicos, além de ser também dificilmente replicáveis; por outro, metodologias nas quais a produção dos informantes é mais controlada podem produzir dados mais facilmente comparáveis e ajudar a circunscrever aspectos específicos da língua.

Palavras-chave: pragmática linguística; *corpus*; metodologias de coleta de dados; *role play*.

1. *Corpus* e língua falada

Realizar uma pesquisa a partir de um *corpus* de língua falada pressupõe decisões importantes sobre a metodologia de coleta dos dados, pois, antes mesmo de iniciar o planejamento do trabalho, é preciso avaliar com extrema atenção vantagens e desvantagens de cada uma das possibilidades. Se o objetivo da pesquisa for, por exemplo, estudar a língua falada sob diferentes pontos de vista (da fonética, da fonologia, da prosódia, do léxico, da morfologia, da sintaxe *etc*), será essencial dispor de material linguístico que seja diastraticamente e diafasicamente o mais variado possível, para que se possam fazer afirmações que, mesmo dizendo respeito às amostras de língua coletadas, possam "representar" o todo. Quando, ao contrário, o pesquisador estabelece metas mais detalhadas e pretende se dedicar a fenômenos específicos da língua falada, pode ser necessário utilizar metodologias que deem subsídios de outra natureza para a análise a ser desenvolvida.

Pretendemos aqui discutir brevemente algumas das alternativas que se colocam para o pesquisador, pensando, em especial, nas escolhas feitas para um estudo realizado com o italiano contemporâneo, que se insere no âmbito das pesquisas em pragmática linguística e procura investigar dois atos de fala específicos, a saber, pedidos e pedidos de desculpas, a partir de gravações em áudio e vídeo.

Poderia ser útil – e trazer ainda outras questões, entre as quais nem sempre há consenso entre os pesquisadores – analisar também as definições de *corpus*, inclusive colocando-as em relação com os objetivos de pesquisa e com o tipo de análise a ser realizado. No entanto, não faremos isso aqui e iremos nos concentrar em considerações relativas à validade interna e externa dos dados coletados, para podermos refletir sobre as diferentes abordagens e metodologias que impedem ou, ao contrário, permitem a execução de um determinado tipo de pesquisa.

2. Validade externa e validade interna

Em primeiro lugar, cabe explicar o que entendemos quando falamos de validade externa e de validade interna dos dados. Começando pela validade externa, podemos dizer que esta se julga dada, quando é possível generalizar os resultados de uma pesquisa, que, a partir das amostras escolhidas, podem ser considerados válidos para a língua em análise como um todo. Para tanto, reputa-se imprescindível gravar os informantes em situações que eles não sintam como "estranhas", isto é, que não sejam distantes de sua habitual prática linguística.

A validade interna, ao contrário, refere-se à interpretabilidade da pesquisa e deve permitir dizer se as variações presentes nos dados podem ser tratadas como uma consequência das variáveis analisadas. A validade interna está relacionada aos fatores que podem influenciar diretamente os resultados e é avaliada levando em conta se as diferenças encontradas na variável dependente (que medimos para ver quais são os efeitos da variável independente sobre ela), se relacionam diretamente com a variável independente (aquela que pode "causar" o resultado). A validade interna implica, portanto, que os dados sejam mais controlados, e precisa de instrumentos de coleta que permitam isolar variáveis de modo a garantir sua adequada avaliação separadamente e em sua interação com outras.

Há muitos fatores que podem comprometer a validade interna dos dados de uma pesquisa, entre os quais, por exemplo, as características e o comportamento dos participantes, o equipamento utilizado, a atitude do pesquisador que coleta os dados e a situação em que isso é feito.

Além disso, é importante não esquecer que, em geral, estudos com elevada validade externa sofrem em relação à validade interna, porque o respeito à integridade do contexto impede que sejam controladas as variáveis –

como é possível fazer, por exemplo, com um protocolo experimental – assim que afirmações de natureza causal ou o estabelecimento de relações entre os dados serão sempre problemáticos ou até impossíveis.

Tendo isso em vista, no caso de uma pesquisa que visa a estudar a realização de atos de fala específicos na interação entre dois falantes e que se propõe a descobrir eventuais relações entre variáveis, para poder entender o funcionamento de uma determinada língua natural em uso, uma vez controladas as características ligadas aos informantes, será necessário abdicar, ao menos em parte, da validade externa da fala espontânea não controlada e procurar metodologias de coleta dos dados que permitam também análises causais, que relacionem os dados.

3. *Corpora* no estudo da pragmática linguística

Em muitas pesquisas que se propõem a constituir *corpora* para a investigação da pragmática linguística, especialmente quando ligadas à mais ortodoxa análise da conversação[1], atenta-se principalmente para a validade externa dos dados a serem estudados, isto é, procura-se coletá-los de modo que haja a maior correspondência possível entre os fenômenos observados ao longo da investigação e os que acontecem, ou se presume aconteçam, na vida real. Em outras palavras, os dados considerados de maior relevância para o estudo da pragmática das línguas, principalmente em contextos cotidianos, são os ditos "dados naturalísticos", coletados de preferência sem que o informante tenha ciência, no momento em que os fornece, de participar de uma pesquisa e, possivelmente, sobretudo no caso de gravações só em áudio, com os aparelhos escondidos, de modo que o informante nem mesmo saiba que sua fala está sendo gravada. É o caso das ditas "gravações secretas", nas quais se procuram voluntários dispostos a colaborar nas pesquisas, que, em geral, gravam conversações das pessoas de seu convívio, revelando só depois de concluída a gravação sua participação no projeto.

Não citaremos aqui as questões éticas e legais que procedimentos como esses envolvem (para isso, sugerimos, por exemplo, a leitura de Bazzanella, 1994). Embora isso não seja considerado admissível por alguns, pois alteraria por si só a validade e a confiabilidade dos dados, basta que os informantes sejam avisados e aceitem ser gravados – como acontece nas gravações que chamaremos "consentidas" – para que essa dificuldade seja superada.

Mesmo assim, os dados produzidos a partir desse tipo de metodologia podem ser, segundo alguns, menos "naturais", pois os informantes, ao saberem que estão sendo gravados, alterariam sua fala. É preciso, contudo, lembrar que a própria definição de dado naturalístico não é isenta de problemas.

De fato, é suficiente pensar nas observações sobre o "paradoxo do observador" de Labov (1970) ou nos questionamentos de Ochs (1979) sobre a impossível neutralidade do processo de transcrição, para concluir que a realidade linguística não poderá nunca ser colhida em toda a sua complexidade e que o pesquisador sempre irá intervir para recortar do material coletado as partes mais significativas para o seu projeto de pesquisa, eliminando em alguns casos o contexto e produzindo, assim, alterações que também deveriam ser levadas em conta.

Desta forma, do nosso ponto de vista, pode ser por vezes desmedida a atenção dada à definição do que se pode considerar fala espontânea ou semi-espontânea: se acreditarmos que a gravação e a transcrição em si já alteram o contexto da fala e precisariam, portanto, ser levadas em conta na hora de analisar os dados, deveríamos também relativizar a rigidez que muitas vezes acompanha o julgamento das maneiras como foram coletados.

Não obstante, é claro que há distinções a serem feitas entre as possíveis maneiras de eliciar dados e que é necessário ter consciência de quais são, sempre atentando também para os objetivos de cada pesquisa. Como dizíamos no início, diferente será, por exemplo, coletar um *corpus* com o objetivo de estudar fenômenos gerais da língua, não ligados a situações peculiares e importantes pela sua recorrência em diferentes contextos comunicativos; ou tentar delimitar e fixar na gravação o mesmo fenômeno que se repete diversas vezes, de modo que suas manifestações, em contextos de partida idênticos, possam ser comparadas e estudadas.

Na hipótese em que se queira, como no exemplo da pesquisa de que falamos aqui, verificar se o mesmo pedido é realizado com o uso de formas linguísticas distintas, caso intervenha uma determinada variável, será necessário controlar a variável escolhida e comparar o maior número possível de ocorrências realizadas a partir do mesmo *input*. É evidente que isso só poderá ser feito se os dados forem coletados com metodologias que prevejam o controle das variáveis e será praticamente impossível com gravações "livres".

Visando a contribuir para uma maior clareza sobre as diferenças na coleta de dados para o estudo da língua falada, há estudiosos que prepararam listas e propuseram hierarquizações das metodologias, colocando-as em uma ordem que vai do menor ao maior grau de controle sobre a produção dos dados, isto é, da maior validade externa à maior validade interna (se veja, Pallotti, 2001).

Tentaremos fazer aqui algo parecido, refletindo, em especial, sobre as pesquisas relativas ao estudo da pragmática linguística, intercultural e interlinguística.

Citamos acima dois procedimentos que se propõem a "capturar" a realidade linguística *assim como ela é* e que se impõem para esse fim várias e, muitas vezes, rígidas limitações metodológicas.

Pensando ainda em termos de validade externa e interna, podemos observar que no outro extremo em relação às metodologias mencionadas acima, em especial quando a perspectiva é a da pragmática intercultural ou interlinguística, é prática comum coletar os dados utilizando instrumentos que possuem um elevado grau de

---

[1] Ver, entre outros, Briz e Grupo Val.Es.Co. (2002).

controle sobre as variáveis. São, de fato, frequentes os casos nos quais, para a coleta dos dados, se escolhem DCT (*Discourse Completion Tests*) escritos, nos quais os informantes, utilizando-se da escrita para fornecer dados que deveriam pertencer à oralidade, escrevem o que diriam em determinadas situações (ver, por exemplo, Hudson, Detmer & Brown, 1995) ou até realizam atividades de escolha múltipla, em que o informante deve apenas assinalar qual das alternativas apresentadas considera mais adequada para responder à situação comunicativa descrita.

Controle mínimo sobre as produções dos informantes (elevada validade externa)		
	Gravação secreta	A interação não é guiada. Os informantes não sabem que estão sendo gravados.
	Gravação consentida	A interação não é guiada. Os informantes sabem que estão sendo gravados.
	Gravação participante	A interação não é guiada, mas há a participação do pesquisador.
	Role play **aberto**	A interação não é guiada. Os turnos de fala e sua duração não são pré-determinados.
	Role play **semi-aberto**	A interação é parcialmente guiada, pois há um *input* que indica a situação. Os turnos de fala e sua duração não são pré-determinados.
	Role play **fechado**	O roteiro da interação é pré-estabelecido. O número de falas é pré-fixado (em geral, trata-se de apenas um turno).
	Discourse Completion Test **oral**	A fala de um dos interlocutores é dada. O informante completa oralmente.
	Discourse Completion Test **escrito**	A fala de um dos interlocutores é dada. O informante completa por escrito.
	Escolha múltipla	São apresentadas várias respostas possíveis para uma determinada fala. O informante precisa apenas escolher entre elas.
Controle máximo sobre as produções dos informantes (elevada validade interna)		

Tabela 1: Algumas metodologias para coleta de dados

Com essas metodologias as produções dos informantes são muito controladas e, além disso, os dados assim coletados requerem um baixo dispêndio de tempo e energias, pois não precisam de equipamentos de gravação em áudio e vídeo e podem ser gerados em grande número até em uma única sessão.

É evidente que metodologias dessa natureza afastam demasiadamente os dados coletados daqueles que se considerariam "naturais". De fato, tentando reproduzir por escrito aquilo que diriam nas situações dadas os informantes eliminam completamente os traços característicos da língua falada (falsas partidas, reformulações, hesitações etc) e "limpam" suas manifestações linguísticas de todos os elementos que caracterizam a fala. Além disso, *escrever* no lugar de *dizer* significa eliminar completamente a interação oral entre dois indivíduos; e dispor de um tempo maior antes de produzir o dado o priva da imediatez característica da língua falada, na qual se reage a um estímulo oral, sem ter a chance de refletir ou de se preparar.

Citamos acima apenas as metodologias mais livres, de um lado, e mais controladas, do outro. A seguir, apresentamos nossa proposta de uma escala de metodologias para a coleta dos dados, pensada a partir das escolhas mais frequentes feitas para estudos de pragmática linguística. As metodologias foram hierarquizadas de modo que a menos controlada e com maior validade externa foi colocada na parte superior da tabela, enquanto a mais controlada e com menor validade externa na parte inferior.

4. Um *corpus* de italiano falado para o estudo de pedidos e pedidos de desculpas

Para a constituição de um *corpus* de italiano falado que se propõe a analisar pedidos e pedidos de desculpas, optamos por uma metodologia de coleta dos dados que se coloca na posição intermediária da tabela apresentada acima. Trata-se do *role play* semi-aberto que, em relação a outras opções controladas de coleta de dados, possui, para começar, a vantagem de criar uma verdadeira interação oral entre dois interlocutores, mantendo, portanto, as características da língua falada, embora a interação não seja a consequência de uma necessidade real dos interlocutores e seja induzida pelo pesquisador.

Em geral, a distinção se faz apenas entre *role play* aberto, que envolve a interação entre dois ou mais indivíduos, reagindo a uma determinada situação; e *role play* fechado, nos quais é apresentada aos participantes uma situação específica, à qual devem responder, em geral, com um único turno de fala. Considera-se que o *role play* fechado pode não refletir dados que poderiam ocorrer naturalmente, enquanto o aberto os refletiria mais exatamente, por prever a interação e uma reação "livre" à situação dada[2]. Para o tipo de *role play* utilizado na nossa pesquisa preferimos utilizar a categoria "*role play* semi-aberto", pois aos interlocutores foi pedido que reagissem a uma situação comunicativa específica e em um contexto dado.

---

[2] A esse respeito Mackey & Gass (2005: 91) afirmam: *"Open role plays, on the other hand, involve interaction played out by two or more individuals in response to a particular situation. [...] Closed role plays suffer from the possibility of not being a reflection of naturally occurring data. Open role plays reflect natural data more exactly [...]."*

As instruções eram dadas por escrito a apenas um dos dois participantes ao qual cabia também iniciar a interação. Isso foi feito com o objetivo de o input ser idêntico para todos os participantes da pesquisa que receberam as instruções com o mesmo papel e, portanto, com exatamente as mesmas palavras e do mesmo modo. Além disso, essa escolha permitiu que um dos dois informantes reagisse diretamente à fala do outro, sem saber antes a que situação teria que reagir.

A decisão de como transformar em palavras e em interação com o outro a situação descrita no papel recebido era "livre" quanto às formas linguísticas e sem limites de tempo ou turnos de fala limitados. Procuramos ainda, sempre que possível, recriar o contexto (*setting*)[3], para que os informantes pudessem mais facilmente evocar as rotinas linguísticas utilizadas em situações do mesmo tipo. Assim, as situações que aconteciam "na rua" foram efetivamente gravadas na rua e assim também foram gravadas em casas as situações do contexto "casa de outrem". Com um fim parecido, procuramos também definir contextos e situações nos quais todos os informantes poderiam se encontrar na vida real, de modo a não levá-los a representar um "papel" no qual dificilmente se encontrariam na vida real. Mais um corretivo aplicado ao *role play* é que os interlocutores não mudaram sua relação da vida real e se trataram no *role play* como se tratariam fora dele.

Com essa metodologia poderiam ser controladas as variáveis independentes. Brown e Levinson (1987: 76) identificam três variáveis para os atos de fala: a *distância social* entre os interlocutores que cria um eixo horizontal; o *poder relativo* entre eles, a partir do qual se estabelece um eixo vertical; e o *grau de imposição* de um ato de fala, ou seja, a relação custo/benefício que a realização do ato representa para os interlocutores.

No nosso caso, se é verdade que a escolha quanto a respeitar a identidade e a relação real entre os informantes limitou ou até impossibilitou a seleção de situações com claras diferenças de poder relativo e distância social (para imaginá-las teria sido necessário pensar em contextos como o ambiente de trabalho, nos quais isso é mais evidente), é também verdade que o grau de imposição, variável que pode produzir notáveis diferenças nos atos de fala, pôde ser incluído. De fato, a um maior grau de imposição corresponde em geral um aumento da atenuação, dos modificadores, da necessidade de justificar um pedido ou de procurar reparar o prejuízo provocado no caso de pedidos de desculpas.

Procuramos, portanto, organizar as situações previstas para os *role plays* em pares, nos quais sempre havia uma situação com um baixo grau de imposição (– I), isto é, com um pedido ou um pedido de desculpas que previa um ônus baixo para o interlocutor, e outra, no mesmo contexto, com alto grau de imposição (+I), ou seja, com um ônus elevado. Cabe dizer que, para garantir que fosse claro o diferente grau de imposição entre os pares de situações no mesmo contexto, procuramos escolher

---

pedidos e pedidos de desculpas em que as diferenças fossem muito marcantes. Assim, por exemplo, no contexto "casa de outrem", o pedido –I do informante que chega à casa de outra pessoa é um copo d'água, enquanto o pedido +I é poder tirar a roupa molhada e tomar um banho, porque a pessoa que chega foi surpreendida por um forte temporal e estava sem guarda-chuva.

Os contextos em que as situações foram colocadas eram ao todo três – a rua, o trem e a casa de outrem – e havia, para cada um deles, dois pedidos e dois pedidos de desculpas, chegando-se assim a 12 situações gravadas pelas 30 duplas de informantes que participaram da pesquisa e realizaram interações orais a partir do mesmo *input*.

Vale acrescentar que foi considerado na elaboração dos *role plays* que haveria diferentes graus de familiaridade entre os participantes e foi assim decidido dividi-los em duas grandes categorias, tratando como um grupo os que declararam ter um grau de conhecimento de 1 a 5 (desconhecidos, conhecidos, pessoas que acabaram de se conhecer), e como um segundo grupo os com um grau de conhecimento de 6 a 10 (amigos ou parentes).

Apenas para os pedidos realizamos também gravações em estabelecimentos públicos e comerciais, de três diferentes cidades italianas, nos quais pudemos contar com a participação das pessoas que habitualmente atendem o público. Para essas gravações foi dada aos informantes uma instrução oral reduzida ao essencial para que pudessem realizar a ação prevista (do tipo: "entre na loja e compre um presente").

Além de permitir o controle das variáveis e, portanto, uma validade interna elevada que possibilita um estudo sistemático das ocorrências, o *corpus* coletado é caracterizado pela replicabilidade e pela possibilidade de ser ampliado. Pretendemos, de fato, constituir um *corpus* com as mesmas características para o português brasileiro que possibilite a realização de estudos de pragmática intercultural, comparando a realização dos mesmos atos de fala por falantes nativos de italiano e de português brasileiro. Foram iniciadas também coletas de dados e pesquisas com aprendizes brasileiros de italiano que poderão representar a base para analisar a pragmática interlinguística, isto é, como um aprendiz brasileiro desenvolve sua competência pragmática em italiano, que tipo de relação essa competência possui com os conhecimentos gramaticais, e se as instruções explícitas podem ter efeitos reconhecíveis.

5. Conclusões

Para um projeto dessa natureza, o *role play*, realizado com os corretivos antes mencionados, representou uma forma de coletar dados que, por um lado, permitiu conservar as peculiaridades da língua falada e, por outro, ofereceu a possibilidade de isolar variáveis e analisar as alterações que poderiam ser provocadas por cada uma delas. Isso significou criar um *corpus* com características homogêneas, capaz de fornecer primeiros dados comparáveis para o estudo de pedidos e pedidos de desculpa. Os *role plays* gravados em áudio e vídeo pelo

---

[3] Sobre a relevância do contexto em pragmática, cf. Nickel (2006).

mesmo par de informantes, mas com diferentes graus de imposição nos permitem observar a maior e menor presença, por exemplo, de modificadores e atenuadores, ou a presença/ausência de uma justificativa para um pedido ou para um pedido de desculpas e isso pode nos ajudar a reconduzir as escolhas a variáveis pré-determinadas, dando-nos assim a possibilidade de identificar as prováveis "causas" de específicas manifestações linguísticas e fornecendo-nos dados para criar relações entre o mundo e a língua.

6. Agradecimentos

7. Bibliografia

Bazzanella, C. (1994). *Le facce del parlare.* Firenze: La Nuova Italia.

Briz, A., Grupo Val.Es.Co. (2002). *Corpus de conversaciones coloquiales.* Anexo 1 de Oralia. Madrid: Arco Libros.

Brown, P., Levinson, S.C. (1987). *Politeness. Some universals in language use, 2.* ed., Cambridge: Cambridge University Press.

Hudson, T., Detmer, E. and Brown, J.D. (1995). *Developing Prototypic Measures of Cross-Cultural Pragmatics.* Honolulu: Second Language Teaching & Curriculum Center, University of Hawai'i.

Labov, W. (1970). The study of language in its social context. *Studium Generale*, 23, pp. 30--87.

Mackey, A. & Gass, S. M. (2005). *Second Language Research. Methodology and Design.* Malwah, New Jersey: Lawrence Erlbaum Associates.

Nickel, E. L. (2006). "Interlanguage Pragmatics and the Effects of Setting". In K. Bardovi-Harlig, J.C. Félix-Brasdefer and A.S. Omar. (Eds.), *Pragmatics & Language Learning,* vol 11, Honolulu: University of Hawai'i, pp. 253--280.

Ochs, E. (1979). Transcription as theory. In E. Ochs, B. Schieffelin (Eds.), *Developmental Pragmatics.* New York: Academic Press.

Pallotti, G. (2001). L'ecologia del linguaggio: contestualizzazione dei dati e costruzione di teorie. In F. Albano Leoni, E. Stenta Krosbakken, R. Sornicola and C. Stromboli (Eds.), *Dati empirici e teorie linguistiche.* Atti del XXXIII Congresso Internazionale di Studi della Società di Linguistica Italiana, Roma: Bulzoni, pp. 37--57.

SPEECH TECHNOLOGY AND DATA BASES

SweDia 2000 – A Swedish dialect research database

Anders ERIKSSON

Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg, Box 200, SE-405 30 Gothenburg, Sweden
anders.eriksson@ling.gu.se

Abstract

The SweDia 2000 dialect database (SweDat as we refer to it in our daily work) is a speech database containing recordings of Swedish dialects from all over Sweden and Swedish speaking communities in Finland. The database contains recordings of at least 12 speakers per dialect from 107 locations. A little over 1300 speakers have been recorded and the total recording time is about 800 hours. Each dialect is represented by two generations of speakers, an older generation 55–75 years of age and a younger generation 20–35 years of age. Each age group is represented by an equal number of male and female speakers. The data is organised in two separate databases – one publicly available database containing four short samples from each dialect and primarily intended for educational purposes, and a research database containing the entire material but with access rights limited to researchers. In this paper we will describe the criteria behind the selection of locations, speech types etc., the collection of data, the linguistic structure and properties of the database, examples of how the material is used and finally what we are presently doing to preserve the data for future generations of researchers.

Keywords: dialect; database; e-science.

1. Introduction

The SweDia 2000 database (we often refer to it as SweDat) is the result of two project efforts. The first project, *SweDia 2000 – Phonetics and phonology of the Swedish dialects around the year 2000*, was funded by the *Bank of Sweden Tercentenary Foundation* (grant 1997-5066:01/02 and ran between 1998 and 2003. During this period all the data was collected and a first version of the database set up. The goal of the present work on the database is to update data formats and to make the database available to the research community over the Internet. This is done within a follow-up project, *SweDia 2000 – A Swedish dialect database*, funded by *The Swedish Research Council* (grant 825-2007-7432) for the period 2007–2011).

In the following we will describe the considerations behind the selection of recording sites and the data collection procedure itself. The general properties and linguistic structure of the database will then be described and finally a description of the state of development we are in now and examples of the many different uses of the database for education and research.

2. General considerations

The goal was not, as is often the case in traditional dialectology, to find the most archaic samples of the selected dialects, but to collect samples representative of the linguistic varieties used in the daily lives of socially active people in the selected speech communities.

The chosen recording sites are evenly distributed over Sweden and the Swedish speaking communities in Finland taking into account both geographical dispersion and population density. The selection was done in close co-operation with dialect experts at the Swedish and Finnish dialect archives. Where there was more than one site that fulfilled the above mentioned two criteria, the site was chosen based on the amount of earlier material available in the dialect archives in order to maximize the possibility of historical comparisons. Only rural dialects were considered, no major towns are included in the data. The reason behind this decision was that the driving forces behind language change are quite different in the rural communities and major cities. In the cities change is driven by the influx of new inhabitants from other linguistic areas whereas the situation in the rural communities is almost the opposite, here non-mobility has been the major factor. Another consideration was the fact that rapid linguistic levelling is going on in many smaller communities and we wanted to capture the situation before that levelling had gone too far.

3. Data collection

The bulk of recordings were made during the summer of 1999. But a number of preliminary recordings were made already during 1998. These recordings were made to test the procedures with respect to recording techniques, interview types and logistics (travel arrangements, lodging facilities, time consumption etc.).

We were also not quite sure how many recordings per site would be necessary in order to control for inter-speaker variation. So the choices were more recordings per site and fewer sites or fewer recordings and more sites. Our goal was to collect data from two age groups, young adults aged 20–35 years of age and an older generation 55–75 years of age, and an equal number of male and female speakers in each age group. In the trial round we tested two alternatives, 5 or 3 subjects per group, that is a total of 20 or 12 speakers per location. Subsequent analyses indicated that 12 speakers per location would be sufficient. When all relevant factors had been considered the decision was to collect data from 107 different recording locations including the ones recorded in the trial round (see Appendix!).

As was mentioned above the goal of the project was to collect data representative of the speech used by socially active people in their daily lives. We therefore required that the participants should either still be working or should take active part in the social activities in their communities in some other way. For the younger

informants we also required that they should be second generation native speakers of the dialect. This was not a formal requirement for the older generation, but it turned out that most of them met the requirement anyway.

The plan was to record most of the dialects during the summer holidays of 1999. To be able to accomplish this, very careful preparations were of the essence. According to the time plan, a site should be completed in one work week and there was really no margin of error. For this to be possible everything had to be prepared well in advance. Data collection was made by linguistics students at the universities of Umeå, Stockholm and Lund. They were recruited in the beginning of 1999 and spent the spring term of 1999 planning the work. Informants were recruited via municipal organisations, social clubs etc. When the recordings began, 12 speakers per location had already been contacted and agreed to participate. We also had a few extra contacts in case anyone should be unable to participate, for example due to illness.

The field teams were recruited mainly among the students who had been responsible for the preparations. A team consisted of two students working together, taking turns in interviewing and handling the recording equipment. They had a rented car at their disposal, a credit card for expenses, Digital Audio Tape (DAT) recorders with lapel microphones, a mobile phone to manage contacts and a lap top computer for making notes.

They had all been thoroughly trained for the task both in terms of interviewing techniques and handling of the equipment. The students performing the field work were not generally native speakers of the dialects of the informants but often spoke some similar dialect. For some of the more deviant dialects we chose, however, to recruit students who were themselves native speakers of the dialect.

4. General properties of the database

The SweDia 2000 database has some properties, which as far as we are aware, are not common in otherwise comparable databases.

Synchronicity: All recordings were made within a narrow and precisely defined time slice. They therefore represent the dialectal variation at a precisely defined moment in time.

Consistency: The material has three well controlled parts that represent three fundamental, phonological properties – the quantity system, the accent system, and the phoneme inventory. It is thus possible to analyze and compare speech material of identical types for all dialects.

Completeness: The recorded material also contains about 30 minutes of spontaneous speech per speaker. This gives us additional information about how observed phonological rules are realized in everyday speech. It may also be used for other types of studies; for example studies of syntax and morphology (see below!).

5. Linguistic structure of the database

Linguistically, the database may be divided into two major parts – structurally controlled material and semi spontaneous speech.

The data in the controlled material consist of words or phrases repeated 3–5 times exemplifying the phoneme inventory of the dialect, the phonetic realization of quantity, and certain prosodic properties (word stress, tonal accent and phrasal focus).

The part intended for phoneme inventory analyses contained everyday words which could be assumed to have existed in the dialect (albeit not necessarily with the same pronunciation) for a very long time (several hundred years). The word lists were constructed in close co-operation with experts on historical dialectology from the departments of Swedish and researchers at the national dialect archives in Sweden and Finland.

The quantity word lists consist of minimal word pairs differing in quantity only. Old Swedish had a four-way quantity system (V:C:, V:C, VC:, VC). In modern Swedish only two of the contrasts are still used (V:C and VC:). There are, however, several dialects that still have a three-way system where VC is also contrastive. Many such examples exist in the recordings of semi spontaneous speech but unfortunately the quantity word list did not include such examples.

Swedish is not a tone language in the strictest sense of the term, but has nevertheless a contrastive tonal accent. Examples of tonal accent as well as word stress and phrasal stress may be found in the prosody part of the elicited material.

In order to influence the pronunciation of the target words and phrases in the controlled parts as little as possible, crossword-like word games were used to elicit the intended targets.

Most of the spontaneous material consists of informal interviews where the interviewer had been instructed to interfere as little as possible. In some cases, dialogues between two speakers of the dialect were used as an alternative.

6. Further development of the database

Maintaining the data and making it accessible for research is of course an important factor. This may seem as a fairly trivial task, but it is not. Sound format standards, for example, are changing over time. At the time of creating the database, we used an analysis package called ESPS/Waves. Neither the sound file format nor the format of the time aligned transcriptions are commonly used anymore and before not too long they will be completely outdated. It is therefore necessary to regularly update the file formats used in the database. There is simply no other way of long term preservation than regularly migrating the whole database to the currently favoured formats.

As mentioned above, the data consists of audio recordings and time aligned transcriptions. The original data in the ESPS/Waves format have now been converted to the currently most widely used formats – wav for the sound files and Praat TextGrid for the time aligned transcriptions.

Basic data about the speakers recorded for the database may be of great value for certain types of

linguistic studies. Minimally these data should contain information about speaker sex and age, educational level, vocational training and work experience. Some of the information presented in this paper about recording techniques, project descriptions (background and financing) and addresses to the people responsible for maintaining the database and monitoring access right should also accompany the database, ideally in the form of a meta database directly connected to the recorded data. There exists a now partly outdated version of such a meta database in the IMDI format developed by the Max Planck Institute. We are currently working on updating this database. Whether we will stay with the IMDI format has not been decided at the time of writing. We are also considering a move to a more modern and somewhat more flexible type of meta database, CMDI, also developed by Max Planck Institute.

7.   Preserving the data for future generations

In the previous paragraph we have described one problem connected with maintaining digital databases – continuous format changes. Other factors influencing accessibility is constant technological change and mobility among the people involved. If we want to preserve the data for future research and guarantee its availability, the data must be secured in a way that does not depend on specific individuals, formats, server locations etc. Trying to solve this problem is one of our main concerns at the present stage. Fortunately we are not the only ones who are actively looking for solutions to these problems. There is considerable activity going on in this field.

For the purpose of long term preservation only, a copy of the database will be hosted by the *Swedish National Data Service (NDS)*. But we are also working on a more advanced solution providing services specifically designed to service the speech research community. This service, *Speech & Language Repository (SLDR)* is hosted at the Aix-Marseille University in France.

8.   Examples of research based on the material in the database

Intonation as a function of dialect has been studied for Swedish for a long time. The first study appeared already in the thirties (Meyer, 1937). This has been followed by many more studies over the years. Based on data in the SweDia database, a group of researchers at Lund University have developed models to simulate the prosodic variation among Swedish dialects. This work has been done within a project called SIMULEKT and the results have been described in a number of publications (e.g. Bruce *et al.*, 2007, 2008).

Helgason has studied preaspiration in Nordic languages based, among other data, on material from the SweDia database (e.g. Helgason 2002, 2003).

Many more examples of studies using data from the SweDia database may be found in the publications list from the SweDia project (see. link at the end of this paper!).

9.   Language variation from a somewhat different angle

Traditionally, the driving forces behind language variation and change are considered to be geographical dispersion and isolation of groups of speakers as well as renewed contact as a result of migration. These factors are no doubt important, but if that were all there is, the observed variation is likely to be more chaotic than what we seem to observe. A basic tenet in the SweDia project is the belief that although there is certainly a random element involved in language change it is primarily rule governed. One way of approaching this question is to look for coherence, or clustering of phonological properties within the entire speech community rather than assuming any specific areal distributions.

Promising results along these lines have been obtained by approaching the description of regional distribution from an angle that does not assume any geographically based constraints at all. In three studies (Leinonen, 2010; Lundberg, 2005; Shaeffler, 2005) based on the SweDia 2000 data, cluster analysis has been used as a means of creating dialect "areas" based only on acoustically grounded phonological properties. In those studies, geographical areas are defined by dialects whose properties cluster together. This approach could, in principle result in a very scattered picture with no obvious geographical coherence. This did not, however, turn out to be the case. On the contrary, dialects grouped into geographical areas that in many cases closely resemble those suggested in traditional dialectology. If the clustering had been based on the same considerations as in the traditional analyses this would have to be seen as a rather trivial finding, but this is not the case at all. In all the above studies, cluster analyses were based solely on acoustic properties like formant frequencies (Leinonen; Lundberg) or segment durations (Shaeffler) never considered in traditional dialectology. The results in a study by Livijn (2010) on the articulation of coronals, using a similar approach but without using cluster analysis, point in the same direction. Moreover, there is considerable overlap between the areas resulting from these studies. This lends support for the assumption that dialectal change is rather strongly constrained by the compatibility of internal factors.

10. Additional uses of the database

In addition to the research database, there is also a limited version of the database developed for educational purposes in university courses on dialectology, secondary schools and study groups of interested individuals. This database contains speech samples from all dialects represented by short sound files (30–50 seconds) from one speaker per category (age/sex) together with simplified phonetic-like transcriptions and translations to standard Swedish. This database may be accessed over the Internet. At present the interface exists only in Swedish. There are no immediate plans to translate the

interface.

A group of researchers at Lund University are using material from the database for studies of dialect syntax. They are part of a Nordic network of dialect syntax researchers (ScanDiaSyn). Studies of this kind were not envisaged when our data were collected, but we are pleased to see that the data can be fruitfully used also for such studies. To support their efforts we supply the ScanDiaSyn database hosted at the University of Oslo with data for their studies.

Although the data were collected for the primary purpose of studying language variation and change in the phonological domain, the usefulness is not necessarily limited to that area. As mentioned above, the data is now used also for the study of dialect syntax.

The database contains data from speakers of ages ranging from 20 years of age up to 75 years of age for both male and female speakers. That means that in addition to language variation data the database can be used to study speaker variation as a function of age. This has been done in a series of studies by Schötz. In her doctoral dissertation (2006) she studied the variation of parameters such as fundamental frequency, formant frequencies, jitter, shimmer and speech rate as a function of age. These results were then used as a basis for a model that could be implemented in speech synthesis to simulate speaker age. This has been further developed in later studies (e.g. 2007).

Another successful use of the data is as a reference database for automatic speaker recognition for forensic purposes. This has been described in Lindh and Eriksson (2009).

11. Summary

In this paper we have presented the SweDia project and the database created and developed within the project and in the last paragraphs we have given many examples of various uses of the data, not only uses which are primarily in the area of dialectology or even linguistics in a restricted sense. This may be seen as an example of what is often referred to as e-science, that is re-using existing data for new research, not envisioned when the data was collected but made possible because the data now exist.

12. Acknowledgements

13. References

Bruce, G., Granström, B., and Schötz, S. (2007). Simulating Intonational Varieties of Swedish. In *Proceedings of ICPhS XVI*, Saarbrücken, Germany, pp. 1237--1240.

Bruce, G., Schötz, S., Granström, B., and Enflo, L. (2008). Modelling intonation in varieties of Swedish. In *Proceedings of Speech Prosody 2008*, Campinas, Brazil, pp. 571--574.

CMDI home page. Available at: <www.clarin.eu/cmdi>.

Helgason, P. (2002). *Preaspiration in the Nordic Languages: Synchronic and Diachronic Aspects*. Doctoral Dissertation. Department of Linguistics, Stockholm University.

Helgason, P., Stölten, K and Engstrand, O. (2003). Dialectal and sociophonetic aspects of preaspiration. In *Proceedings of ICPhS XV*, Barcelona, Spain, pp. 17--20.

IMDI home page. Available at: <www.mpi.nl/imdi/>.

Leinonen, T. (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects*. (Doctoral Dissertation), Groningen University, Groningen Dissertations in Linguistics, 83.

Lind, J., Eriksson, A. (2009). The SweDat Project and Swedia Database for Phonetic and Acoustic Research. In *Proceedings of e-Science 2009*, Oxford, UK, pp. 45--49.

Livijn, P. (2010). *En perceptuell och akustisk studie av svenskans koronaler i ett dialektperspektiv*. Doctoral Dissertation. Department of Linguistics, Stockholm University.

Lundberg, J. (2005). *Classifying Dialects Using Cluster Analysis*. Master's Thesis in Computational Linguistics, Department of Linguistics, University of Gothenburg.

Meyer, E. A. (1937). *Die Intonation im Schwedischen I: Die Sveamundarten*. Studies in Scandinavian Philology Nr. 10. Stockholm University.

Reference list of Swedia 2000 publications. Available at: <www.ling.gu.se/~anders/SWEDIA/publ_eng.html>.

Schaeffler, F. (2005). *Phonological Quantity in Swedish Dialects*. Doctoral Dissertation, Umeå University: PHONUM 10 – Reports in phonetics.

Schötz, S. (2006). *Perception, Analysis and Synthesis of Speaker Age*. Doctoral Dissertation, Lund University, Department of Linguistics and Phonetics, Centre for Languages and Literature.

Schötz, S. (2007). Analysis and Synthesis of Speaker Age. In *Proceedings of ICPhS XVI*, Saarbrücken, Germany, pp. 1841--1844.

Speech & Language Repository (SLDR). Available at: <http://crdo.up.univ-aix.fr/>.

SweDat project presentation. Available at: <http://swedat.ling.gu.se>.

Swedish National Data Service. Available at: <http://snd.gu.se/en>.

The educational database. Available at: <http://swedia.ling.gu.se>.

14. Appendix

Överkalix
Arjeplog
Kalix
Nederluleå
Sorsele
Piteå
Vilhelmina
Burträsk
Frostviken
Vindeln
Strömsund
Fjällsjö
Bjurholm
Nysätra
Anundsjö
Åre
Aspås
Munsala
Ragunda
Vörå
Storsjö
Berg
Kramfors
Vemdalen
Torp
Indal
Närpes
Lillhärdal
Färila
Delsbo
Särna
Älvdalen
Orsa
Ovanåker
Skog
Dalby
Leksand
Ockelbo
S, Finnskoga
Årsunda
Gräsö
Saltvik
Brändö
Houtskär
Borgå
Malung
Grangärde
Nora
Kyrkslätt
Gräsmark
Skinnskatteberg
Skuttunge
Skillingmark
Villberga
Dragsfjärd
Snappertuna
Gåsborn
Köla
Järnboås
Haraker
Kårsta
St. Mellösa
Bengtsfors
Viby
Länna
Skee
Sorunda
Torsö
V. Vingåker
Frändefors
Tjällmo
Korsberga
Orust
Floby
Rimforsa
S:t Anna
Östad
Fårö
Kärna
Asby
Ankarsrum
Fole
Järsnäs
Öxabäck
Sproge
Stenberga
Böda
Frillesås
Burseryd
Bredsätra
Årstad-Heberg
Hamneda
Väckelsång
Våxtorp
Broby
Torsås
Segerstad
Bjuv
Össjö
Jämshög
N. Rörum
Hällevik
Torhamn
Löderup
Bara

Figure 1: The geographical distribution of recording sites

Easyalign for Brazilian Portuguese: a (semi) automatic segmentation tool under Praat

Jean-Philippe GOLDMAN[1], Maíra Avelar MIRANDA[2], Cirineu Cecote STEIN[3], Antoine AUCHLIN[1]

[1]University of Geneva, Switzerland; [2]Federal University of Minas Gerais, Brazil; [3]Federal University of Paraíba, Brazil
jeanphilippegoldman@gmail.com, mairavelar@gmail.com, cirineu.stein@cchla.ufpb.br, Antoine.Auchlin@unige.ch

Abstract

This communication presents an automatic phone-text alignment system, EasyAlign, in its latest adaptation to Brazilian Portuguese. Automated steps are crucial in large corpora prosodic investigation As opposite to time-consuming human alignment, they are both more consistent, and reproductible. They also are open to adaptation and improvements. One issue is the tool's precision in alignment at phone level.

Keywords: automatic segmentation; automatic alignment; Brazilian Portuguese; EasyAlign.

1. Introduction

Phonetic alignments (or phonetic segmentation) purpose is to determine the time position of phone, syllable, and/or word boundaries in a speech corpus of any duration, on the basis of the audio recording and its orthographic transcription. Resulting aligned corpora are widely used in various speech applications, such as automatic speech recognition, speech synthesis, as well as prosody and phonetic research.

Conducting fully manually an accurate segmentation would require as many as 800 times real-time; i.e. 13 hours for a one-minute recording (Schiel & Draxler, 2004). Processing time is a major drawback for manual labeling, especially facing with very large spontaneous speech corpora. This is why an automatic phonetic alignment tool is highly desirable. Such an automatic approach, besides, is not only consistent (i.e. has the same precision throughout the corpus), it also is reproducible (i.e. can be repeated, within a short time interval, and many times). An alignment tool can save time, but speech, especially spontaneous speech, presents unpredictable phonetic variations that can decrease process' accuracy. Even with precise computational tools and data preparation, automatic systems can make errors that a human would not. Thus, manual or automatic post-processing detection of major segmentation errors is needed to improve accuracy. Automatic approaches are never fully automatic nor straightforward and instantaneous as claimed by existing systems. It is a matter of balance between time, expected precision and computational skills. The determination comes from the needed degree of accuracy; i.e. a corpus-based text-to-speech (TTS) system needs a high precision, but other studies (at syllable level) require a lesser precision.

For automatic phonetic alignment, several methods have been designed: some borrow techniques from the automatic speech recognition (ASR) domain. But, the alignment task is much easier than speech recognition as the alignment tool does not need to determine what the segments are but only their position in time. For that, HMM (Hidden Markov Models)-based ASR systems are used as a forced-alignment process for segmentation like HTK in (Young & Woodland, 2000). Other approaches combine a TTS system and a Dynamic Time-Wrapping (DTW) algorithm. In this case, the orthographic transcription is used to synthesize speech, which is compared to the authentic speech to segment as in (Malfrère, 2003). The DTW algorithm finds the best temporal mapping between the acoustic features of the two enunciations. A dual system based on these two approaches (first HMM then TTS+DTW) is presented in (Sérgio & Oliveira, 2004), with better results. Finally in Van Santen and Sproat (1999), contour detection techniques are borrowed from image processing, providing relevant results. Although these systems are usually freely available and give good results, it should be noted that they are not ready to use, as a training of the acoustic models is required.

The presented system, named EasyAlign, relies on HTK (for HMM ToolKit), a well-known HMM package. It can be seen as a friendly layer within Praat software (Boersma & Weenink, 2009), which does the whole alignment process as it is provided with a grapheme-phoneme conversion system and embeds already trained acoustic models.

2. EasyAlign

EasyAlign (Goldman, 2011) is a plugin developed for Praat. It produces semi-automatically a multi-tier annotation with a phonemic, syllabic, word and utterance segmentation from a sound recording and the corresponding orthographic and phonetic transcriptions. The plugin is made of Praat scripts, but it also includes two external components: a grapheme-to-phoneme conversion system and a segmentation tool for the alignment at the phone level. Consequently, the whole procedure is a succession of 3 automatic steps in between which some manual adjustments may be necessary. The 5 resulting tiers are grouped in TextGrids and named as phones, syll, words, phono and ortho as illustrated in Figure 1.
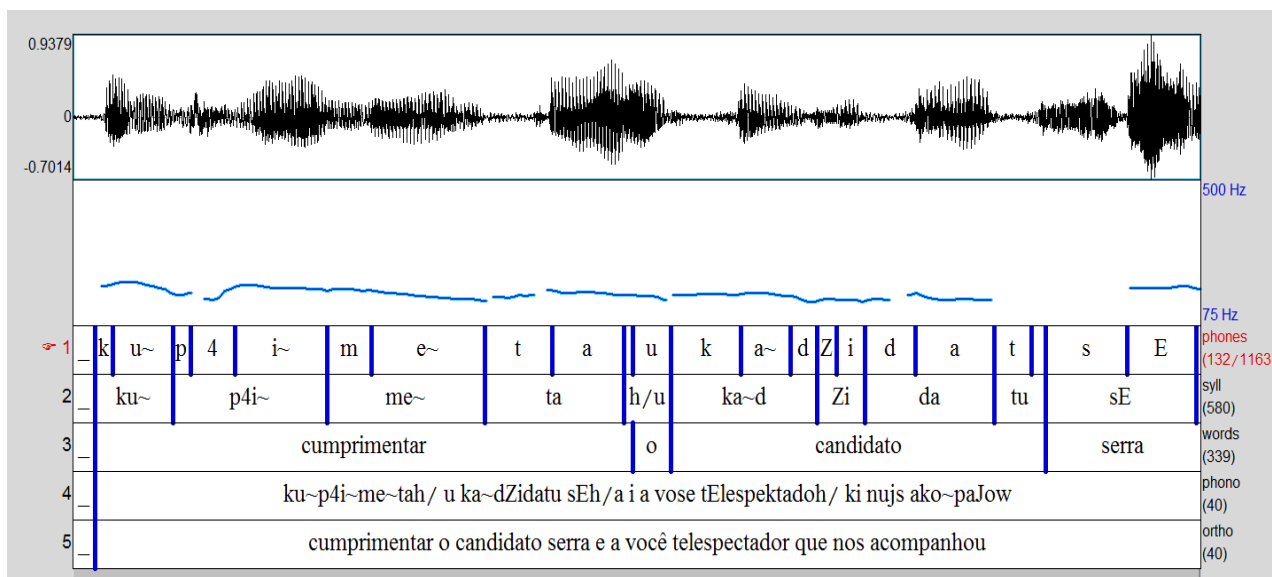
Figure1: Full resulting TextGrid with 5 tiers from bottom to top:
ortho, phono, words, syllables, phones for the utterance "cumprimentar o candidato serra"

The tool is already designed for French, Spanish and Taiwan Min and has been recently adapted to Brazilian Portuguese (BP). It is freely available and works on Windows only. It is distributed as a self-installable plug-in, and comes with the already trained acoustic models of the phones.

The segmentation of a speech file occurs as follows: from a speech audio file and its corresponding orthographic transcription in a text file, the user has to go through three automatic steps; manual verifications and adjustments can be done in-between to ensure even better quality. More precisely, these three steps are as mentioned in (Goldman, 2011).

2.1   Macro-segmentation at utterance level

The first automatic step of EasyAlign creates a macro-segmentation as a TextGrid with one tier named ortho, on the basis of the previously loaded Sound object (the sound file to segment) and Strings object (the transcription). After this step, the Sound and the new TextGrid are opened. The internal algorithm computes a heuristics based on signal duration and utterance transcription length to estimate utterance duration and also relies on pauses.

2.2   Grapheme-to-phoneme conversion

The second step duplicates the ortho tier into a phono tier (i.e. with the same boundaries) but replaces the orthographic transcription by a phonetic transcription according to the SAMPA phonetic alphabet. The grapheme-to-phoneme module of a full text-to-speech system named eLite and developed at Faculté Polytechnique de Mons (Belgium), does a linguistic analysis of the orthographic transcription to produce a phonetic transcription based on a phonetic dictionary and pronunciation rules.

2.3   Phone segmentation

The third step aims at creating the phones, syll and words tiers. For each utterance, the orthographic and phonetic transcriptions are used by a well-known speech recognition engine named HTK (HMM ToolKit) set to a "forced alignment" mode to catch the temporal boundaries of phones and words. The syllabification is rule-based. Two main principles are: 1. there is one and only one vowel per syllable; and 2. the sonority principle is used to split the consonant clusters. The pauses are also used as syllabic boundaries.

2.4   Result

The result is a multi-tier TextGrid, the annotation format within Praat, with phones, syllables, words and utterance segmentation, as showed in Figure 1. It is important to highlight that, before performing each of these steps, it is necessary to make some manual adjustments, as showed in Figure 2.

As it can be observed, at first, we have a preliminary manual step (if the transcription is in a paragraph format and/or without punctuation), in which the user has to reformat the transcription file with one utterance per line. After that, an utterance segmentation script is run, which creates a TextGrid with an interval tier ortho containing the transcription. The user, then, manually verifies the utterance boundaries. Next, it is done the automatic grapheme-to-phoneme conversion: the script duplicates the ortho tier to the phono tier, generating a phonetic transcription with major variations. At this point, the user may validate the phonetic transcription to ensure sporadic phonological variant of pronunciation. This optional time-consuming task might be skipped. Finally, the phone segmentation is automatically performed: the script is run and generates the phones and words tiers, then the syllables tier.
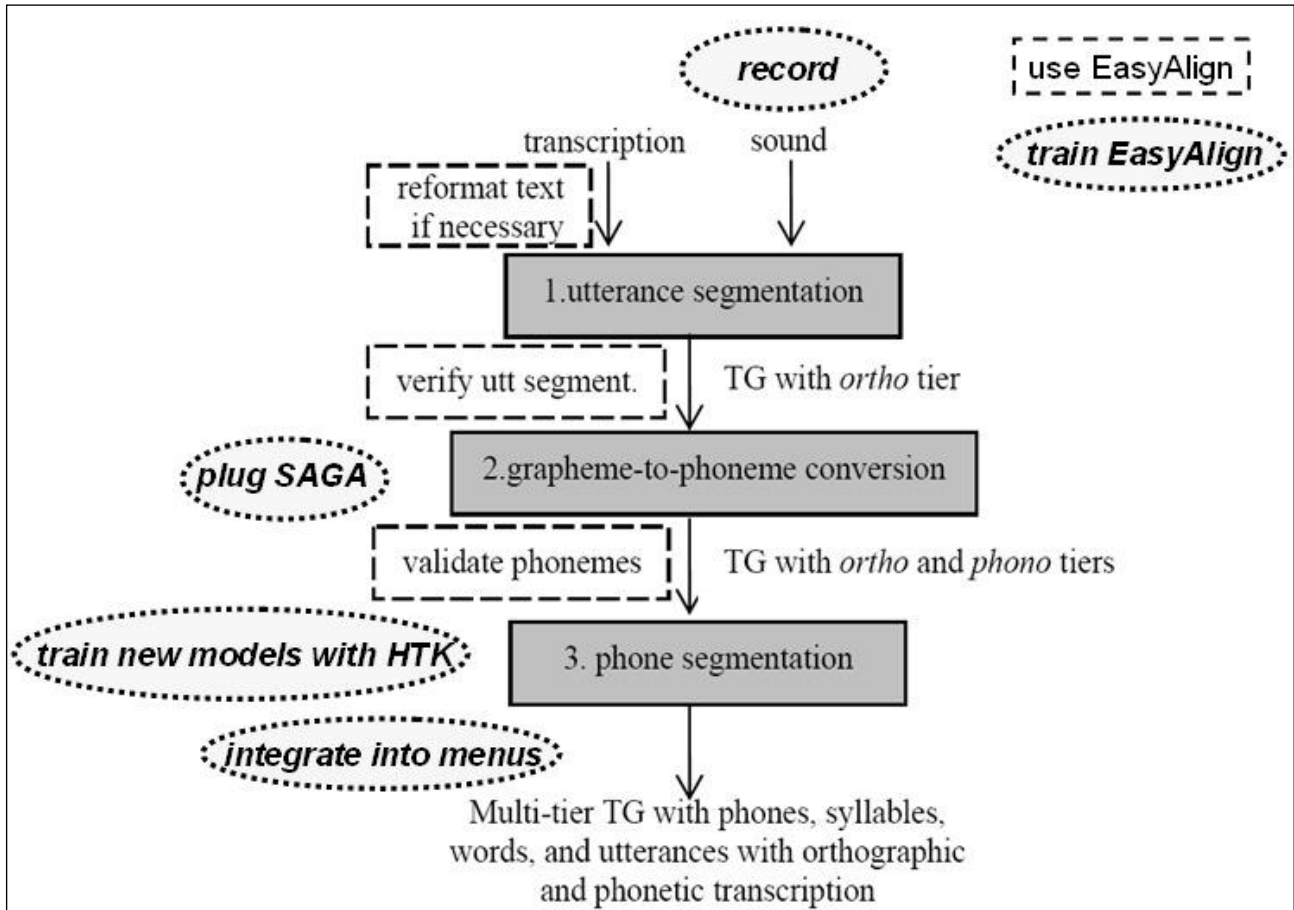
Figure 2: EasyAlign usual process in square boxes as in Goldman (2011) and the adaptation steps in oval shapes

3.   EasyAlign for Brazilian Portuguese

3.1   Development

EasyAlign adaptation to a new language needs: some speech data and a grapheme-to-phoneme conversion system. First of all, we selected two audio samples, that have a total duration of 20 minutes (10 minutes produced by a male reader and 10 minutes by a female one), from a corpus (Barbosa *et al.*, 2004) composed by 4 subjects, 2 males and 2 females, reading sentences. The corpus was manually aligned. Then, we integrated the "Conversor Grafema-fone v1.6" (Grapheme to Phone conversor 1.6) phonetizer within EasyAlign to convert the orthographic transcription to the phonetic one.

This phonetizer was developed by Fala Brasil team (http://www.laps.ufpa.br/falabrasil/) at the Federal University of Pará (UFPA) Brazil (Siravenha *et al.*, 2008). Despite its good performance, some problems must be solved, such as the specification of a dictionary of exceptions for open vowels (the ones which are not predicted by phonological rules, or are no longer predicted through orthography, due to the latest official orthographic changes in Portuguese). That improvement is on course.

As shown in the *phono* tier of Figure 1, the grapheme-phoneme conversion tool provides a phonetic transcription, in SAMPA alphabet, on the basis of the orthographic transcription. The phonetic transcription of each utterance was manually checked so as to exactly match the produced utterance.

In the end, a HTK-based stochastic training was performed with this speech material and their phonetic transcription. The result is a collection of acoustic models. The Figure 2 shows the necessary steps to train then to use EasyAlign.

3.2   Evaluation

According to (Goldman & Schwab, 2011), "the evaluation of a semi-automatic system can be seen in two ways: i) its automatic performance, i.e. how robust and accurate the automatic tool is, and ii) its ergonomics, i.e. how the whole process is made easier and how many times real-time it takes".

The automatic performance has been evaluated on the basis of a corpus of twenty-four minutes, that was manually annotated by one phonetic expert (reference alignment) and was compared to the automatic alignment. Among the speakers, two were "internal" speakers, used in the training corpus and two were new "external" speakers, taken from political debates broadcasted by Record TV channel. The internal corpus was composed by twenty minutes (10 minutes produced by a male speaker and 10 composed by a female one) and the

external corpus was composed by four minutes (2 minutes produced by a male speaker and 2 by a female one).

Evaluation was performed according to three approaches: a boundary-based, a duration-based and a segment-based approach. In each of the evaluations, the pauses were discarded, and only phones were taken into account. As some segments might be very short, especially in spontaneous speech, the evaluation was done with two thresholds: the 20ms (as mentioned above) and a narrower one set at 10ms.

3.3 Boundary-based evaluation

In this first evaluation, we computed the absolute difference (in ms), for each phone (n = 11650), between the automatic and the manual initial boundaries. Results showed that 43.2% of the differences between automatic and manual boundaries lie within 10ms, and 73.1% within 20ms for the internal corpus.

Internal evaluation

Differences	Total
Within 10 ms	43, 2%
Within 20 ms	73,1%

Table 1: Boundary-based evaluation for the internal corpus

External evaluation

Differences	Total
Within 10 ms	33, 2%
Within 20 ms	42,5%

Table 2: Boundary-based evaluation for the external corpus

According to the
Table 1: Boundary-based evaluation for the internal corpus
Table 2: Boundary-based evaluation for the external corpus, 73% of the automatic boundaries of the interval corpus are less than 20 ms from the correspondent manual boundary. As for the external evaluation, there are slight differences between 10 and 20 ms. This similarity demonstrates that the acoustic training is not broader enough to make generalizations.

3.4 Duration-based evaluation

For each phone, we looked at the difference between the automatic and manual segment durations.

Internal evaluation

Duration	
Mean	0.017
Sdev	0.034

Table 3: Duration-based evaluation for the internal corpus

External evaluation

Duration	
Mean	0.014
Sdev	0.070

Table 4: Duration-based evaluation for the external corpus

The mean value is not very significant, whereas the standard deviation explains the variation of the error (duration difference). Again, the internal corpus gives better results than the external corpus.

3.5 Segment-based evaluation

According to Goldman and Schwab (2011), in the segment-based evaluation, we computed, for each phone, the Overlapping-rate, a speech-rate independent measure (Sérgio & Oliveira, 2004), which represents the ratio between the common part of the automatic and manual segment and the maximal duration of the segment considering initial and final boundaries of both automatic and manual segmentations. A rate of 0 means that there is no overlap between the automatic and manual segments, while a rate of 1 means that the overlap is total. According to Sérgio and Oliveira (2004), a segment with an overlapping rate of 0.75 is considered well segmented.

Internal evaluation

OVR	
Mean	0.671
Sdev	0.239

Table 5: OVR evaluation for the internal corpus

External evaluation

OVR	
Mean	0.377
Sdev	0.374

Table 6: OVR evaluation for the external corpus

In summary, the mean value is much higher for the internal corpus than for the external corpus, which indicates a better overlapping rate for the internal corpus, and thus the need of a better training.

4. Conclusion

The 3 kinds of evaluation were done – boundary-based, duration-based and segment-based. All of them showed promising results, with a good training of the training corpus. On the other hand, the external evaluation corpus was under-represented and, consequently, generated poor results. We need, then, to increase the size of the training corpus to obtain a more accurate training and make good generalizations from the external evaluation corpus.

To sum up, EasyAlign appears as an friendly and

efficient tool which helps aligning speech from an orthographic transcription within Praat. The tool is freely available online and is complemented by a demo mode and a tutorial. It can be downloaded from this link:

http://latlntic.unige.ch/phonetique/easyalign

To our knowledge, such a tool was not, until now, available for Brazilian Portuguese.

5. Acknowledgements

6. References

Barbosa, P., F. Resende Jr., L. Couto, and J. A. Moraes. (2004). Obtenção de 1000 frases foneticamente balanceadas para o português brasileiro utilizando a abordagem de algoritmos genéticos. In *Anais da Semana Eletrônica 2004.* Rio de Janeiro: UFRJ.

Boersma, P., Weenink, D. (2009). Praat: doing phonetics by computer. *Computer program version 4322.* Available at: <www.praat.org>.

Goldman, J-P. (2011). EasyAlign: an Automatic Phonetic Alignment Tool under Praat. In *Proceedings of Interspeech*, pp. 3233--3236.

Goldman, J-P, Schwab, S. (2011). EasyAlign Spanish: an (semi-)automatic tool under Praat. In *V Congreso de Fonética Experimental*. Cáceres.

Malfrère, F. (2003). Phonetic alignment: speech synthesis-based vs. Viterbi-based. In *Speech Communication* 40 (4): pp. 503--515. doi:10.1016/S0167-6393(02)00131-0. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0167639 302001310>.

Van Santen, J., Sproat, R. (1999). Highaccuracy automatic segmentation. *System*. Available at: <citeseer.ist.psu.edu/vansanten99highaccuracy.html>.

Schiel, F., C, Draxler. (2004). The Production of Speech Corpora Bavarian Archive for Speech Signals.

Serridge, Castro. (2008). Faster time-aligned phonetic transcriptions through partial automation. In *ExLing*.

Siravenha, A.C., Neto, N., Macedo, V. and Klautau, A. (2008). Uso de Regras Fonológicas com Determinação de Vogal Tônica para Conversão Grafema-Fone em Português Brasileiro. In *7th International Information and Telecommunication Technologies Symposium*.

Sérgio, P., Oliveira, L. (2004). Automatic Phonetic Alignment and Its Confidence Measures. In *4th EsTAL*, pp. 36--44. Springer Verlag.

Young, S., Woodland, P. (2000). *The HTK Book*. Ed. Microsoft Corporation. *Network*. Vol. 2. Cambridge: Cambridge University Press. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=1 0.1.1.124.3972&rep=rep1&type=pdf>.

DB-IPIC: an XML database for informational patterning analysis

Lorenzo GREGORI, Alessandro PANUNZI

Università di Firenze
E-mail: lorenzo.gregori@unifi.it, alessandro.panunzi@unifi.it

Abstract

DB-IPIC is a linguistic web resource for the analysis of spoken language based on the Informational Patterning Theory of E. Cresti and M. Moneglia. The corpora stored inside the database take parts of the C-ORAL-ROM and C-ORAL-BRASIL projects and enrich them with informational and PoS tagging. This paper focuses on DB-IPIC's construction, from the annotation processes of acoustic sessions to the retrieval capabilities of the web interface. In the first part we give a short overview of the theoretical framework on which the database has been structured and we describe the annotation procedure of speech sessions. In the second part we explain the XML data model and the conversion process from annotated data to XML. Finally, we describe the steps that have been followed to build DB-IPIC itself, along with its querying capabilities; in particular we'll describe the web interface and its features for extracting information patterns and analysing results.

Keywords: DB-IPIC, XML database; information patterning; C-ORAL-ROM.

1. Introduction

DB-IPIC is a database of transcribed and annotated spoken language: in this paper we are going to describe this resource, focusing on the data types comprising the database and on the tools provided by the web interface to query it. At the moment, the database stores a corpus of 74 spoken Italian language texts chosen from the Informal section of Italian C-ORAL-ROM (Cresti, Panunzi & Scarano, 2005). The whole corpus has been tagged with respect to the informational structure and exploited to build a queryable XML database for the study of linear relations among Informational Units in spoken language (Panunzi & Gregori, 2012). In addition to this we inserted a subset of the C-ORAL-BRASIL (20 texts; Raso & Mello, 2012) corpus and provided an Italian collection with the same size for comparison with the Brazilian one (Mittmann & Raso, 2012).

Besides the database, DB-IPIC includes a web interface that provides an easy means of extracting complex data from the corpora. With this tool it's possible to query the database, crossing different kinds of information stored in different logical levels (the logical model is explained in paragraph 2). The DB-IPIC web interface is specifically designed for the search and analysis of information patterns and the comparison of the informational values and prosodic profiles of linguistic structures (Mittmann et al., in this volume). Beyond this, DB-IPIC provides more search features, such as part-of speech filtering and communicative context restriction.

1.1 Theoretical framework

DB-IPIC is built in accordance with Language into Act Theory and Informational Patterning Theory (Cresti, 2000; Cresti & Moneglia, 2010). These two theoretical models form a framework that can be productively applied to the annotation of spoken language.

The framework identifies two different pragmatic levels in oral production: the first one is macro-pragmatic and deals with Speech Act production (Austin, 1962; Cresti, 2000), while the second is micro-pragmatic and deals with the informational structure. Both of these levels are governed by prosody which splits the speech flow into terminated sequences and tone units using terminal and non-terminal breaks, respectively. These breaks are pragmatically defined as perceptually relevant prosodic variations in the speech flow (Cresti & Moneglia, 2005: 17) and acoustic source analysis has revealed a link between prosodic breaks and F0 behaviour.

At the macro-pragmatic level, the oral performance is structured into Utterances, which correspond to the pragmatic reference unit for spoken language. An Utterance is a sequence of words that can be pragmatically interpreted and corresponds to a Speech Act. On the prosodic side, an Utterance is included in a Terminated Sequence (TS), which ends with a perceptually identifiable terminal break. So, at this level, we have the TS that is prosodically recognizable by the terminal break and even pragmatically interpretable, since it achieves an Utterance.

At the micro-pragmatic level, Utterances can also be divided into sub-elements that are coherent with respect to the information value they carry. These elements are called Information Units (IU); on the prosodic side, IUs are segmented by non-terminal breaks, which split the TS into a sequence of Tone Units (TU). The Comment is the core IU of an Utterance and corresponds to the expression of an illocutionary force, being necessary as it ensures pragmatic interpretability. The Comment can be surrounded by other IUs, each with a specific informational value. The IUs can be divided into two main classes: the textual units that participate in the construction of the semantic content of the Utterance (Topic, Appendix, Parenthesis, Introducer), and the dialogical units that are devoted to the successful pragmatic performance of the Utterance in the communicative context (Incipit, Phatic, Allocutive, Conative, Connector, etc.). The full tagset with descriptions is available in Table 1 in appendix.

Within the proposed theoretical framework, each Utterance consists of a pattern of IUs that is roughly isomorphic to a pattern of TUs (*informational patterning principle*). Therefore, there is a strict connection in the

definition of the two pragmatic levels of spoken language: (a) An Utterance is defined as the linguistic expression of a Speech Act, but it can be also viewed as a pattern of IUs; (b) the informational pattern necessarily contains a Comment IU, which properly accomplishes the illocution and corresponds to a single TU.

There are "exceptions" in which the theoretical principles explained previously cannot be applied. Two cases in particular should be mentioned:

- Illocutionary patterns: these structures occur within an Utterance when the Comment carrying the illocution doesn't "fit" inside a unique TU; in these cases a Multiple Comment is produced which consists of a pattern of two or more TUs (linked together through a compositional informational/prosodic model) with an overall illocutionary value;
- Stanzas: these structures are oral performances in which there is more than one Comment unit in sequence with weak illocutionary force. These sequences do not form compositional units, but they are produced by progressive adjunctions, outside of any informational/prosodic model. Stanzas correspond to a linguistic "activity" whose primary intention is the production of an oral text.

Thus, three types of Comment unit are defined:

- Comment (COM): the standard Comment IU, which accomplishes the illocutionary force of the Utterance and corresponds to a single TU;
- Multiple Comment (CMM): a complex IU composed of two or more TUs and forming an *illocutionary pattern*;
- Bound Comment (COB): occurring within Stanzas and corresponding to a non-patterned sequence of Comments with weak illocutionary force.

In short, there are two referring units to which a TS can correspond: the Utterance, which mostly aims at an interactive exchange with the interlocutor (Speech Act performance), and the Stanza, which intends the construction of a "text" by the speaker. Within the database, these units are distinguished with different tags (different attributes of the TS element). Furthermore, single and multiple Comments are recognizable inside the data model because of the different labels applied to the IUs in these two structures (COM vs. CMM).

1.2 Annotation procedure

Following the sketched theoretical framework, the prosody drives the annotation procedure. Since the prosodic segmentation of speech flow is strictly connected to the pragmatic features, the first annotation step consists of marking terminal and non-terminal breaks. This is done manually and occurs in parallel with the transcription procedure. For this step the annotators used WinPitch (Martin, 2005), a tool which allows one to listen an audio recording and carry out the text-sound alignment of a transcription, as well as to analyse the acoustic features of the source (in particular, an F0 real-time examination is required to safely determine the breaks).

The second annotation step, also performed manually, consists of tagging the IUs and exploits the *informational patterning principle*: once the prosodic boundaries of a TS and the internal pattern of the tonal units have been detected, it is possible to mark each TU with its informational value to get the informational pattern (Scarano, 2009; Moneglia, 2011). This is divided into two stages: first the Comment unit is identified and the TS type determined (Utterance or Stanza), and then the other TUs are tagged.

Finally, general session metadata, comprehending details about the audio source, the participants features, and communicative context are added to the annotation. Data and metadata are written in a CHAT-like format (MacWhinney, 2000; Moneglia & Cresti, 1997).

The result of this work is a set of sessions with audio, transcription, text-sound alignment, prosodic annotation, and information structure annotation. A corpus created following this multi-level annotation procedure is a resource rich in data that can be used as reference for studies on spoken language. But, as we will see in the chapter below, the original structure of the corpus does not make it effectively accessible to the scientific community: for these reasons, the production of an integrated database has been necessary.

2. Annotation tree and XML model

As a result of the manual annotation procedure, we have three files for each spoken session: an RTF file containing metadata, a transcription, and annotation tags, the WAV audio file of the recording, and the WinPitch XML file containing the text-sound alignment information. For the purpose of building a queryable resource, this representation format has several problems: data are sparse, RTF is not a real standard, annotations and transcriptions are written in a non machine-readable format, and all information is inserted inline into the text file without considering the dependence among different annotation levels. For these reasons a new representation model was developed.

As mentioned previously, we have two main structures involved in the segmentation of speech flow, Terminated Sequences and Tone Units, and one is superordinate to the other (the informational features can simply be added as labels belonging to the elements).

We then have high level metadata specifying session features and low level data that includes transcription and prosodic annotation. A peculiarity of this multi-level annotation is that it is structured as a tree, in which logical levels are linked in a hierarchical data model. This is one of the reasons that led us to use XML as the standard format for the corpus representation (Gregori, 2011).

XML has many good features causing it to be widely

used, especially for encoding and sharing information throughout the web, but, generally, corpora with multi-level annotation cannot be easily stored in an XML model. Commonly, each annotation level is independent from the other ones and a tree is too rigid for representing the data structure: we can say that it's typically difficult to encode a multi-level annotated corpus into XML without losing human-readability feature, as more than one file is needed per session. Otherwise, our collection fits well into an XML tree and the features of this language make it a good choice for storing DB-IPIC: firstly, the XML format allows an efficient standardization of the annotated data and formal validation. Moreover, XML is able to encode information that requires different kinds of representation (category, structural and relational information) and its elements are organized into a hierarchical model. Finally, the XML "family" comprehends query languages directly applicable to the annotated texts. The necessity to find a representation format for the IPIC collection led to the development of an XML data model and of software for automatic data migration in the new format.

An additional feature we decided to insert into the corpus is PoS-Tagging, so another annotation level has been inserted into the XML model. This additional information is derived automatically using TreeTagger. TreeTagger execution runs automatically inside our internal software that converts the corpus into XML format.

Each session of the corpus is composed of the following data types:

- an audio stream, containing the audio recording;
- general metadata, containing details about the session (audio quality, communicative context, etc.);
- transcription, consisting of a text that reports the word sequence;
- prosodic annotation, containing speech flow segmentation;
- information structure annotation, specifying the informational rule of any TU;
- morpho-syntactic annotation, automatically induced by TreeTagger;
- text-sound alignment, generated at the transcription procedure time by the WinPitch software.

All these data has been structured into an XML model according to the theoretical framework and considering the following relational rules among the levels:

- transcription data are at the lower level of the annotation tree; each transcription element is qualified depending on its nature (word, break, fragment, paralinguistic);
- part-of-speech and lemma are properties of words;

- TUs are superordinate to transcription elements and IUs are isomorphic to them, so informational values are properties of the TU elements;
- TSs are superordinate to TUs and they have a number and a type that depends on the relative reference unit: Utterance or Stanza. Alignment data are also a property of TSs, since it specifies their start and end times;
- general metadata is independent from the annotation tree: it depends only on the session;
- the session is the root level and includes all the other data.

This model has been translated into XML: objects and properties have been transformed into elements and attributes, preserving their logical difference. Figure 1 in appendix shows the structure.

3.   DB-IPIC resource

As the IPIC collection is stored in XML files, we decided to use an XML database to index it and make it queryable. Even if this kind of storage technology is not as efficient as common relational databases, the choice is justified by the fact that we have a unique data format for representation and querying. In addition to this, the corpus size is adequate enough to yield a good response time for any query.

We chose eXist-db, which is an open source software that runs as a server and can be queried via web protocols using the standard query language "XQuery". A user-friendly web interface has been developed in PHP to allow the extraction of informational patterns from the database (Figure 2). With this tool it's possible to query the corpus at different levels in relation to the logical structure of the data set. In particular DB-IPIC can operate on five levels:

1. data source: it is possible to query the whole corpus or to specify a subset of sessions; different corpora can be managed in DB-IPIC;
2. metadata: sessions can be filtered by their properties, specifying the communicative context (familiar or public) and the interaction type (monologue, dialogue, or conversation);
3. informational patterns: the user can select the TSs by specifying their IU pattern;
4. information units: it's possible to search TSs containing or not containing specific IUs independently of their informational pattern;
5. words: finally, the user can refine their search by including or excluding words with a specific form, PoS, or lemma.

As mentioned, the main purpose of the DB-IPIC resource is the search of information patterns and for this reason it provides advanced features for searching the objects inside the corpus. The following actions are allowed:

- definition of multiple sequences of IUs at once by using regular expressions: each element of the IU pattern can be extended to a variety of IUs using the W3C regular expressions syntax (Peterson et al., 2012);
- selection of the linear relation among the IUs of the pattern by specifying what IUs can optionally interrupt the sequence. There are five possible choices, from the more rigid, in which the IUs must be adjacent, to the freest, in which there are no restrictions about the IUs that can interrupt the sequence;
- specifying the content of each IU of the pattern in terms of word form, part-of-speech, and lemma. For this feature we developed a lightweight CQL[1] parser and a graphical tool that helps the user to write the restriction in the correct syntax.

In addition to the information pattern definition, DB-IPIC allows one to make complex queries through the intersection of the five logical levels described above.

Figure 2: DB-IPIC web interface

We can take Figure 2 as an example of the search capabilities of the resource: we decided to retrieve the dialogues in public context and excluding Stanzas (*General filter* section) from the Italian corpus (*Source selection* section); each Utterance must contain an Appendix of Comment and the lemma "essere" and cannot contain a Multiple Comment (*Utterance restrictions* section); Utterances must include the Topic-Comment pattern, in which the Topic contains a noun and is the first IU of the Utterance (*Search for Information Pattern* section).

You can see in this query that all five logical levels described previously are involved. The results are shown in Figure 3 in appendix. Query results are displayed in the CHAT format: the interface shows the list of Utterances matching the query parameters. Audio is directly accessible, through the exploitation of the alignment data and the three buttons located on the right of each entry correspond to the functions available for each TS: online audio playing, audio file download (in WAV format), and opening of the acoustic stream with WinPitch for deep analysis. Finally, it's possible to download all the search results in a format compatible with spreadsheet applications (CSV file), by clicking the icon in the upper right of the page.

The DB-IPIC web resource is available at the project's homepage [2] and freely usable. Though it's possible to query the corpora using the XQuery language by following the public XML Schema, this approach is not recommended due to the complexity of the XML model. DB-IPIC is already designed to support data retrieval at different levels, from general metadata to words in the transcription.

4.   Conclusions

In closing, we want to remark that the annotation is based on prosodic features that are perceptually relevant. The inter-annotator agreement of such an annotation has been proved by a statistic analysis done for C-ORAL-ROM, which points out an agreement of more than 95% in the distinction of breaks (Moneglia et al., 2005). The high reliability of these data is an important quality of the corpus and, in general, of the whole annotation procedure that is founded on a universally agreed feature of speech. Moreover, this validates the choice to consider TSs and TUs as the structural elements of our data model.

On the other hand, we don't have statistics about the accuracy of the informational tagging, because the full revision of the corpus has not yet been done. A corpus validation session is necessary for informational data, requiring an inter-rater agreement approach, and can lead to data alterations in the database. On this point we want to underline that data inside DB-IPIC are easy to modify: this is an important feature that we obtained from the creation of a structured data model and the usage of an XML database.

We also note that information about parts-of-speech and lemmas is induced automatically with software that uses a probabilistic model: with this approach errors are frequent, especially in a spoken language context. A manual revision of PoS-tagging would be desirable and would allow us to produce a gold standard for the informational annotation of spoken language resources.

---

[1] CQL (Corpus Query Language) is a language developed from the University of Stuttgart to make lexical queries on corpora.

[2] http://lablita.dit.unifi.it/ipic

5.   References

Austin, J.L. (1962). *How to do things with words*. Oxford: Oxford University Press.

Cresti, E. (2000). *Corpus di italiano parlato*. Firenze: Accademia della Crusca.

Cresti, E., Moneglia, M. (Eds.) (2005). *C-ORAL-ROM. Integrated reference corpora for spoken romance languages*. Amsterdam/Philadelphia: John Benjamins.

Cresti, E., Moneglia, M. (2010). Informational patterning theory and the corpus-based description of spoken language. The compositionality issue in the topic-comment pattern. In M. Moneglia & A. Panunzi (Eds.), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Firenze: FUP.

Cresti, E., Panunzi, A. and Scarano, A. (2005). The Italian corpus. In E. Cresti & M. Moneglia (Eds.), *C-ORAL-ROM: integrated reference corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins, pp. 71--110.

Gregori, L. (2011). Database XML per annotazione multilivello del corpus di parlato spontaneo LABLITA. MA thesis, Università degli Studi di Firenze.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Martin, Ph. (2005). WinPitch Corpus: a text-to-speech analysis and alignment tool. In E. Cresti, M. Moneglia (Eds.), *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins, pp. 40--51.

Mittmann, M., Panunzi, A., Cresti, E., Moneglia, M., Mello, H. and Raso, T. (this volume). Information patterning strategies in spontaneous speech: a cross-linguistic study.

Mittmann, M., Raso, T. (2012). The C-ORAL-BRASIL informationally tagged mini-corpus. In H. Mello, A. Panunzi & T. Raso (Eds.), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*, Firenze: FUP, pp. 151--183.

Moneglia, M. (2011). Spoken corpora and pragmatics. In H. Mello & S. Gries (Eds.), *Brasilian Journal of applied linguistics/Revista brasileira de linguìstica aplicada*,11(2), pp. 479--519.

Moneglia, M., Cresti, E. (1997). Intonazione e criteri di trascrizione del parlato. In U. Bortolini & E. Pizzuto (Eds.), *Il progetto CHILDES Italia, vol. II*. Pisa: Il Cerro, pp. 59--90.

Moneglia, M., Fabbri, M., Quazza, S., Panizza, A., Danieli, M., Garrido, J.M. and Swerts, M. (2005). Evaluation of consensus on the annotation of terminal and non-terminal prosodic breaks in the C-ORAL-ROM Corpus. In E. Cresti & M. Moneglia (Eds.), *C-ORAL-ROM. Integrated reference corpora for spoken romance languages*. Amsterdam/Philadelphia: John Benjamins, pp. 257--276.

Panunzi, A., Gregori, L. (2012). DB-IPIC. An XML database for the representation of information structure in spoken language. In H. Mello, A. Panunzi & T. Raso (Eds.), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*, Firenze: FUP, pp. 133—150.

Peterson, D., Gao, S., Malhotra, A., Sperberg-McQueen, C.M. and Thompson, H.S. (2012). W3C XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes. Available at: <http://www.w3.org/TR/2012/REC-xmlschema11-2-20120405/>.

Raso, T., Mello, H. (2012). *C-ORAL-BRASIL I. Corpus de referência para a fala espontânea informal do português do Brasil*. Belo Horizonte: UFMG.

Scarano, A. (2009). The prosodic annotation of C-ORAL-ROM and the structure of information in spoken language. In L. Mereu (Ed.), *Information structures and its interfaces*. Berlin/New York: Mouton de Gruyter, pp. 51--74.

Treetagger. Available at: <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>.

WinPitch. Available at: <http://www.winpitch.com/>.

eXist-db. Available at: <http://exist.sourceforge.net/>.

6.   Appendix
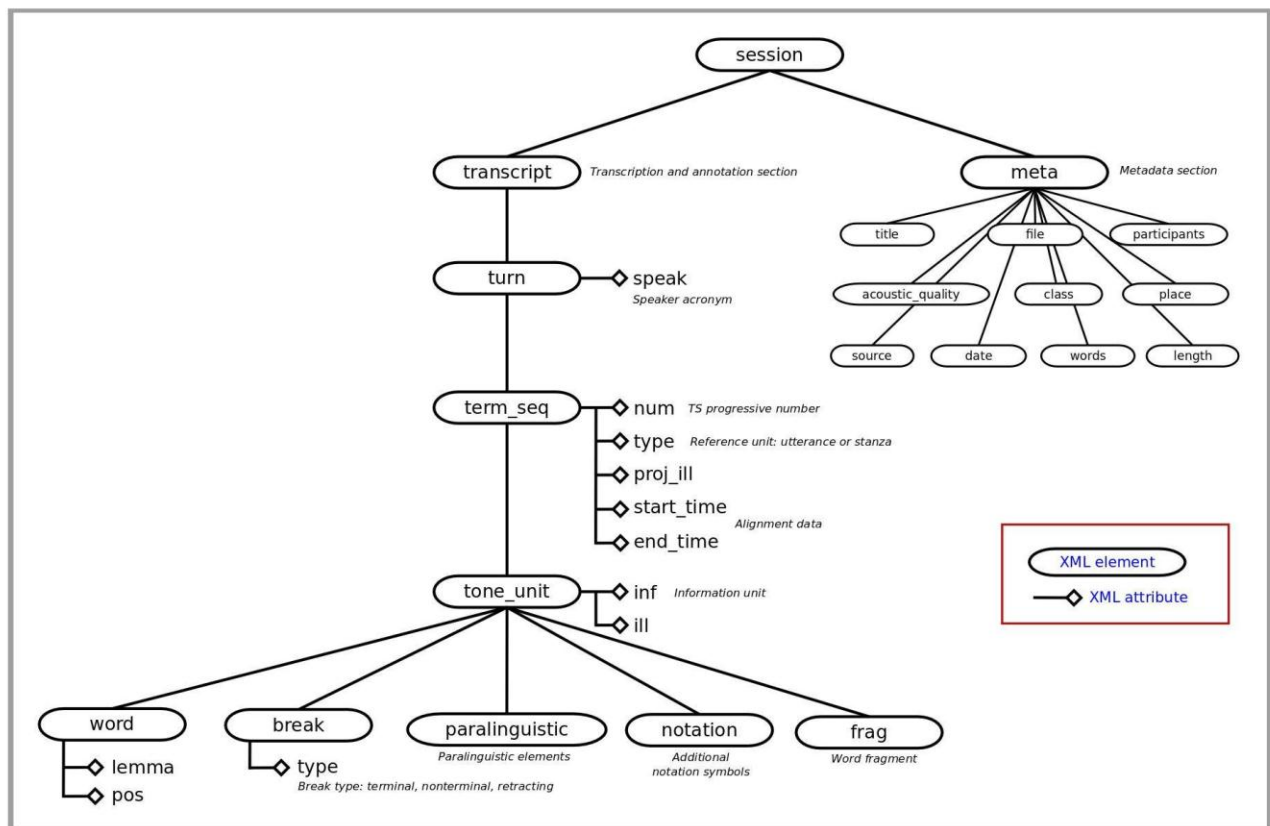
Figure 1: The XML data model

Figure 3: DB-IPIC results page

Class	Name	Tag	Definition
Comment Units	Comment	COM	Comment IU accomplishes the illocutionary force of the Utterance, and it is therefore necessary and sufficient to perform an utterance
	Multiple-Comment	CMM	A complex IU comprised of two or more Comments, forming an *illocutionary pattern*
	Bound Comment	COB	A sequence of Bound Comments with weak illocutionary force, produced by progressive adjunctions following the flow of thought, out of any model of informational patterning
Textual Units	Topic	TOP	It identifies the domain of application for the illocutionary act expressed by the comment, providing the Speech Act with a cognitive reference and allowing the Utterance displacement from the actual context
	Topic List	TPL	A chain of Topics forming a pattern of Topics
	Appendix of Comment	APC	It integrates the text of the Comment and concludes the Utterance
	Appendix of Topic	APT	It gives a delayed integration of the information given in the Topic adding specification for the addressee
	Parenthesis	PAR	It adds information to the utterance with a meta-linguistic value having "backward" or "forward" scope; always bears a modal value.
	Locutive Introducer	INT	It is used for introducing a sequence of IUs that have a strong and unitary "point of view", as in reported speech and reported thought
Dialogic Units	Incipit	INP	It opens the communicative channel bearing a contrastive value starting a dialogic turn or an utterance
	Conative	CNT	It pushes the listener to take part in the dialogue in an adequate way, or stops his non collaborative behavior
	Phatic	PHA	It is dedicated to controlling the communicative channel, ensuring its maintenance; it stimulates the listener to the social cohesion needed by the dialogical exchange and/or ensures the reception of the utterance
	Allocutive	ALL	It specifies to whom the message is directed keeping his attention. Simultaneously it plays a cohesive and empathic function, bringing the interlocutor to share the point of view of the utterance
	Expressive	EXP	It works as an emotional support. It stresses the sharing of a common social affiliation with the interlocutor, searching for social cohesion.
	Discourse Connector	DCT	It zips different parts of the discourse (e.g. utterances within a turn), signaling to the addressee that the discourse is going on and that the entity which follows holds a relation with the previous ones.
Non-informative Units	Scanning	SCA	It occurs when the corresponding prosodic unit has no informational function and its locutive content is part of a larger IU (by default occurring on its right)
	Interrupted	EMP	Interrupted units which cannot be evaluated
	Time Taking	TMT	Time taking units for programming needs
	Unclassified	UNC	Unclassified Units

Table 1: Information units

NURC digital: uma proposta de preservação dos dados do projeto NURC

Miguel OLIVEIRA Jr.

Universidade Federal de Alagoas

Maceió, Alagoas, Brasil

miguel@fale.ufal.br

Resumo

O presente artigo descreve um projeto de pesquisa tem por objetivo central propor um modelo de informatização de um dos corpora mais influentes na pesquisa linguística do Brasil: o corpus do Projeto NURC. Partindo de recomendações de órgão internacionais especializados em práticas de codificação e transmissão de dados digitais, um corpus de dados representativos do Projeto NURC será organizado e apresentado aos coordenadores de todas as capitais brasileiras que sediam o Projeto NURC como possível modelo a ser adotado para a informatização, preservação e disponibilização de seu acervo, que atualmente se encontra em sério risco de deterioração devido à ação do tempo.

Palavras-chave: NURC; preservação; dados orais.

1. Introdução

O Projeto da Norma Urbana Linguística Culta teve seu início em 1969, tendo sido proposto como uma extensão do *Proyecto de Estudio Coordinado de la Norma Lingüística Culta de las Principales Cidades de Iberoamérica y de la Península Ibérica*, de que participavam países de língua espanhola da América Latina. A proposta inicial do Projeto era documentar e estudar a norma falada culta de cinco capitais brasileiras: Recife, Salvador, Rio de Janeiro, São Paulo e Porto Alegre. A seleção dessas capitais foi feita a partir dos seguintes critérios: ter a cidade pelo menos um milhão de habitantes e estratificação social suficiente para atender às exigências do projeto.

Os dados que fazem parte do acervo do Projeto NURC têm sido utilizados para a elaboração de um grande número de trabalhos acadêmicos, incluindo dissertações de mestrado, teses de doutorado, artigos publicados em periódicos nacionais e internacionais, e trabalhos apresentados em encontros científicos ao redor do mundo. A Gramática do Português Falado (Castilho, 1990; Castilho, 1993; Castilho & Basílio, 1996; Ilari, 1992; Kato, 1996; Koch, 1996; Neves, 1999; Abaurre & Rodrigues, 2002), grande e ambicioso projeto nacional que envolveu entre 1988 e 2002 cerca de cinquenta pesquisadores na área da linguística, resultou em uma série de volumes, todos contendo análises de materiais extraídos dos dados do Projeto NURC. É, pois, incontestável a importância do material pertencente ao arquivo do Projeto NURC.

Lamentavelmente, os registros magnéticos dos inquéritos do Projeto NURC, feitos em fita de rolo, estão em sério risco de deterioração. Na verdade, muitos desses registros já se encontram irremediavelmente destruídos pela ação do tempo. Assim, por exemplo, as chuvas de abril/maio de 2011 inundaram a sala do Projeto NURC/Recife, e ainda não se sabe a dimensão dos estragos que foram provocados por esse incidente, no que diz respeito ao material ali arquivado. É imprescindível, portanto, que este valioso material seja resgatado o quanto antes, mediante a transposição de seus dados analógicos para formatos digitais que garantam a sua preservação e utilização no futuro.

O objetivo central do projeto de pesquisa aqui descrito é desenvolver uma metodologia e práticas específicas para gestão de registros sonoros resultantes das pesquisas do NURC, bem como de estratégias de migração para formatos digitais, curadoria e preservação digital do acervo. Esta pesquisa deve indicar meios que poderão ser utilizados pelo Projeto NURC em todas as capitais em que está sediado para a preservação e a disponibilização mais efetiva de seus corpora. Para isso, a iniciativa proposta pretende digitalizar e anotar um corpus representativo de inquéritos pertencentes ao acervo do NURC Recife, mediante técnicas de digitalização e de arquivamento recomendadas por órgãos internacionais especializados em arquivamento de dados digitais.

2. Justificativa

Entende-se por corpus, nos estudos linguísticos, uma "coletânea de porções de linguagem que são selecionadas e organizadas de acordo com critérios linguísticos explícitos, a fim de serem usadas como uma amostra da linguagem" (Percy *et al*., 1996: 4). O corpus do Projeto NURC é uma coletânea de dados de fala de informantes com formação universitária completa (chamados cultos), organizada para servir de estudo da modalidade oral da língua portuguesa culta falada no Brasil. O material do Projeto NURC foi – e tem sido – largamente utilizado para o estudo de diversas características da oralidade, que vão desde aspectos discursivos, tais como a análise de narrativas inseridas na conversação (Oliveira Jr., 1999) e de questões discursivas e ideológicas presentes nas diversas modalidades de gravações feitas pelo Projeto (Cunha, 2003), até aspectos mais formais, tais como a análise de elementos argumentativos e pragmáticos, da intertextualidade e da organização interacional e sintática presentes no texto oral (Sá, 2004).

A maior parte dos estudos desenvolvidos a partir dos dados do Projeto NURC deriva de uma série de publicações feitas com transcrições de material selecionado pelos grupos de pesquisadores atuantes em cada uma das capitais em que o Projeto era desenvolvido. Essas coletâneas de transcrições publicadas a partir da década de 80 ficaram conhecidas por Materiais Para o Seu

Estudo: Castilho e Preti (1986, 1987), Preti e Urbano (1990), Callou (1992), Callou e Lopes (1993, 1994), Motta e Rollemberg (1994), Hilgert (1997), Sá *et al.* (1996, 2005). Os estudos feitos a partir dessas publicações desconsideravam, em sua grande maioria, o registro de áudio, baseando-se exclusivamente nas transcrições aí presentes. Essa não era, evidentemente, uma opção dos estudiosos. Tratava-se mesmo de uma questão de dificuldade de acesso aos dados gravados. Todas as gravações feitas pelo Projeto NURC utilizaram, como meio, fitas magnéticas de rolo, que, se por um lado garantia a qualidade das gravações, por outro dificultava o acesso às mesmas, uma vez que reprodutores de fita de rolo eram equipamentos caros e pouco comuns.

Uma outra dificuldade que a utilização do material do Projeto NURC apresentava aos estudiosos era – e continua sendo, em grande parte – a não disponibilização dos dados transcritos em formato digital. Assim, o processo de análise a partir dos textos publicados em formato impresso era – e continua sendo – necessariamente demorado e eventualmente falho, uma vez que não se podia contar com buscas automatizadas de fenômenos linguísticos particulares.

Com o advento da tecnologia, tem-se cada vez mais incentivado a disponibilização de dados linguísticos em formato digital, que possam ser acessados por humanos e máquinas. A simples digitação de dados é apenas um primeiro passo para a criação de um corpus digital. Há, na verdade, uma série de medidas recomendadas por especialistas na área da construção de corpora eletrônicos que precisam ser consideradas, se o objetivo for construir um corpus que seja também legível por máquinas (Sardinha, 2000). A vantagem de se construir um corpus com essa característica é mesmo a de facilitar as análises linguísticas feitas a partir dele, automatizando certos aspectos da análise. À análise linguística que toma por base corpora informatizados para deles fazer considerações probabilísticas tem-se comumente referido como linguística de corpus (Sardinha, 2000).

Já houve tentativas isoladas de informatização de dados do Projeto NURC (Castilho *et al.*, 1995). Assim, por exemplo, muitos dos dados do Projeto NURC do Rio de Janeiro foram digitalizados e disponibilizados na internet (http://www.letras.ufrj.br/nurc-rj/home.htm). A despeito de ser essa uma empreitada louvável, a metodologia empregada para a disponibilização desses dados on line não levou em consideração uma série de recomendações metodológicas bastante importantes no processo de elaboração de bancos de dados digitais. Desse modo, apesar de agora pesquisadores interessados em aspectos da oralidade poderem ter acesso aos arquivos de áudio a que se referem algumas transcrições, e poderem fazer buscas bastante rudimentares no corpus disponibilizado pelo NURC-RJ, não poderão, entre outras coisas, proceder, por exemplo, a uma análise automatizada de frequência de ocorrência de traços linguísticos de várias ordens (lexicais, sintáticos, semânticos, discursivos, etc.), ou a uma possível análise acústica, devido à não-observação das já referidas recomendações metodológicas.

A área da linguística que tem se preocupado em estabelecer bases teóricas para a construção de corpora linguísticos digitais é chamada linguística documentativa (Himmelmann, 2006). A linguística documentativa emergiu como uma resposta para uma necessidade urgente de se fazer registros duradouros de línguas em risco de extinção, utilizando-se o aparato tecnológico disponível na atualidade. Entretanto, a sua área de atuação hoje em dia vai além da documentação de línguas em risco de extinção. A linguística documentativa se ocupa em indicar métodos e ferramentas para a elaboração de registros de qualquer língua natural, ou de variedades de uma língua, que sejam representativos, duradouros e que permitam múltiplos usos. Para isso, é fundamental que um corpus seja acompanhado não apenas de uma transcrição, mas de metadados contendo informações relevantes acerca do contexto e do uso do material, e de anotações multiníveis que garantam o seu amplo uso.

Assim, os procedimentos estabelecidos para a construção de um corpus linguístico digital permitem a sua utilização não apenas em diversas áreas da linguística, tais como a fonologia, a fonética, a morfologia, a sintaxe, a semântica, a análise do texto e do discurso, a sociolinguística, a tipologia, etc., mas também em áreas afins, como a história (história oral), a antropologia (aspectos culturais, questões acerca da interação), a sociologia, a poética (aspectos musicais e métricos da literatura oral), e a educação (estudo de gêneros da oralidade em sala de aula), por exemplo.

Além disso, a observância desses procedimentos metodológicos garantirá a preservação do valioso material do Projeto NURC, de forma que o mesmo possa ser utilizado mais eficazmente não apenas no presente, mas por futuras gerações de pesquisadores.

3. Objetivos

O principal objetivo do presente projeto de pesquisa é propor uma metodologia de organização de um corpus representativo do acervo do Projeto NURC, em formato digital, que servirá como possível modelo a ser adotado para a informatização de todo o material pertencente ao arquivo do Projeto NURC. Para isso, serão levados em conta procedimentos internacionais estabelecidos para a construção de corpus linguístico digital. Este projeto representa, assim, um importante passo no processo de preservação do valioso acervo do Projeto NURC, que atualmente se encontra em sério risco de deterioração ocasionada pela ação do tempo. Além disso, os resultados provenientes da execução do projeto aqui proposto beneficiará diretamente a comunidade científica, que passará a ter disponíveis para consulta otimizada dados – anteriormente de difícil acesso – em formato digital de alta qualidade, devidamente catalogados, etiquetados e transcritos.

Como objetivos específicos, o projeto aqui proposto pretende:

i. contribuir para a formação de pesquisadores nas áreas da documentação linguística, da linguística

de corpus e da análise da oralidade;

ii. digitalizar todo o acervo do Projeto NURC/Recife, originalmente gravado em formato analógico, respeitando os padrões recomendados pelos órgãos internacionais de codificação e transmissão de dados digitais;

iii. catalogar e armazenar em formato digital todas as informações referentes ao material de áudio digitalizado;

iv. informatizar os dados de transcrição referentes a parte do material de áudio digitalizado (o corpus compartilhado do Projeto NURC/Recife, tornando-os alinhados, o que propiciará uma utilização mais proveitosa dos mesmos;

v. propor um sistema de anotação/etiquetagem multi-nível para os dados do Projeto NURC;

vi. anotar / etiquetar um corpus representativo dos dados do Projeto NURC, com informações multi-níveis;

vii. arquivar os dados informatizados em bancos de dados internacionais, assegurando assim a sua preservação;

viii. elaborar um documento com proposta de digitalização, preservação e anotação dos dados do Projeto NURC, elaborada a partir de discussão com todos os coordenadores do Projeto NURC, levando-se em conta as recomendações de órgão internacionais especializados em arquivamento de dados digitais;

ix. republicar os Materiais para o Seu Estudo em formato digital, contendo todos os dados do corpus compartilhado (transcrição, anotação e áudio);

x. editar um volume *Estudos*, composto de artigos feitos a partir do corpus compartilhado do NURC/Recife;

xi. disponibilizar o corpus compartilhado digitalizado e anotado para a elaboração de trabalhos os mais variados (artigos, capítulos de livro, dissertações e teses), dentro do âmbito do projeto.

4. Metodologia

O presente projeto de pesquisa tem por objetivo informatizar um corpus representativo do material do Projeto NURC, com o propósito de sugerir uma metodologia padrão, baseada em recomendações feitas por órgãos internacionais de codificação e transmissão de dados digitais, para ser adotada no Projeto NURC como um todo, preservando, assim, o seu precioso acervo, e permitindo que ele seja utilizado de maneira mais eficiente no futuro. Todo o acervo do Projeto NURC/Recife será digitalizado. Parte deste acervo será também anotado. O material a ser anotado corresponderá ao corpus compartilhado do Projeto NURC Recife. Justifica-se a escolha desse material pelo fato de ser o proponente deste projeto pesquisador do Projeto NURC Recife desde 1990, tendo, portanto, acesso ao acervo

daquela capital. Além disso, cumpre notar que a sala do Projeto NURC Recife foi recentemente inundada, devido às fortes chuvas de abril/maio de 2011 naquela região. Ainda não se tem ideia da proporção dos estragos causados por esse incidente no que diz respeito ao material ali arquivado. Entretanto, o incidente por si só já justifica a necessidade – e mesmo a urgência – de se estudar uma estratégia de arquivamento mais eficiente para o acervo do Projeto NURC em geral, e do acervo do Projeto NURC/Recife em particular.

Os dados de áudio do corpus compartilhado do Projeto NURC/Recife – material selecionado para compor o corpus representativo deste projeto – serão digitalizados observando-se as recomendações propostas pelo *Open Archival Information System* (OAIS), que é um modelo de referência, com padrão ISO (14721:2003), adotado pelos bancos digitais de dados linguísticos mais recentes, e pelo Comitê Técnico da IASA para objetos digitais (Bradley, 2009; Von Arb & Lars, 2005). As informações referentes aos arquivos de áudio e às transcrições (metadados) serão registradas seguindo o padrão *Dublin Core* e o protocolo da *Open Archives Initiative Protocol for Metadata Harvesting*, também adotados por bancos de dados internacionais. As transcrições dos dados serão registradas no aplicativo ELAN, que possibilita o seu alinhamento com os arquivos de áudio a que se referem, além de permitir que áudio, transcrições e metadados sejam pesquisáveis local e virtualmente. Durante toda a fase de digitalização e tratamento do material do Projeto NURC, backups regulares serão realizados em lugares diferentes do local onde os dados primários estarão custodiados, garantindo assim a preservação dos mesmos.

Os inquéritos do Projeto NURC foram gravados em condições variadas. Em geral, as gravações eram realizadas com microfones dinâmicos omnidirecionais, apoiados em uma mesa. Todos os inquéritos foram registrados em fita de rolo. A depender do tipo de inquérito, as gravações eram realizadas em salas específicas, em salas de aula, em auditórios e, em algumas casos, nas casas dos informantes. Portanto, a qualidade acústica das gravações do Projeto NURC é bastante heterogênea, não sendo possível descrever um perfil das gravações como um todo em termos de relação sinal-ruído. Diante deste cenário, não é viável que se aponte como objetivo do presente projeto disponibilizar arquivos sonoros com qualidade suficiente para análises acústicas sofisticadas, embora, em alguns casos, a depender das condições da gravação, isso será perfeitamente possível.

Como indicado acima, todos os cuidados metodológicos, recomendados por órgãos internacionais especializados em arquivamento de dados digitais serão considerados no processo de digitalização dos arquivos de áudio, procurando-se, na medida do possível, preservar as características originais do sinal analógico. Quando necessário, técnicas automatizadas de redução de ruído (como, por exemplo, ruídos de *pitch* fixo – *hum* e *whistles* –, associados geralmente a gravações analógicas em fitas magnéticas) serão empregadas. Entre as técnicas mais

comuns de redução de ruídos associados a fitas magnéticas estão a utilização de filtros de frequências.

Experiência prévia de digitalização de arquivos do Projeto NURC, como, por exemplo, a realizada pelo Projeto NURC do Rio de Janeiro, com apoio financeiro do CNPq, demonstra que a proposta aqui apresentada é exequível.

A anotação / etiquetagem do corpus compartilhado do Projeto NURC/Recife será feita a partir da utilização esquemas previamente utilizados com sucesso para o português brasileiro, como, por exemplo, o tagset proposto pelo Núcleo Interinstitucional de Linguística Computacional (NILC), o NILC Tagset (Aires *et al.*, 2000), e o etiquetador morfossintático MXPOST (Ratnaparkhi, 1996).

5.   Contribuições da Proposta

O corpus informatizado será arquivado localmente, nos servidores da Universidade Federal de Pernambuco e da Universidade Federal de Alagoas, em um site dedicado ao Projeto NURC/Recife, para livre consulta pela comunidade científica, e depositado em bancos internacionais, tais como o do IMDI (http://www.lat-mpi.eu/archive/), com o intuito de garantir a sua preservação.

Uma vez constituído e devidamente arquivado, o corpus digitalizado será apresentado aos atuais coordenadores do Projeto NURC, em todas as capitais, como modelo a ser discutido e, eventualmente, adotado, para a informatização e preservação de todo o material coletado por este importante projeto na área da linguística.

6.   Referências

Abaurre M.B.M., Rodrigues, Â.C.S. (eDs.) (2002). *Gramática do português falado*. Campinas: Editora da Unicamp.

Aires, R.V.X., Aluísio, S.M., Kuhn, D.C.S., Andreeta, M.L.B. and Oliveira Jr., O.N. (2000). *Combining Multiple Classifiers to Improve Part of Speech Tagging: A Case Study for Brazilian Portuguese*. (SBIA'2000) Atibaia, SP, November, pp. 20--22.

Callou, D.M.I. (Ed.) (1992) *A Linguagem Falada Culta na Cidade do Rio de Janeiro. Materiais para seu estudo*. Rio de Janeiro: UFRJ/FJB, vol. I, Elocuções Formais.

Callou, D.M.I., Lopes, C.R. (Eds.) (1993). *A Linguagem Falada Culta na Cidade do Rio de Janeiro. Materiais para seu estudo*. Rio de Janeiro: UFRJ/CAPES, vol. II, Diálogo entre Informante e Documentador.

Callou, D.M.I., Lopes, C.R. (Eds.) (1994). *A Linguagem Falada Culta na Cidade do Rio de Janeiro.Materiais para seu estudo*. Rio de Janeiro: UFRJ/CAPES, vol. III, Diálogos entre dois informantes.

Castilho, A. (Ed.) (1990). *Gramática do português falado*. Campinas: Editora da Unicamp; São Paulo: Fapesp.

Castilho, A. (Ed.) (1993). *Gramática do português falado*. Campinas: Editora da Unicamp; São Paulo: Fapesp.

Castilho, A. (2007). Fundamentos teóricos da Gramática do português culto falado no Brasil. *Alfa*, São Paulo, 51 (1): pp. 99--135.

Castilho, A., Basílio, M. (Eds.) (1996). *Gramática do português falado*. Campinas: Editora da Unicamp; São Paulo: Fapesp.

Castilho, A. *et al*. (1995). Informatização de acervos da Língua Portuguesa. *Boletim da Associação Brasileira de Lingüística* 17: pp. 143--154.

Castilho, A., Preti, D. (Eds.) (1986). *A Linguagem Falada Culta na Cidade de São Paulo. Materiais para seu estudo*. São Paulo: TAQ/Fapesp, vol. I, Elocuções Formais.

Castilho, A., Preti, D. (Eds.) (1987). *A Linguagem Falada Culta na Cidade de São Paulo. Materais para seu estudo*. São Paulo: TAQ/Fapesp, vol. II, Diálogos entre dois informantes.

Cunha, D.A.C. (2003). A produção de sentido na fala e na escrita. *Revista do GELNE*. UFC. V.3. pp. 27--32.

Hilgert, J.G. (Ed .) (1997). *A Linguagem Falada Culta na Cidade de Porto Alegre*. Passo Fundo: Ediupf / Porto Alegre: Ed. Universidade/Ufrgs, vol. I, Diálogos entre informante e documentador.

Himmelmann, N.P. (2006). Language documentation: what is it and what is it good for? In *Essentials of Language Documentation*, eds. Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel. Berlin: Mouton de Gruyter, pp. 1--30.

Bradley, K. (2009). IASA Technical Committee, Guidelines on the Production and Preservation of Digital Audio Objects. International Association of Sound and Audiovisual Archives.

Ilari, R. (Ed.) (1992). *Gramática do português falado: níveis de análise lingüística*. Campinas: Editora da Unicamp.

Janssen, M., T. Freitas (2008). Spock – a spoken corpus client. In Proceedings of LREC VI, Marraqueche.

Kato, M. (Ed.) (1996). *Gramática do português falado: convergências*. Campinas: Editora da Unicamp; São Paulo: Fapesp.

Koch, I.G.V. (Ed.) (1996). *Gramática do português falado*. Campinas: Editora da Unicamp; São Paulo: Fapesp.

Motta, J., Rollemberg, V. (Eds.) (1994). *A Linguagem Falada Culta na Cidade de Salvador. Materiais para seu estudo*. Salvador: Instituto de Letras da UFBa, vol. I, Diálogos entre Informante e Documentador.

Neves, M.H.deM. (Ed.) (1999). *Gramática do português falado*. São Paulo: Humanitas; Campinas: Editora da Unicamp.

Oliveira Jr, M. (1999). The Function of Self-Aggrandizement in Storytelling. *Narrative Inquiry*, 9 (1): pp. 25--47.

Percy, C.E. *et al*. (Eds.) (1996). Synchronic Corpus Linguistics. *Papers from the sixteenth International Conference on English Language and Research on Computerized Corpora* (ICAME 16). Amsterdam/Atlanta, GA: Rodipi.

Preti, D., Urbano, H. (Eds.) (1990). *A Linguagem Falada Culta na Cidade de São Paulo. Materiais para seu*

estudo. São Paulo: TAQ/Fapesp, vol. III, Diálogos entre o Informante e o Documentador.

Preti, D., Urbano, H. (Eds.) (1990). *A Linguagem Falada Culta na Cidade de São Paulo: Estudos*. São Paulo: T. A. Queiroz.

Sá, M.P.M. (2004). Estrutura e natureza da narrativa na conversação. Boletim Informativo, n.32. Maceió: UFAL. Apresentado no *XIX Encontro Nacional da Anpoll*, na Universidade Federal de Alagoas.

Sá, M.P.M. *et al*. (Eds.) (1996). *A Linguagem Falada Culta na Cidade do Recife*. Recife: Universidade Federal de Pernambuco, Programa de Pós-Graduação em Letras e Lingüística, vol. I: Diálogos entre informante e documentador.

Sá, M.P.M. *et al*. (Eds.) (2005). *A linguagem falada culta na cidade do Recife: elocuções formais*. Recife: Universidade Federal de Pernambuco/Programa de Pós-Graduação em Letras e Lingüística.

Sardinha, T.B. (2000). Linguística de Corpus: Histórico e Problemática. In *D.E.L.T.A.*, Vol. 16, No. 2, pp. Pp. 323--367.

Von Arb, J., Gaustad, L. Guidelines on the Production and Preservation of Digital Audio Objects – optimalizing quality access through digital preservation practice" in World Library and Information Congress: 71th IFLA General Conference and Council "Libraries - A voyage of discovery" June 17, 2005, Oslo, Norway. Available at: <http://www.ifla.org/IV/ifla71/Programme.htm>.

Analyzing (-r) with R

Livia OUSHIRO

University of Sao Paulo

Av. Prof. Luciano Gualberto, 403 - Sala 16 - Cidade Universitária - 05508-010 - São Paulo - SP

livia.oushiro@usp.br

Abstract

This paper presents a script written for the free software R (Gries, 2009; Hornik, 2011), which has been employed in the analysis of variable (-r) in Paulistano Portuguese (Oushiro, 2012a) to automatically (i) identify and mark tokens of the variable; (ii) extract tokens into a spreadsheet file precoded with social factors; and (iii) extract a balanced subsample of a specific number of tokens per speaker (Wolfram, 1993). It describes the tasks to be performed by R, and discusses the script's advantages and existing shortcomings. The script seems to work better with phonetic and morphological variables, and naturally does not exempt the researcher from a thorough qualitative analysis of their corpus (for example, for identifying possible exclusions). On the other hand, the script can be adapted to studying a number of variables, its different tasks can be performed separately, it allows the researcher to handle data in a more consistent manner and, by reducing the time spent in preparing the token file, it allows more time to perform statistical analyses and interpret results.

Keywords: variable (-r); software R; data handling; Paulistano Portuguese; variationist sociolinguistics.

1. Introduction

Quantitative analyses of sociolinguistic variation (Guy, 1993; Bayley, 2002) often involve handling hundreds or thousands of tokens of a variable, especially in studies of phonetic variation. Analyses in softwares such as GoldVarb X and RBrul should be preceded by the identification, isolation, coding, and extraction of variants within a variable context. These tasks are mechanical, time-consuming, tiresome, and subject to errors due to lapses of attention on the part of the researcher. In fact, there have recently been a number of initiatives for automatizing certain tasks of sociolinguistic quantitative analyses (see e.g. Cieri & Strassel, 2010; Rosenfelder & Labov, 2010).

This paper presents a script written for the free software R [1] (Gries, 2009; Hornik, 2011), which was employed in the analysis of variable coda (-r) in Paulistano Portuguese (Oushiro, 2012a) in a corpus of 102 hour-long sociolinguistic interviews (about 1.5 million words), which yielded 63,994 tokens of the variable.

The software R allows researchers to perform a number of tasks, including corpus linguistics data handling, statistical analyses, plotting graphs (Gries, 2009). The aforementioned tasks for handling such a number of tokens were greatly minimized by the use of the software R, which was employed to automatically: (i) identify tokens of variants in the speech of informants; (ii) extract tokens with preceding and subsequent context into a precoded spreadsheet file; and (iii) extract a balanced subsample of a specific number of tokens per speaker (Wolfram, 1993).

The scripts are largely based on Gries (2009) and the internet discussion list "CorpLing with R" (https://groups.google.com/group/corpling-with-r). In the scripts below, the relevant functions are in bold. Although it is possible to simply copy the scripts and substitute the relevant variables marked below as "X," the reader is also advised to consult R manuals, such as Gries (2009), since most functions are not described in details here. Section 2 presents the full scripts and discusses some of their main functions; Section 3 discusses their applicability to other sociolinguistic variables and some of its present shortcomings.

2. Full Scripts

2.1 Identifying tokens

This script identifies tokens of a variable in transcript files and marks them with "<>". Take the excerpt below as an example:

(1)
S1: não foi assim que eu escolhi a Mooca acho que a Mooca me escolheu [risos] eu [hes.] não foi assim pensado ah eu quero **morar** naquele bairro **porque** eu nem/ a minha **irmã** mora aqui há muitos anos mas eu vinha aqui só a passeio né? mas [hes.] é depois que você muda para cá aí você não **quer** mais **sair** não **quer** mais **mudar** {sabe}?
D1: {ah que legal}
S1: é bem [hes.] é bem gostoso aqui
D1: então assim [hes.] a senhora diz que morou a maioria/ a *maior parte* do tempo em São Mateus
S1: é próximo a São Mateus

The task R was to perform was finding the tokens of coda (-r) (e.g. in the words *morar*, *porque*, *irmã* etc.) in the speech of informants (S1), marked here in bold, but not the tokens in the speech of the interviewer (D1) (e.g. *maior*, *parte*), marked in italics. The desired output is shown in (2):

(2)
S1: não foi assim que eu escolhi a Mooca acho que a Mooca me escolheu [risos] eu [hes.] não foi assim pensado ah eu quero morar <> naquele bairro

---

[1] Available at: <www.r-project.org>.

porque <> eu nem/ a minha irmã <> mora aqui há muitos anos mas eu vinha aqui só a passeio né? mas [hes.] é depois que você muda para cá aí você não quer <> mais sair <> não quer <> mais mudar <> {sabe}?
D1: {ah que legal}
S1: é bem [hes.] é bem gostoso aqui
D1: então assim [hes.] a senhora diz que morou a maioria/ a maior parte do tempo em São Mateus
S1: é próximo a São Mateus

To do this, the script requires the user to follow the steps 1 to 5 below and, after that, copying and pasting the whole script into R ("#" below is a comment character, which means that R will not read what follows it as functions). The trickiest part is certainly defining the variable (Step 4) in a way that the software R will correctly identify all instances of relevant variants. Identification of variable (-r), for example, employed the following specification:

(3)
```
\\b.*?[aáâeéêiíoóôuú]r[bcçdfgjklmn
pqstvwxz)\\}/\\.!\\? ].*?\\b
```

The line in (3) instructs R to look for instances of a vowel [aáâeéêiíoóôuú], followed by the grapheme "r" and a consonant or end of a word [bcçdfgjklmnpqstvwxz)\\}/\\.!\\?]. It further instructs R to respect word boundaries (\\b), because the intended output should be, e.g., `porta <>` and not `por<>ta`. Finally, the symbols `.*?` indicate that there can or cannot be other characters before or after the (vowel-r-consonant) sequence within the same word.

In Script 1, R identifies all .txt files (`dir`) in the directory specified in Step 2 as the working directory (`setwd`); creates an empty list (`list`) into which all the transcript files will be loaded (`scan`); marks the relevant tokens of the variable specified in Step 4 (`gsub`) in the speech of the informant S1 with <>; and saves (`cat`) the transcript files with variant markings in a new directory (`setwd`) specified in Step 3.

```
#####################
# SCRIPT 1
# Step 1. Create 2 folders:
# (i) for the original transcript files (copy files there);
# (ii) for the files which will contain marked tokens.

# Step 2. Specify the directory where the original transcript files are by substituting "X" below for their complete path.
# e.g. "C:/Users/Documents/Transcriptions/"
# Transcript files should be in plain text format (.txt).
originalfiles<-paste("X")

# Step 3. Specify the directory where transcripts with token identification are to be stored by substituting "X" below for their complete path.
```

e.g. "C:/Users/Documents/Markedfiles/"
```
markings<-paste("X")

# Step 4. Specify the variable by substituting "X" below for a "syntactic definition" of the variable.
thevariable<-paste("X")

# Step 5. Copy and paste this script in R (from "SCRIPT 1" to "END OF SCRIPT 1").
#####################

setwd(originalfiles)
files<-dir(path=originalfiles,pattern=".txt",all.files=F)

all.corpus<-list()
for (i in files) {
    all.corpus[[i]]<-scan(i, what="char", sep="\n", skip=0)
    }

for (i in 1:length(files)) {
    tokens<-gsub(thevariable, "\\1 <> ", all.corpus[[i]], ignore.case=T)
    naos<-paste("(^[a-rt-z].*|S[2-9].*|S1: 2.*)<>")
    so.S1<-gsub(naos, "\\1", tokens, ignore.case=T)
    so.S1<-gsub(naos, "\\1", so.S1, ignore.case=T)
    so.S1<-gsub(naos, "\\1", so.S1, ignore.case=T)
    so.S1<-gsub(naos, "\\1", so.S1, ignore.case=T)
    so.S1<-gsub(naos, "\\1", so.S1, ignore.case=T)
    so.S1<-gsub(naos, "\\1", so.S1, ignore.case=T)
    so.S1<-gsub(naos, "\\1", so.S1, ignore.case=T)
    so.S1<-gsub(naos, "\\1", so.S1, ignore.case=T)
    so.S1<-gsub(naos, "\\1", so.S1, ignore.case=T)
    so.S1<-gsub(naos, "\\1", so.S1, ignore.case=T)
    markedfiles<-gsub(" +", " ", so.S1)
    setwd(markings)
    cat(markedfiles, file=files[i], sep="\n")
    }

#END OF SCRIPT 1
```

2.2   Extracting tokens

Script 2 finds tokens of variables (previously identified by "<>") and extracts them into a spreadsheet file with 4 words of the preceding context and 4 words of the subsequent context in the interview. Information of the speaker's social profile is extracted from the transcript file header, according to our research group's transcription norms:

(4)
```
#cab
S1: 2009F63MUEO0xM
P: SC, M: MG
LucianaM
Perdizes
D1: Livia Oushiro
Duração total: 01h14min49seg
Início: 00h00min00seg
```

Fim: 01h08min07seg
Comentários:
#

For instance, the second line of this header informs: the year the interview was recorded (2009), the speaker's sex (F), age (63), level of education (M, for "Ensino Médio" or high school), type of school (U for "public"), area of residence (E for "expanded central area"), zone of residence (O for "zona oeste" or West Zone), generation of the family in the city (0, i.e. none of the parents are from São Paulo), family's place of origin (x for "mixed"), and mobility (M for "has lived in different zones of the city"). The other lines respectively inform the speaker's parents' place of origin, the speaker's pseudonym, neighborhood, interviewer, total time of recording, at which point the transcription starts and ends, and additional comments.

Other researchers should naturally be aware of their own transcription norms in their corpora and make the necessary adjustments.

Similarly to Script 1, in Script 2 R first sets the working directory specified in Step 1 (setwd) and identifies all .txt files there (dir) to be loaded into a list (list) by the (scan) function. R then reads the second line of each transcript file as the speaker's social profile (all.tokens[[i]][2]->spkprofile) and breaks this line into characters (substr) to identify the codes for year of recording (characters 5-8 in the second line), sex (character 9), age (characters 10-11) etc. These pieces of information are stored in separate vectors, which will be called later on. After that, R breaks the transcript files into words (gsub("(\\W)","\\1", all.tokens[[i]])) and looks for instances of the variants specified in Step 2 (grep). Finally, R creates a new file specified in Step 3, which outputs each token in a different line. Each line also contains the respective speaker's social characteristics separated by a tab stop (\\t), four words from the preceding context (tokens2[max(0,i-5):max(0,i-2)]) and four words from the subsequent context (tokens2[(i+1):min(i+4,length(tokens2)+1)]). When opening the newly created file in a spreadsheet program such as Excel or Calc, the tab stops can be read as different columns, thus separating each social variable.

```
#####################
#SCRIPT 2

# Step 1. Specify the directory where the marked transcript files are by substituting "X" below for their complete path.
# e.g. "C:/Users/Documents/Markedfiles/"
markings<-paste("X")

# Step 2. Specify what should be extracted to the spreadsheet file.
# if you wish R to extract all tokens, don't change the command line below
```

```
# if you prefer to extract tokens of specific variants, substitute "<" for the code of the variants between <> and separate variants by "|":
# e.g. "<R>|<T>"
token<-paste("<")

# Step 3. Define a name for the coding spreadsheet file by substituting "X" below for the name of the file; don't delete ".txt".
# e.g. "DataR.txt"
codingfile<-paste("X.txt")

# Step 4. Copy and paste this whole script in R (from "SCRIPT 2" to "END OF SCRIPT 2").
# The coding file will be in the directory specified in (1) above, with the name specified in (3).
# To open the file in a spreadsheet program (such as Excel or Calc), right-click on the file and choose "Open with... ".
#####################

setwd(markings)
files2<-dir(markings, pattern=".txt", all.files=F)

all.tokens<-list()
for (i in files2) {
 all.tokens[[i]]<-scan(i, what="char", sep="\n", skip=0,
 encoding="UTF-8")
 }

for (i in 1:length(all.tokens)) {
 all.tokens[[i]][2]->spkprofile
 substr(spkprofile, 5,8)->year
 substr(spkprofile, 9,9)->sex
 substr(spkprofile, 10,11)->age
 substr(spkprofile, 12,12)->leveled
 substr(spkprofile, 13,13)->typeschool
 substr(spkprofile, 13,13)->region
 substr(spkprofile, 14,14)->zone
 substr(spkprofile, 15,15)->generation
 substr(spkprofile, 16,16)->porigin
 substr(spkprofile, 17,17)->mobility
 all.tokens[[i]][3]->origin
 all.tokens[[i]][4]->inf
 all.tokens[[i]][5]->neighborhood
 all.tokens[[i]][6]->intr

 tokens1<-gsub("(\\W)","\\1", all.tokens[[i]])
 tokens2<-unlist(strsplit(tokens1," +"))
 (matches.in.corpus<-grep(token,tokens2, ignore.case=T))

  for(i in matches.in.corpus) {
  cat("\t", substr(tokens2[i], 2,2),
  "\t", year,
  "\t", inf,
  "\t", intr,
  "\t", sex,
```

```
    "\t", age,
    "\t", leveled,
    "\t", typeschool,
    "\t", region,
    "\t", zone,
    "\t", neighborhood,
    "\t", generation,
    "\t", origin,
    "\t", porigin,
    "\t", mobility,
    "\t", tokens2[max(0,i-5):max(0,i-2)],
    "\t", tokens2[max(0,i-1):max(0,i-0)],
    "\t", tokens2[(i+1):min(i+4,length(tokens2)+1)],
    file=codingfile, append=T, "\n")
    }
}
```

#END OF SCRIPT 2

2.3   Sampling

Script 3 employs the R package NCStats to randomly select a number of tokens per speaker from the complete coding file, in order to reduce the time spent on coding linguistic factor groups and to avoid data biasing by possible idiosyncratic speakers (Wolfram, 1993). In the analysis of variable coda (-r), we selected 50 tokens per speaker, which yielded a final coding file containing 5,100 tokens (50 tokens x 102 speakers), an amount which can be much more easily handled than the original 63,994 tokens.

In this script, after setting the working directory specified in Step 2 and reading the coding file (`read.table`), R is asked to show the data structure (`str`). It invokes the package NCStats (`library`), which will be required because of its `srsdf` function. It then identifies all speakers in the sample, which had been specified in column 4 in the coding file, by making a list of all the different values (`unique`); then, for each speaker, R randomly (`srsdf`) selects a subset of tokens (`subset`), given the number specified in Step 4, and creates a new coding file with the name specified in Step 5. The .txt file can be opened in a spreadsheet program by right-clicking on it and selecting "Open with...".

```
####################
#SCRIPT 3
```

# **Step 0**. Requires package NCStats. It can be installed by copying the following line in R when connected to the Internet:
source("http://www.rforge.net/NCStats/InstallNCStats.R")
#It is only required the first time this script is run.

Step 1. Make sure the coding file does not contain empty cells; save the coding file in "tab delimited" format.

Step 2. Specify the name of the complete coding

file by substituting "X" below for the name of complete coding file; don't delete ".txt".
codingfile2<-paste("X.txt")

# **Step 3**. Specify the directory where the complete coding file is, by substituting "X" below for the complete path.
# e.g. "C:/Users/Documents/Markedfiles/"
codingfilefolder<-paste("X")

# **Step 4**. Specify the number of tokens per speaker you wish to extract by substituting X for the number of tokens.
# e.g. 50
numbertokens<-X

# **Step 5**. Specify the name of the new coding file with randomly selected tokens by substituting "X" below for the name of the new coding file; don't delete ".txt".
# e.g. "DataR-50.txt"
sampledcodingfile<-paste("X.txt")

Step 6. Copy and paste this script in R (from "SCRIPT 3" to "END OF SCRIPT 3").

```
############################
setwd(codingfilefolder)
alldata<-read.table(file=codingfile2,    header=T,
sep="\t",
quote="", na.strings="NA", comment.char="")
attach(alldata)
str(alldata)
library(NCStats)
spks<-unique(alldata[,4])
SPKS<-paste(spks[1:length(spks)])

for (i in SPKS) {
    write.table(srsdf(subset(alldata,
    INF==i),numbertokens),
    file=sampledcodingfile, append=T,
    quote=F, sep="\t", row.names=F, col.names=F)
    }
```

#END OF SCRIPT 3

3.   Applicability

The scripts shown above can be adapted to the study of other sociolinguistic variables. This can be done essentially by creating new "syntactic definitions" for new variables, in Step 4 of Script 1. For instance, analyses of diminutives (e.g. *menininha* 'little girl') (Mendes, 2011) and variable nasal /e/ (Oushiro, 2012b) in Brazilian Portuguese have employed the following definitions respectively:

(5)
\\b.*?inh[oa]s?\\b
(6)

```
\\b.*?[eéê][nm][bcçdfgjklpqrstvwxz
)\\}\\.,:;\\!\\? \\].*?\\b
```

The line in (5) specifies that R should look for instances of words ending in "-inho," "-inha," "-inhos" or "-inhas". The line in (6) instructs R to look for sequences of the vowel "e", followed by "n" or "m", and by a consonant or end of a word.

The automatization of the task of identifying tokens does not exempt the researcher from a thorough qualitative analysis of his/her corpus. As an example, the definition specified in (5) above correctly identifies tokens such as *copinho* 'little cup' and *bonitinhas* 'pretty', but will also mark instances of the words *minha* 'my, mine', *tinha* 'had', *vinha* 'came', which are not diminutives. In this case, the researcher can run Script 1 first as a general survey, make a list of all unwanted tokens, and further exclude words that should not be included by employing the `gsub` function:

(7)
```
clean.files<-gsub((minhas?|tinha|v
inha)   <>,\\1,   all.corpus[[i]],
ignore.case=T)
```

The command line in (7) instructs R to look for instances of `minha <>`, `minhas <>`, `tinha <>` and `vinha <>` and to substitute them only by the first element (`\\1`) (`minha, minhas, tinha, vinha`), i.e. to delete `<>` in these sequences.

In general, it seems that it is easier to create "syntactic definitions" for phonetic and morphological variables than for syntactic and discourse variables, if one is working with an unannotated corpus. This is because such analyses will probably require further information about the syntactic or discourse function of words in a given corpus. Take, for instance, an analysis of the variable use of Wh-interrogatives in Brazilian Portuguese (Oushiro, 2011), which has four variants: (i) *Onde você mora?*; (ii) *Onde que você mora?*; (iii) *Onde é que você mora?*; and *Você mora onde?* 'Where do you live?' Finding wh-words (*o que* 'what', *que* 'what, which', *quem* 'who', *qual* 'which', *quando* 'when', *onde* 'where' etc.) is not a difficult task, in principle; however, these words do not only function as wh-words, but can also be relative pronouns (Cf. *a casa onde morei* 'the house where I lived,' *o homem que gosta de TV* "the man who likes TV") or complementizers (Cf. *ele disse que viria* 'he said (that) he would come'). An unannotated corpus would miss these distinctions.

On the other hand, these different scripts can be employed independently from one another. If a researcher already has transcript files marked with the relevant tokens, Script 2 can be applied to extract them into a spreadsheet file regardless of Script 1 having been used. If a researcher simply wants to produce a subsample of a larger token file, Script 3 can be applied independently from scripts 1 and 2.

4. Conclusion

The software R enables the sociolinguist to automatize certain mechanical tasks in preparing a coding file for quantitative analyses. The scripts presented here allow the researcher to handle data in a more consistent manner and, by reducing the time spent in preparing the token file, it allows more time to perform statistical analyses and interpret results. Adapting the scripts should allow the analyses of a number of sociolinguistic variables, especially phonetic and morphological ones.

5. Acknowledgements

6. References

Bayley, R. (2002). The quantitative paradigm. In J.K. Chambers, P. Trudgill and N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change*. Malden, MA: Blackwell, pp. 117--141.

Cieri, C., Strassel, S. (2010). Towards best practices in sociophonetics: Robust, digital, empirical, reproducible sociolinguistic methodology. Paper presented at NWAV40, Washington DC. Available at: <http://papers.ldc.upenn.edu/>.

Gries, S.Th. (2009) *Quantitative Corpus Linguistics with R*. New York: Routledge.

Guy, G.R. (1993). The quantitative analysis of linguistic variation. In D. Preston (Ed.), *American Dialect Research*. Amsterdam/Philadelphia: John Benjamins, pp. 223--249.

Hornik, K. (2011). R FAQ. Available at : <http://cran.r-project.org/doc/FAQ/R-FAQ.html>.

Mendes, R. B. (2011). How feminine do diminutives sound? Paper presented at NWAV 40, Georgetown University, USA. Available at: <http://nwav40.georgetown.edu/Panel2.pdf>.

Oushiro, L. (2011). Uma análise variacionista para as Interrogativas-Q. Master's dissertation. Available at: <http://www.teses.usp.br/teses/disponiveis/8/8139/tde-14102011-135709/publico/2011_LiviaOushiro.pdf>.

Oushiro, L. (2012a). Fatores sociais na pronúncia variável de /r/ na cidade de São Paulo. Paper presented at II Congresso Internacional de Linguística Histórica. Avaiable at: < http://www.cilh.fflch.usp.br/node/73>.

Oushiro, L. (2012b). PhD Qualifying Report. Ms.

Rosenfelder, I., Labov, W. (2010). New methods for large scale automatic vowel analysis. Workshop at NWAV 40, Washington DC. Available at: <http://nwav40.georgetown.edu/Bill%20Labov%20and%20Ingrid%20Rosenfelder.pdf>.

Wolfram, W. (1993). Identifying and interpreting variables. In D. Preston (Ed.), *American Dialect Research*. Amsterdam/Philadelphia: John Benjamins, pp. 193--221.

ILLOCUTIONS AND ATTITUDES

Prosodic cross-linguistic perception of social affects in Mandarin Chinese by native, French and Vietnamese listeners

Yan LU[1], Véronique AUBERGÉ[1], Albert RILLIARD[2], Wentao GU[3]

[1]GIPSA-Lab; Grenoble France; [2]LIMSI; Orsay France; [3]Institute of Linguistic Science and Technology, Nanjing Normal University, Nanjing, China

yan.lu@gipsa-lab.grenoble-inp.fr, veronique.auberge@gipsa-lab.grenoble-inp.fr, albert.rilliard@limsi.fr, wtgu@njnu.edu.cn

Abstract

Social affects play an important role in the face-to-face interaction and are implied in the realization of speech acts. The prosody is a main vector of social affects and its cross-language variability is a challenge for language description as well as for foreign language teaching. The present work aims at examining the perception of Chinese social affects in an intra-cultural perceptual experiment and the influences of tones on the perception of these social affects in another inter-cultural perceptual experiment. A speech corpus was designed with the variation of length, tone location and syntactic structures of utterances, and has been incorporated with 19 social affects. For each experiment, a specific sub-corpus was selected. The tests results show that the social affects were globally recognized over chance level by native and non-native listeners; "declaration" is the attitude which attracted the most confusions; all subject groups separated the 19 Chinese social affects in two subsets: a subset of "assertive" attitudes (represented by "declaration") and a subset of "interrogative" attitudes (represented by "question" and "doubt"); more similarities were found between French and Vietnamese listeners in inter-cultural perception experiment.

Keywords: social affects; prosodic perception; tones; Mandarin Chinese; French; Vietnamese.

1. Introduction

The affects expressed in interactive speech imply two different levels of the speaker's cognitive processing (Aubergé, 2002): the involuntarily controlled expressions of affects (so-called "emotions"), and the intentionally controlled expressions expressed through audio-visual prosody (so-called "social affects" or "attitudes"). Prosodic attitudes, -functions of the speaker's opinion, beliefs or knowledge (Wichmann, 2000), are an integral part of the language interaction building and are performed through the audio-visual prosody. They need to be learned in infancy and would benefit to be explicitly taught in foreign language teaching. In the present work, some values of social affects, which potentially reveal the speaker's opinion or some social and situational cues, e.g. the speaker-hearer relationship, were selected for two perceptual experiments in order to investigate the prosodic perception of social affects in Mandarin Chinese by native and non-native listeners.

Different hypotheses have been set up about the typologies of attitudinal expressions (Martins-Baltar,1977; Wichmann, 2000; de Moraes *et al.*, 2010; Gu *et al.*, 2011), and we propose to classify social affects into three categories: first, the attitude, intention or opinion of the speaker about what he says (even if he does not express any attitude by performing a simple declaration or question, it is then considered as the attitude to give no information on his own attitude, -Aubergé, 2002); second, some expressions characterising the social relation implied in the interaction, e.g. politeness, authority; third, the expressions depending on the socio-cultural context of interaction, typically for intimacy, infant-directed speech and seduction.

Mandarin Chinese (also referred to as Putonghua or Standard Chinese) has four tones which were defined customarily according to the characteristics of their fundamental frequency curve as: high level (tone 1), rising (tone 2), dipping (tone 3) and falling (tone 4). Belonging to different families of languages, Mandarin Chinese, Vietnamese and French have their own specific linguistic structures. Both Mandarin and Vietnamese are tonal languages; French is not tonal (and not stressed). Compared to French, and from the prosodic and cultural point of view, Vietnamese could be considered as closer to Chinese. Therefore, it is supposed that Chinese lexical tones could influence to some extent the prosodic perception of Chinese social affects by subjects of different language backgrounds.

Hence, an intra-cultural perceptual experiment was designed to examine how prosodic social affects in Chinese can be perceived by native Chinese, and another inter-cultural perceptual experiment was required to investigate how these social affects can be perceived by French and Vietnamese listeners and if the effect of tones can be shown on the perception of social affects outside of any morphosyntactic and semantic influences.

2. Corpus

2.1 Speech corpus design

In order to compare the parameters implied in the variability of prosody, a dedicated and controlled corpus was built to convey different social affects.

The corpus was designed with consideration of utterances' length (in syllables), of tones location and of syntactic structure, which were systematically varied in order to analyze further the variation of one parameter in the same context for the others. As the social affects could not be produced without reference to context, a dedicated context of interaction was described for each social affect, in order to help the speaker to express them

as naturally as possible. All utterances were constructed to bear a literally neutral meaning (i.e. not conveying any meaning which implies a specific social affect nor emotion) but in the same time could be expressed with all the social affects studied. The complete corpus contains 152 utterances performed with 19 attitudes, i.e. 2888 stimuli.

2.2   Selected social affects

Some social affects (attitudes) in different language have been studied in by Fujisaki & Hirose (1993), Aubergé (1998), Mac *et al*. (2010), Gu *et al*. (2011) and Lu *et al*. (2012). In this work, 19 social affects, which are commonly encountered in daily conversation, were selected. Table 1 shows the 19 social affects and their abbreviations, grouped in three categories.

	Social affects and abbreviation
Attitudes	Declaration (DECL) Question (QUES) Admiration (ADMI) Confidence (CONFI) Irritation (IRRI) Resignation (RESI) Contempt (CONT) Irony (IRON) Doubt (DOUB) Obviousness (OBVI) Disappointment (DISA) Neutral surprise (NEU-S) Positive surprise (POS-S) Negative surprise (NEG-S)
Social parameters	Politeness (POLI) Authority (AUTH)
Social context	Seduction (SEDU) Infant-directed speech (IDS) Intimacy (INTI)

Table 1: Classification of social affects and their abbreviation

2.3   Corpus recording

One native Mandarin female from Shaanxi province of China took part in the recording. She is teacher of French as a foreign language in a Chinese college, and speaks unmarked standard Mandarin Chinese. The recording was conducted in a sound proof room at GIPSA-Lab in Grenoble, France, both in video and audio modalities. To make the attitudinal expressions consistent, the sentences sharing the same attitude were recorded in one session after the speaker had understood and had got familiar with the situational context of the given affect. 19 social affects were conveyed one by one in the same way. Another native Chinese from the same area of China as the speaker was also present during the recording to supervise the performance of the speaker.

3.   Native perceptual validation

3.1   Description of the experiment

To test the validity of the attitudinal speech corpus and to look into the perception and the recognition of attitudes, we designed this perceptual experiment with a sub-corpus of 21 utterances conveying the 19 social affects, i.e. 399 stimuli. The listening subjects were composed of 30 native Mandarin Chinese, from different areas of

China: 15 males and 15 females with an average age of 25.2 years. They're almost all postgraduate students or PhD students in Grenoble, France (except one male subject who works as computer programmer in an IT company in Grenoble), and none of them reported any listening and understanding disorder.

All 399 target stimuli were presented to the subjects through headphones in a quiet room and were introduced by a presentation of the experiment and a description of each social affect with examples of situations in which such social affects can happen. The listeners had the written instructions in their native language at their disposal during the experiment. They listened only one time each stimulus and had to choose the perceived attitude amongst the 19 proposed labels, written in Chinese. The presentation order of the stimuli was randomized for each subject.

3.2   Analysis and results

An analysis of variance (completely randomized three-factorial design) was carried out on the data. The three fixed factors were the subjects' gender (G, 2 levels), the presented attitudes (A, 19 levels) and the sentences length (L, 4 levels). Each cell of this design contained at least 60 observations. The significance level was set at 0.01. Table 2 shows the results of the analysis of variance for each factor.

The factors "Attitude", "Length" and the interaction between "Attitude" and "Length" have significant effect; "Attitude" has the highest observed strength of effect (η^2). Factors "Gender" and "Length" are significant at the 1% level, but does only explain a small part of the variance observed.

	Sum Sq	Df	F value	Pr(>F)	η^2
A	**253.43**	**18**	**86.2218**	**0.0000**	**0.693**
G	**1.97**	**1**	**12.0648**	**0.0005**	**0.005**
L	**16.19**	**3**	**33.0582**	**0.0000**	**0.044**
A*G	5.57	18	1.8960	0.0122	0.015
A*L	**80.30**	**54**	**9.1061**	**0.0000**	**0.220**
G*L	0.13	3	0.2556	0.8574	0.000
A*G*L	7.87	54	0.8929	0.6958	0.021

Table 2: ANOVA's results – significant effects in bold

Through the mean recognition rate of 19 social affects and the mean recognition rate of social affects distinguished by stimulus's length and gender presented in figure 1, it is observed that for native Chinese listeners, almost all of the social affects were recognized above chance level, except "confidence" and they can be classified in the decreasing order (cf. Figure 1, top). The identification of social affects varies with the stimuli's length: according to the confusion matrix of attitudes by length, there is a clear separation between the 1-syllable stimuli and the longer ones. The 1-syllable stimuli received lower recognition scores while the 4-syllable stimuli received the highest (the 9-and 2-syllable stimuli are just under the 4-syllable ones). The graph of the

mean recognition rate for social affect by length (figure 1, bottom) shows that "infant-directed speech" and "irritation" don't' follow this trend. For "infant-directed speech", the 1 and 2-syllable stimuli were better recognized than the 4 and 9-syllable ones (who were mixed up with "seduction"). For "irritation", the 2-syllable stimuli were not well perceived, in comparison with other lengths, and were confused with "declaration" and "confidence".
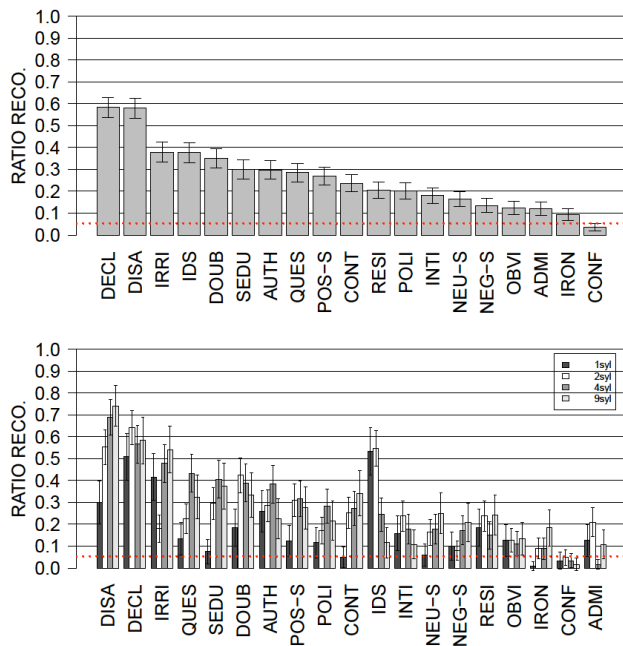
Figure 1: Recognition rate for the 19 social affects: rate per attitude (top), detailed per stimuli's length (bottom)

4.   Cross-cultural perception

4.1   Description of experiment

This perception test was aimed to study French and Vietnamese listeners' perception of these Chinese social affects and to measure a potential interaction between attitudes and tones. A sub-corpus of 16 utterances was selected with a systematic variation of tones values and location. There are no morpho-syntactic nor semantic variations in the sub-corpus. 15 French (6 females, average age of 33 years) and 15 Vietnamese (8 females, average age of 27 years) took part in the experiment. All of them work or study in Grenoble, France. None of the 30 subjects reported any listening disorder. The test's paradigm was the same as for the first experiment.

4.2   Analysis and results

An analysis of variance (completely randomized three-factorial design) was carried out on the data. The three fixed factors were the presented attitudes (A, 19 levels), the sequence of tones (T, 16 levels) and the native language of subjects (L, 2 levels). The significance level was set at 0.01. Table 3 shows the general results of the analysis of variance for each factor. The factors

"Attitude", "Tones sequence" and the interaction between "Attitude" and "Tones" show significant effects. Both "Attitude" and the "Attitude & Tones" interaction have the most important effect size (cf. the η2 column of Table 2), and thus are the most influencing factor on listeners' answers.

	Sum Sq.	Df	F value	P	η²
A	**81.85**	**18**	**35.8**	**0.000**	**0.070**
L	0.11	1	0.8	0.362	0.000
T	**4.61**	**15**	**2.4**	**0.001**	**0.004**
A*L	**10.36**	**18**	**4.5**	**0.000**	**0.009**
A*T	**68.53**	**270**	**2.0**	**0.000**	**0.060**
L*T	3.11	15	1.6	0.058	0.003
A*L*T	37.72	270	1.1	0.131	0.034

Table 3: Global ANOVA results – significant effects in bold

Two separated ANOVAs on French and on Vietnamese subjects were run (table 4). Results show that the effect of "Tones" is significant for French subjects while it is not significant for Vietnamese subjects (cf. mean results on fig. 2), although there is a significant interaction between "Attitude" & "Tones".

	Df	French		Vietnamese	
		F	p	F	P
Attitude	18	**20.2**	**0.000**	**20.1**	**0.000**
Tones	15	**2.4**	**0.002**	1.6	0.054
Attitude*Tones	270	**1.4**	**0.000**	**1.4**	**0.000**

Table 4: Separated ANOVAs by language – significant effects in bold
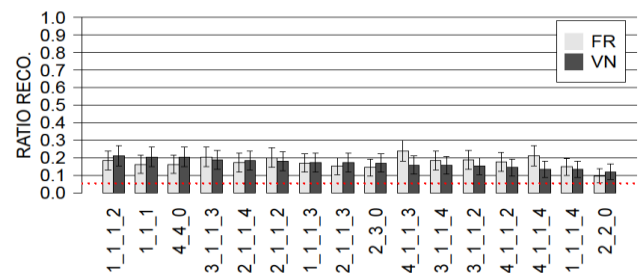
Figure 2: Mean recognition rate for each tone sequence, per language background
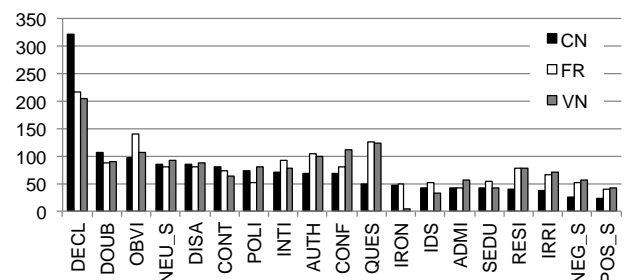
Figure 3: Mean recognition for the 19 social affects, per language background

Figure 3 shows the mean recognition of the 19 social affects by French and Vietnamese subjects. Almost all of the social affects were identified above chance, except "contempt", "irony" and "confidence" for French and "irony" for Vietnamese subjects.

5.   Discussion and conclusions

A comparison of results obtained in both experiments allows to analyse three aspects: the mean recognition rate of the 19 social affects; the attractivity of individual social affects and the different clustering made by native and non-native subjects.

5.1   Native and non-native results

In the intra-cultural test, almost all of the social affects were recognized over chance, except "confidence", and "Declaration" was the best recognized attitude. In the inter-cultural test, almost all of the social affects were identified above chance, except "contempt" "irony" and "confidence" for French and "irony" for Vietnamese. "Declaration" is the best-recognized attitude by French, while it is "disappointment" for Vietnamese. Native listeners received higher recognition scores than non-native listeners. For "seduction" and "authority", French listeners show the highest recognition scores, and for "confidence", Vietnamese listeners did the best (cf. figure 1 (top) and figure 3). Concerning the less identified attitudes in this audio modality, they have been supposed to rely strongly on the visual modality (Shochi, 2008). Hence, another multimodal perceptual experiment will be carried out to investigate how the 19 audio-visual prosodic attitudes will be perceived by native and non-native subjects.

Analysis of the cross-cultural experiment also showed that the tones have some influences on the perception of several social affects and that the tonal effect is more important for French subjects than for Vietnamese ones. As it was commonly accepted that there are cross-cultural similarity in the uses of F0 to signal affect, intention, or emotion (Ohala, 1994), in order to validate the findings, our future works will focus on the acoustic analysis of the social affects, with an emphasis on the F0 contour of tones which is the primary acoustic parameter for Mandarin tones (Allard *et al.*, 2006).

5.2   Attractivity of Chinese social affects

The attractivity of attitudes – the sum of all confusions attributed to a given attitude (cf. fig. 4) – shows some interesting results. For native listeners, the attitude attracting most answers is "declaration", which is mainly used when judges cannot identify any attitude. This result is coherent to common behaviors of perceiving his language (de Moraes *et al.*, 2010; Diaferia, 2002; Mac *et al.*, 2010; Shochi *et al.*, 2009). Moreover, recognizing a perceived stimulus amongst 19 attitudinal labels is a cognitively complex task. Thus, choosing "declaration" is a way to avoid false or uncertain answers without specifying any information about attitude. French and

Vietnamese listeners show, to a lesser degree, the same preference for "declaration", but with quite clear second choice: "question" for Vietnamese and "obviousness" for French judges. "Irony" was not well recognized by Vietnamese judges, nor did it attract any attitude.
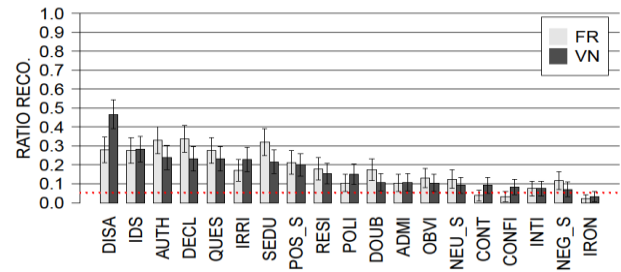
Figure 4: The attractivity – cumulated percentage of confusions from others attitudes to each one, per language background

5.3   Clustering of Chinese social affects

In order to measure the perceptive distances between each stimulus and to identify the higher perceptual categories for Chinese, French and Vietnamese subjects, as well as the perceptual differences between the three groups, a hierarchical clustering analysis was run on the dispersion matrix. Distances were expressed as the correlation (r) between rows (1-r is used as the distance). From these perceived distances, a hierarchical clustering algorithm was applied, which allowed the observation of the main clusters of attitudes for each language group (cf. figure 5). The three groups clustered the attitudes almost in the same way and all have separated the attitudes in two subsets: a subset of "assertive" attitudes (represented by "declaration") and a subset of "interrogative" attitudes (represented by "question" and "doubt"). Meanwhile, in observing closely the clustering, we found that French and Vietnamese listeners have grouped the perceived social affects in the same eight clusters - that differ to some extend from the seven groups made by Chinese subjects. This result is contrary to our hypothesis in which there should be more similarities between Chinese and Vietnamese listeners in respect to cognitive processing of social affects. An evaluation of the classification of the concepts of Chinese and French social affects will be carried out in order to measure the cognitive distances between the attitudinal concepts and propose a cognitive clustering of social affects in daily life.
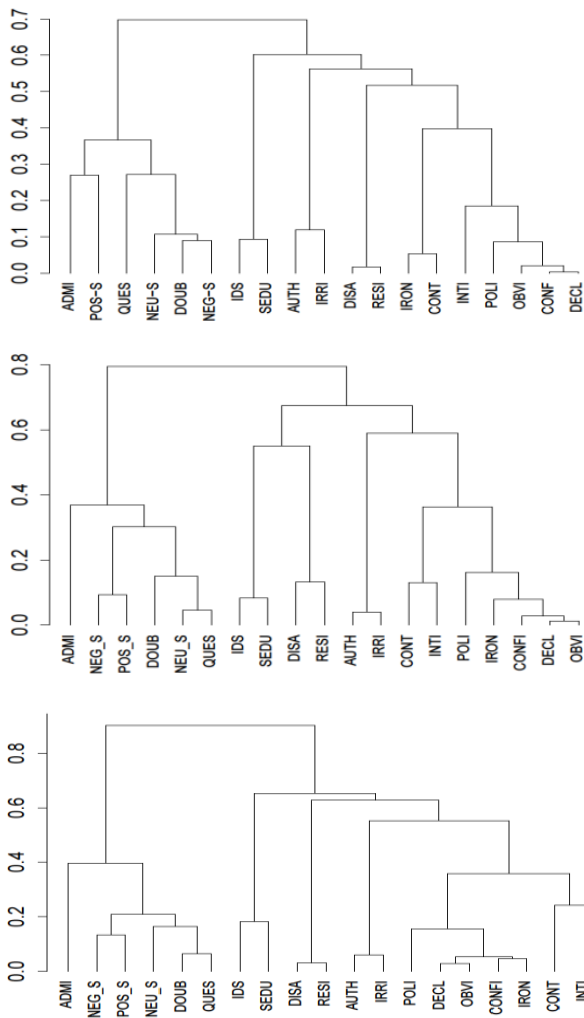
Figure 5: Hierarchical clustering of perceived social affects, based on R complete grouping criterion. The grouping done by Chinese subjects is shown on top, by French subjects in middle, by Vietnamese subjects on the bottom

6.   Acknowledgements

7.   References

Allard, J., Wang, Y., Corinne, B.M. and Joan, A.S. (2006). Perception and Production of Mandarin Chinese Tones. In E. Bates, L.H. Tan, and O.J.L Tzeng (Eds.), *Handbook of Chinese Psycholinguistics*. Cambridge University Press, Cambridge.

Aubergé, V., Grépillat , T. and Rilliard, A. (1997). Can we Perceive Attitudes before the End of Sentences? The Gating Paradigm for Prosodic Contours. In *EuroSpeech'97*, Rhodes, Greece, pp.871-- 877.

Aubergé, V. (2002). A Gestalt Morphology of Prosody Directed by Functions. In *Speech Prosody 2002 Proc.* Aix en Provence, France, pp. 151--154.

de Moraes, J.A., Rilliard, A., Alberto, B. and Shochi, T. (2010). Production and perception of attitudinal

meaning in Brazilian Portuguese. In *Speech Prosody 2010 Proc.* Chicago, USA.

Diaferia, M.L. (2002). Les Attitudes de l'Anglais: Premiers Indices Prosodiques. Master thesis in Cognitive Science. National Polytechnique Institut of Grenoble, France.

Gu, W., Zhang, T. and Fujisaki, H. (2011). Prosodic Analysis and Perception of Mandarin Utterances Conveying Attitudes. In *Proceedings of Interspeech 2011*. Firenze, Italy, pp. 1069--1072.

Lu, Y., Aubergé, V. and Rilliard, A. (2012). Do you hear my attitude? Prosodic perception of social affects in Mandarin. In *Speech Prosody 2012 Proc.* Shanghai, China.

Mac, D.K., Aubergé, V., Rilliard, A. and Castelli, E. (2010). How prosodic attitudes can be recognized and confused: Vietnamese multimodal social affects. *In* SLTU, Penang, Malaysia.

Martins-Baltar, M. (1977). *De l'énoncé à l'énonciation: une approche des fonctions intonatives.* Didier, Paris.

Ohala, J.J. (1994). The frequency codes underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols and J.J. Ohala Eds), *Sound symbolism*. Cambridge University Press, Cambridge, pp. 325--347.

Rilliard, A., de Moraes, J.A., Erickson, D., Shochi, T. (2012). Prosodic analysis of Brazilian Portuguese attitudes. In *Proceedings of the Speech Prosody 2012.* Shanghai, China, pp. 677--680.

Shochi, T. (2008). Prosodie des affects socioculturels en japonais, français et anglais: à la recherche des vrais et faux-amis pour le parcours de l'apprenant. PhD thesis in Language Science. University of Grenoble III - Stendhal, France.

Shochi, T., Rilliard, A., Aubergé, V. and Erickson, D. (2009). Intercultural Perception of English, French and Japanese Social Affective Prosody. In S. Hancil [Ed], *The role of prosody in Affective Speech*. Linguistic Insights 97, Peter Lang AG, Bern, pp. 31--59.

Wichmann, A. (2000). The attitudinal effects of prosody, and how they relate to emotion. In *Proceedings of the ISCA Workshop on Speech and Emotion.* Newcastle, North Ireland, pp. 143--148.

Can the tones influence the acoustic perception of the Vietnamese attitudes by French listeners? Some evidences for global vs. local processing of prosody

Dang-Khoa MAC[1,2], Véronique AUBERGÉ[2], Albert RILLIARD[3,] Eric CASTELLI[1]

[1] International Research Institute MICA, HUST-CNRS/UMI 2954 Grenoble INP, Hanoi, Vietnam; [2] Laboratory of Informatics of Grenoble (LIG), CNRS, France; [3] LIMSI-CNRS, Orsay, France

dang-khoa.mac@mica.edu.vn, eric.castelli@mica.edu.vn, veronique.auberge@imag.fr, albert.rilliard@limsi.fr

Abstract

Attitudes or social affects are strongly implied in the interaction processing, and specifically into the socio-cultural aspects of language. The prosody has been shown as a main vector for expressing attitudes in different languages. In tonal language, the lexical access function is also implemented by the parameters of prosody. This paper presents a study of attitudinal expressions in Vietnamese, a tonal language, under the light of cross-cultural perception. Sixteen Vietnamese attitudes, performed on sentences including with tonal variation, were used in a perception experiment with French listeners. The result of French subjects on the utterances with tones and non-tone allow us to explore the influences of tones on the different Vietnamese attitudes in non-tonal language speakers.

Keywords: attitude; social affect; tone; global prosodic patterns; cross-cultural perception.

1. Introduction

The attitudes, and more generally the social affects, are an important part of the face-to-face interaction and are linked to the language through the socio-culture. These expressions are clearly social: they carry the intentions and points of view of the speaker (e.g. surprise, confirmation, etc.) and can give the social context on the interaction (e.g. intimacy, politeness). When the speaker does not express any attitude in his speech act (in the case of a declaration or a "simple" question), she/he expresses that she/he has no opinion on this utterance or that she/he does not want or cannot express any attitude (Aubergé, 2002).

Even if many such social affects are universal in their values or in their prosodic forms, some prosodic implementation and even some attitudinal values are specific to the culture and the language (Scherer et al., 2001; Shochi et al., 2007). Anyway, the attitudes are built inside each culture and language, and they are acquired by children inside their culture or learned by the learners of second language (Shochi et al., 2010). The understanding of this phenomenon may benefit from cross-cultural studies (Scherer et al., 2001; Shochi et al., 2010).

The attitudes or social affects are supposed to be involved into voluntary cognitive controls, whereas emotions are involuntary controls (Aubergé, 2002). The prosody has been shown as a main vector for expressing attitudes in different languages (Wichmann, 2000; Aubergé, 2002). The "classical" prosodic parameters (F0, intensity, timing), are strongly implied in the expression of attitudes (Fónagy, 1983; Wichmann, 2000; Aubergé, 2002). Campbell & Mokhtari (2003) proposed the voice quality as a 4th dimension of prosody; it has been also shown as a fundamental parameter for emotions (Banse & Scherer, 1996; Audibert, 2005) and is used in some attitudes (Shochi et al., 2007). Many different functions are implemented by prosody by using the same acoustic parameters (F0, intensity, timing and voice quality).
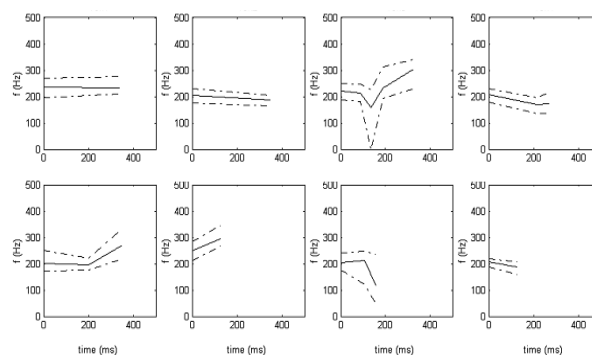
Figure 1: Examples of contours of 8 Vietnamese tone representations from a female subject (Pham et al., 2002). From the left to right, top to bottom: tone 1, 2, 3, 4, 5, 5b, 6, 6b

In tonal languages such as Vietnamese, a part of the lexical access function is implemented by F0. The Vietnamese language has 6 tones: level (1), falling (2), broken (3), curve (4), rising (5) and drop (6) as shown in Figure 1. Tone 5b and 6b correspond to tone 5 and 6 on a syllable ended by a stop consonant. Moreover the Vietnamese tonal system can employ some changes of voice quality, with the F0 variations, with co-occurrence of glottalization during the production of tone 3 and tone 6. Tone 3 is accompanied with a harsh voice quality due to a glottal stop (or a rapid series of glottal stops) around the middle of the vowel. Tone 6 has the same kind of harsh voice quality as tone 3; however, it is distinguished by dropping very sharply and it is almost immediately cut off by a strong glottal stop (Do et al., 1998).

The domain of the tonal function is the syllable, which represent a local domain of variation compared to the length of a complete utterance. The attitudinal function concerns the utterance unit, and the prosody of attitude can be described as a global contour related to the utterance (Aubergé, 2002). Modification of F0 values due to either the global attitudinal function or the "local" tonal function seems to be clearly differentiated by native tonal language speakers, but the question of the perceptive

processing of such functional variations by speakers of a non-tonal language, could inform on the cognitive mechanisms of this social signal.

This work aims at exploring possible perturbations by the tonal system on the perception of Vietnamese attitudes by French speakers (i.e. a non-tonal language): will they be able to perceptively extract and separate, from the same acoustic parameters, the tonal values from the attitudinal information? That is to process the lexical access function, attached to word domain, within the attitude function, attached to the whole utterance domain, but morphologically implemented by prominences. May the local tonal variation interfere with the decoding by speakers of a non-tonal language of the utterance-length variations of attitudinal prosody? How does behave such local vs. global cues, described by Gestalt theories of prosodic morphology (Aubergé, 2002).

To answer this question, in this paper, after presenting the construction of the corpus of Vietnamese attitudes, we describe the perceptual experiment of attitudes with tones variation designed for French listeners. The perception results are analyzed, and compared with previous results (Mac *et al.*, 2010) of attitudinal perception on non-tonal (tone 1) Vietnamese utterances by French listeners. The results allow us to answer the question of whether the non-tonal language listeners are able to extract and separate a tone's lexical F0 value from the attitudinal information. This paper concludes with some discussions and perspectives.

2. Corpus

2.1   Vietnamese attitude corpus

Based on research on some attitudes studied in Vietnamese (Le, 1989) and in other languages (Diaféria, 2002; Shochi, 2008, Rilliard *et al.*, 2009), 16 attitudes have been selected for Vietnamese in our corpus (Table 1).

Declaration	DEC	Irritation	IRR
Interrogation	INT	Sarcastic irony	SAR
Exclamation of neutral surprise	EXo	Scorn	SCO
Exclamation of positive surprise	EXp	Politeness	POL
Exclamation of negative surprise	EXn	Admiration	ADM
Obviousness	OBV	Infant-directed speech	IDS
Doubt-Incredulity	DOU	Seduction	SED
Authority	AUT	Colloquial	COL

Table 1: 16 selected Vietnamese attitudes, with their abbreviations

To observe the effects of tone and tonal co-articulation on attitudinal expression, the corpus contains 8 sentences of one-syllable length, corresponding to the 8 types of Vietnamese tone, and 72 sentences of two-syllable length, which correspond to all combinations of two tones among the 8 Vietnamese tones. The remainder of the corpus is based on 45 sentences of 3-

to 8-syllable length and systematically varied in their syntactic structure: single word, nominal group, verbal group and a simple structure "subject-verb-object". That means the corpus is built on 125 sentences without specific affective meaning, produced with all the 16 attitudes and balanced in terms of tone position. These sentences were recorded (both in audio and video, but audio only is focused in this paper) by one male speaker native of the Hanoi dialect (standard pronunciation). The whole corpus thus contains 2000 sentences corresponding to more than 90 minutes of speech after post-processing.

2.2   Sub-corpus for tone variation experiment

From this corpus, a subset was selected with a systematic variation of tones in different syntagmatic and paradigmatic locations. Nineteen sentences of 2- and 3-syllable length were chosen from the corpus for the test. The tones were set at different positions (at the first, middle and last syllable), as shown in Table 2. The selection was done on 2- and 3-syllable length sentences, since they are short enough to avoid syntactic complexity.

Tone sequence	Utterance in Vietnamese	English
1_1	anh ta	him
2_1	người ta	them
3_1	đã xong	finished
4_1	thủy tinh	glass
5_1	chúng ta	us
6_1	chị ta	her
5b_1	héc ta	hectare
6b_1	tốp ca	choral
1_2	rau cần	celeri
1_3	dây kẽm	steel wire
1_4	cây cảnh	home plant
1_5	y tá	male nurse
1_6	danh bạ	year book
1_5b	công tác	mission
1_6b	sa mạc	desert
4_1_1	bảy mươi ba	73
1_5_1	hai chúng ta	both of us
6_5b_3	hợp tác xã	cooperation
1_4_6	em bảo chị	you tell me

Table 2: Sub-set of tonal variation
for 2 and 3 syllables length

3.   The perception protocol

The perception experiment was carried out to study the influence of Vietnamese tones at varied location on the perception of the 16 Vietnamese attitudes. Twenty French listeners who have no experience with the Vietnamese language took the experiment. The perception test was carried out in a quiet room, using a high-quality headset at a comfortable hearing level. The program interface gave the label and the explanation of the 16 attitudes (in the native language of the listener). No listener expressed any

difficulty in understanding the concepts of these 16 attitudes. All subjects listened to each stimulus only once. After each stimulus, they were asked to indicate the perceived attitude among the 16 attitudes and to indicate the intensity of its expressiveness on a scale ranging from "hardly perceptible" (encoded as 1) to "very marked" (encoded as 100). The score 0 was assigned to the 15 other attitudes.

4.   Result analysis

A cross-cultural perceptual experiment has already been performed with Vietnamese attitudes on utterances using only the "neutral" (flat) tone (Mac *et al.*, 2010). This experiment was carried out to have a reference of the non-native perception of attitudes, without tonal variations. For the comparison with the Vietnamese listener's performances, cf. Mac *et al.* (2010b)

4.1   Effect of factors

The results of the perception test were first analysed with a repeated measure ANOVA, in order to evaluate the relative effect of the tones and their position on the listener's perceptual responses. First, the ANOVA of neutral tone sentences (Table 3) show a main effect of the presented attitudes for both Vietnamese and French listeners without a sentence length effect. This result confirms the choice of the 2- vs. 3-syllable length utterances for the experiment on the tonal sentences.

	df	Vietnamese		French	
		F	p	F	p
Attitude	**15**	**47.804**	**0.000**	**33.100**	**.000**
Sentence Length	2	3.735	0.024	1.655	.191
Attitude*SentenceLength	**30**	**3.542**	**0.000**	**3.007**	**.000**

Table 3: Output of ANOVA (on the percentage of attitude recognition) for Vietnamese and French subjects and phrase without tone. Significant effects at the 1% level are set in bold face

		% recognition		Confidence Rating	
	df	F	p	F	P
Attitude	**15**	**15.790**	**.000**	**15.419**	**.000**
Tone	7	1.582	.136	1.321	.236
TonePosition	**2**	**8.301**	**.000**	**10.007**	**.000**
Attitude * Tone	**105**	**1.976**	**.000**	**1.950**	**.000**
Attitude * TonePosition	**30**	**2.064**	**.001**	**2.005**	**.001**
Tone * TonePosition	6	2.519	.020	2.038	.057
Attitude * Tone * TonePosition	**90**	**3.528**	**.000**	**3.260**	**.000**

Table 4: Output of ANOVA (on the percentage of attitude recognition and level of confidence rating) for French subjects and phrase with tone. Significant effects at the 1% level are set in bold face

For the perception of French subjects on tonal sentences, the ANOVA results (Table 4) show that attitude has a significant effect on perception. There are also significant effects of the interactions between attitude, tones and tone positions. The tone has no significant effect on the perception result. However, the interaction between attitudes and tones is significant. That creates the appearance of the perturbation by tone prosody of some salient cues that are decisive information for some given patterns of attitudes. It must be further verified if it happens only when the local cues can be acoustically confused with salient cues of another global pattern. However the global confusions between attitudes are not changed by tones (see Figure 4 and 5).

4.2   Tones vs. Non-tone structures

4.2.1.   Attitude identification

Figure 2 shows the mean recognition rate (in %) for French listeners with 8 representations of Vietnamese tones. The attitude recognition results for the French listeners on the tone variable sentences are not so different. This is verified by the ANOVA result: globally, the tone variation has no effect on attitude perception. That means the non-native listeners can separate the (local) tonal effects and the (global) attitudinal effects.
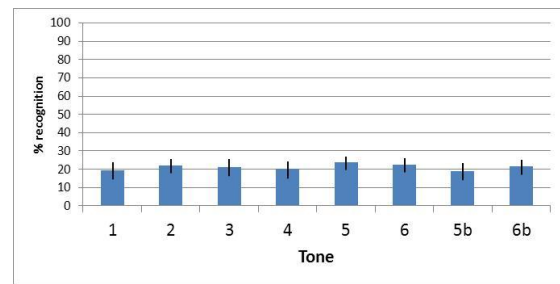
Figure 2: Mean recognition rate (%) for French listeners for each of the 8 tones presented

Figure 3 shows the perception differences between Vietnamese and French subjects. Globally, most attitudes were recognized above chance level, and native listeners have higher recognition scores than non-native French listeners (except the case of EXn), averaging for tone variation or for the neutral tone sub-corpus. It has to be noted that for French subjects, the neutral tone utterances are better recognized than the non-neutral tone utterances, except the cases of SCO and POL.
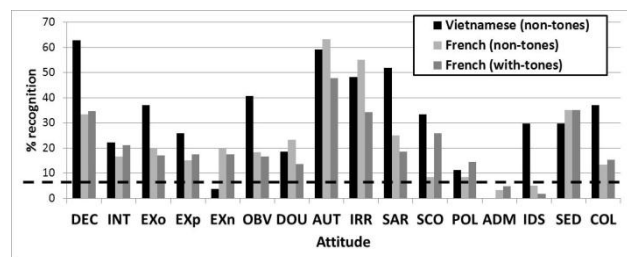
Figure 3: Recognition rate of 16 attitudes on non- tone and tonal sentences with Vietnamese and French listeners. The dash line: chance level

4.2.2. Attitude confusion

Figures 4 and 5 show the confusions matrices between attitudes for the varied tone sub-corpus and for the neutral tone sub-corpus. The two most clear results are: (1) on varied tone stimuli, the mean degree of confusion increases; (2) the confusions share the same tendencies in both sub-corpus. Only one new confusion bewteen DOU and INT (that are conceptually close) appears quite clearly for the varied tone stimuli. It means that the local perturbation by tones increases the complexity of the processing of global cues, but does not imply a re-organization, nor a clear misunderstanding by perturbing salient local cues. This result needs to be further explained by studying the similarity in prosodic characteristics of Vietnamese tones and attitudes through the French prosodic patterns.

Perceived attitudes

Presented Attitudes	DEC	INT	EXo	EXp	EXn	OBV	DOU	AUT	IRR	SAR	SCO	POL	ADM	IDS	SED	COL
DEC	33	5	7	0	7	25	2	7	0	2	7	3	0	2	0	2
INT	28	17	2	0	2	15	8	3	3	3	3	7	2	0	0	7
EXo	7	25	20	10	7	7	15	0	0	2	2	3	2	0	0	2
EXp	0	0	15	15	42	2	10	2	10	0	0	5	0	0	0	0
EXn	13	3	13	5	20	3	18	0	0	3	0	5	3	7	2	3
OBV	17	10	12	7	5	18	10	0	5	3	2	0	3	0	0	8
DOU	5	10	18	3	17	2	23	0	2	7	2	2	2	3	2	3
AUT	7	0	2	0	0	3	0	63	13	3	3	3	0	0	0	2
IRR	2	3	2	0	3	2	2	25	55	0	7	0	0	0	0	0
SAR	5	3	3	5	3	8	22	0	8	25	5	3	2	2	0	5
SCO	10	0	10	0	3	17	8	2	0	10	8	10	2	7	7	7
POL	27	5	12	3	0	17	5	0	0	0	0	8	5	8	0	10
ADM	12	7	7	3	2	3	2	0	0	0	2	7	3	25	5	13
IDS	5	5	10	7	7	13	15	3	0	15	3	2	5	5	0	5
SED	8	13	5	0	0	5	2	0	0	0	0	10	8	8	35	5
COL	13	3	8	0	0	13	0	0	0	2	2	8	7	7	23	13

Figure 4: Confusion matrix for French on neutral tone variation Vietnamese attitudes (Mac *et al*., 2010b)

Perceived attitudes

Presented Attitudes	DEC	INT	EXo	EXp	EXn	OBV	DOU	AUT	IRR	SAR	SCO	POL	ADM	IDS	SED	COL
DEC	36	7	5	0	1	20	4	5	2	4	4	5	1	0	2	4
INT	32	21	5	0	0	11	8	2	1	0	1	6	2	2	1	7
EXo	7	22	15	6	7	9	13	4	1	0	1	6	3	1	0	5
EXp	0	0	4	16	41	2	4	15	10	1	2	0	2	1	0	2
EXn	7	10	8	3	15	4	10	0	0	0	0	9	13	9	5	4
OBV	13	19	11	2	2	18	7	5	1	2	1	7	2	1	2	7
DOU	4	25	7	5	5	2	14	1	2	2	0	10	7	7	1	7
AUT	8	8	2	1	0	8	1	49	13	1	5	1	0	0	0	2
IRR	1	2	3	2	11	1	2	31	35	1	11	0	0	0	0	1
SAR	3	2	5	6	2	6	16	0	12	19	8	3	11	1	2	3
SCO	9	4	4	2	1	17	7	4	6	8	25	3	3	3	2	3
POL	31	7	6	0	1	11	3	0	0	0	0	14	2	7	5	11
ADM	11	3	4	0	0	7	5	0	0	1	3	10	4	19	18	16
IDS	4	7	3	1	2	9	11	4	7	31	8	5	6	2	0	1
SED	12	5	2	0	1	4	4	0	0	1	0	9	6	13	31	12
COL	14	1	1	1	0	6	1	1	0	1	2	11	4	19	20	16

Figure 5: Confusion matrix for French on tonal variation Vietnamese corpus

4.3 Effects of the interaction between attitudes, tones and tone positions

Figure 6, in appendix, shows the recognition results of 16 attitudes for each tone located on the first and last syllables of the sentences. As shown by the ANOVA, there is no global effect of the tone, but there is a significant effect of the interaction between attitudes, tones and tone positions. For example, the DEC is poorly recognized if tone 4 is located on the first syllable or tone 5 on the final syllable. For AUT, the tone 2 (falling) inhibits recognition when located at the first syllable, but not if located on the final syllable, while tone 3 (broken) impedes recognition of this attitude if located at the last syllable. EXn, DOU, SAR and SED are well recognized in the neutral tone sub-corpus, with a special effect of tone 6 for EXn in varied tone at the last syllable location. SCO is better recognized in the varied tone corpus than in the neutral tone corpus. In particular, tone 5b is efficient only if located on the first syllable, while the opposite is true of tone 5. INT is better recognized in the varied tone corpus, especially with tone 2 on the first syllable and tone 5b on the last syllable.

5. Conclusion

This work aims at studying the cross-cultural perception of Vietnamese social affects, a tonal language where a "neutral tone" can be used. The question of the prosodic influence of the local cues of tones on the global processing of attitudinal prosody can be asked. Some attitudinal stimuli with varied tones were presented to French listeners, who have no experience with lexical tone processing. The main experimental result is that the French listeners can globally separate the tone (local) processing from the attitude (global) processing. The tone processing can be considered as an increased cognitive load for French listeners that reinforces the degree of confusions between attitudes. However, interactions between the tone type, the tone location, and the attitude value indicate that the local cues of tones and the salient cues of global patterns (Aubergé, 2002) could be confused, but depending on the coinciding morphologies of the global and local patterns Thus these results need to be verified by further appropriate acoustic analysis to find out the acoustical parameters that lead to the perception of these social affects.

6. References

Aubergé V. (2002). A Gestalt Morphology of Prosody Directed by Functions: the Example of a Step by Step Model Developed at ICP. In *Proceedings of Speech Prosody*.

Audibert N., Aubergé V. and Rilliard A. (2005). The Relative Weights of the Different Prosodic Dimensions in Expressive Speech: A Resynthesis Study. In *Affective Computing and Intelligent Interaction*, pp. 527--534.

Banse, R., Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. In *Journal of Personality and Social Psychology*, 70, pp. 614--636.

Campbell, N., Mokhtari, P. (2003). Voice Quality: the 4th Prosodic Dimension. In *Proceedings of 15th International Congress of Phonetic Sciences*, Barcelona, Spain, pp. 2417--2420.

Diaféria, M.-L. (2002). Les Attitudes de l'Anglais: Premiers Indices Prosodiques. Master Thesis of Science Cognitives, Institut National Polytechnique de Grenoble, France.

Do, T.D., Tran, T. H., Boulakia G. (1998). Intonation in vietnamese. D. Hirst and A. Di Cristo (Eds.), *Intonation systems: A survey of 22 languages*. Cambridge University Press, pp. 395--416.

Fónagy, I. (1983). La vive voix. Essais de psycho-phonétique, Paris, Payot.

Gobl, C., Ni Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. Speech Communication 40(1-2): pp. 189--212.

Le, T.X. (1989). Etude contrastive de l'intonation expressive en français et en vietnamien. Linguistic and Phonetic, Université Paris 3. PhD

Mac, D.K., Aubergé, V., Rilliard, A. and Castelli, E. (2010). Cross-cultural perception of Vietnamese Audio-Visual prosodic attitudes. In *Proceedings of Speech Prosody* 2010, Chicago, USA.

Mac, D.K., Aubergé, V., Rilliard, A. and Castelli, E. (2010b). Vietnamese multimodal social affects: how the prosodic attitudes can be recognized and confused. In *Proceedings of International Workshop on Spoken Languages Technologies for Under-resourced languages* (SLTU 2010), Penang, Malaysia, 24-28.

Pham, T.N.Y., Castelli, E. and Nguyen, Q.C. (2002). Gabarits des tons vietnamiens. In *Proceedings of JEP*, Nancy, France, pp. 23--26.

Rilliard, A., Shochi, T., Martin, J.-C., Erickson, D. and Aubergé, V. (2009). Multimodal indices to Japanese and French: Prosodicaly expressed social affects. In *Language and Speech*: pp. 223--243.

Wichmann, A. (2000). The attitudinal effects of prosody, and how they relate to emotion. In *Proceedings of ITRW on Speech and Emotion*, Newcastle, Northern Ireland, UK.

Scherer, K.R., Banse, R. and Wallbott, H.G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. In *Journal of Cross-Cultural Psychology* 32: pp. 76--92.

Shochi, T., Aubergé, V. and Rilliard, A. (2007). Cross-listening of japanese, english and french social affect: about universals, false friends and unknown attitudes. In *Proceedings of 16th ICPhS*, Saarbrücken.

Shochi, T. (2008). Prosodie des affects socioculturels en japonais, français et anglais : à la recherche des vrais et faux-amis pour le parcours de l'apprenant. PhD Thesis de Science du Langage, Universites Grenoble III, France.

Shochi, T., Gagnié, G, Rilliard, A., Erickson, D. and Aubergé, V. (2010). Learning effect of prosodic social affects for Japanese learners of French language. In *Proceedings of Speech Posody*, Chicago, USA.

7.   Appendix
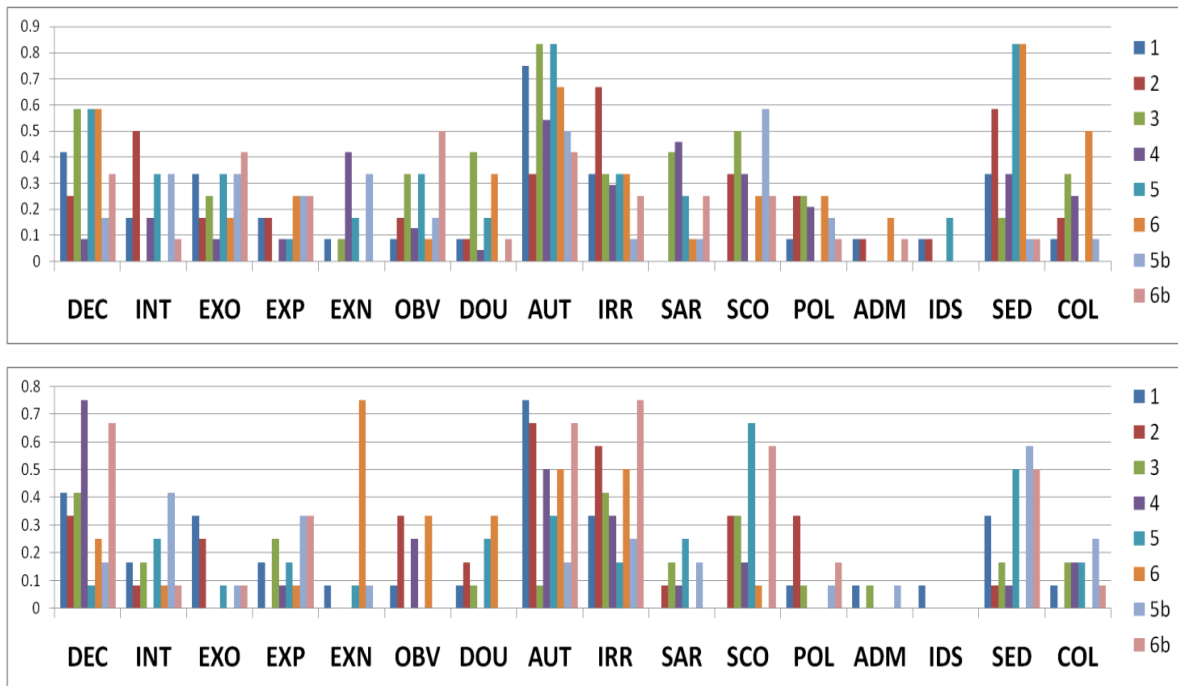
Figure 6: Recognition rate per attitude for each tone (1,2,3,4,5,6, 5b and 6b) located at the first (top) and the last (bottom) syllable of the sentences. Others syllables in sentences bear the neutral tone (tone 1)

A entonação e a força ilocucionária como pistas da atitude do locutor em atos de fala diretivos

Horácio dos Santos QUEIROZ

Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627 - Belo Horizonte/MG - Brasil
hsqueiroz@yahoo.com.br

Resumo

O presente estudo relaciona aspectos entonacionais e pragmáticos na expressão de atitudes do locutor a partir do exame do comportamento da curva de *f0*, na produção de atos de fala diretivos do português brasileiro da região metropolitana de Belo Horizonte, Minas Gerais, quais sejam, o *pedido*, a *súplica* e a *ordem*. No nível pragmático, o trabalho explora a Teoria dos Atos de Fala, comparando-se características entonacionais desses atos diretivos à noção de força ilocucionária, sobretudo aos critérios operacionais que a alteram (Vanderveken, 1991). Os resultados demonstram que diferentes estratégias entonacionais estão relacionadas à interpretação do modo de realização efetivo do ato de fala, e que as atitudes do locutor podem ser inferidas, pelo menos em um primeiro instante, com base nas operações que modificam a força ilocucionária.

Keywords: entonação; força ilocucionária; atitude do locutor; atos de fala diretivos.

1. Introdução

A súplica, o pedido e a ordem são atos diretivos que se materializam na comunicação através da forma sintática das sentenças imperativas. A distinção entre esses modos de realização linguística no Português do Brasil (PB) é feita principalmente através da entonação (Rizzo, 1981; Moraes, 1984; Bodolay, 2009; Colamarco, 2009; Queiroz, 2011). Apesar de a literatura concordar sobre o assunto, há certa carência de estudos que desenvolvam mais detalhadamente aspectos referentes à lógica ilocucionária. Por conseguinte, o objetivo do presente estudo é, por um lado, caracterizar padrões entonacionais do português brasileiro para os atos diretivos com modos de realização de pedido, súplica e ordem, mas igualmente privilegiar, com base na Teoria dos Atos de Fala (TAF), aspectos pragmáticos passíveis de envolvê-los, como uma alternativa que possa auxiliar nos estudos voltados para o papel da entonação na expressão de atitudes em atos de fala.

2. Entonação e força ilocucionária

De maneira geral, a entonação pode ser considerada como um dos mecanismos utilizados na distinção tipológica de atos de fala. Todo ato de fala pressupõe uma força ilocucionária, um conteúdo proposicional e suas condições de sucesso e satisfação subjacentes à lógica ilocucionária. No entanto, a força ilocucionária mostra-se como um elemento intimamente associado à interpretação do ato de fala, pois é a principal responsável por determinar o *modo de realização efetivo* do ato de fala (Vanderveken, 1990-91), deduzido com base no 'vigor' de sua força ilocucionária, que possui graus variáveis numa mesma dimensão do propósito ilocucionário (Searle, 1995). Alterando-se a força ilocucionária, altera-se necessariamente o modo de realização do ato de fala. Se, por um lado, a entonação é um dos elementos empregados na distinção de atos de fala, por outro, os critérios operacionais que alteram a força ilocucionária fornecem indícios de relações existentes entre configurações entonacionais específicas e modos de realizações específicos.

2.1 Operações que alteram a força ilocucionária

Pela lógica ilocucionária, as operações que alteram a força ilocucionária (Vanderveken, 1991) se resumem em seis e somente seis: (i) restrição o modo de realização do ponto ilocucionário, pela imposição de um modo de realização especial; ii) adicionar um novo conteúdo proposicional particular; iii) acrescentar novas condições preparatórias; iv) acrescentar novas condições de sinceridade; v) e vi) aumentar ou diminuir o grau de intensidade das condições de sinceridade. O *ponto ilocucionário* é o principal componente da força ilocucionária e determina a direção de ajuste, no caso dos diretivos: fazer o mundo corresponder às palavras; a cada diretivo é imposto um *modo de realização especial*, com características prosódicas distintas (padrão melódico, duração, amplitudes das variações). A *condição de conteúdo proposicional* da força ilocucionária é determinada pelo ponto ilocucionário, cujo propósito nos diretivos é sempre levar o alocutário a realizar uma ação futura; é basicamente analisável pela boa formação e consistência sintáticas. A *condição preparatória* dos diretivos consiste nas pressuposições que o locutor faz sobre a situação e seu interlocutor; o locutor pressupõe (ou toma como verdade) que o alocutário seja capaz de realizar a ação futura e que este possa recusar ou não satisfazê-la; as *condições preparatórias adicionais* estão relacionadas ao acréscimo de elementos adicionais que transcendem de alguma forma a característica autorreferencial do ato de fala diretivo (desejo do locutor), como transmitir um desejo e ao mesmo tempo comunicar que a ação futura a ser realizada será vantajosa (ou não). A *condição de sinceridade* dos diretivos é o *desejo* ou *vontade*; a força ilocucionária é modificada quando há *condições de sinceridade adicionais*, revelando uma atitude particular, como um *desejo* somado a uma *insatisfação*. As condições de sinceridade definem os modos de realização do ponto com diferentes forças ilocucionárias e diferentes

graus de intensidade, por exemplo, quem suplica expressa (estado psicológico expresso) com mais vigor o seu desejo do que quem pede. Em suma, essas seis operações lógicas colocam em xeque a força ilocucionária. Ao investigador cabe estabelecer relações existentes entre os tipos de diretivos e as especificidades entonacionais a fim de identificar o modo efetivo do ato de fala.

3. Métodos

3.1 *Corpus*

Dez atores profissionais do sexo masculino da região metropolitana de Belo Horizonte, Minas Gerais. O total de enunciados analisados é parte do *corpus* inicial de 900 atos diretivos, divididos em três grupos - das súplicas, dos pedidos e das ordens (Cf. Queiroz, 2011). A amostra quantitativamente analisada foi de 299 enunciados: 230 enunciados do *grupo dos pedidos*, divididos em três categorias (Cf. Teoria da polidez, Brown & Levinson, 1987): *pedido conciso* (136 enunciados), *pedido autoritário* (35 enunciados) e *pedido com polidez positiva* (59 enunciados); 36 enunciados do *grupo* das *súplicas*; 33 enunciados do *grupo das ordens*.

3.2 Coleta de dados

Para coleta dos dados, foram elaboradas dez sentenças *imperativas de base*, contendo de quatro a sete sílabas, para posteriormente serem proferidas como os atos de fala diretivos propostos. Aos informantes foram explicados os objetivos da pesquisa e o que se pretendia dos atores, de maneira simples e objetiva. A estratégia foi elaborar situações hipotéticas em que a súplica, o pedido e a ordem ocorressem com auxílio de esboços que generalizassem as relações entre o locutor (L) e o alocutário (A) nessas situações hipotéticas, como abaixo:
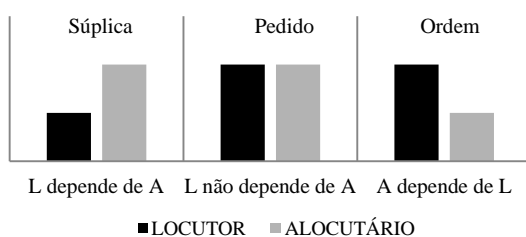
Figura 1: Relação Hierárquica Situacional

Quem suplica deseja muito algo e depende daquele a quem dirige seu desejo. Aquele que pede deseja algo e está numa situação de igualdade relativa com o alocutário. Aquele que ordena não depende do outro, ao contrário, está numa posição de autoridade. Para as gravações, foram automatizadas três apresentações distintas de *slide*s, contendo as sentenças a serem proferidas conforme o tipo de diretivo. Ressalta-se que não foi dado ou sugestionado qualquer padrão melódico que pudesse servir de modelo, a fim de que os informantes não mecanizassem padrões melódicos, ficando a critério do conhecimento internalizado dos atores informantes os padrões a serem

reproduzidos. Cada uma das dez sentenças de base foi produzida pelos informantes três vezes, em três etapas distintas, com duas etapas livres, sem nenhuma orientação, e uma etapa orientada por uma situação hipotética (Cf. Queiroz, 2011).

3.3 Caracterização pragmática

A atribuição de rótulos para os pedidos baseou-se na Teoria da Polidez (Brown & Levinson, 1987): i) pedido conciso (PdCon), e*stratégia de polidez aberta e direta (bald on-record);* ii) pedido com polidez positiva (PdPol+), e*stratégia de polidez aberta e indireta com polidez positiva (positive politeness)*; iii) pedido autoritário (PdAut), *estratégia com ações que ameaçam a imagem (face-threatening acts)*. A ordem foi considerada como prototípica, em razão da literatura na descrição da entonação de atos de fala do PB (e.g. Moraes, 2011; Bodolay, 2009; Colamarco, 2009), e do português europeu (Falé & Faria, 2007). O mesmo se dá no caso da súplica.

3.4 Caracterização melódica

As sentenças foram analisadas através do software *Praat* (Boersma & Weenink, 1992-2008). As configurações melódicas foram obtidas pela segmentação dos eventos locais (ou eventos-chave): *f0* inicial (*f0i*); pico de *f0* (*pf0* ou *pf0/ton1*, quando coincide com a 1ª sílaba tônica não nuclear); sílaba pretônica (*preT*), antecede imediatamente a nuclear; sílaba tônica proeminente ou nuclear (*TonP*).

4. Resultados

4.1 Configuração melódica dos diretivos

4.1.1. Pedido conciso
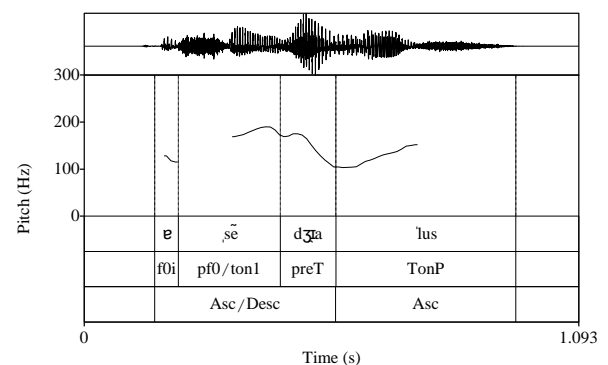
Figura 2: Padrão do pedido conciso

A configuração da curva de *f0* no enunciado "*Acende a luz*" do pedido conciso apresenta um movimento prenuclear (*f0i→preT*) ascendente/descendente, com o início em um nível relativo médio e pico de *f0* (*pf0*) sobre a primeira sílaba tônica. A configuração intrassilábica da sílaba nuclear (*Tonp*) descreve um movimento ascendente, exibindo um alinhamento "*tardio*" (H*>), localizado na porção final da sílaba [ᶚlus].

4.1.2.   Pedido com polidez positiva
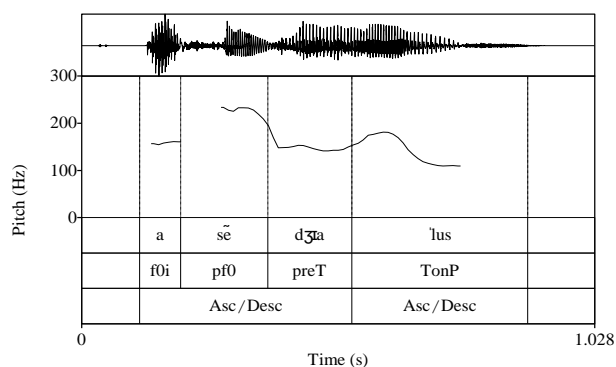
Figura 3: Padrão do pedido com polidez positiva

No pedido com polidez positiva, o contorno prenuclear *f0i→preT* é ascendente/descendente, com o início de *f0* em um nível relativo médio e pico de *f0* localizado sobre a primeira sílaba tônica do enunciado. A configuração intrassilábica da proeminente (*Tonp*) descreve também um movimento ascendente/descendente, com o pico alinhado à porção mais inicial da vogal da sílaba proeminente (alinhamento adiantado).

4.1.3.   Pedido autoritário
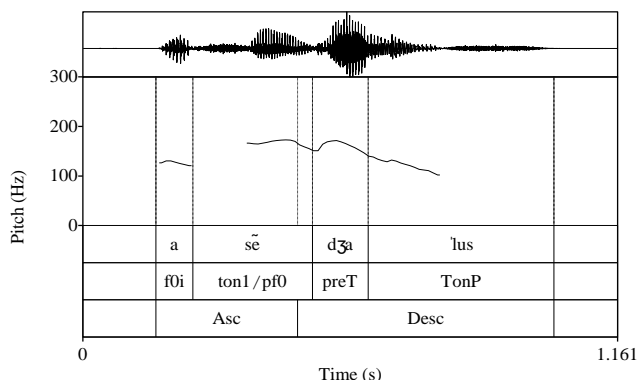
Figura 34: Padrão do pedido autoritário

O padrão melódico global do pedido autoritário é ascendente/descendente. A *f0* inicial está situada num nível relativo médio e a curva melódica descreve um movimento ascendente até atingir o pico de *f0*, localizado no final da vogal da primeira sílaba tônica (*ton1*). Após o ponto mais alto da curva de *f0*, a melodia descreve uma suave descida até o final do enunciado, nível relativo mais baixo de *f0*, com um padrão intrassilábico descendente por toda extensão da sílaba tônica final [ˈlus].

4.1.4.   Súplica
A configuração melódica global da súplica do enunciado *"Acende a luz"* é similar à configuração do pedido com polidez positiva. No entanto, o início (*f0i*) situa-se em um nível relativo significativamente (Cf. Queiroz, 2011) mais elevado (médio-alto). A nuclear descreve o mesmo padrão, mas o pico tende a se alinhar sobre a porção

central da sílaba. A duração da sílaba nuclear [ˈlus] é significativamente mais elevada na súplica.
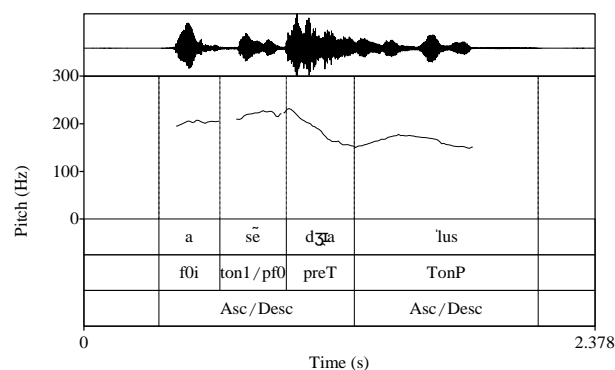
Figura 5: Padrão da súplica

4.1.5.   Ordem
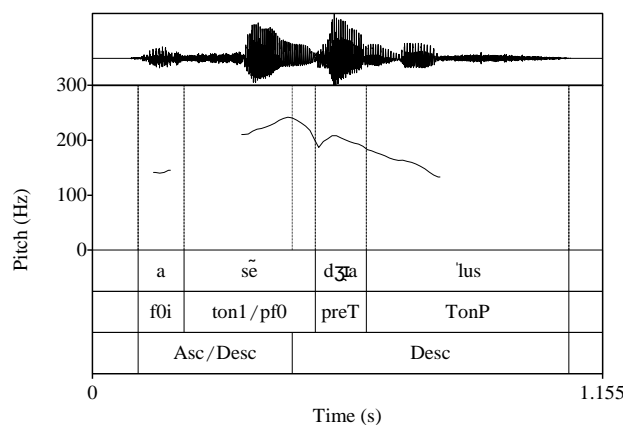
Figura 6: Padrão da ordem

Para a ordem, a configuração global é similar ao padrão encontrado para o pedido autoritário. No entanto, a ordem se distingue do pedido autoritário principalmente no que se refere às amplitudes das variações locais de *f0* (e.g. ataque e sílaba tônica proeminente), bem como pelo nível do registro, mais elevado na ordem do que no pedido autoritário. A ordem apresenta sobre o evento nuclear variação melódica significativamente mais alta, com uma taxa de elocução mais elevada e queda mais abrupta de *f0* do que no pedido autoritário, que apresenta uma melodia que declina suavemente (Cf. Queiroz, 2011).

4.2   Interpretação pragmática
A interpretação pragmática considera a relação entre as características dos diretivos e os seis critérios operacionais que alteram a força ilocucionária.

A Tabela 1, na página seguinte, caracteriza os atos diretivos conforme as operações que modificam a força ilocucionária e estão sintetizadas como a seguir.

Ato Operação	PdCon	PdPol(+)	PdAut	Súplica	Ordem
i) Restrição do modo de realização	Sim	Sim	Sim	Sim	Sim
ii) Conteúdo proposicional adicional	Não	Não	Não	Não	Não
iii) Condição preparatória adicional	Sim/Não	Não	Não	Sim	Sim
iv) Condição de sinceridade adicional	Não	Não	Sim	Sim	Sim
v) Grau de intensidade (condições de sinceridade)	Sim/Não	Não	Não	Sim	Sim

Tabela 1: Síntese da caracterização pragmática

i) *Restrição do Modo de Realização* - A cada um dos pedidos é imposto um modo de realização especial, cinco modos de realização do ponto ilocucionário; cinco maneiras distintas, com características entonacionais também distintas: o padrão melódico, a duração e as amplitudes das variações nos movimentos da curva de *f0* são pistas importantes para definição do modo efetivo de realização de cada um dos diretivos. Nos pedidos, as diferenças nas configurações melódicas são claras, sobretudo no que se refere às configurações intrassilábicas sobre a sílaba nuclear (TonP). No caso do pedido com polidez positiva e da súplica, embora possuam semelhanças configuracionais, distinguem-se quanto ao nível do ataque e quanto a duração sobre a sílaba nuclear. De modo análogo, o pedido autoritário e a ordem possuem semelhanças melódicas, mas o registro é mais elevado na ordem por toda extensão dos enunciados analisados; o movimento melódico da nuclear é descendente em ambos, mas a ordem apresenta sobre o evento maior variação de *f0,* taxa de elocução mais elevada e queda mais abrupta do que no pedido.

ii) *Conteúdo Proposicional Adicional* - A condição do conteúdo proposicional é determinada pelo ponto ilocucionário: toda força ilocucionária de ponto diretivo tem como condição que o conteúdo proposicional represente o desejo do locutor de uma ação futura do alocutário. Todos os cinco tipos possuem o mesmo conteúdo proposicional, sem acrescentar nenhum conteúdo proposicional novo.

iii) *Condição Preparatória Adicional* - A condição preparatória da força ilocucionária consiste nas pressuposições que o locutor faz sobre a situação e seu interlocutor. Em todos os casos o locutor pressupõe (ou toma como verdade) que o alocutário seja capaz de realizar a ação futura e que o alocutário pode recusar ou não a realizá-la. No caso do pedido com polidez positiva e o pedido autoritário não há indícios fortes de condições preparatórias adicionais, embora não seja esta uma interpretação estanque. Já o pedido conciso possibilita outra interpretação: o locutor toma como certo que o alocutário seja capaz de realizar a ação futura, mas

adiciona a condição preparatória de que esta ação será benéfica ou favorável ao alocutário, ou pelo menos para si mesmo, pois o pedido conciso pode ser interpretado, dependendo do contexto, como uma "sugestão". Trata-se de um caso representativo em que a relação entre *forma* e *função* não se estabelece de maneira exclusiva, pois o padrão melódico do exemplo do enunciado "*Acende a luz*" (Figura 2, item 4.1.1) caberia confortavelmente em situações nas quais o locutor adicionasse a condição preparatória adicional de que fosse melhor que a luz estivesse acesa. Nessas situações hipotéticas, a interpretação do enunciado com o padrão do pedido conciso seria preferencialmente algo do tipo "*Acende a luz... é melhor*". Aliás, o padrão melódico funcionaria de maneira semelhante para situações que requeiram interpretações similares: "*Fecha a porta... é melhor*", "*Vai tomar banho... é melhor*", etc. Tanto na súplica quanto na ordem há a adição de condições preparatórias. No caso da súplica, a força ilocucionária é alterada pela condição preparatória adicional de a ação futura ser favorável, pelo menos, e mais geralmente, para o locutor, que toma como certo que está numa situação de dependência, em termos de relação de forças com o alocutário (item 3.2). Na ordem, o locutor toma como verdade que o alocutário pode recusar ou não a obedecê-la, mas adiciona a condição preparatória de ser ruim para o alocutário, caso não a obedeça, pois a relação de forças é desfavorável ao alocutário.

iv) *Condição de Sinceridade Adicional* - Pelas condições de sinceridade o locutor expressa (ou manifesta) os estados mentais intencionais do locutor, os quais são dirigidos para, ou acerca de objetos e estados de coisas no mundo (SEARLE, 1995) e revelam certos estados psicológicos do locutor. No caso do pedido com polidez positiva e do pedido conciso não há presença clara de índices que os caracterizem como tendo sido modificados pelo acréscimo de condições de sinceridade adicionais. Nos dois tipos, o locutor manifesta abertamente sua intenção, mas os dois tipos sejam diferentes à luz da teoria da polidez (BROWN & LEVINSON, 1978). O pedido conciso é uma estratégia de polidez aberta e direta. O locutor mostra claramente sua intenção, envolve fazê-lo do modo mais direto possível e não há intuito de neutralizar um dano potencial ou conflito que ponha em perigo a própria imagem ou a do alocutário (*face-threatening acts*). Trata-se, portanto, de um ato passível de ser ameaçador à face do (e.g. um pedido conciso a uma pessoa que mal se conhece, desvaloriza a face do alocutário, criando um dano potencial, mas também a face positiva do locutor, que pode ser visto como uma pessoa grosseira). No pedido com polidez positiva, o locutor mostra abertamente sua intenção, no entanto, a estratégia de polidez positiva é orientada em direção à face positiva do alocutário. Assim, uso do tipo com polidez positiva a uma pessoa que mal se conhece, ao contrário do pedido conciso, é uma estratégia que valoriza a face do alocutário, pois o locutor demonstra uma atitude mais cortês, mais consideração pela face do interlocutor do que no caso do pedido conciso. Já no caso do pedido

autoritário, diferentemente dos outros dois tipos de pedidos, a condição de sinceridade adicional ocorre porque o locutor expressa sua vontade, mas adiciona a condição de não estar satisfeito acerca do estado de coisas. O locutor não tem intenção de neutralizar um dano potencial ou conflito, como no pedido conciso, mas difere deste por valorizar face positiva do locutor, ao passo que desvaloriza a face positiva do alocutário (e*stratégia com ações que ameaça a imagem*), o que pode ser interpretado socialmente como um ato de fala ríspido, grosseiro, autoritário etc. Além da impolidez, o pedido autoritário pode indicar impaciência, irritação ou humor momentâneo do locutor. Na súplica, o locutor expressa intencionalmente um estado mental dirigido ao estado de coisas que desejaria que estivessem de outra maneira. O locutor valoriza a imagem do outro, de quem sabe hierarquicamente depender (condições preparatórias adicionais), e, por isso, pode sugerir atitudes de submissão ou auto-humilhação, expressando um estado psicológico que desvaloriza a própria imagem, ao mesmo tempo em que busca valorizar a imagem daquele de quem depende, embora possam indicar outros atributos psicológicos, como impaciência, irritação, insistência etc. No caso da ordem, o locutor expressa seu desejo, adicionando a condição de sinceridade de não estar satisfeito com o estado de coisas. O modo psicológico é expresso com uma atitude autoritária, de modo a *impor* o seu desejo, valendo-se de uma condição de sinceridade adicional.

v) *Grau de Intensidade das Condições de Sinceridade* - Os estados mentais são expressos com diferentes graus de intensidade (*degree of strength*), dependendo da força ilocucionária. Para que o ato de fala seja perfeito em quantidade (Grice, 1975), o locutor deve expressar sua posição de modo que o ato não seja nem mais nem menos intenso em relação ao seu propósito. Para os pedidos, o grau de intensidade das condições de sinceridade da força ilocucionária é relativamente o mesmo. O desejo que o alocutário faça a ação futura não sinaliza fortemente para graus mais ou menos intensos de desejo, embora, dependendo do contexto, seja possível estabelecer alguma diferenciação, como no caso do pedido conciso, interpretado como sugestão (e.g. "*Acende a luz... é melhor*"), visto que a força ilocucionária da *sugestão* é derivada da força primitiva diretiva, diminuindo-se o grau de intensidade, pois sugerir é uma tentativa mais branda para que o alocutário faça a ação futura, do que pedir, suplicar ou ordenar. O grau de intensidade das condições de sinceridade da súplica, por sua vez, é mais intenso do que quem pede, porque quem suplica expressa um desejo mais intenso do que quem pede ou sugestiona. A ordem apresenta também características que indicam que o desejo do locutor seja mais intenso do que nos pedidos, o grau de intensidade das condições de sinceridade é geralmente expresso através da entonação, logo, an increase in the degree of strength of the intonation contour serves in general to increase the degree of strength of the sincerity conditions" (Vanderveken, 1991: 119).

5. Conclusão

A entonação fornece informações importantes para definição do modo de realização efetivo dos diretivos analisados, evidenciando que o pedido, a súplica e a ordem não são categorias estanques, como exemplifica o caso do ato de pedir, que pode ser feito, pelo menos no dialeto mineiro, de duas maneiras diferentes, dois modos de realização, mas com a mesma intenção comunicativa, apesar de possuírem forças ilocucionária diferentes. Aliás, o pedido conciso, sua regularidade (136 ocorrências no total de 300) e sua força ilocucionária revelam modos de organização social, visto que, dependendo do contexto, não seria apropriado dirigi-lo, numa situação formal, a alguém que mal se conhece ou acabou de se conhecer, pois o modo pelo qual as coisas são socialmente organizadas exige, a sua maneira, outro comportamento entonacional. Os resultados demonstram que alguns padrões melódicos são mais difíceis de serem relacionados à expressão de atitudes, como nos caso do pedido conciso e do pedido com polidez positiva, considerados padrões entonacionais mais "neutros", em comparação com os demais diretivos analisados. Os demais tipos são melodicamente marcados, como a súplica, cujo padrão entonacional pode ser relacionado à atitude de *submissão* ou *auto-humilhação*, e a possibilidade da sobreposição de atitudes como *insistência* e *submissão*. Como o pedido autoritário, pelo qual o locutor expressa sua *insatisfação* ou *descontentamento* com o estado de coisas, podendo ser associado às atitudes *impaciência*, *irritação*, à *impolidez* e mesmo outros estados afetivos, como, o humor momentâneo do locutor. Ou ainda pela ordem, em que o locutor impõe sua vontade de maneira autoritária. Enfim, nos casos em que a entonação não contribui como um forte índice das atitudes, estas podem ser interpretadas com base em fatores internos e externos ao sistema linguístico, que incluem aspectos sintáticos, semânticos e pragmáticos, bem como as noções de estado psicológico expresso, de conteúdo proposicional, adição de condições preparatórias e de sinceridade adicionais.

6. Referências

Brown, P., Levinson, S. (1987). *Politeness: Some Universals in Language usage*. Cambridge University Press, Cambridge.

Bodolay, A.N. (2009). Pragmática da entonação: a relação prosódia/contexto em atos diretivos no Português. Tese de Doutorado. Faculdade de Letras, Universidade Federal de Minas Gerais.

Colamarco, M. (2009). A expressão das emoções em atos de fala no português do Brasil: produção e percepção. Dissertação de Mestrado. UFRJ. Rio de Janeiro.

Falé, I., Faria, I.H. (2007). Imperatives, orders and request in European Portuguese. In *Proceedings XVI*[th] *ICPhs*. Saarbrücken.. August. pp. 1041--1044.

Grice, P.H. (1975). Logic and conversation. In Cole, Morgan (Eds.), *Syntax and Semantics*. pp. 41--58.

Moraes, J.A. (1984). *Recherches sur l'intonation modale du portugais brésilien parlé à Rio de Janeiro.* Thèse de Doctorat. Université de la Sorbonne Nouvelle, Paris III.

Moraes, J.A. (2011). A entoação de atos de fala diretivos no PB. In *Resumo III Colóquio Brasileiro de Prosódia da Fala*. FALE-UFMG, Belo Horizonte.

Queiroz, H.S. (2011). A contribuição da prosódia e da qualidade de voz na expressão de atitudes do locutor em atos de fala diretivos. Tese de Doutorado. Faculdade de Letras, Universidade Federal de Minas Gerais.

Rizzo, J. (1981). O Papel da entoação do português brasileiro na descrição de atos de fala. Dissertação de Mestrado. IEL. Unicamp.

Searle, J.R. (1995). *Expressão e significado: estudos da teoria dos atos de fala.* São Paulo: Martins Fontes.

Vanderveken, D. (1991). *Meaning and speech act; .principles of language*. Cambridge University Press.

Facial gestures in the expression of prosodic attitudes of Brazilian Portuguese

João Antônio de MORAES[1,2]**, Luma da Silva MIRANDA**[1]**, Albert RILLIARD**[3]

[1]Universidade Federal do Rio de Janeiro (UFRJ), Brazil; [2]CNPp; [3]LIMSI-CNRS, Orsay, France

jamoares3@gmail.com, lumah.miranda@gmail.com, albert.rilliard@limsi.fr

Abstract

This paper presents the results of (i) an identification test of Brazilian Portuguese prosodic attitudes based on visual cues and (ii) a preliminary analysis of the facial gestures involved in its expression. Eleven attitudes, separated between social and propositional categories, performed by two native Brazilian speakers, were audio-visually recorded and analyzed in terms of Ekman's Action Units,in order to correlate the speaker's intention and the objective manifestations of facial expressions. Results show the importance of these gestures for the recognition of attitudes as well as the consistency between the two subjects in their use of facial gestures.

Keywords: visual prosody; attitudes; facial gestures.

1. Introduction

In the last decade, several studies in visual prosody have been undertaken to explore in a given language how audio and visual features combine to express the so called intonational meaning, either properly linguistic (Kendon, 2004; Wollermann & Schröder, 2009; Wollermann *et al*., 2012) or attitudinal (Rilliard *et al*., 2009; Tanaka *et al*., 2010). It seems rather obvious that in face-to-face interactions, attitudes are expressed and perceived within a multimodal paradigm, integrating audio and visual elements (Barkhuysen *et al*., 2007). However, while several studies on prosodic attitudes have been carried out, most of them have analyzed the acoustic modality only. The multimodal approach has yet to be fully explored. The audiovisual expression of attitudinal meanings is, in a large extent, conventionally encoded within a particular culture and a particular language. They are learned by the speaker and are produced during face-to-face communication, which implies that the manifestation of these attitudes may be ambiguous or even not recognized by foreign speakers.

The importance of facial gestures for the recognition of prosodic attitudes in Brazilian Portuguese (BP) was shown in a previous study (Moraes *et al*., 2010). In this study, after presenting the main results of these identification tests, we will focus on the description of the gestures involved.

2. Method

2.1 Corpus

A semantically neutral declarative sentence: "*Roberta dançava.*" ("*Roberta was dancing.*") was produced by two BP speakers with eleven different attitudes. These attitudes were grouped in two categories: (i) propositional attitudes, that refer to speaker's attitudes towards the propositional content of the sentence and (ii) social attitudes, which represent the speaker's attitudes towards its interlocutor. Five propositional attitudes were performed: doubt (DOU), irony (IRO), incredulity (INC), obviousness (OBV) and surprise (SUR); and six social attitudes: arrogance (ARR), authority (AUT), contempt (CON), irritation (IRR), politeness (POL) and seduction

(SED). A "neutral" attitude was also produced, characterized by the absence of any special affect. Each of the 12 attitudes was performed in assertive mode.

2.2 Perceptual validation

The stimuli have been presented in three modalities (audio-only, visual-only, audio-visual) to 29 native BP listeners who had to recognize, in a forced-choice paradigm, the performed attitudes, among the possible attitudes in a given category, propositional or social. Each attitudinal label was completed by a longer description, in order to ease its identification by the listeners.

Each stimulus was played/showed twice on each run. Subjects had to give their answers by selecting on a slider the relative intensity of the perceived attitude. The scale ranged from "barely marked attitude" to "very marked attitude".

2.3 Description of facial gestures

To describe the facial movements present in the expressions of attitudes a simplified version of the Facial Action Coding System (FACS) proposed by Ekman and colleagues (2002) was adopted. An Action Unit (AU) is defined as a muscular activity that produces momentary changes in the facial features in various areas of the speakers' face. The facial topography is divided into two principal areas. The first area is the upper face which affects the eyebrows, forehead, and eyelids; the second area is the lower face, which includes movements such as up/down, horizontal and oblique motions of the head, the shoulder and/ or the jaw. Using Ekman's system of facial mapping, the following 15 Action Units were selected for our analysis:

(a) Eyebrow raiser (Inner + Outer brow raiser) AU 1+2
(b) Eyebrow lowerer AU 4
(c) Lid tightener AU 7
(d) Upper lid raiser   AU5
(e) Blink AU 45
(f) Lip corner depressor AU 15
(g) Lip corner puller AU 12
(h) Upper lip raiser AU 10
(i) Jaw drop AU 26

(j) Cheek raiser AU 6
(k) Up and down head movement AU 85
(l) Right and left head movement AU 51+52
(m) One side tilt movement AU 55/56
(n) Head up AU 53
(o) Shoulder shrug AU 82

Three researchers separately analyzed each video, marking the emergence of AU's related to the upper face, lower face and head positions based on appearance changes according to the FACS Manual (Ekman *et al.*, 2002), and reached a consensus in case of disagreement; the intensity of the appearance change was not scored.

3.   Results

3.1   Identification test

The results of perceptual recognition tests (Moraes *et al.*, 2010, 2011) indicated that the overall recognition rate increases when both audio and visual channels were combined, but when we have access to only one channel, the visual one is generally more effective for the recognition of the speaker's attitude than the audio channel. There is, nevertheless, a significant difference: while for propositional attitudes the performance of each channel separately is relatively close, for social attitudes the difference is striking (Figure 1).
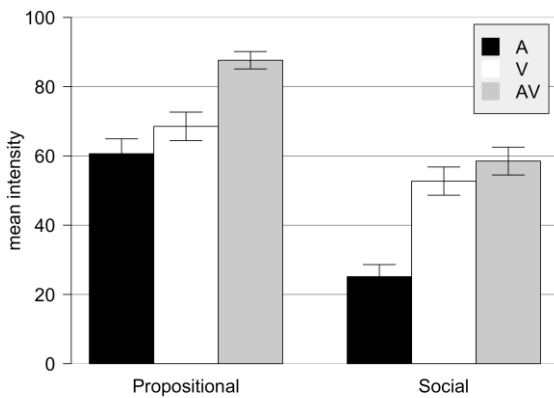
Figure 1: Mean intensity rating in each modality, forboth types of attitudes

By looking at the identification of each propositional attitude (Figure 2), it is clear that for most attitudes the visual information prevailed (although the observed difference was not very pronounced); with the exception of incredulity, in which audio information was dominant, each of these channels was itself actually very effective, showing a recognition rate of the speaker's intention far above the simple chance.

For social attitudes (Figure 3), however, the presence of visual information is crucial. In some attitudes such as arrogance, contempt and authority the audio information is poorly recognized, near the chance level. This is probably due to the fact that among social attitudes there are not prosodic patterns clearly distinct, as

occurs with propositional attitudes (Moraes *et al.*, 2011).
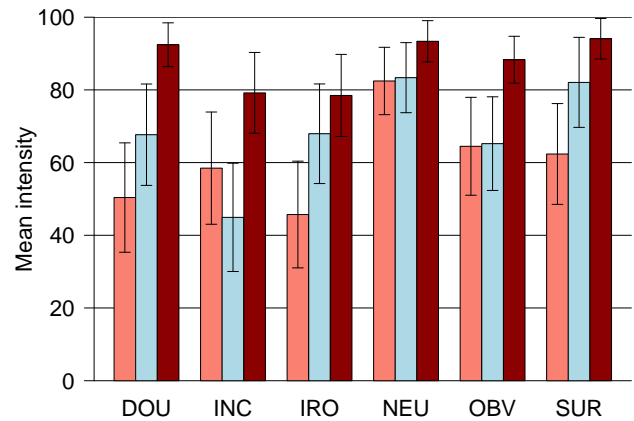
Figure 2: Mean intensity rating for the identification of propositional attitudes. Results for audio-only in pink (1st column), visual-only in blue (2nd col.) and audio-visual in brown (3rd col.), both speakers

percep. produc.	DOU	OBV	INC	IRO	NEU	SUR
doubt	**44**	4	8	1	0	1
obviousness	2	**47**	2	3	4	0
incredulity	10	1	**33**	10	1	3
irony	1	1	10	**46**	0	0
neutral	0	3	0	0	**55**	0
surprise	1	0	7	0	1	**49**

Table 1: Confusion matrix of visual stimuli in propositional attitudes (both speakers)

If we examine the confusion matrices concerning these attitudes, it can be seen that there were few confusions between the production and the perception, indicating that the gestures were sufficiently distinct in general. Among propositional attitudes (Table 1) confusions based on visual information are rather rare: they occur basically between incredulity *vs.* doubt and between incredulity *vs.* irony, in both directions, what can be explained by the fact that these attitudes are semantically close. Interestingly, the visual recognition of incredulity was offset by the audio channel, which receceived better scores.

Social attitudes were also generally well recognized visually, although in a somewhat less effective way, with confusions for arrogance, interpreted as contempt (quite similar attitudes) and politeness, interpreted as neutral.
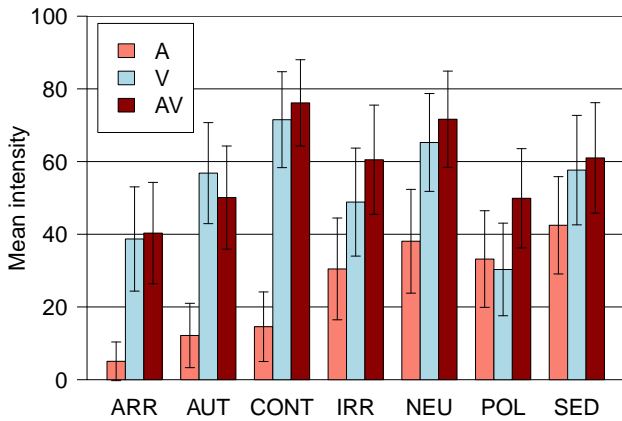
Figure 3: Identification of social attitudes. Results for audio-only in pink (1st column), visual-only in blue (2nd col.) and audio-visual in brown (3rd col.), both speakers

percep. produc.	AR	AU	SED	CON	IRR	NEU	POL
arrogance	29	3	0	25	1	0	0
authority	8	43	0	1	5	0	1
seduction	0	0	40	0	1	2	15
contempt	9	1	0	48	0	0	0
irritation	6	7	0	6	36	2	1
neutral	5	4	0	0	2	45	2
politeness	2	3	1	0	2	23	27

Table 2: Confusion matrix of visual stimuli in social attitudes (both speakers)

3.2 Facial gestures

The preliminary findings in this study disclosed discrete categories formed by the AU's for the facial expressions of each attitude performed; each attitude was distinguished from the others by the set of its AUs (table 3 in appendix). On average 3.8 AUs were employed in the expression of each attitude, with virtually no difference between the number of gestures present in propositional (3,9) and in social (3.7) attitudes. It is noteworthy that the male subject has used on average more AUs (4.4) than the female one (3.2). Although some attitudes were occasionally conveyed using different strategies between the subjects (irritation, for instance, has no AUs in common between the subjects), the overall similarity of the gestures employed by them is striking, which can be verified by a simple visual inspection of selected photos put side by side (Figures 4 to 15), which illustrate the attitudes expressed by each subject (for a closer and more effective examination, the attached video set can be seen).

3.3 Propositional attitudes

Figure 4: Doubt

Figure 5: Irony

Figure 6: Incredulity

Figure 7: Obviousness

Figure 8: Surprise

Figure 9: Neutral

3.4   Social attitudes

Figure 10: Arrogance

Figure 11: Authority

Figure 12: Contempt

Figure 13: Irritation

Figure 14: Politeness

Figure 15: Seduction

Some AUs, such as (a) AU 1+2 and (b) AU 4 (eyebrow movements) and the four different head movements here considered are rather frequent and very productive in discriminating attitude groups. Others, on the contrary, have an occasional, limited participation, such as (d) upper lid raiser, (i) jaw drop, (j) cheek raiser and (o) shoulder shrug, and are frequently associated to specific attitudes. Thus (d) AU 5 (upper lid raiser) is typical of surprise, and so is (i) AU 26 (jaw drop); (n) AU 53 (Head up) denotes arrogance, and (o) shoulder shrug correlates with obviousness (and also contempt, for the male subject). It is worth noting that (h) AU 10 (upper lip raiser) and in a large extent, (e) AU 45 (blink), are basically dedicated to the expression of social attitudes. Interestingly, the (f) lip corner depressor (AU 15), which appears in four different attitudes, was used only by the male subject, while the (h) upper lip raiser (AU 10) was used only by the female subject; their use in the attitudes of arrogance and contempt seems to suggest that they are individual (or may be gender) gestural variants in the expression of the same set of attitudes.

It can be observed, finally, that pairs of attitudes that are semantically close, such as arrogance/contempt, or politeness/seduction were expressed by a similar set of gestures: they were distinguished from each other by a small number of AUs.

On the other hand, semantically distant propositional attitudes, such as incredulity and obviousness, can be also visually quite similar, which did not prevent them from being clearly identified visually, probably due to the presence of the distinctive shoulder shrug in obviousness, and the difference in head orientation.

4.   Conclusions

The results of this study confirm that listeners rely upon the visual channel to better understand what attitudes a speaker is communicating in face-to-face speech, and the facial mapping here undertaken provides a preliminary framework for identifying and interpreting which facial features communicate which particular attitude.

Because of the limited size of this study, the results are not yet conclusive. Additional research with a greater number of native Brazilian Portuguese speakers will be required to confirm the accuracy of these findings and to address other Brazilian Portuguese prosodic attitudes.

5.   References

Barkhuysen, P., Krahmer, E. and Swerts, M. (2007). Cross-modal perception of emotional speech. In *16th International Congress of Phonetic Sciences*, Saarbrucken, Germany, pp. 2133--2136.

Ekman, P., Friesen, W. and Hager, J. (2002). *Facial Action Coding System*. Salt lake City, UT: The Manual on CD ROM., Research nexus division of Network Information Research Corporation.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Moraes J.A., Rilliard A., Mota B. and Shochi T. (2010).

Multimodal perception and production of attitudinal meaning in Brazilian Portuguese. In *Speech Prosody*, Chicago.

Moraes J.A., Rilliard A., Erickson D. and Shochi T. (2011). Perception of attitudinal meaning in interrogative sentences of Brazilian Portuguese. In *17th International Congress of Phonetic Sciences* Hong-Kong, China.

Rilliard, A., Shochi, T. Martin, J-C, Erickson, D. and Aubergé, V. (2009). Multimodal Indices to Japanese and French prosodically expressed social affects. In *Language and Speech*, 52 (2/3): pp. 223--243.

Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E. and de Gelder, B. (2010). I feel your voice: cultural differences in the multisensory perception of emotion. In *Psychological Science 21* (9): pp. 1259--1262.

Wollermann, C. and Schröder, B. (2009). Effects of Exhaustivity and Uncertainty on Audiovisual Focus Production. In *Proceedings of International Conference on Auditory-Visual Speech Processing*, Norwich, UK, pp. 145--150.

Wollermann, C., Schröder, B. and Schade, U. (2012). Pitch accent, eyebrows ad context: effects on the exhausty of answers? In *Proceedings of 6th International Conference Speech   Prosody 2012*, Shanghai.

6.   Appendix

AUs / Attitudes	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
doubt		XY	XY		Y					XY	XY	XY			
irony		Y	XY				XY				XY		XY		
incredulity	Y					Y	X						XY		
obviousness	Y					Y					XY		XY		XY
surprise	XY			XY					XY						
arrogance	XY		XY		X	Y		X				X	Y	XY	
authority		Y			Y						XY				
contempt	XY		X		Y	Y		X			Y		XY		Y
irritation		Y	Y					X		Y		X			
politeness	Y				XY						XY		Y		
seduction	Y				XY		XY				Y		XY		

Table 3: AUs in propositional (in red) and social (in blue) attitudes for female (X) and male (Y) speakers; the letters (a) to (o) correspond to the 15 AUs listed in 2.3

Acoustic analysis of a corpus of Brazilian Portuguese attitudes

João Antônio de MORAES[1], Albert RILLIARD[2], Donna ERICKSON[3], Takaaki SHOCHI[4]

[1]Federal University of Rio de Janeiro (UFRJ), Brazil; [2]LIMSI-CNRS, Orsay, France; [3]Showa Music University, Japan; [4]Bordeaux 3University, France

jamoares3@gmail.com, rilliard@limsi.fr, ericksondonna2000@gmail.com, shochi38@gmail.com

Abstract

This paper presents the prosodic analysis of a corpus of Brazilian Portuguese attitudes. Attitudes are separated between the social and propositional categories, and performed either with an assertive or an interrogative modality. Previous studies show the particular relevance of prosodic cues for propositional attitudes, while visual cues are more relevant for social ones. This paper shows that this greater relevance of prosody for propositional attitudes is also observed on the prosodic parameters' variations – and enhance particularly the clearly different and prototypical F0 contours that distinguished such expressions.

Keywords: prosodic attitudes; Brazilian Portuguese.

1. Introduction

The expression of a speaker's opinion, belief and knowledge to his interlocutor is partly performed through the use of prosodic attitudinal expressions (Wichmann, 2000). The use of such prosodic strategies constitutes an important part of the speaker's engagement in his speech (Daneš, 1994) and may contribute for an important part of the semantic content of utterances. For example, a sentence produced with an ironic tone of voice will certainly not carry the same meaning than a more its more neutrally performed counterpart. Such prosodic attitudes differ from emotional expressions in that they are voluntarily produced during the interaction, in a given social setting where the attitudes are conventionally encoded for a language and a culture, and may vary with them (Rilliard *et al*., 2009).

Typologies of attitudinal expressions vary with authors and their points of interest (e.g. Martins-Baltar, 1977; Gu *et al*., 2011). The present study is based on a separation between two categories of attitudes (already used by Martins-Baltar, 1997 and Fónagy *et al*.,1984): propositional and social attitudes. The propositional ones address the propositional content of the sentence (e.g. doubt, obviousness, irony), while social ones refer to the interpersonal relationship between the speaker and the receiver (e.g. politeness, irritation, arrogance). Wichmann (2000) proposes a similar distinction between what she calls propositional and behavioural categories of attitudes.

This study describes the prosodic analysis of a corpus of such attitudes in Brazilian Portuguese (BP). The attitudes have been perceptually validated in previous studies (Moraes *et al*., 2010, 2011), and the present paper will focus on the prosodic parameters relevant to such a perception. After describing the corpus of BP attitudes, the process of prosodic analysis is detailed, and the main results observed on the corpus are given.

2. Method

The set of attitudes used in this study is based on the distinction between propositional and social attitudes introduced above, with a supplementary distinction between the assertive or interrogative modes of the carrying sentences.

The attitudes described here are the following:

Assertive mode:
- *Social*: arrogance (ARR), authority (AUT), contempt (CONT), irritation (IRR), politeness (POL) and seduction (SED);
- *Propositional*: doubt (DOU), irony (IRO), incredulity (INC), obviousness (OBV) and surprise (SUR).

Interrogative mode:
- *Social*: arrogance (ARR), authority (AUT), contempt (CONT), irritation (IRR), politeness (POL) and seduction (SED);
- *Propositional*: confirmation (CONF), incredulity (INC), rhetoricity (RET) and surprise (SUR).

The labels and the number of attitudes vary according to the sentence's mode (11 attitudes for assertion, 10 for interrogation), as some attitudes are incompatible with some modes (e.g. obviousness with interrogation).

BP sentence	L.	Stress	Translation
Tá	1	oxytone	OK
Vaidançar	3	oxytone	(s)he is going to dance
Dançava	3	paroxytone	she/he danced
Roberta vai dançar	6	oxytone	Roberta is going to dance
Roberta dançava	6	paroxytone	Roberta danced

Table 1: Sentences used for the attitudes, with their length (L., in syllables), the position of the lexical stress on their last word and an English translation

All attitudes were performed by two native BP speakers (a female and a male), on a set of five sentences from 1- to 6-syllable long and with varying lexical stress position (cf. Table 1). The sentences don't have any particular meaning in relation to the attitudes nor the modes. Their performances were audio-visually recorded

using high quality equipment.

These attitudes (including the neutral assertive and interrogative sentences), performed on the 5 sentences, were recorded (in the audio and video modality) in three repetitions by the two speakers, resulting in 690 stimuli. The recordings were phonetically aligned by hand, using Praat (Boersma & Weenink, 2011).

3.   Perceptual validation

In order to assess the pertinence of the speakers' performances, one repetition of each attitude from the last sentence of Table 1 were chosen in order to perform perception tests, separately for the assertive and interrogative modes. These attitudes have been presented in three modalities (audio-only, visual-only, audio-visual) to native BP listeners who had to recognize the performed attitudes, among the possible attitudes in a given mode and category (propositional or social). The perception results are fully described in Moraes *et al.* (2010, 2011), and provide a validation of the pertinence of the above-described prosodic parameters. Figure 1 presents the mean recognition scores obtained by each attitude, in each three modality, for the two modes and for propositional or social attitudes.
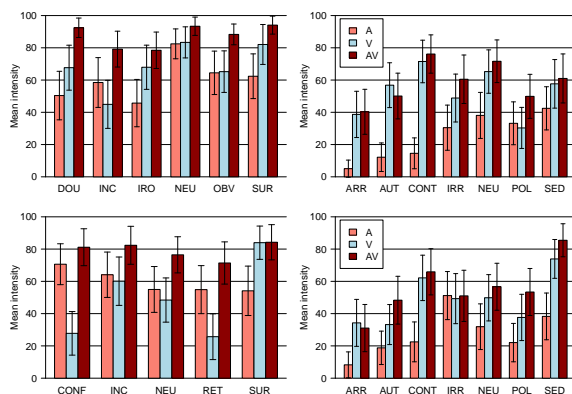
Figure 1: Identification of propositional assertive (top left), social assertive (top right), propositional interrogative (bottom left) and social interrogative attitudes (bottom right). First column (pink), only audio stimulus, second column (blue), visual, third column (brown), audio-visual

The most important result that was learned from these perception tests concerns the relative importance of visual and audio modality to the recognition of the two categories. While the visual cues clearly outperformed the audio cues for the recognition of social attitudes, it seems that audio cues are generally more important than the visual ones for the propositional attitudes (mostly for propositional interrogatives).

This primary use of audio cues for signalling information relating to the propositional content of utterances rather than information relating to the interpersonal relationship during a face-to-face interaction is interesting and led us to a complete analysis of the prosodic variation of this attitudinal corpus.

4.   Prosodic analysis

From each of the 690 stimuli, the following prosodic parameters were extracted: the fundamental frequency (F0, expressed in semitones), the intensity (in dB), and the phonemic duration expressed in z-score, following Campbell (1993) method. Both F0 and intensity where measured on each vowel, at three points (at 10, 50, 90% of the vowel's length).

4.1   Means values over sentences

As it has been claimed by e.g. Gu *et al.* (2011), the mean distribution of pitch over sentences already gives indication on the type of attitude: a high or low pitch – regarding to the speaker's mean laryngeal frequency, constitute a first kind of indices.

Category	Spk.	F0 (st)	Z-duration	Intensity
Propositional	F	93.4 (5.1)	0.39 (1.31)	68.9 (6.5)
Social	F	92.1 (3.4)	-0.25 (0.61)	68.8 (6.2)
Propositional	M	83.8 (6.3)	0.38 (1.26)	68.5 (4.9)
Social	M	81.6 (4.7)	-0.24 (0.68)	66.8

Table 2: Mean (standard deviation) of F0, Z-duration and intensity observed for each category of attitude, and for each speaker (Female & Male).

Figure 2 (in appendix) presents the distributions of F0 values for each attitude over all sentences. In each category of attitude, different patterns of distributions are observed: attitudes with high mean pitch and wide distribution (e.g. CONF and DOU), attitudes with a low and flat pitch (e.g. INC), etc. – supporting the above hypothesis.

A comparison of propositional and social attitudes shows a tendency to a wider distribution of the measured parameters in the case of the former, supporting the perceptual result: a higher importance of prosodic cue for these attitudes (cf. Table 2). This is mainly marked for F0 and Z-duration parameters.

4.2   Prototypical contours

The means and distributions of prosodic parameters can hardly distinguish between a complex set of attitudes. The evolution of these parameters across time and with respect to the carrying sentence's morphosyntactic structure shall also play a role. To assess such an importance of prosodic contours, they have also been inspected for all three prosodic parameters (cf. Figures 3 and 4 in appendix for the F0 contours in the interrogative mode).

Interestingly, the shapes of contours for propositional attitudes are characteristically different for each one, while the shapes of social attitudes tend to be more similar. For example, for the 1-syllable long sentence (first columns in Figures 3 and 4) propositional attitudes show a large diversity of contours (rising,

falling, flat-rising…), while the contours observed for social attitude are all rising – with small differences of pitch mean. The increase of sentences length shows the evolution of the global contours' shapes that tend to conserve a similar shape, whatever the length (under some constraints of minimal length). Such an observation is in line with Morlec *et al*. (2001) principle of "prosodic movement expansion" shown on French prosodic attitudes.

Visual inspection also shows the influence of the linguistic constraints of prosody on the global contours of attitudes: a main difference between Morlec *et al*. (2001) description and BP attitude is linked with the importance and varying position of lexical stress in BP. Whereas lexical stress in French always occurs at the final syllable, the described corpus proposes a systematic variation of oxytone and paroxytone words at the end of sentences. So, the two 3- and 6-syllable long sentences (respectively oxytone at the 2nd and 4th columns, and paroxytone at the 3rd and 5th columns of Figures 3 and 4) have a different morphosyntactic constraint that imposes a varying position of the main F0 peak and lengthening. This is especially clear for the CONF attitude (2nd line on the figures), where the final slope occurs on the stressed syllable – for both speakers. Such a phenomenon can also be seen for other attitudes. The other parts of the contours remain similar across sentences.

For the segmental duration, similar phenomena are observed. Figure 5 (in appendix) shows the large lengthening of the stressed syllables observed for IRON (5th row) that differ completely from the strategies used by this speaker to perform the others propositional attitudes. In a similar fashion, social attitudes' duration patterns tend to be more comparable across attitudes.

5.   Discussion & conclusion

This paper has presented a prosodic analysis of the variation induced by attitudinal expressions into the prosodic parameters of a set of BP sentences. These modifications affect the speaker's mean register, pitch range and rhythm. To rate the efficiency of mean prosodic patterns to convey attitudinal expressions would require perceptual tests based on a gating paradigm to check whether e.g. a high start followed by a slope at the beginning of a sentence will be systematically perceived as an expression of rhetoric question (cf. Shochi *et al*., 2009, for such an experiment on Japanese attitudes).

The modifications also affect the sentences' prosodic contours. Prototypical strategies have been observed for each propositional attitude, and are reproduced in a similar fashion over speakers for several attitudes – but not for all. The CONF attitude show a rise until the last stressed syllable for the female speaker, while the male speaker tend to produce a high plateau, but both make a steep slope on the stressed syllable. This shows that several communication strategies may coexist

in a same language, with common grounds. This variation may be accounted for by gender differences, but more investigation is required to confirm this hypothesis. Preceding perception results have also shown such inter-speakers differences, with higher performances obtained by either the female or the male speaker on several attitudes. To describe the possibility of strategic variations inside a given attitudinal expression would require a larger set of speakers to be recorded and analysed.

6.   Acknowledgements

7.   References

Boersma, P., Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.1.05) [Computer program].

Campbell, N. (1993). Automatic detection of prosodic boundaries in speech. In *Speech Communication*, 13, pp. 343--354.

Daneš, F. (1994). Involvement with language and in language, In *Journal of Pragmatics,* 22, pp. 251--264.

Gu W., Zhang T. and Fujisaki, H. (2011). Prosodic Analysis and Perception of Mandarin Utterances Conveying Attitudes. In *Proc. of Interspeech 2011*, Firenze, Italy, pp. 1069--1072.

Fónagy, I., Bérard, E. and Fónagy, J. (1984). Clichés mélodiques. *Folia Linguistica*, 17, pp. 153--185.

Martins-Baltar, M. (1977). *De l'énoncé à l'énonciation : une approche des fonctions intonatives*, Paris: Didier.

Moraes, J.A., Rilliard A., Mota, B. and Shochi, T. (2010). Multimodal perception and production of attitudinal meaning in Brazilian Portuguese. In *Proc. of Speech Prosody 2010*, Chicago.

Moraes, J.A., Rilliard A., Erickson, D. and Shochi, T. (2011). Perception of attitudinal meaning in interrogative sentences of Brazilian Portuguese. In *Proc. of ICPhS 2011*, Hong-Kong.

Morlec, Y., Bailly, G. and Aubergé, V. (2001). Generating prosodic attitudes in French: Data, model and evaluation. In *Speech Communication*, 33, pp. 357--371.

Rilliard, A., Shochi, T., Martin, J.C., Erickson, D. and Aubergé, V. (2009). Multimodal indices to Japanese and French prosodically expressed social affects. In *Language & Speech,* 52(2/3), pp. 223--243.

Shochi, T., Aubergé, V., Rilliard, A. (2009). Nihongo bogowasha wa hatsuwa no ichibu wo kiite taido wo chikaku dekiruka?. In *Bulletin de l'Enseignement du Japonais en France*, 4, pp. 65--75.

Wichmann, A. (2000). The attitudinal effects of prosody, and how they relate to emotion. In *Proceedings of the ISCA Workshop on Speech and Emotion*, Newcastle, North Ireland, pp. 143--148.

8.   Appendix
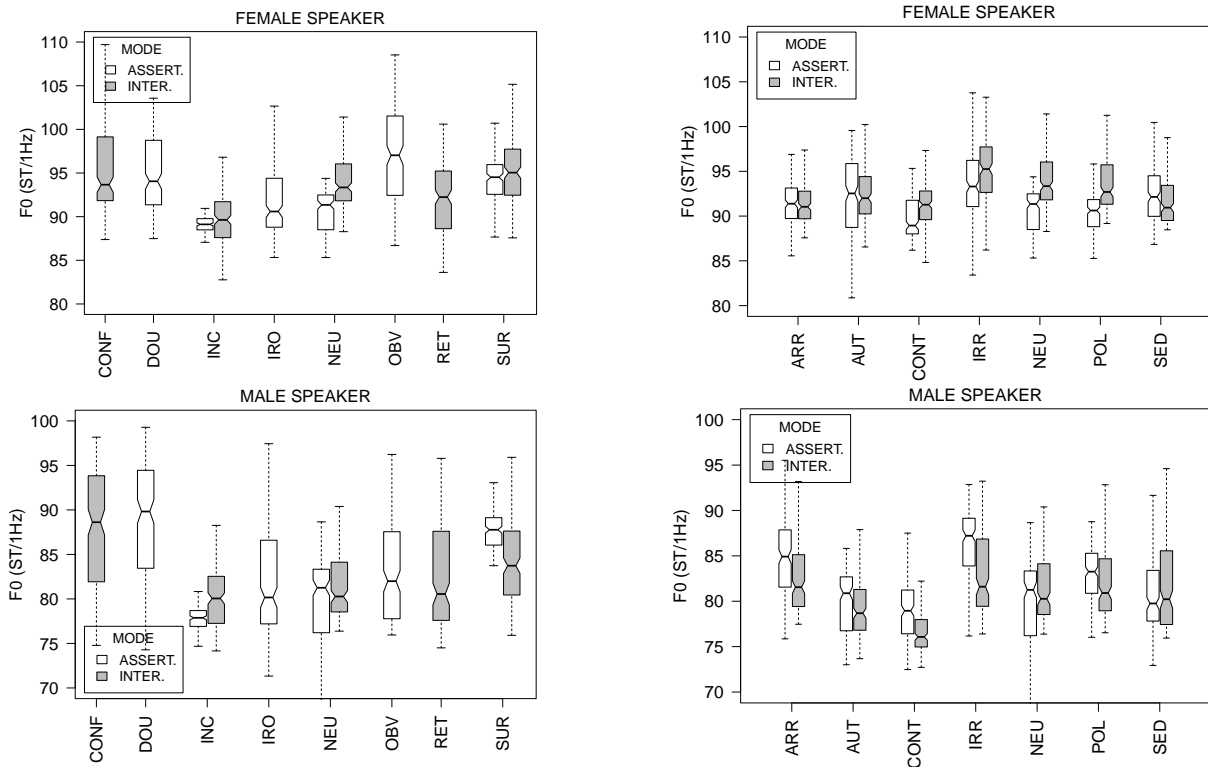
Figure 2: Dispersion of the F0 values measured for the female and male speakers on the propositional (left) and the social (right) attitudes, for both assertive and interrogative modes
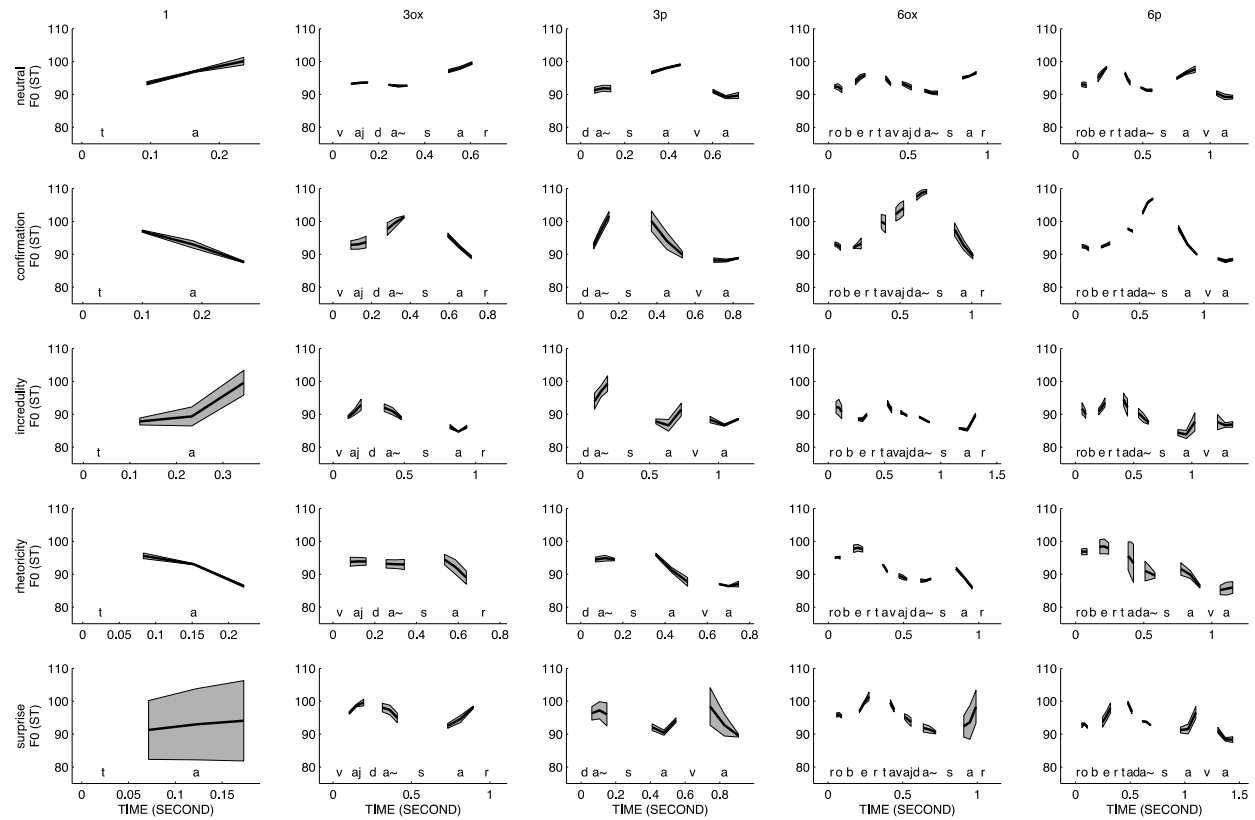
Figure 3: F0 contours (mean of 3 repetitions in black, standard deviation in gray) for the 5 interrogative sentences (in columns) with the 4 propositional attitudesplus the neutral interrogation (first row), as performed by the femalespeaker.
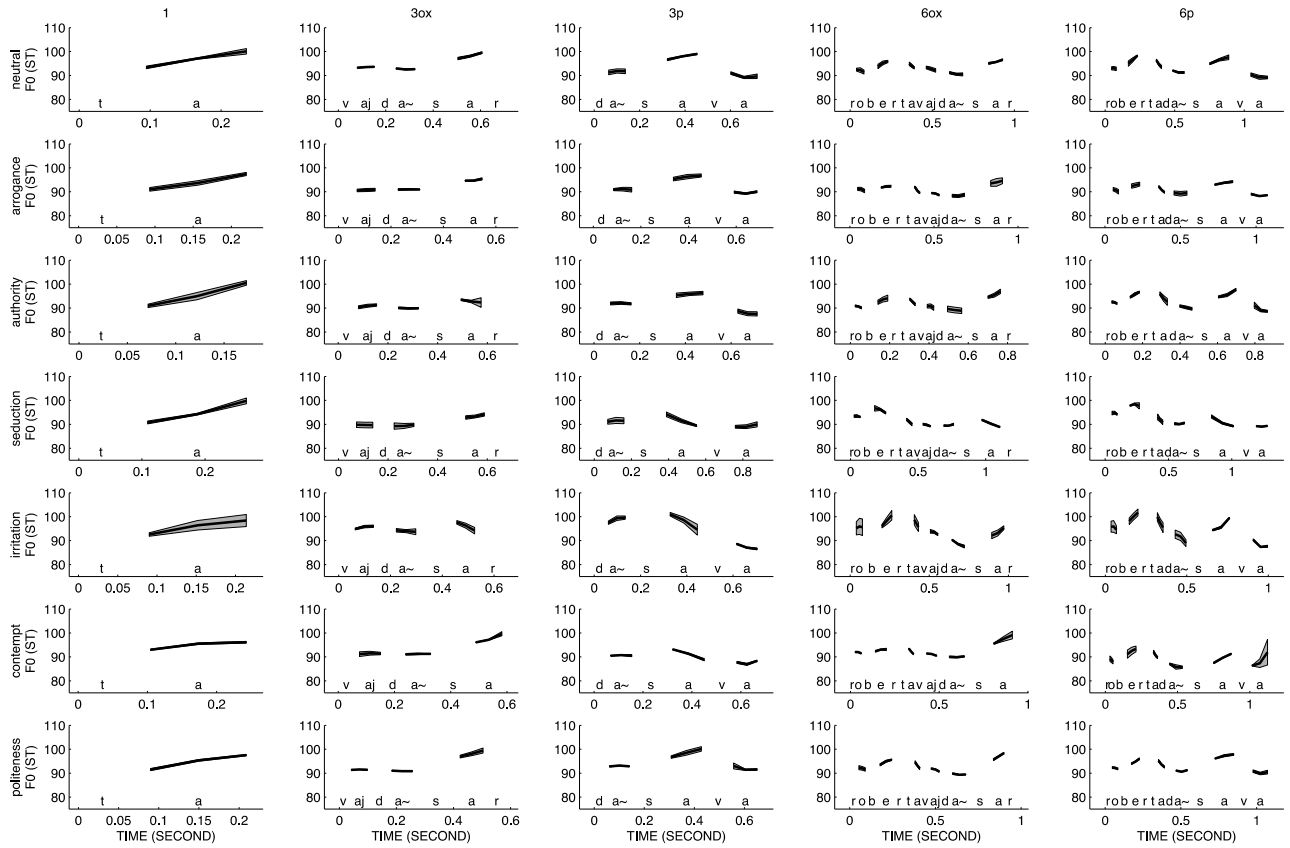
Figure 4: F0 contours (mean of 3 repetitions in black, standard deviation in gray) for the 5 interrogative sentences (in columns) with the 6 social attitudes, plus the neutral interrogation (first row), as performed by the female speaker.
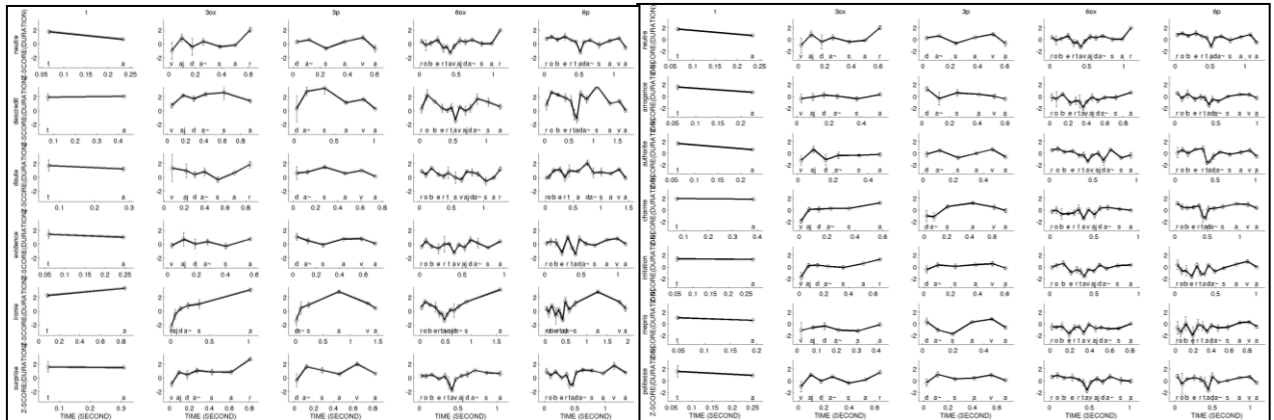
Figure 5: Z-duration contours (mean of 3 repetitions, with standard deviation) for the 5 assertive sentences (in columns) with the 5 propositional attitudes (left, in rows) plus the neutral declaration (first row), and for the 6 social attitudes (right panel, in rows) plus the neutral declaration (first row), as performed by the male speaker.

Developmental perception of polite & impolite non-verbal behaviours in Japanese

Albert RILLIARD[1], **Takaaki SHOCHI**[2], **Donna ERICKSON**[3], **João Antônio de MORAES**[4]

[1]LIMSI-CNRS, Orsay, France; [2]University Bordeaux 3, France; [3]Showa Music University, Japan; [4]Federal University of Rio de Janeiro (UFRJ), Brazil

albert.rilliard@limsi.fr, Takaaki.Shochi@u-bordeaux3.fr, ericksondonna2000@gmail.com, jamoares3@gmail.com

Abstract

This paper uses a corpus containing a set of prosodic attitudes, encoded by Japanese culture and language to express politeness and impoliteness. Three expressions of politeness are used: courtesy-politeness, sincerity-politeness and kyoshuku. A neutral declarative expression and an impolite attitude of arrogance complete the set of expressions. The question addressed here is twofold: first can young Japanese children perceive the expressive differences conveyed by these 5 attitudes in a way similar to that of native adults? And second, can a pair comparison paradigm be used with young (6 to 10 years) children still unfamiliar with the written language? Results show the progression with age of children's perceptual spaces towards adults' perception. The perceptions of audio and visual modalities are also compared.

Keywords: prosodic attitudes; Japanese; (im)politeness; perceptual development.

1. Introduction

During face-to-face interactions, speakers are *involved* in their speech (cf. Daneš, 1994: 253). They convey their message through their lexical and syntactic choices, as well as through gestures, facial expressions and prosodic variations. Emotional expressions may be seen as the most typical example of involvement in speech – as they are always part of an utterance. Such a continuous variation of emotional phenomenon is described by Russell & Barrett (1999) as "core affects" – elementary affective feelings always present and continuously varying. They separate such affects from "prototypical emotional episodes", which are the rare instances of full-blown basic emotions, conceptualized in language through lexical items. Such prototypical emotional episodes correspond to so-called "emotions". Their descriptions and the labels naming them may be refined hierarchically up to fine conceptual differences (cf. Golan et al., 2006 for such a list of labels). As any conceptually constructed object, such emotions may be described by scripts, able to capture subtle differences and similarities across cultures (Wierzbicka, 1986; Russell, 1991).

Widden & Russell (2003) describe the acquisition process leading to a diversification of the use of such emotional labels by children of 2 up to 5 years old: children master on average 1 emotional label at the age of 2, and 6 at the age of 5. In a similar fashion, the classification of emotions proposed by Zinck & Newen (2008) postulates an increase of affect types' complexity with the cognitive development of children, and their different physiological stages. The most complex affects of their classification are coined "secondary cognitive emotions" (such as shame or pride); they are related to cultural norms, and demand an experience of social relationships. Such kinds of affects are strongly linked to the culture and the language in which there are conceptualized. Such "social affects", as well as other kinds of expressive behaviour such as irony or politeness, are described by Wichmann (2000) as *attitudinal* expressions – because they allow the speaker to express his attitude towards what he says or towards his interlocutor in a given interaction context. Such attitudes are part of the speakers' communicative strategies, and to be efficient, they must observe linguistic and cultural norms.

This work is based on a corpus that contains a set of such prosodic attitudes, typical of the Japanese language and culture (cf. Shochi *et al.*, 2009a). A subset of this corpus, grouping 5 attitudes of politeness or impoliteness has been selected. Three politeness expressions are used: courtesy politeness (PO), sincerity politeness (SIN) and a typically Japanese expression of *kyoshuku*[1] (KYO). A neutral declarative expression (DC) and an impolite expression of arrogance (AR) complete this set. Detailed definitions may be found in Shochi et al. (2009b). This work aims at measuring on one hand if young Japanese children perceive the expressive differences encoded by these 5 attitudes in a fashion similar to that of native adults; and on the other hand if a pair comparison paradigm can be successfully applied with groups of children still not accustomed to written language.

2. Methodology

Shochi et al. (2009b) measured the ability of children (who could read Japanese) to judge the degree of politeness of these five prosodic attitudes. Judgements were made on a politeness scale (ranging from "impolite" to "polite", with neutral in the middle). The results acknowledge the position of arrogance on one hand and courtesy politeness and sincerity politeness on the other hand to each end of the scale. Meanwhile, both declarative and *kyoshuku* expressions were placed close to

---

[1] This Japanese word, without English equivalent is described by Sadanobu as "*a mixture of suffering ashamedness and embarrassment, [which] comes from the speaker's consciousness of the fact that his/her utterance of request imposes a burden to the hearer*" (2004: 34).

"neutral". This result was interpreted as a strong hint of the multidimensional nature of such expressions, which cannot be constrained on this one-dimensional polite-impolite scale. Another limitation of this preceding study is linked to the mandatory use of written instructions describing complex concepts such as politeness. Children below 9 years old do not have sufficient reading skills to adequately perform such a task. But oral presentation of these concepts may also raise difficulties with younger children.

The work of Romney *et al*. (2000) and Moore *et al*. (2002) propose an experimental methodology allowing a precise evaluation of the multidimensional nature of a semantic domain, and an evaluation of the cross-cultural variation of this structure. Their work mostly applies to lexical entries (e.g. emotional or kinship terms), but their methodology may be applied also to the comparison of prosodic expressions. Most interestingly, the statistical methods developed by these authors allow a precise (and quantified) comparison of the specificities of different groups of subjects, as well as the quantification of the amount of shared knowledge between groups. Graphical representations of the main dimensions that structure the semantic space under investigation is an additional interest of this approach. The methodology is based on the evaluation by subjects of a perceived distance between stimuli. Such an evaluation of distances between pairs of stimuli is quite straightforward to explain, and does not require complex conceptual definitions or understanding. Such a pair-comparison paradigm was thus selected to test young children at different ages, and compare their results to native adults'.

2.1   Stimuli

Stimuli presented to subjects consist of the 5 attitudes, spoken on one sentence-type by a native Japanese speaker, a teacher of Japanese as a foreign language, trained to play such expressions in front of students. The same sentence, which has no connotation linked with any (im)politeness attitude, is used to produce all the attitudes. The speaker's performance was recorded with a high quality microphone and a DV camera in a sound proof booth; individual sentences were segmented by hand.

The prosodic and behavioural performances of the speaker show some characteristic differences between the five attitudes that are summarized here. Expressions of sincerity politeness and *kyoshuku* have a faster mean syllabic rhythm with a more flat F0 contour (especially for *kyoshuku*) than the others. They show a limited F0 and intensity register, around the speaker's mean. Courtesy politeness and arrogance, like declaration, show a comparatively wider F0 and intensity slope over the sentence; the F0 slope is more pronounced and linearly decreases in the case of courtesy politeness. The voice quality of each of the five attitudes can be heard as clearly different. Declaration and courtesy politeness use modal voice, while sincerity politeness uses a breathy phonation, which softens the speaker's voice. *Kyoshuku* is performed with a characteristic tense, creaky voice. For arrogance,

the speaker uses a nasalized phonation (cf. d'Alessandro, 2006, for a description of voice quality).

The facial expressions linked with these five attitudes vary, although very little specific information is shown for declaration (such a "lack" of information may well be typical of such a neutral expression). Courtesy politeness and sincerity politeness show a similar slight rising of the brow with a small movement up and down of the head. Arrogance and *kyoshuku* are much more marked: while expressing arrogance, the speaker turns his head to his left and raises his brow. For the *kyoshuku* attitude, the speaker makes a grimace mimicking suffering with a strong frown, wrinkling his nose, and shutting his eyes, and then makes a pronounced bowing.

2.2   Subjects

The 96 subjects of this experiment, all native Japanese speakers, are grouped into four groups of age level.

- 40 adults (AD: 28 females; mean age of 21.6)
- 19 children attending the 4th grade classes (4th grade: 9 females; mean age of 9.5)
- 19 children attending the 2nd grade classes (2nd grade: 13 female; mean age of 7.4)
- 18 children attending the 1st grade classes (1st grade: 11 female; mean age of 6.1)

The adults group is seen as the reference of competent native speakers. The performance of each children groups of growing age will be compared to this adult group.

2.3   Experimental paradigm

All pairs, composed of two different attitudes amongst the 5, are presented to subjects in a random order. Pairs of stimuli are presented in different modalities: audio-only (A), visual-only (V) and audio-visual (AV). The presentation order of these modalities is balanced amongst subjects: half of them took the A modality first, then V and finally AV, while the second half took V first, then A, followed by AV. For each pair, subjects have to judge the perceived difference between the two performances, on a 1 to 9 scale. A pair is only presented once to a subject.

3.   Results

Results are analysed following the methodology described by Romney *et al*. (2000). Details may be found there and in references herein; more specific references will be made for specific points. Statistical methods are tuned for a measure of similarity, thus the obtained judgements of distances are expressed as similarity scores by taking 10 minus the obtained distance score for each pair of different attitudes, and a 10 for the pairs of identical attitudes (not presented to listeners). A 5x5 similarity matrix is obtained for each subject in each modality – with a row containing the similarity scores for an attitude toward each of the possible 5 attitudes. These

matrices are stacked for all subjects and modalities, giving a 1440x5 large matrix X (based on 5 attitudes x 3 modalities x 96 subjects).

3.1 Perceptual distribution of attitudes

A correspondence analysis (CA) is then applied to X, with the row scores standardized using the Kumbasar *et al*. (1994) method, in order to neutralize possible differences in the use of the judgement scale by subjects. The results of the CA give a cloud of points for each attitude as perceived by each subject in each modality, in the 4-dimensional space of the CA. The first two dimensions of the CA explain more than 70% of the observed variance, thus in subsequent graphs only the first two dimensions will be used to display that attitudes' distributions. Individual points obtained for each subject are regrouped according to the 4 age groups, and the mean position of attitudes is displayed, surrounded by a 97.5% confidence interval ellipsis, for each group and each modality. Figure 1 presents the observed dispersion of attitudes in each modality, comparing differences between age groups.

The main tendency observed on these graphs is the overall similarity of the attitudes' distributions over age groups – and to a lesser degree, over modalities. The 2 politeness expressions are on the top right corner, close to declaration, while arrogance is situated on the very left part of the plots. These four attitudes are more or less linearly distributed on an oblique dimension going from arrogance to politeness. This dimension, and the placement of attitudes on it across modalities and age groups, is close to the "impolite-polite" dimension observed by Shochi *et al*. (2009b). By contrast, the *kyoshuku* expression is situated on the bottom right corner (in all modalities), not at all on the same "impolite-polite" dimension. However, if we consider the data from the viewpoint of a one-dimensional paradigm, then *kyoshuku* becomes situated on the same orthogonal line somewhere between polite and impolite, giving this attitude a position close to those obtained by Shochi *et al*. (2009b). This result confirms the similarity of the tasks performed by subjects in this experiment and in the preceding one. Moreover, it shows that in an open evaluation test as this one, *kyoshuku* is clearly differentiated by all listeners from other expressions of politeness.

A detailed observation still shows clear differences, between modalities as well as between age groups. The AV modality takes up the wider space, while the audio-only one defines a more restrained one – but with clear distinctions between each attitude (for adults). The visual-only modality, which occupies a space quiet similar to the AV one, only makes differences between three sets of attitudes: *kyoshuku*, arrogance, and a cluster grouping declaration with the two politeness expressions. Differences between age groups show a progressive extension of the space taken up by the 5 attitudes, from the more limited one with the 1st grade group, expanding with age toward that of the adult's.
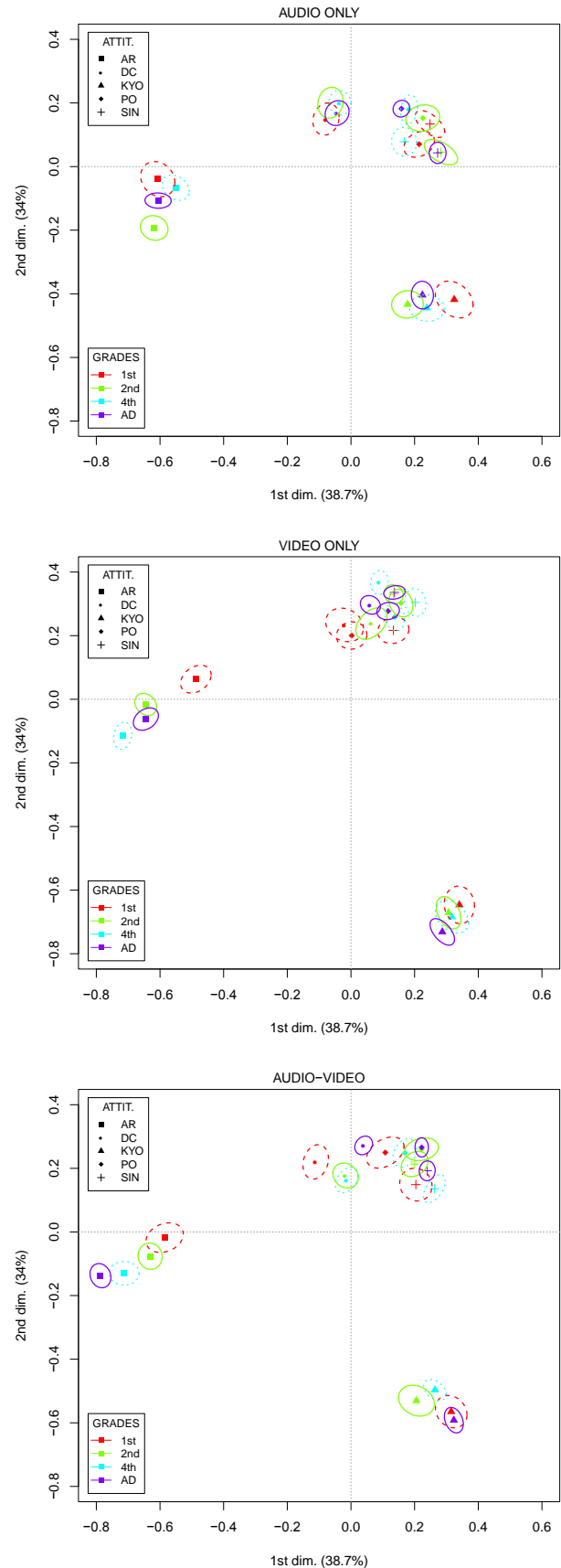
Figure 1: position of each attitude in the 2 first dimensions of the CA, plotted separately for each modality (from the top: A, V, AV)
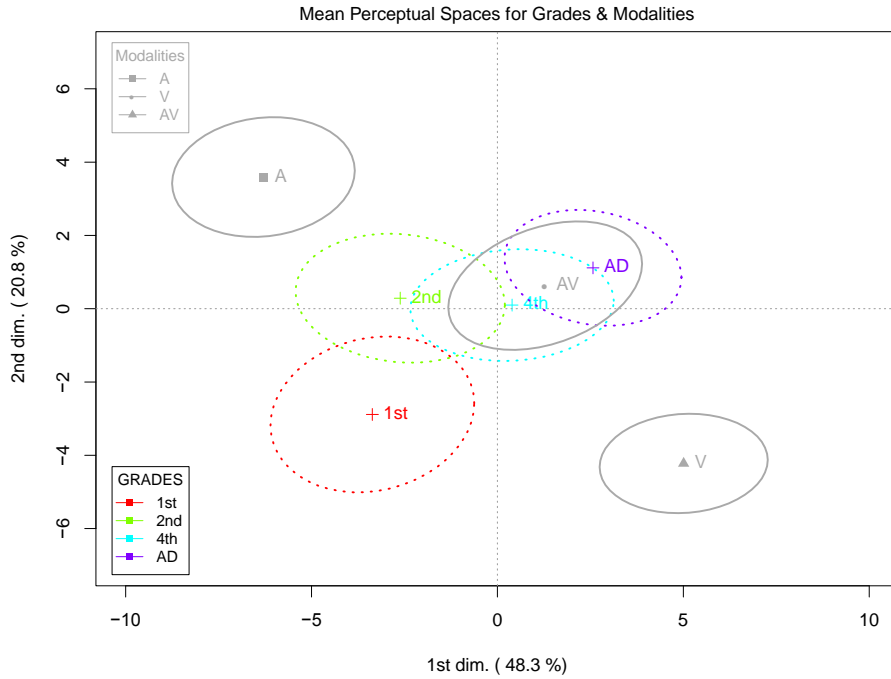
Figure 2: Coloured points represent the mean placement for the four age groups, all modalities averaged;
grey points represent the mean placement for the three modalities (averaged for age).
Ellipses correspond to the 97.5% confidence limit from the means

3.2 Quantification of observed variations

The differences subjectively described above may be quantified to obtain a measure of the differences between groups. Romney *et al*. (2000) propose to use the set of Euclidean distances between points representing the 5 attitudes in the 4 dimensions of the CA to compare the shapes of perceptive structures obtained for each subject (i.e. the shape of the distribution of the 5 attitudes obtained by the CA). The distances between each pair of attitudes compose a vector. The vectors obtained from each subject and modality are then compared via a correlation measure (cf. Rao & Suryawanshi, 1996), resulting in a 288x288 correlation matrix (96 subjects x 3 modalities). The square root of these correlations is used as a measure of the "shared knowledge" between two subjects, or two groups of subjects, by taking the mean of square root correlations (details on this point can be found in Romney *et al*., 2000). A principal component analysis (PCA) was run on this correlation matrix, in order to observe the differences of shape captured by the paradigm, across groups of modalities and groups of ages.

Figure 2 shows the results of the PCA: the place of subjects in the PCA represents the similarity of their perceptual dispersion of attitudes. These positions are averaged either by groups of the same modality (the grey dots indicating the A, V and AV modalities, surrounded by 97.5% confidence ellipses), or averaged by age groups (the coloured dots indicating the 1st, 2nd, 4th grades and adult groups, surrounded by 97.5% confidence ellipses). It is clear from this figure that the audio and visual modalities constitute the factors introducing most

variance in the subjects' answers, and that the audio-visual presentation, the least, suggesting that perception of attitudes is enhanced by information from both modalities. Moreover, the progressive evolution with age of these perceptive structures toward the adults' reference is clear – 1st grade subjects showing a maximally different perceptual shape from that of the adults'.

The mean square roots of observed correlations within- and between-age groups (cf. table 1) indicate the average shared knowledge among subjects of that category. The progressive increase of this shared knowledge with age is a clear indication of the acquisition by native children of the proposed attitudinal expressions, from about 6 to 10 years old. This result reinforces the idea of a progressive construction by children of cultural conceptual spaces with age – particularly in the case of such prosodic social affects.

	1st	2nd	4th	Adults
1st	**.55**	.44	.45	.60
2nd	.44	**.58**	.55	.59
4th	.45	.55	**.63**	.65
Adults	.60	.59	.65	**.71**

Table 1: square roots of the correlations obtained from age groups; within group correlations (in bold) and between groups correlations are given

4. Conclusions

Since results obtained from this perception test corroborate previous findings, it can be assumed that they validate the use of such a pair comparison paradigm for at least three purposes: testing young children with none or few reading skills, investigating the multidimensional distribution of prosodic expressive performances, and measuring the evolution with age of children's understanding of their mother language's social affective strategies.

Such kinds of perception tests are still difficult to run with the youngest group of 1st grade children. Additional information on the children's understanding of the stimuli is also important. For example, Shochi *et al.* (2009b) asked subjects questions about the interlocutor, specifically, what kind of interlocutor may be addressed in that way? Informal discussions with subjects during this test gave interesting answers: 1st grade children described the *kyoshuku* expression as "crying", while 2nd grade children perceived it as "suffering" – a description closer to Sadanobu's (2004) definition. Similar, more accurate descriptions by older children were also observed for arrogance, described by 1st grade children as a "sleeping person", while 2nd grade children thought the speaker was "sulking". Such informal descriptions may also be an interesting path to follow in order to acquire a deeper understanding of children's developing capabilities in their social relationships. The main drawback of such experiments is their complexity.

5. Acknowledgements

6. References

Daneš, F. (1994). Involvement with language and in language. *Journal of Pragmatics*,22, pp. 251--264.

Golan, O., Baron-Cohen, S. and Hill, J. (2006). The Cambridge Mindreading (CAM) Face-Voice Battery: Testing Complex Emotion Recognition in Adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders*, 36(2), pp. 169--183.

Moore C.C., Romney A.K. and Hsia T.-L. (2002). Cultural, Gender, and Individual Differences in Perceptual and Semantic Structures of Basic Colors in Chinese and English. *Journal of Cognition and Culture*, 2, pp. 1--28.

Rao, C.R., Suryawanshi, S. (1996). Statistical analysis of shape of objects based on landmark data. In *Proceedings of the National Academy of Sciences*, 93, pp. 12132--12136.

Romney A.K., Moore C.C., Batchelder W.H. and Hshia T.-L. (2000). Statistical methods for characterizing similarities and differences between semantic structures. *Proceedings of the National Academy of Sciences*, 97 (1), pp. 518--523.

Russell J.A. (1991). Culture and the Categorization of Emotions. *Psychological Bulletin*,110 (3), 426--450.

Russell J.A., Barrett L.F. (1999). Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant. *Journal of Personality and Social Psychology*, 76 (5), pp. 805--819.

Sadanobu T. (2004). A natural history of Japanese pressed voice. *Journal of the Phonetic Society of Japan*, 8 (1), pp. 29--44.

Shochi T., Rilliard A., Aubergé V. and Erickson D. (2009a). Intercultural Perception of English, French and Japanese Social Affective Prosody. In S. Hancil (Ed.), *The role of prosody in Affective Speech*, Linguistic Insights 97, Peter Lang AG, Bern, pp. 31--59.

Shochi T., Erickson D., Sekiyama K., Rilliard A. and Aubergé V. (2009b). Comparison between Japanese children and adults perception of prosodic politeness expressions. *Proc. of Meetings on Acoustics*, ASA, 6 (1), 062001.

Wichmann A. (2000). The attitudinal effects of prosody, and how they relate to emotion. *Proc. ISCA Workshop on Speech and Emotion*, Newcastle, North Ireland, pp. 143--148.

Widden S.C., Russell J.A. (2003).A Closer Look at Preschoolers' Freely Produced Labels for Facial Expressions. *Developmental Psychology*, 39 (1), pp. 114--128

Wierzbicka, A. (1986). Human Emotions: Universal or Culture-Specific? *American Anthropologist*, 88 (3), pp. 584--594.

Zinck A., Newen A. (2008). Classifying emotion: a developmental account. *Synthese*, 161, pp. 1--25.

SPEECH AND INFORMATION STRUCTURE

A comment on comment clauses: data from European Portuguese

Aida CARDOSO, Sandra PEREIRA, Sandra ANTUNES, Rita VELOSO

Centro de Linguística da Universidade de Lisboa

Av. Professor Gama Pinto, 2, 1649-003 Lisboa, Portugal

aidacard@gmail.com, spereira@clul.ul.pt, sandra.antunes@clul.ul.pt, rita.veloso@clul.ul.pt

Abstract

In this study, we analyze the prosodic realization of comment clauses in European Portuguese in a corpus of spontaneous speech: the Portuguese C-ORAL-ROM corpus. Focusing our analysis on comment clauses involving the verb 'dizer' ('to say'), our main goal is to find if there is a pattern in the prosodic realization of similar comment clauses. Building on regular patterns found for the prosodic structure of these constructions, we discuss systematic relations between prosody and discourse structure in terms of semantic-pragmatic meaning. Our data evidences some regularities in the behaviour of comment clauses involving the verb 'dizer' ('to say'), but we also found asymmetries between the prosodic realization of comment clauses constructed with different verb forms (the conditional form 'diria' – 'I would say' – and the subjunctive form 'digamos' – 'let's say'). We discuss these results considering three main points: (i) the results that have been described for parentheticals (and especially comment clauses) for other languages, (ii) the relation between prosodic structure and scope disambiguation, and (iii) the role of the concept of 'cline of grammaticalization' (Dehé & Wichmann, 2010) in the understanding of the status of comment clauses in the informational structure of the sentence.

Keywords: comment clauses; parentheticals; prosody-discourse interface; European Portuguese.

1. Introduction

Recently, parentheticals have been receiving a special attention in literature and have been studied from different perspectives. Nevertheless, establishing a typology of parenthetical structures or even describing its features can be challenging. One of the reasons is the fact that the designation 'parenthetical' covers a wide variety of structures that are heterogeneous in their nature.

Despite the complexity of the topic, many recent studies are relevant contributes towards a better understanding of the syntactic, semantic, prosodic and pragmatic features of parentheticals (e.g., Dehé & Kavalova, 2007). Moreover, as has been proved by the perspective adopted in several studies, parentheticals provide a very interesting subject for interface studies.

In this paper we focus our attention in a particular type of parentheticals – comment clauses (CC). Specifically we describe data from European Portuguese obtained from the prosodic analysis of CCs formed by verb 'dizer' ('to say') in a corpus of spontaneous speech. The discussion of the results of our prosodic analysis will take into account the relation between prosody and discourse. Our data allow us to indentify some patterns in the prosodic realization of the CCs analyzed and, thus, present some hypothesis regarding the relation between prosody and semantic-pragmatics, specifically in terms of scope disambiguation and grammaticalization.

2. Theoretical Background

Parentheticals have been traditionally described, considering the relation between syntax and prosody, as having some specific characteristics regarding phrasing and intonation, namely that they are separated by pauses from the rest of the utterance (e.g., Nespor & Vogel, 1986; Frota, 2000) and that they are most commonly produced with a lower pitch than the rest of the utterance (e.g., Crystal, 1969; Bolinger, 1989). Authors such as

Wichmann (2000), Dehé (2007, 2009), Dehé & Wichmann (2010), on the contrary, argue that there is no one-to-one relation between syntax and prosody and present data (in particular data from spontaneous speech), showing that parentheticals are not obligatory set off by pauses and that they can be associated with different intonation contours.

In the case of European Portuguese, a few studies have described some prosodic features of parentheticals. Frota (2000, in press) describes parenthetical clauses as forming a major intonational phrase (set off by pauses) independent from the rest of the utterance. The author also indicates that parentheticals are associated with the intonation contour L*+H H%. In a study specifically about vocatives, Abalada, Cabarrão & Cardoso (2011) argue that these parenthetical elements do not always form a major intonation phrase and that both the phrasing and the intonation reflect a close relation between syntactic distribution, pragmatic value and prosodic realization of the vocatives. For example, the authors observed that initial vocatives had a stronger tendency to form major intonational phrases than the non-initial (media or final) vocatives and that there were differences in the intonation contours associated with initial and non-initial vocatives.

Regarding CCs, they are often analyzed grouped with other elements, and, accordingly, their characterization is made on a par with other types of parentheticals. Therefore, the prosodic features referred above have been applied to CCs as well. Moreover, the definition of CCs presents some challenges, since it is not always clear where to draw a boundary between them and other parentheticals, such as discourse markers or reporting verbs, as pointed out by Kaltenböck (2007) and Dehé (2009). Both authors present definitions of CCs based on syntactic and semantic criteria: the former identifies CCs with "asyndetic clauses (…) linked to the host in that they contain a syntactic gap (typically the complement of the verb) which is filled conceptually by

the host clause" (Kaltenböck, 2007: 4) and the latter defines CCs as consisting "of a first-person pronoun and a verb of knowledge, belief or conjecture or a corresponding adjectival construction" (Dehé, 2009: 14).

Furthermore, in what concerns the prosodic features of CCs and the prosody-pragmatics relations, the results discussed in studies such as Peters (2006), Kaltenböck (2007), Dehé (2007, 2009), and Dehé & Wichmann (2010) show that CCs tend to not form a major intonational phrase, being accentuated or not. In fact, these authors mention that there are several factors that can influence the prosodic phrasing of these elements, namely the length, the syntactic complexity and even the semantic-pragmatic scope of the parenthetical element.

Secondly, CCs seem to be associated to various intonation contours. Lowered pitch, higher pitch and rising contours are some of the prosodic realizations of parentheticals described by authors as Bolinger (1989), Wichmann (2000), Dehé (2009), Dehé & Wichmann (2010).

Finally, it is important to mention that Kaltenböck (2007) and Dehé & Wichmann (2010) take into account the interface between prosody and semantic-pragmatics meaning in their analysis. Kaltenböck (2007) focuses on the role of prosody in the disambiguation of the semantic-pragmatic scope of the CCs. In this context, the level of juncture between the CC and the sentence is a key factor to determine the scope of the first one and to decide whether the scope of a CC is clausal or phrasal. On the other hand, Dehé & Wichmann (2010) propose an analysis of 'cline of grammaticalisation', where the prosodic properties of CCs, along with their semantic-pragmatic status, place CCs in a continuum between 'propositional' and 'formulaic' meaning. Hence, the authors argue that prosodic separation and prominence are indicators of CCs with a 'propositional meaning', but that CCs associated with disfluency and hesitations have more of a 'formulaic meaning'. In an intermediate position of this continuum, we can find CCs with prosodic integration and deaccentuation, which have "discursal, interactional and interpersonal purposes" (Dehé & Wichmann, 2010: 39).

3. Methodology

For this study, we analyzed the Portuguese C-ORAL-ROM corpus (Bacelar do Nascimento *et al.*, 2005), a multimedia corpus of spontaneous spoken speech, in a total of approximately 300,000 words. This spoken corpus represents real communication acts collected among sociolinguistically diverse speakers and it is composed by 153 recordings, in a total of 30 hours. Each text/recording comprises: (i) the acoustic source; (ii) the orthographic transcription in CHAT[1] format and enriched with the tagging of terminal and non terminal prosodic breaks, and (iii) session metadata containing essential information of speakers, recording situation and contents of each session; (iv) text-to-sound synchronization, based

on the alignment with the acoustic source of each transcribed utterance; (v) a second orthographic transcription with lemma and PoS tags of each form in the transcribed texts, and (vi) frequency lists of forms and lemmas.

This corpus is constituted by different types of informal and formal speech acts, as shown in Table 1, below.

INFORMAL REGISTER			
Family / / Private	Conversations	24,449	
	Dialogs	62,738	
	Monologs	46,005	**133,192**
Public	Conversations	1,817	
	Dialogs	23,119	
	Monologs	7,710	**32,646**
TOTAL			**165,838**

FORMAL REGISTER			
Natural Context	Business	10,215	
	Conferences	9,750	
	Law	6,315	
	Political Debate	8,923	
	Prof. Explanation	6,473	
	Preaching	6,127	
	Political Speech	8,649	
	Teaching	9,822	**66,274**
Media	Interviews	14,570	
	News	1,859	
	Reportages	10,762	
	Scientific Press	9,923	
	Sport	5,676	
	Talk Shows	17,396	
	Weather Forecast	1,930	**62,116**
Telephone	Private		**24,365**
TOTAL			**152,755**

Table 1: Portuguese C-ORAL-ROM corpus constitution

In order to extract our sample of CCs from this corpus, we adopted a definition of CC along the same lines as what has been described in the literature referred above (Kaltenböck, 2007; Dehé, 2009). Then, we selected a sample of 30 occurrences of CCs involving the verb 'dizer' ('to say'), namely the forms 'diria' ('I would say') – 1st person singular of the conditional – and 'digamos' ('let's say') – 1st person plural of the subjunctive present. This sample includes 26 CCs in interpolated contexts and 4 in final contexts. In what concerns the number of syllables, it must be mentioned that the CCs have a minimum of 3 syllables and a maximum of 6 syllables. This variation in the number of syllables is related with some slight differences in the composition of the CCs analyzed. Hence, it is worth noting that, in the case of the 1st person singular of the conditional form, the CCs can be formed: (i) by the verb form – 'diria' –, since European

---

[1] http://childes.psy.cmu.edu/manuals/CHAT.pdf.

Portuguese is a null subject language; or (ii) by the verb form plus the 1st person singular of the personal pronoun – 'eu' ('I') –, or (iii) by the verb form, the 1st person singular of the personal pronoun, and the adverb 'quase' ('almost'), as in 'quase diria eu' ('I would almost say'). In the case of the 1st person plural of the subjunctive present, on the other hand, the CCs included in our sample are formed either by the verb form only – 'digamos' – or by the verb form followed by the adverb 'assim' ('this way'), as in 'digamos assim' ('let's say it this way').

Regarding the prosodic annotation, we used Praat (Boersma & Weenink, 2009) and our analysis focused on two aspects: (i) the break indices on the left and right boundaries of the CCs, and (ii) the nuclear pitch accent and boundary tone of each CC.

In the annotation of our data, we adopted an autossegmental perspective, accordingly with what is described in Pierrehumbert & Hirschberg (1990) and Beckman, Hirschberg & Shattuck-Hufnagel (2005). Hence, we followed the conventions defined by Viana *et al.* (2007) in the annotation system *Towards a P_ToBI* and took into consideration their description of pitch accents and boundary tones for EP. In what concerns break indices, we annotated the juncture level between the CC and the sentence using the break index values described in ToBI (Beckman *et al.*, 2005) – 0, 1, 3, 4 – in which 0 represents the maximum level of juncture between words, 1 represents a normal level of cohesion inside of a prosodic constituent, 3 represents a minor intonational phrase boundary (in EP), and 4 represents a major intonational phrase boundary.

4. Data

Regarding the data, our analysis reveals important regularities in the prosodic realization of the sample of CCs considered in this study.

Firstly, it is worth discussing the level of juncture between the CCs and the utterance. Hence, the data revealed that the analyzed CCs do not tend to form a major intonational phrase, since only 10% of the totality of our CCd formed a major intonational phrase independent from the sentence. These results enable us to compare our data with some findings reported for other languages: the fact that a syntactic parenthesis does not obligatory correspond to a prosodic parenthesis points to the non existence of a one-to-one relation between syntax and prosody, as has been stated before by Dehé (2007, 2009) or Dehé & Wichmann (2010). Furthermore, and taking into account the number of syllables of the CCs, we hypothesized that variables like the length of the parenthetical also play a role in the prosodic phrasing of the CCs analyzed in this study, in the same line as what is argued by Peters (2006) and Dehé (2009).

Additionally, our data can be related with the results found for vocatives in European Portuguese (Abalada *et al.*, 2011), in terms of prosodic integration, in the sense that, despite of having a different nature than CCs, vocatives are also short parenthetical elements and do not always form a major intonational phrase, especially vocatives in medial and final position.

Nevertheless, we did find a high percentage of CCs that form a minor intonational phrase (73,3%), which suggests that, although CCs are more likely to not form an independent tone unit, this does not necessarily translates in a total prosodic integration of the CC in relation with the host sentence. In fact, we observed, particularly in what concerns the CCs formed by the conditional form ('diria'), some differences in the strength of the break index on the left and right boundaries. As Kaltenböck (2007) remarked, the level of juncture between the utterance and the CC can be related to informational structure, which may represent a clue to identify the semantic-pragmatic scope of the CC. In our data, we also noticed that the phrasing differences referred to above can be related with the fact that the CC has a clausal or phrasal scope. Example (1) illustrates a case in which the phrasing evidences that the CC has a clausal scope, and not a phrasal one, since the break index on the left boundary of the CC ('eu diria') is stronger – [4] – than the one identified on the right boundary – [3].

(1) Os três outros evangelistas [4] eu diria [3] têm características tão salientes e tão próprias (...).

(The other three evangelists [4] I would say [3] have such evident and unique features (…).)

By contrast, CCs formed by the subjunctive verb form ('digamos') evidence a greater level of juncture in relation with the utterance and, significantly, it is on the right boundary of these CCs that we find a higher frequency of break indices of level 0 and 1.

Similarly to what has been described for phrasing, there are also some relevant aspects regarding intonation that provide some clues to a better understanding of the prosodic behavior of the two types of CCs analyzed. First of all, is should be mentioned that there is a high percentage of CCs (86,6%) that are accented. Nevertheless, this percentage is higher in the case of CCs with the conditional verbal form 'diria'. In fact, 18,8% of the CCs formed by the subjunctive form 'digamos' are un-accented (as shown in Table 2).

Regarding the distribution of pitch accents (cf. Table 2), we identified five pitch accents associated with the CCs included in our data. The fact that these parenthetical elements are characterized by various pitch accents allows us to draw a comparison between our data and what has been stated for other languages, namely English, by authors such as Wichmann (2000), Dehé (2009b), Dehé & Wichmann (2010). In spite of the importance of the non-existence of a obligatory association of these parenthetical elements to a certain intonation contour, it is also relevant that, considering both types of CCs ('diria' and 'digamos'), there is a higher percentage of CCs associated with low pitch accents (L*), followed by the rising pitch accent L+H* and by the high pitch accents (H*).

Pitch Accents	Comment Clauses	
	'diria' ('I would say')	'digamos' ('let's say')
H*	21,4%	12,5%
L+H*	21,4%	18,8%
H*+L	7,1%	-
L*+H	7,1%	-
H+L*	14,3%	-
L*	21,4%	50%
Un-accented	7,1%	18,8%
TOTAL	100%	100%

Table 2: Distribution of pitch accents

Once again, though, the data reveals some differences in the prosodic realization of CCs with the conditional form and with the subjunctive form of the verb 'dizer'. As can be observed in Table 2, whereas CCs formed by the conditional verb form 'diria' are characterized by a greater variety of pitch accents (cf. Figure 1), CCs with the subjunctive verb form 'digamos' are associated with three distinct pitch accents. Moreover, in the case of 'digamos', we observe that the L* pitch accent corresponds to 50% of the totality of the occurrences (cf. Figure 2).

Figure 1: CC with the conditional form 'diria' ('I would say'), which forms a minor intonational phrase and has a rising intonation contour (L+H* H%)

Figure 2: CC with the subjunctive form 'digamos' ('let's say'), which does not form an independent intonational phrase and has a low pitch accent (L*)

In what concerns boundary tones, Table 3 shows that in both types of CCs we found a higher percentage of low boundary tones, but the subjunctive form 'digamos' has a higher percentage of cases with no boundary tone, accordingly to what has been previously discussed about the prosodic integration of CCs with this verb form.

Boundary Tones	Comment Clauses	
	'diria' ('I would say')	'digamos' ('let's say')
H- / H%	35,7%	25%
L- / L%	57,1%	43,8%
No boundary tone	7,1%	31,3%
TOTAL	100%	100%

Table 3: Distribution of boundary tones

We think that the results presented above can be interpreted along the lines of what Dehé & Wichmann (2010) have described as 'cline of grammaticalization'. On the one hand, the fact that the prosodic realization of CCs can play a role in scope disambiguation and that CCs do not evidence a tendency to total prosodic integration seems to indicate that the CCs included in our sample do not have a 'formulaic meaning'. On the other hand, we found differences between CCs with two different forms of the verb 'dizer'. As a result, some of the prosodic characteristics of the subjunctive form 'digamos' contrast with what can be observed for the conditional form 'diria': (i) the former does not seem to play such an important role in scope disambiguation as the latter; (ii) the subjunctive form shows a greater tendency for prosodic integration; (iii) there is a higher percentage of low pitch accents associated with the subjunctive verb form, and (iii) there is a higher percentage of un-accented occurrences of CCs with the subjunctive form. Considering these results, we hypothesize that the two types of CCs are in different stages of a grammaticalization continuum. Hence, whereas CCs with the conditional form seem to have more of a propositional meaning, CCs with the subjunctive form are possibly closer to an intermediate stage between propositional and formulaic meaning, characterized pragmatically has having "discursal, interactional and interpersonal purposes" (Dehé & Wichmann, 2010: 39), and prosodically by prosodic integration and deaccentuation.
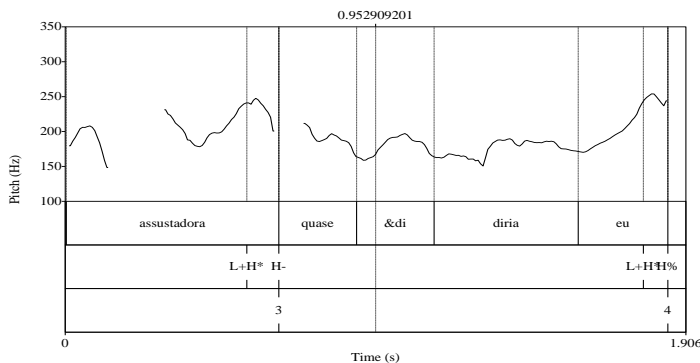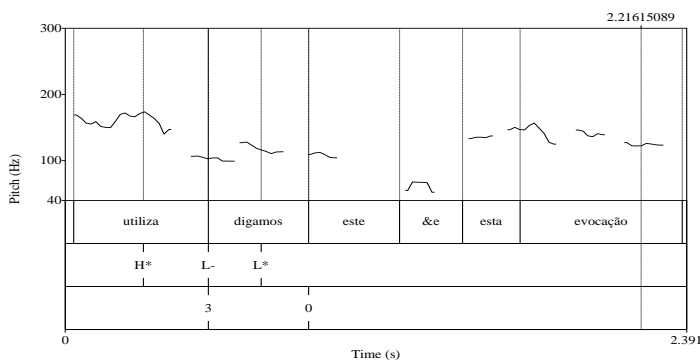
5. Conclusion

The results discussed in this paper are a starting point to the study of CCs in European Portuguese. By studying a sample of CCs formed by the same verb – 'dizer' ('to say') – we were able to detect patterns in the prosodic realization of these parenthetical elements.

The fact that CCs do not always form an independent tonal unit and that they are not obligatory associated with a single intonation contour is in agreement with the idea that (i) syntactic parenthesis do

not necessarily correspond to prosodic parenthesis, as argued by Dehé (2007), and (ii) parenthetical elements can have intonation contours other than a lowered pitch accent, as have been shown in studies such as Wichmann (2000), Dehé (2009), and Dehé & Wichmann (2010).

On the other hand, we also found some asymmetries in the prosodic behaviour of CCs with different verb forms, namely the conditional form 'diria' ('I would say') and the subjunctive form 'digamos' (let's say'). We interpreted such asymmetries in relation with CCs' semantic-pragmatic meaning, in terms of scope disambiguation and grammaticalization. More specifically, our data suggested that the conditional verb form evidences more features associated with a propositional meaning than the subjunctive verb form.

6.   Acknowledgements

7.   References

Abalada, S., Cabarrão, V. and Cardoso, A. (2011). O Vocativo em Português Europeu: Estudo de Parâmetros Prosódicos em Vocativos com Diferentes Distribuições. In *Textos Seleccionados do XXVI Encontro Nacional da Associação Portuguesa de Linguística 2010.*

Bacelar do Nascimento, F. *et al.* (2005). "The Portuguese Corpus". In E. Cresti & M. Moneglia (Eds.), *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages.* Amsterdam/Philadelphia: John Benjamins Publishing Company, Studies in Corpus Linguistics, nº 15 (with DVD).

Beckman, M.E., Hirschberg, J. and Shattuck-Hufnagel, S. (2005). The Original ToBI System and the Evolution of the ToBI Framework. In S.-A. Jun (Ed.), *Prosodic Typology. The Phonology of Intonation and Phrasing.* Oxford: Oxford University Press.

Boersma, P. , Weenink, D. (2009). Praat: doing phonetics by computer**.** In http://www.praat.org/.

Bolinger, D. (1989). *Intonation and its uses. Melody in grammar and discourse.* London: Edward Arnold.

Crystal, D. (1969). *Prosodic systems and intonation in English.* Cambridge: Cambridge University Press.

Dehé, N. & Kavalova, Y. (Eds.). (2007). *Parentheticals.* Amsterdam: John Benjamins.

Dehé, N. (2007). The relation between syntactic an prosodic parenthesis. In Dehé, Nicole & Yordanka Kavalova (Eds.) *Parentheticals.* Amsterdam: John Benjamins.

Dehé, N. (2009). Clausal parentheticals, intonational phrasing, and prosodic theory. In *Journal of Linguistics* 45 (3).

Dehé, N., Wichmann, A. (2010). The multifunctionality of epistemic parentheticals in discourse: prosodic cues to the semantic-pragmatic boundary. In *Functions of Language* 17 (1).

Frota, S. (2000). *Prosody and Focus in European Portuguese.* New York & London: Garland Publishing.

Frota, S. (in press). The intonational phonology of European Portuguese. In S.-A. Jun (Ed.) *Prosodic Typology II: The Phonology and Intonation of Phrasing.* Oxford: Oxford University Press.

Kaltenböck, G. (2007). Position, prosody, and scope: the case of English comment clauses. In *Vienna English Working Papers* 16 (1), pp. 3--38.

Nespor, M., Vogel, I. (1986). *Prosodic phonology* (Studies in Generative Grammar 28). Dordrecht: Foris.

Pierrehumbert, J., Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen *et al.* (Eds.), *Intentions in communication.* Cambridge: MIT Press.

Peters, J. (2006). Syntactic and prosodic parenthesis. *Proceedings of the International Conference on Speech Prosody*, Dresden.

Viana, M. do C. *et al.* (2007). "Towards a P_ToBI". Towards a P_ToBI. PAPI2007. Workshop on the Transcription of Intonation in Ibero-Romance. Universidade do Minho.Available at: <http://www.ling.ohio-state.edu/~tobi/>.

Wichmann, A. (2000). *Intonation in Text and Discourse: Beginnings, Middles and Ends.* Harlow: Longman/Pearson Education.

Topic-Focus and Comment-Focus in the Language into Act Theory

Emanuela CRESTI

LABLITA – Dipartimento di Italianistica

P.za Savonarola, 1 – Firenze

elicresti@unifi.it

Abstract

This paper presents a new semantic definition of Focus within the LAcT's assumptions and on the basis of corpus evidence. We discuss current theoretical frames based on the concept of Common Ground, such as Alternative Semantics, following which Focus is given a semantic definition inside a question-answer model depending on the context. According to the LAcT's pragmatic interpretation of the information structure, the information unit (IU) of Comment, which is devoted to the accomplishment of the illocution, is considered its centre. Comment, as a pragmatic entity, and Focus, as a semantic entity, must be distinguished given that their respective levels of activity are illocution and locution in Austinian terms. Within the locutionary act, Focus is a semantic entity signalling the apex of a domain functioning in the illocutionary act such as an IU. Moreover, we claim the existence of two kinds of Foci: Topic- Focus and Comment-Focus, characterized by two different semantic values. We derive this from the condition that a Focus must be marked by a prosodic prominence, and from corpus observations showing that a Topic-Comment information pattern is necessarily performed with two prosodic prominences.

Keywords: information structure; pragmatics; comment; focus; prosody; corpus data.

1. The pragmatic nature of Comment

In the framework of the Language into Act Theory (LAcT)[1] the information structure of the utterance is pragmatically based. An utterance can be compounded of many information units (IU), developing different information functions and each must be identified by a prosodic unit (PU). Alternatively, it can be simple i.e. compounded of only one IU, necessary and sufficient, named Comment, whose specific function is the accomplishment of the illocutionary force (Austin, 1962). According to LAcT the starting point of the information patterning (IP) is the accomplishment of the illocutionary force by the specific Comment IU, because the action of the speaker, given its affective nature, has a subjective and internal origin and is continuously changing. In accordance, the information about the type of action the speaker will utter is expected as necessary and unforeseeable.[2] This perspective on the structure of information departs from the track of traditional assumptions[3] in relation to one particular feature: the

pragmatic origin of information, which they do not consider, ignoring its illocutionary definition. Generally speaking, they also share two other aspects diverging from LAcT: the semantic nature of Focus, which is substantially identified on the basis of its novelty with respect to context and represents the only key to explaining the information structure, and the fact that Topic derives from the context. In that way the entire information organization of the utterance results is conditioned by the context.

The reason for these differences is that no distinction is foreseen between different activities (illocution and locution) accomplished by the speaker simultaneously (Austin, 1962), but which diverge in their nature (affective/pragmatic vs cognitive). Given the lack of the illocutionary notion of Comment, in the literature there is no distinction between the semantic concept of Focus and the pragmatic one of Comment, and Comment and Focus are employed like terminological variants. But Comment develops a pragmatic role and cannot be defined in semantic terms, such as is done for Focus.

Traditional semantic definitions foresee that Focus represents the "most important" or "new" information in an utterance. However, importance is a vague aspect and can hardly be verified. As regards the feature of novelty, on the basis of our corpus data we have already shown that a Comment can record old semantic content from a contextual point of view (becoming "new" for the illocutionary accomplishment), and that a Topic can record new semantic content (Cresti & Moneglia, 2010). If a Topic can be new and a Comment can be old, are importance and novelty opposing values? These questions don't seem to have clear solutions.

We will discuss current theoretical frames based on the concept of Common Ground, such as Alternative Semantics, and assuming that Focus must be prosodically marked we will also argue against the perspective of

---

[1] See Cresti (2000; 2006; 2012a).

[2] Actually, in the last decade the analysis carried out by LABLITA has led to the identification of a larger set of about 90 speech act types (Cresti & Firenzuoli, 1999; Firenzuoli, 2003; Cresti, 2006; Moneglia, 2011), found empirically, whose classification criteria imply pragmatic identification and definition, lexical and prosodic correlations, and frequency data. This means that spontaneous speech is systematically characterized by a rich pragmatic variation.

[3] The first attempts to study and explain the structure of spoken language and its information organization date back to the middle of 1800, with concepts such as point de départ and but du discours by Weill (1844). Jumping to the second half of 1900, we can remember the most relevant frameworks with the translation of the Praguian concepts of *theme* and *rheme*, imported into the USA with terms like *topic* and *comment* by Hockett (1958) and (Chafe, 1970; Gundel, 1977) and transformed into topic and focus (Chomsky, 1971; Jackendoff, 1972), as well as other approaches proposing given and new (Halliday, 1976), figure and ground (Talmy, 1975), and source

and goal (Langacker, 1991).

Contrastive Focus on the basis both of our pragmatic framework and corpus data evidence. We will propose the existence of two Foci within an utterance corresponding to a Topic-Comment IP: a Topic-Focus and a Comment-Focus with a new semantic definition .

2.   The model of Common Ground

Some acknowledged research on information structure employs the concept of Common Ground (CG) in the place of context. The concept was formulated by Stalnaker (1974) and can be described as «a way to model the information that is mutually known to be shared, which is continuously modified in the course of communication». But if a pragmatic perspective is adopted, no "mutually shared information" can exist. The fact that the context is real does not mean that it is an independent entity, knowable in its whole as a logic universe. Everybody knows it subjectively, following his mood and giving attention to what is interesting for his own attitude in that moment. There are no mandatory information prominences in the context, but only those inputs which are prominent for the speaker's attention in that moment. Moreover, there is no determination from contextual inputs to the performance of a specific illocutionary speech act, because of the internal affective and mental origin of the latter. The speaker's next speech act is unforeseeable despite every kind of contextual prominence. Mutually shared information could exist only in a platonically semantic or logic context existing outside of the speakers and in spite of their living actions.

In some sense a more concrete definition of Focus seems to be given within the framework of Alternative Semantics (Rooth, 1992; Krifka, 2006).

> "Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions. […] This distinction is relevant for information packaging, as the CG changes continuously, and information has to be packaged corresponding to the CG at the point at which it is uttered".

This assumption could seem reasonable and of use, but the claim that information can be packaged «corresponding to the CG at the point at which it is uttered» seems to lead again to a semantic dependence of the information structure on the context. It means that some specific objective features in the context, identifying a point in the CG, condition Focus.

Advancing along these lines, Krifka explicates that the prominent use of Focus is the identification of context-questions in answers:

> "The idea is that the meaning of a question identifies a set of alternative propositions, the answer picks out one of these, the Focus within the answer signals the alternative propositions inherent in the question".

In substance, following Alternative Semantics, the core of an assertion i.e. the part adding a novelty to the CG, should be the answer chosen by the speaker among the possible ones, given a certain open question in the CG, that may be optionally reported in the theme/Topic.

A relevant extension of the question-answer model is due to theories assuming that a *coherent discourse* is structured by implicit questions (van Kuppervelt, 1994; Büring, 2003) and by Focus on the answers. The concept of implicit questions foresees that context is characterized by features that by themselves can constitute or suggest questions for the addressee. In this sense the activity of speech is reduced to answering in a coherent way the questions suggested by the world and the operation could be reduced to a logic schema. The semantic question-answer model transforms the context into an open variable and assumes its satisfaction in the answer, ensuring a result which is characterized by a propositional form. No pragmatic value of the utterance is even hypothesized and this ends the claim of equivalence between utterance and proposition and the allowance of the analysis of the former in the semantic terms of the latter.

However, corpus data supports the fact that real spontaneous spoken activity does not occur in this way[4]. The framework of Alternative semantics, defining the pragmatic use of Focus as the point marking an alternative in an answer to an overt or covert CG question, does not seem adequate for explaining corpus data. Analyzing a stretch of spontaneous dialogue will demonstrate as impossible the continuous discovery of elements for consideration as the origin of covert questions in the CG, so that they are adequate input for speech behavior (Cresti, 2012). The normal manner of human spoken communication is about the context but has its origin in the speakers' thoughts and in the affective dynamics among these speakers. They are not determined by the context and they continue on with subjective actions and reactions.

For instance, what could be the context-question generating the accomplishment of a specific illocutionary act like an alternative question, or an instruction, or an expression of obviousness? Given that at least 40% of the illocutionary values of utterances in spontaneous speech are not assertive (Moneglia, 2011), it is not clear what the covert question in the context could be, being the adequate input of one of these specific speech acts. In conclusion, a constant aspect of every utterance derives from its pragmatic nature and from its illocutionary types which very rarely can be connected in an incontrovertible way to an objective/contextual input, and on the contrary ties to an internal affective disposition.

---

[4] It's likely that research carried out on map-task data, or call center conversations, or other kinds of ruled spoken exchanges will allow a different perspective, because within a shared and limited context the task of the participants is exactly that of posing questions and giving appropriate answers. However, even in these instances it is easy to find continuous counter-examples.

3.   The prosodic Prominence

At this point it must be stressed that real speech must also be studied with consideration for its sound counterpart and, especially, some prosodic cues like terminal and non-terminal breaks, prosodic forms with illocutionary values, prosodic prominences necessarily signaling focus, etc. In accordance with these premises, it is assumed by the greater part of the literature that Focus must correlate with a phonetic-prosodic prominence[5]. The taking into account of this prosodic cue causes new contradictions, because it cannot be ignored, too, that there are utterances bearing two prosodic prominences. Thus, the occurrence in the same utterance of a first prominence and a second one, corresponding to semantic Foci, are a phenomenon it becomes necessary to explain.

In reality, systematic controls on the corpus carried out in LABLITA make us certain that not only are the *root*[6] PUs performing a Comment characterized by a prosodic prominence, but that also *prefix* PUs performing Topics are mandatorily concluded by a perceptual prominence, sometimes more relevant than that in the Comment. This means that the Topic-Comment IP is always performed with a *prefix* PU and a *root* PU, each of them recording a prominence, corresponding to the prosodic nucleus of the PU. In conclusion, every utterance corresponding to a Topic-Comment information pattern is characterized by two Foci.

Facing the case of the two Foci utterances, scholars have been, in some sense, obliged to make the hypothesis of a Contrastive Focus (Büring, 2003). This has been explained within the context question-answer model through the hypothesis of a double question which should motivate the double Focus (*who stole what?*)[7].

It is obvious that if the finding of a mandatory question input in the context to explain a Focus in the answer hardly appears acceptable, the hypothesis that the context questions had to double to also explain a Contrastive Focus seems even less so. It must be considered, moreover, that corpus data records about 10% of non-simple topicalisation phenomena i.e. the IP of a lot of utterances is not composed of a Topic-Comment information pattern, but of a Topic-Topic-Comment, or a Topic-Topic-Topic-Comment, or of a List of Topics and a Comment. In this case each of the *prefix* PUs performing the respective Topic bears its own prosodic prominence, marking a Focus. Thus, according to the question-answer model there has to be a new Contrastive Focus every time there is a Topic, and by consequence a multi-multi covert questions input has to be found in the context to justify that result[8].

For instance, for (1) below, how do we formulate a triple covert questions input, implying a covert question for the first Contrastive Focus in Topic, one for the second Contrastive Focus in the second Topic, and finally one for the Focus in the assertive Comment?

(1)   *MAA: la **maggior parte** /TOP […] quelli che hanno portato **Pinocchio** /TOP va proprio bene quello **che hanno** //COM '*the major part... those who brought Pinocchio, what they have is all right* ' %ill: assertion [ipubcv02][9]

We don't see how it could be possible to justify as input such a triple covert question in the context, which seems to be a totally *ad hoc* solution.

In conclusion: in a lot of influential literature the notion of Focus is strictly semantic and has been considered the central point for the information structure of the utterance. The concept has traditionally been defined according to vague notions of importance and novelty. Starting from the assumption of Common Ground within the model of context question-answering, more recent approaches have proposed the function of Focus as highlighting a semantic alternative in the answer and have hypothesized the existence of Contrastive Foci to explain the occurrence of utterances with two Foci.

4.   The LAcT definition of Focus

In the LAcT perspective the importance of the concept of Focus is strongly rescaled because the information structure is not conceived as a semantic entity with a propositional size/form, whose Focus has to be the center. Information patterning does not depend on it, but on the pragmatic accomplishment of an illocution by the Comment, and on the pattern of Topic-Comment. The overall structure is not semantic but is still informative.

Focus remains a semantic concept in LAcT too, but its domain spreads only to the boundary of a textual IU of Comment or Topic. Expressions are conceived to develop an information function of Comment or Topic within the illocutionary act. Simultaneously, the same expressions, produced with an information function, are performed, from a syntactic and semantic point of view, as islands within the locutionary act. The performance of a Topic-Comment IP constitutes the accomplishment of an utterance, whose definition is pragmatic, but it does not correspond to the performance of a semantic proposition or of a syntactic sentence at the simultaneous locutionary level (Cresti & Moneglia 2010).

Specifically, the semantics both of Topic and of Comment record kinds of relations regarding regency, quantification, modality, and Focus. Focus is a high semantic level of composition occurring both in Comment and Topic IUs. So, even if Focus is still a

---

[5] See studies on prosodic Focus (Avesani & Vayra, 2003; D'Imperio, 2001).
[6] For this terminology see ('t Hart et al.,1990; Firenzuoli, 2003).
[7] It must be noted that the hypothesis of Contrastive Focus does not assume that there is a Focus in the Topic but only that there are utterances with two Foci, one of which is considered "contrastive".
[8] On the contrary, with our information perspective the speaker can duplicate or triplicate the field of application of the illocutionary force, Topic, just adding linguistic details.

[9] Examples are taken from Cresti & Moneglia, 2005 and from the LABLITA archive and are transcribed with an implemented version of the CHAT format (McWhinney, 2000; Moneglia & Cresti, 1997)

semantic notion, its domain is related not to an entire utterance, or presumed proposition, but to the semantics of a Topic or a Comment IU, which often copes only with phrasal constituents from a syntactic point of view.

Thus a general semantic definition of Focus in LAcT:

> "A Focus signals the apex of a semantic domain which develops a Topic or a Comment information function within the information pattern of an utterance".

The semantics of the domain behaving like a Topic or a Comment is conditioned by the information function that the expression is developing: in the case of Topic that of a field of application of an illocutionary force (T-Focus) and in the case of Comment that of the expression of an illocutionary force (C-Focus).

The Foci of the two IUs are apexes of semantic domains which systematically diverge for their semantic content and their respective lexical and morpho-syntactic composition: for Topic 75% of the linguistic content corresponds to Noun phrases and Prepositional phrases and for Comment 61% to Verb phrases. Generally speaking, T-Focus has a semantic identification function within a non-action domain and C-Focus has a semantic specification function within an action domain.

Therefore we claim that there are Topic Focus (T-Focus) and Comment Focus (C-Focus).

4.1 Topic-Focus (T-Focus)

For T-Focus, following its function must lead to the apex of a domain adequate in identifying the field of application of the illocutionary force. In an utterance like (2) with a total question force, but with two Topics, each Topic must identify a field of application for the question illocution in the Comment while functioning as a Topic by itself.

(2)   *CEC: di **là** /TOP gli **acidi** /TOP tutto **pronto** ?COM '*there, (for what concerns) the acids, everything ready?*'
       %ill: total question [ifamdl17]
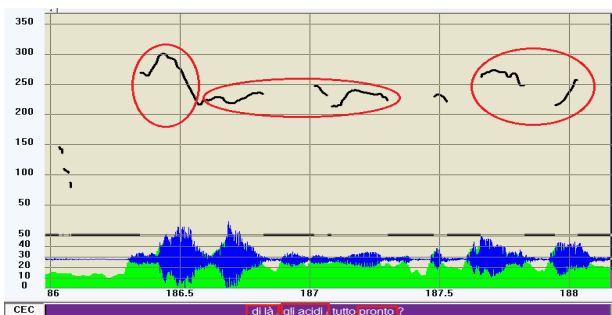
Figure 1: Utterance (2)

The right part of the *prefix* PU is the seat of its prosodic nucleus, containing a prominence, and the majority of times it is performed with a relevant rising or a rising-falling movement. This position copes with the last semantic word of each Topic. So the adverb 'là' (*there*) and the noun 'acidi' (*acids*) can be considered the respective semantic Focus marked by the prosodic prominence.

Very often the expression, functioning as Topic[10], is from a syntactic point of view a well formed phrase (Noun, prepositional, adverbial), whose last word is also the head of the phrase. But it can happen that this coincidence doesn't occur, like in the second Topic of (1), where the proper name 'Pinocchio' is the last word but it is not the head of the noun phrase. It should have been in doubt regarding the semantic or syntactic condition for being the Focus of a Topic domain, but corpus examples allow us to verify that it is always the final seat that correlates with the role of Focus in spite of the syntactic head position.

In speech we have the habit of expecting the end of something in recognizing it as a whole and the signal of ending or starting is given primarily by prosody. As a result, the last semantic word of the Topic marked by a prosodic prominence is recognized by the hearer as the expression closing the domain and identifying it as the semantic entity to be considered in its whole as the application field of the illocution i.e. the Topic.

4.2 Comment-Focus (C-Focus)

On the contrary C-Focus has no fixed seat, even if it occurs very often for a semantic word in the right side of a Comment IU or only at the last word. It depends on the fact that the C-Focus is also marked by the prosodic nucleus of the *root* PU, but this can occur in different seats within the PU depending mostly on the illocutionary type accomplished. Below are some examples with different illocutions where the C-Focus doesn't occur at the last word of the IU.

(3)   *PAO: il **resto** /TOP non voglio **sapere** che cosa pensano //COM '*for the rest, I don't want to know what they think*'
       %ill: refusal [ipubcv01]
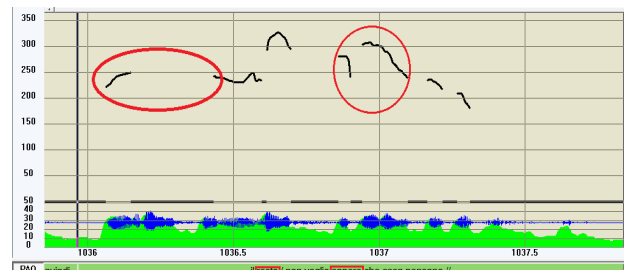
Figure 2: Utterance (3)

---

[10] It must be remembered that there is no syntactic relation between the linguistic filling in both Topics to each other and to that in Comment, so that they can be defined as anacholuta and from a semantic point of view are islands.

(4) *VAL: perché io sono stata nominata /SCA **prima** di' trenta di' giugno //COM '*cause I have been appointed, before the thirtieth of June*'
%ill: answer [ifamvc18]
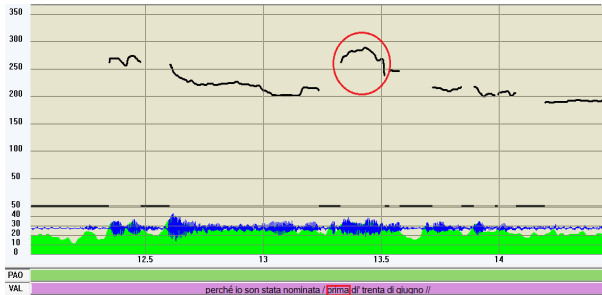
Figure 3: Utterance (4)

For what it has been possible to verify, C-Focus, coping with the prosodic nucleus of the *root* PU, represents the phonetic part necessary to express and specify the illocutionary type of the Comment. This means that the recoverability of the illocutionary type is assured if the prosodic nucleus is the only sound conserved within the *root* PU.

Below are some examples where listening to the bare nucleus of the *root* PU allows the recognition of the illocutionary value. See Figure 2, where the prosodic shape of a total question is clearly recognizable from the last two syllables.

In (5) a partial question is performed.

(5) *PRO: [l'unit linked] /TOP **praticamente** /TOP che **cos'è** ?COM '*the linked unit, actually, what is it?*'
%ill: partial question [ipubdl04]
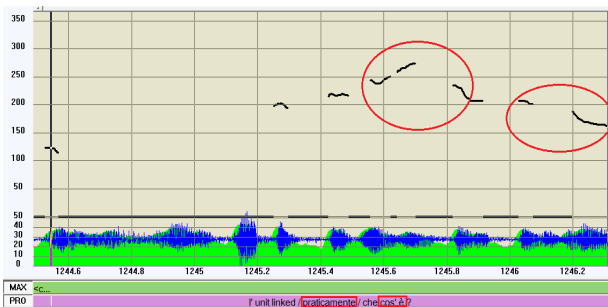
Figure 4: Utterance (5)

(6) is an example of the expressive illocution of Contrast with a high jump on *you*.

(6) *PAO: che tu me l'avevi detto **te** /COM i'cream caramel //APC '*cause it was you that said it to me, the cream caramel*'
%ill: assertion of contrast [ifamdl12]
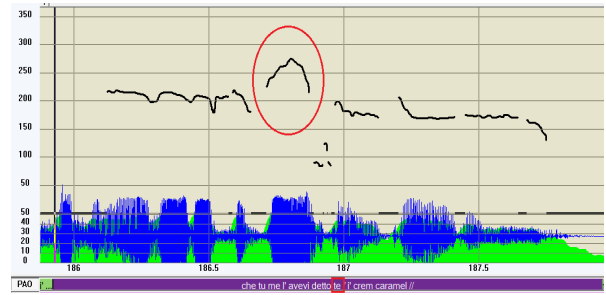
Figure 5: Utterance (6)

Evidently what is relevant to perform with a C-Focus, more than the recoverability of an entire semantic domain (as in the case of T-Focus), is the sense of an expression through which a specific illocutionary act is accomplished. Then the goal of C- Focus emerges as supporting the word (s) and bettering the sense with which a specific illocution can be recognized, and in doing so prompts the addressee's attention to the latter.

C-Focus *marks the expression, allowing us to specify what type of illocution is performed* within the semantic domain, dedicated in its whole to the accomplishment of the illocutionary force.[11]

In conclusion, according to LAcT the IP of the utterance has a pragmatic nature and its origin is in the accomplishment of an illocutionary force by the Comment. The IP does not correspond to a semantic structure whose center is the Focus. IP does not depend on Context. Focus corresponds to a semantic level of composition within the domain both of a Topic and of a Comment IU, and while T-Focus develops the semantic function of allowing the recoverability of the entire field of application for the illocutionary force, in its turn the semantic function of C-Focus is bettering the sense of an expression through which a specific illocution can be recognized. Both are mandatorily signalized by the nuclear prominence of their respective *prefix* PU and *root* PU.

5. References

Austin, J.L. (1962). *How to do things with words*. Oxford: Oxford University Press.

Avesani, C., Vayra, M. (2003). Broad, narrow and contrastive focus in Florentine Italian. In M.J. Solè, D. Recasens and J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences vol 2*. Barcelona: Causal productions Pty, pp. 1803--1806.

---

[11] Often the prosodic prominence marking the nucleus, dealing with the C-Focus, is of little relevance and specifically it can be weaker than that of T-Focus. This is not so strange, depending on the fact that what is necessary is that the *root* PU of Comment must clearly manifest a specific illocution. So this task is accomplished more by the form of the *root* PU than by the scale of the prosodic prominence, while in the Topic the only way to signal the Focus is through the relevance of its prominence.

Büring, D. (2003). On D-trees, beans and B-accents. *Linguistics and Philosophy*, 26, pp. 511--545.

Chafe, W. (1970). *Meaning and the structure of language*. Chicago/London.

Chomsky, N. (1971). Deep Structure, Surface Structure and Semantic Interpretation. In D. Steimberg & L. Jacobovits (Eds.), *Semantics: an Interdisciplinary Reader*. Cambridge: Cambridge University Press, pp. 183--216.

Cresti, E. (2000). *Corpus di Italiano Parlato*, voll. I-II, CD-Rom. Firenze: Accademia della Crusca.

Cresti, E. (2006). Some comparisons between UBLI and C-ORAL-ROM. In Y. Kawaguchi, S. Zaima and T. Takagaki (Eds.), *Spoken Language Corpus and Linguistics Informatics*. Amsterdam: Benjamins, pp. 125--152.

Cresti, E. (2012). The definition of Focus in the framework of the Language into Act Theory (LAcT). In A. Panunzi, T. Raso and H. Mello (Eds.), *Pragmatics and Prosody. Illocution, modality, attitude, information patterning and speech annotation*, Firenze: Firenze University Press, pp. 39--82.

Cresti, E. (2012a). L'unité de suffixe: identification et interprétation des unités de la langue parlé. In S. Caddéo, M.N. Roubaud, M. Rouquier and F. Sabio (Eds.), *Penser les langues avec Claire Blanche-Benveniste*. Presses Universitaires de Provence, pp. 201--213.

Cresti, E., Firenzuoli, V. (1999). Illocution and intonational contours in Italian. *Revue Française de Linguistique Appliquée*, IV(2), pp. 77--98.

Cresti, E., Moneglia, M. (Eds.) (2005). *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: Benjamins.

Cresti, E., Moneglia, M. (2010). Informational Patterning Theory and the Corpus based description of Spoken language. The composiotionality issue in the Topic Comment pattern. In M. Moneglia, A. Panunzi (Eds.), *Bootstrapping Information from Corpora in a Cross Linguistic Perspective*. Firenze: Firenze University Press, pp. 13--46.

C-ORAL-ROM. http://lablita.dit.unifi.it/coralrom/

D'Imperio, M. (2001). Focus and tonal structure in Neapolitan Italian. *Speech communication*, 33(4). Amsterdam: Elsevier Science Publishers, pp. 339--356.

Firenzuoli, V. (2003). Le forme intonative di valore illocutivo nell'italiano parlato. Analisi Sperimentale di un Corpus di Parlato Spontaneo (LABLITA). PhD diss., University of Florence.

Gundel, J.K. (1977). *The Role of Topic and Comment in Linguistic Theory*. Indiana University Linguistics Club.

Halliday, M.A.K. (1976). *Intonation and grammar in British English*. The Hague: Mouton.

't Hart, J., Collier, R. and Cohen, A. (1990). *A Perceptual Study on Intonation. An Experimental Approach to Speech Melody*. Cambridge: Cambridge University Press.

Hockett, C.F. (1958). *A Course in Modern Linguistics*. New York: The Macmillan Company.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge Mass: MIT Press.

Krifka, M. (2006). Basic Notions of Information Structure. In C. Féry, W. Sternfeld (Eds.), *Audiatur vox sapientiae: A festschrift for Achim von Stechow*. Berlin: Akademie Verlag, pp. 287--319.

van Kuppervelt, J. (1994). Topic and comment. In R.E. Asher (Ed.), *The Encyclopedia of Language and Linguistics*. Pergamon Press, pp. 4629--4633.

LABLITA. http://lablita.dit.unifi.it/

Langacker, R. (1991). *Foundation of Cognitive Grammar*. Stanford: Stanford University Press.

McWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Moneglia, M. (2011). Spoken corpora and pragmatics. In H. Mello & S. Gries (Eds.), *Brasilian Journal of applied linguistics* (Revista brasileira de linguìstica aplicada), 11(2), pp. 479--519.

Moneglia, M., Cresti, E. (1997). L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In U. Bortolini, E. Pizzuto (Eds.), *Il progetto CHILDES: strumenti per l'analisi del linguaggio parlato*, vol II. Pisa: Edizioni del Cerro, pp. 57--90.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1, pp. 75--116.

Stalnaker, R. (1974). Pragmatic presuppositions. In K. Milton, M. Munitz & P. Unger (Eds.), *Semantics and Philosophy*. New York: New York University Press, pp. 197--214.

Talmy, L. (1975). Figure and Ground in Complex Sentences. In *Proceedings of the I Annual meeting of the Berkeley Linguistic Society*. Berkeley, pp. 625--649.

Weil, H. (1844). *De l'ordre des mots dans les langues anciennes comparées aux langues modernes. In The order of words in the ancient languages compared with that of the modern languages translation*. Amsterdam: Benjamins.

Prosodic marking of referential status in Brazilian Portuguese: a preliminary study

Alexandre DELFINO, Maria Luiza CUNHA LIMA, Pablo ARANTES

Faculdade de Letras, Universidade Federal de Minas Gerais

Av. Antônio Carlos, 6627 - Belo Horizonte/MG - Brasil

E-mail: aldelfino@gmail.com, marialuiza.cunhalima@gmail.com, pabloarantes@gmail.com

Abstract

Current accounts of referential status (Gundel *et al*, 1993, inter alia) propose that the speaker's mental model is reflected by a variety of linguistic forms during discourse production. In this framework, a scale of numerous informational statuses is related to the degree of givenness of the referents. Languages may mark these different degrees using prosody. For instance, in West Germanic languages such as German and English, it is said that new referents tend to be marked with an intonational prominence, whereas given referents tend to be deaccented. Accessible referents are marked with an intermediate marking, depending on its semantic relationship with previous referents (Baumann, 2006). The aim of this paper is to investigate how different degrees of informational status are acoustically marked along the speaker's discourse in Brazilian Portuguese (BP). This study analyzed word duration, global F0 measures and time-normalized F0 contours of target words in three conditions: new, given and accessible referents. Results show that despite variability across speakers, both duration and F0 are used to mark different statuses. New and given statuses have the most different prosodic patterns and accessible is usually in between the two.

Keywords: information structure; referential status; prosody; Brazilian Portuguese.

1. Introduction

During discourse production, interlocutors refer to entities and events from the real world, and a mental model is built as new information is added and integrated to given, previous information. These entities and events are surfaced as linguistic forms, typically under a form of referential expressions e.g. determiner phrases.

A view widely accepted (Prince, 1981; Gundel *et al*., 1993; Chafe, 1976, 1994; Almor, 1999; Baumann, 2006; Baumann & Riester, 2010) posits referential expressions as taking a whole range of referential statuses, despite a traditional division of referential expressions into given and new information. Consider the following sentence:

> *John had to call the tow service because the engine had broken down on the road.*

In the example above, the referent the engine cannot be taken simply as a given referent. First, a strict morphosyntactic analysis indicates the determiner as typically related to familiar referents yet the referent has not been previously mentioned. On the other hand, it cannot be also considered a new referent, as its meaning can be taken through context. One might conclude that referential forms do not only possess a basic lexical meaning, but also an information status regarding cognitive and contextual factors (Baumann & Riester, 2010, 1). In fact, it seems that 'given' and 'new' information describes both ends of a continuum of referential statuses.

One central question on prosodically-encoded information structure relates to determining which referential statuses can also receive a specific prosodic counterpart. Under a phonological perspective, Baumann & Grice (2006) and Baumann (2006) show that the informational status or activation level of a referent can be either lexically or acoustically marked. In West Germanic languages such as German, a three-way distinction - new, given and accessible - is said to be expressed in terms of prosodic encoding. New referents tend to be marked with a phrasal accent (H*), whereas given referents tend to be deaccented. The acoustic marking of the accessible status is very sensitive to various semantic relationships (e.g. hyponymy, synonymy, meronymy) between the previous item and the referent, and it does not seem to have a well-defined acoustic marking. Baumann (2006) observed that for accessible referents whose semantic relationship to its prime is whole-to-part tend to present an intermediate phrase accent (H+L*). Fowler & Housum (1987) carried out an acoustic analysis of first and second occurrences of words in English and concluded that second occurrence words were shorter than first occurrences.

By means of an electrophysiological measurement experiment (EEG), Schumacher & Baumann (2010) tested how prosodic information can affect reference processing. Two components were analyzed, the N400 and a late-positivity. The results show that reference processing takes prosodic information into account, together with semantic or morphosyntactic marking. 'The data thus show that prosodic information guides the computation of a referent's accessibility and can result in integration costs when less appropriate accent types are encountered' (Schumacher & Baumann, 2010, 620-1). The experiment results also lead to the conclusion that the three-way classification of the referential status is significant not only for production, but also for perception.

The possible acoustic correlates of informational status have not been well studied in Brazilian Portuguese (BP). In a descriptive experimental study of test words nested within noun phrases, Arantes *et al*. (in preparation) show that there are some prosodic differences between new and given referents: (i) new

referents tend to be longer than given ones; (ii) F0 contours of new referents systematically have higher mean F0 values and present wider range and broader standard deviation than given contours and (iii) as a general pattern, contours over the referential expression have two F0 peaks, one NP-initial and other aligned to the stressed syllable of the noun. In new referents the initial peak is usually higher than in the given condition.

In the present experiment we apply the same descriptive tools used by Arantes and collaborators to expand the investigation about the prosodic correlates of referential status in BP beyond the new-given dichotomy. More specifically, we tried to see if the proposed threefold distinction found in German (new, given and accessible) can be found in BP as well. In contrast to Baumann (2006), we focus on the description of the acoustic patterns found and do not provide a phonological interpretation of them.

2.   Experiment

For this study, we designed a corpus of 90 short paragraphs, which were divided distributed into three conditions: given, new and accessible. Each paragraph had one target word, which was embedded in a control phrase.

All target words are four-syllable, penultimate stressed. Relatively long words were chosen as targets because Arantes *et al.* results suggest that prosodic effects due to referential status are more evident when more phonetic material is available to the speaker.

Sentences preceding the control phrase provide context that determine if the target NP is given, new or accessible. The following paragraphs are examples of the three conditions investigated in the experiment. Control phrase in italic and target word in bold italic.

New 🔊

*Um **terremoto** causou destruição* em boa parte da costa leste. Várias cidades não tinham um programa de evacuação, o que deu trabalho para as equipes de resgate.
(*An **earthquake** caused destruction* in a huge part of the East coast. Several cities did not have an evacuation program, which caused problems to the rescue teams)

Given 🔊

O governo decidiu fechar a usina nuclear após o terremoto ocorrido no mês passado. *O **terremoto** causou destruição* no núcleo do reator, aumentando o risco de contaminação. (The government decided to shut down the nuclear plant after the earthquake occurred last month. *The **earthquake** caused destruction* to the reactor nucleus, increasing the risk of contamination.)

Accessible 🔊

Estudiosos da Sismologia têm procurado analisar os dados de tremores para prever novas ocorrências. *O*

***terremoto** causou destruição* sem que ninguém pudesse se prevenir.
(Seismology experts have been trying to analyze the tremors data to predict new occurrences. *The **earthquake** caused destruction* without anyone being able to prepare themselves.)

Four subjects (one male) read the 90 paragraphs, presented one by one on a computer screen in randomized order. Subjects were instructed to read each paragraph silently before reading it out loud to ensure they would be aware of the content of the paragraphs and minimize hesitation. Subjects were recorded in a sound treated room in separate sessions. After each recording session, sound files were edited and labeled. For each sound file the target NP (determiner plus noun) was manually segmented into syllables and the boundaries were stored in Praat metadata files. All acoustical analyses were performed with the help of Praat scripts.

Arantes et al. (in preparation) investigated a wide range of acoustic correlates traditionally linked to prosodic functions in order to find the ones that correlates the best with referential status differences. The authors measured acoustic duration, fundamental frequency, spectral emphasis and long-term average spectrum and suggest that at least for BP duration and fundamental frequency are the best correlates of referential status. Following their suggestion those were the acoustic parameters analyzed in the current study.

Because definite and indefinite determiners have different number of syllables in BP, we decided to measure the duration of the noun instead of the whole NP.   Duration values were extracted and the means across the different values of referential status were then calculated.

Fundamental frequency was analyzed in two ways. First, mean values of central tendency (mean) and variability (standard deviation and range) of the test NPs F0 contours were compared among the referential status values.   For the F0 central tendency analysis, we extracted the mean F0 value of all target NPs (determiner plus noun) and then calculated the mean of the means grouped by referential status value. This measure can be interpreted as the pitch level of the F0 contours. Mean standard deviation and range were obtained applying the same procedure. These two measures were used as estimates of the flatness or "bumpiness" of the F0 contour in each status condition.   Range in semitones was calculated for each contour by applying the formula below. F0max and F0min are respectively the maximum and minimum F0 values in the contour:

$$12*\log_2(\max_{Hz}/\min_{Hz})$$

For the second analysis, individual F0 contours were time-normalized following the procedure described in Arantes (2011), which allows the comparison of F0 contours having different duration. Mean time-normalized contours for each referential status value

were obtained and then visually compared.

The main hypothesis being tested is that 'new' referents, being the most salient, will have longer duration, higher F0 mean, standard deviation and range when compared to 'given' referents. Following Bauman's findings, it's also possible to predict that 'accessible' referents will be in between the two others in terms of the values of the acoustical parameters.

The data generated by the four subjects were analyzed separately. Referential status with three levels (given, new, accessible) was the independent variable for all analysis. Analysis of variance (ANOVA) was used to determine if differences in mean values of the acoustic parameters were statistically different among the levels of the independent variable. An alpha level of 5% was adopted for all analyses. When post-hoc mutiple comparisons were performed the alpha level was adjusted by the Bonferroni correction.

3.   Results

3.1   Duration
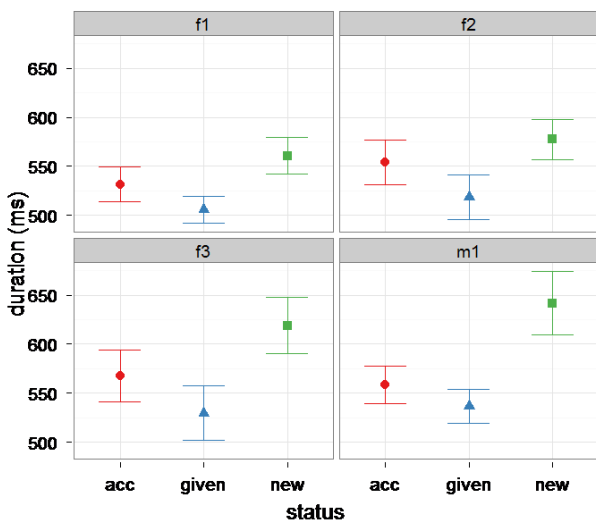
Figure 1: Mean target word duration (in milliseconds) grouped by referential status (accessible, given and new) and subject. Whiskers indicate 95% confidence intervals

Mean target word duration grouped by referential condition and subject is shown in Figure 1. The statistical analysis supports the main hypothesis being tested: for all subjects, 'new' referents were longer than 'given' ones. For subjects f3 and m1, 'new' was also longer than 'accessible'. Differences between 'given' and 'accessible' are never statistically significant. Statistical results by subject are reported below:

- f1: $F(2, 84) = 10.4$ $p < 0.001$; N-G: $p < 0.001$; N-A: p<0.1; A-G: n.s.
- f2: $F(2, 84) = 7$ $p < 0.01$; N-G: $p < 0.01$; N-A: n.s.; A-G: $p < 0.1$.
- f3: $F(2, 84) = 9.9$ $p < 0.001$; N-G: $p < 0.001$; N-A: $p < 0.05$; A-G: n.s.
- m1: $F(2, 84) = 20.5$ $p < 0.001$; N-G: $p < 0.001$;

N-A: $p < 0.001$; A-G: n.s.

3.2   Mean F0

Mean target NP mean F0 grouped by referential condition and subject is shown in Figure 2. The average value of F0 contours of 'new' referents is significantly greater than 'given' and 'accessible' for subjects f1, f2 and m1. There is no difference between 'given' and 'accessible' ones. Subject f3 presented no significant differences among statuses. Statistical results by subject are reported below:

- f1: $F(2, 84) = 6.64$ $p < 0.01$; N-G: $p < 0.01$; N-A: $p < 0.05$; A-G: n.s.
- f2: $F(2, 84) = 7.61$ $p < 0.001$; N-G: $p < 0.01$; N-A: $p < 0.05$; A-G: n.s.
- f3: $F(2, 84) = 0.7$ n.s.
- m1: $F(2, 84) = 123.73$ $p < 0.001$; N-G: $p < 0.001$; N-A: $p < 0.001$; A-G: n.s.
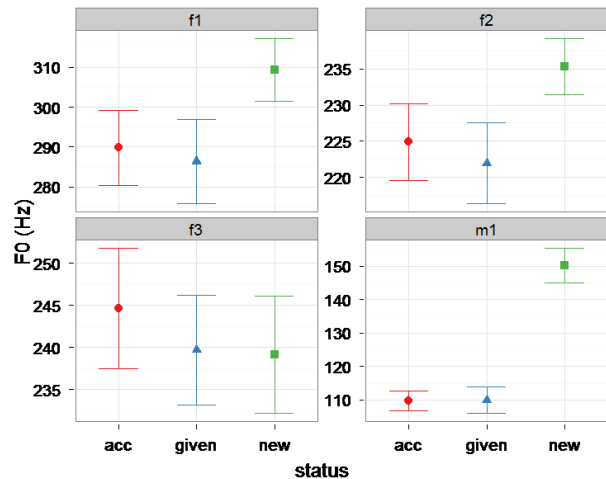
Figure 2: Mean target NP mean F0 (in Hz) grouped by referential status (accessible, given and new) and subject. Whiskers indicate 95% confidence intervals

3.3   Mean SD and Range

Referential status affects standard deviation and range only for one of the speakers, namely m1. Statistical results by subject are reported below:

- f1: SD $F(2, 84) = 3$ $p < 0.1$; range $F(2, 84) = 0.32$ n.s.
- f2: SD $F(2, 84) = 1.8$ n.s.; range $F(2, 84) = 0.03$ n.s.
- f3: SD $F(2, 84) = 0.8$ n.s.; range $F(2, 84) = 0.6$ n.s.
- m1: SD $F(2, 84) = 86$ $p < 0.001$; N-G: $p < 0.001$; N-A: n.s.; A-G: n.s.; range $F(2, 84) = 56$ $p < 0.001$; N-G: $p < 0.001$; N-A: n.s.; A-G: n.s.
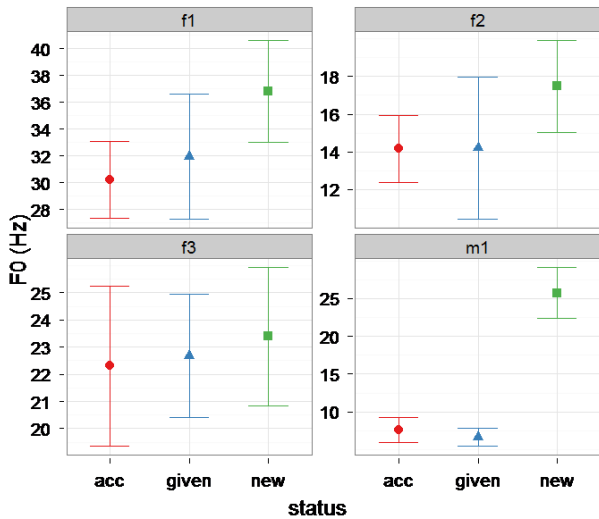
Figure 3: Mean target NP F0 standard deviation grouped by referential status (accessible, given and new) and subject. Whiskers indicate 95% confidence intervals
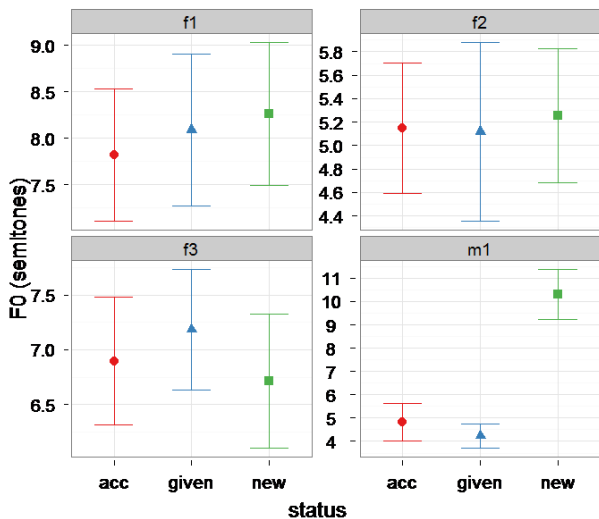
Figure 4: Mean target NP F0 range (in semitones) grouped by referential status (accessible, given and new) and subject. Whiskers indicate 95% confidence intervals

3.4 Time-normalized F0 contours

For all subjects, given and accessible contours overlap significantly. For subjects f1 and f2, new referents contours differ from the other statuses mainly because they present a peak aligned to the first two syllables of the target NP that is absent in the other statuses. Subject m1 given and accessible contours are mostly flat and in a lower register when compared to the new referents, that also have an initial high F0 peak. Subject f3 is unlike the others because referential status does not seem to influence the shape of the F0 contours at all.

On the whole, the results revealed that F0 contours of new referents are different from the given and accessible ones. Despite the individual variability, new referents are characterized by the presence of two major

pitch peaks, one extending over the chain of pre-stressed syllables and other aligned to the stressed syllable. Given and accessible referent contours are very similar to each other.
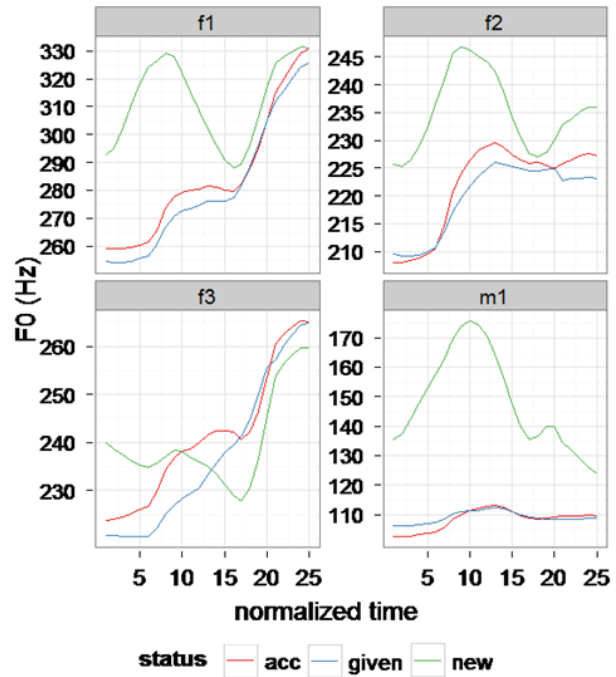
Figure 5: Target NP time-normalized F0 contours grouped by referential status (accessible, given and new) and subject. Syllable boundaries indicated by vertical lines

4. Discussion

In general, there is positive evidence that acoustic parameters are affected by the referential status contrast. Duration seems to be the most robust correlate of the status distinction because it is the only parameter that is affected by the status variable in all subjects.

In addition to duration, mean F0 value is also affected by the referential contrast, with new referents being spoken in a higher register. Except for subject m1, there seems to be little difference in terms of F0 variability between the status categories investigated.

Besides the observed differences in pitch register, the time-normalized contours analysis suggests that the presence of a NP-initial F0 peak can be used as a correlate of the 'new' status in contrast to the other two statuses. In BP, the chain of pre-stressed syllables (including the NP determiner) seems to be an important locus of F0 differences among the status levels.

The lack of a clear distinct acoustic pattern for the accessible status can be evidence that the prosodic marking works, in a general way, associated to other types of information (syntactic position, semantic relationship, register, focus, etc.). Baumann's (2006) results showed that prosodic marking of accessible status were consistently observed in one type of relationship (whole-to-part), and may be of limited use.

5.   Conclusion

The aim of this study is to investigate the relationship between referential status and its prosodic manifestation from the production point of view. Moreover, we intended to observe if any of the analyzed acoustic parameters could set significant differences among the three statuses, i.e. given, new and accessible statuses.

Word duration is the most expressive parameter, followed by the pitch level. Pitch variation and range does not seem to play an important role. New and given status are pretty distinct in most parameters; however, the accessible status is either too sensitive to its semantic relationship to the prime word or it is not relevant for BP speakers as for German speakers.

The current results lead to conclude that there is an interface between referential status and prosodic information, and that relationship is variable in different languages.

6.   Acknowledgements

7.   References

Arantes, P. (2011). Implementação em Praat de algoritmos para descrição de correlatos acústicos da prosódia da fala. Anais da II Jornada de Descrição do Português, pp. 32--38.

Arantes, P., Cunha Lima, M.L. and Barbosa, P.A. (in preparation). Some prosodic correlates of 'new' and 'given' referential status in Brazilian Portuguese.

Baumann, S. (2006). The Intonation of Givenness - Evidence from German. Linguistische Arbeiten, 508. Tübingen: Niemeyer.

Baumann, S., Grice, M. (2006). The intonation of accessibility. Journal of Pragmatics, (38), pp. 1636--1657.

Boersma, P., Weenink, D. (2011). Praat: doing phonetics by computer (Version 4.3.14). Available at: <http://www.fon.hum.uva.nl/paul/praat.htm>

Chafe, W. Givenness, Contrastiveness, Definiteness, Subjects, Topics and Point of View. (1976). In Li, C. (Ed.), Subject and Topic. New York: Academic Press, pp. 25-26.

Chafe, W. (1994). Discourse, Consciousness, and Time. Chicago/London: University of Chicago Press.

Fowler, C., Housum, J. (1987). Talker's signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. Journal of Memory and Language, 26, pp. 489--504.

Gundel, J., Hedberg, N. and Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. Language, 69, pp. 274--307.

Prince, E. F. (1981). Toward a Taxonomy of Given-New Information. In Peter Cole (Ed.), Radical Pragmatics. New York: Academic Press, pp. 223--256.

Schumacher, P., Baumann, S. (2010). Pitch accent type affects the N400 during referential processing. NeuroReport, 21 (9), pp. 618--622.

Topic and Focus marking in an Italian corpus: some results of algorithmic measurement and structural interpretation

Edoardo LOMBARDI VALLAURI[1], Fabio TAMBURINI[2]

[1]Department of Linguistics, University of Roma Tre, Italy; [2]Department of Linguistics and Oriental Studies, University of Bologna, Italy

lombardi@uniroma3.it, fabio.tamburini@unibo.it

Abstract

The results of an experiment on a corpus of spoken Italian suggest a partly new hypothesis on how the main prominence may be interpreted by speakers in the marking of Information Structure (IS). A "topologic" concept of Prominence can be conceived of, as endowed with the function of demarcation between units, beside and before their culmination and characterization. Much of the process by which speakers interpret the IS of utterances may rest upon this, the specific intonational contours of IS units being probably motivated by other functions. In addition, many real utterances seem not always to signal the distinction between Topic-Focus and Broad Focus clearly, remaining rather underspecified in this respect, with no serious effects on communicative dynamism in the subsequent discourse. Such results, obtained by measuring Prominence as a complex entity (not only intonational in nature) strikingly follow the law of least effort. The used algorithm receives confirmation by the fact that automatic measurements and human evaluations of IS patterns show a very high percent of coincidence.

Keywords: information structure; prominence; corpus; spoken italian.

1. Introduction

Acoustic patterns are used to express Information Structure (IS) in linguistic utterances. Adopting the definitions proposed by Cresti (2000) and Lombardi Vallauri (2009), we assume that the Focus is

> "the part of an utterance which carries illocutionary force and realizes the informational purpose of the utterance itself. The Topic, on the contrary, is the part of an utterance that has no illocutionary force, whose function is to allow the comprehension of the Focus with respect to the discourse".

In the present study, Topic and Focus have been located in utterances from two corpora of spoken Italian, by perceptively evaluating acoustic patterns, applying negation tests, and judging which part(s) of utterances convey illocutionary force and New information (Chafe, 1987; 1992). Only three typologies of IS where examined, namely Broad Focus (extending to the whole utterance), Topic-Focus, and Focus-Appendix (i.e. constructions with a Narrow Focus located to the left of the utterance).

Some studies on the matter directly investigate the relations between IS and phonetic phenomena, while others analyse them through an intermediate, phonological level. (e.g. (Ladd, 1996; Pierrehumbert, 1987) and all studies adopting the ToBI labelling scheme (Beckman, *et al.* 2005)). In this second perspective phonological categories are derived from acoustic parameters, mainly considering intonation, i.e. F0 profiles.

Most studies on Italian belong to the Autosegmental Metrical (AM) paradigm, quite often based on read rather than spontaneous speech, and usually examine (typical) tonal profiles, mainly pitch accents, of assertive utterances looking for a specific kind of pitch accent able to mark focalised segments.

Contrastiveness is marked intonationally in Florentine (Avesani & Vayra, 2004), while in Roman (Frascarelli, 2004) and Neapolitan (D'Imperio, 2002b) different pitch accents depend on Focus breadth. It is still unclear whether such differences are due to diatopic variation or to idiosyncrasies of the ToBI transcription scheme. On the one hand ToBI notation seems unable to account for melodic differences clearly perceived by the speakers: Broad Focus of assertive utterances is represented through the same pitch accent although hearers are able to identify the geographic origin of other speakers on the sole basis of intonation (Marotta, 2008). On the other hand, scholars agree on the identification of edge tones and pitch accents, but not about the classification of pitch accents different in nature (Pitrelli, *et al.*, 1994; Syrdal & McGorg, 2000). Disagreement concerns tonal alignment (D'Imperio, 2002a; Gili Fivela, 2002) and tonal target identification, in particular inside plateaux (where a single maximum or minimum cannot be easily discerned) (D'Imperio, 2002a). Information about scaling (i.e. the frequency range within pitch accents) and slope is underestimated, although potentially distinctive (Gili Fivela, 2002).

As suggested in some classical studies (such as Ladd, 1996) and substantiated in more recent investigations (Breen *et al.*, 2010; Lee & Yu, 2010), a focused item might involve a complex combination of different acoustic cues, namely duration, pitch and intensity, and cannot be analysed only through its intonational profile. For these reasons, we will try to investigate the correlation between focused items and phonetic features by considering the concept of prosodic prominence as a complex and rich set of acoustic features combined in a sophisticated way.

2. Prominence definition and automatic detection

Following a common view, we can define prosodic prominence as *a perceptual phenomenon, continuous in*

its nature, emphasizing segmental units with respect to their surrounding context, and supported by a complex interaction of prosodic and phonetic/acoustic parameters.

Due to its methodological rigour, we will primarily refer to (Kohler, 2005) for a description of the interactions between the different prosodic features that determine the perception of prominence. In his view, there are two main 'actors' playing a relevant role in supporting sentence prominence (or sentence accent). The first, *pitch accent,* concerns specific movements in F0 profile. The second, *force accent*, is independent from intonation and is connected with intensity, segmental durations and possibly other parameters. Both phenomena seem to play relevant roles in supporting prominence perception at utterance level (see also Ladd, 1996), reinforcing each other without establishing specific antagonistic or hierarchical roles.

One of the major challenges in predicting syllable prominence is the disentangling of various sources of influence such as fundamental frequency excursions, duration, intensity related parameters and the listeners' linguistic expectancies. At the acoustic level, various studies (e.g. Heldner, 2003; Sluijter & van Heuven, 1996; Streefkerk, 1996) suggest, also cross-linguistically, the dependence of force accents from unit duration and spectral emphasis (spectral tilt or spectral balance), while pitch accents would be supported by specific F0 configurations and by the global intensity inside a particular segmental unit. One of the authors has carried out experiments confirming such relations for some languages (Tamburini, 2005, 2006).

Assuming this view, we can introduce a prominence function which should be able to assign a continuous prominence level to each syllabic nucleus using only acoustic information:

$$Prom^i = W_{FA} \cdot \left[SpEmph^i_{SPLH-SPL} \cdot dur^i \right] + W_{PA} \cdot \left[en^i_{ov} \cdot \left(A^i_{event}(at_M, at_m) \cdot D^i_{event}(at_M, at_m) \right) \right]$$

where $SpEmph_{SPLH-SPL}$ is the spectral emphasis, *dur* is the nucleus duration, en_{ov} is the overall energy in the nucleus and A_{event} and D_{event} are the parameters derived from the TILT model (Taylor, 2000) as a function of the maxima alignment type – at_M – and the minima alignment type – at_m. All parameters are referred to the generic syllable nucleus *i*. See Tamburini (2006) for further details on parameter computation.

The body of the function *Prom* contains nine parameters. Five of them can be considered as supporting the prominence phenomenon from a cross-linguistic point of view ($SpEmph_{SPLH-SPL}$, *dur*, en_{ov}, A_{event} and D_{event}), while the other four, represented in the vector $\mathbf{W} = (W_{FA}, W_{PA}, at_M, at_m)$, can be seen as language specific. In our model, W_{FA} and W_{PA} weigh the contribution of the two different accent types, while at_M and at_m model the different pitch accent alignments specific for each language.

All the parameters involved in the *Prom*-function computation are normalised inside the utterance (using mean and variance), thus the contributions of different speakers and numeric ranges should be factored out. In all the experiments we used $\mathbf{W} = (1.0, 1.0, 2, 2)$.

3.   Experiments

The two experiments presented here were aimed at searching invariancies in position and level of the Main Prominence, identified through the automatic algorithm presented in the previous section, compared to the IS assigned to the utterances by an expert annotator.

The first experiment is a pilot study on a limited corpus of spoken Roman Italian. The second experiment was aimed to verify the results for the same kind of Italian on a different corpus, and to extend the analysis to two further diatopic varieties, namely Florentine and Neapolitan Italian. The annotator identified the mandatory unit of Focus and possible units of Topic and Appendix, if present. He also determined Focus breadth and possible contrastiveness. We will consider here utterances of 3 classes on the basis of IS: (a) TOPIC | FOCUS; (b) BROAD FOCUS; (c) FOCUS | APPENDIX, NARROW FOCUS, CONTRASTIVE FOCUS. The utterances containing retracting, hesitations and speech disfluencies have been discarded.

(a) TOPIC | FOCUS							
Var.-	Main Prominence on the...						No Main Prom
Corp.	LsT	LsF	LsA	IsT	IsF	IsA	
R–B	18	1	-	0	1	-	3
R–C	12	3	-	1	0	-	3
F–C	24	1	-	0	1	-	7
N–C	8	0	-	2	1	-	2
(b) BROAD FOCUS							
Var.-	Main Prominence on the...						No Main Prom
Corp.	LsT	LsF	LsA	IsT	IsF	IsA	
R–B	-	4	-	-	0	-	4
R–C	-	4	-	-	6	-	8
F–C	-	3	-	-	3	-	2
N–C	-	4	-	-	7	-	6
(c) FOCUS | APPENDIX, Narrow F, Contrastive F							
Var.-	Main Prominence on the...						No Main Prom
Corp.	LsT	LsF	LsA	IsT	IsF	IsA	
R–B	-	14	0	-	2	0	0
R–C	-	22	1	-	2	0	2
F–C	-	14	1	-	1	0	2
N–C	-	25	0	-	6	0	0

Table 1: Number of utterances divided by Variety-Corpus pairs (R=Rome, F=Florence, N=Naples; B=Bonvino, C=CLIPS) and configurations (e.g. LsT=Last syl. of Topic, IsF=Internal syl. of Focus). Some combination pairs are not possible; in those cases we have inserted a '-' in the corresponding cells
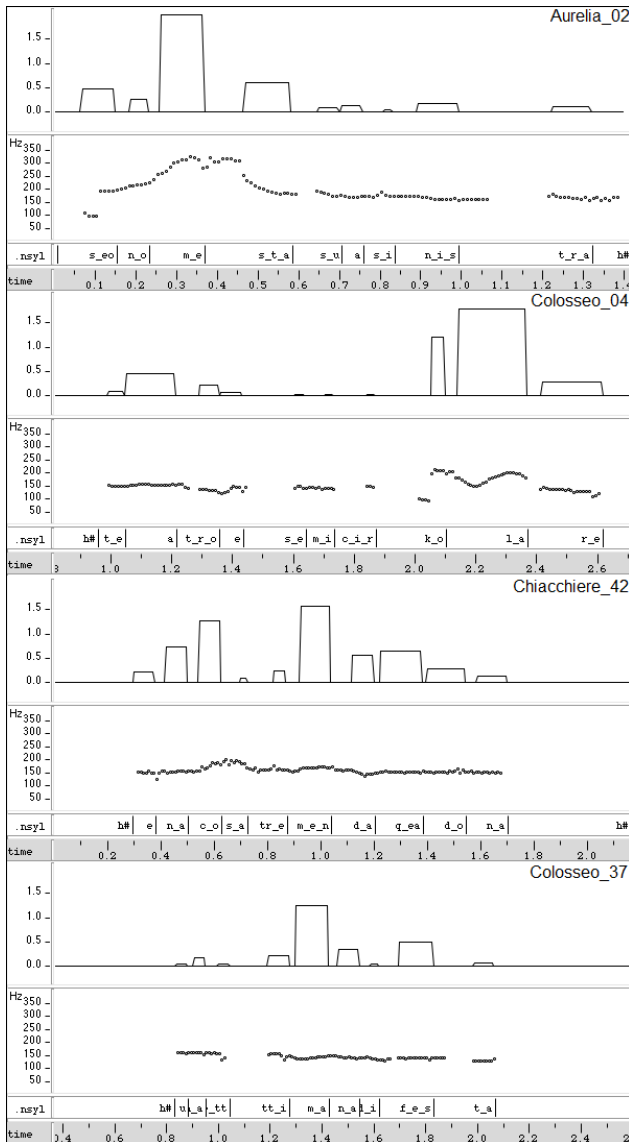
Figure 1: The prominence function profiles – *Prom* – and pitch profiles for some utterances considered in this study.
Aurelia_02: "*Secondo me* T | *stava sulla sinistra* F".
Colosseo_04: "*Il teatro è semicircolare* F".
Chiacchiere_42: "*E' una cosa tremenda* F | *quella donna* A". Colosseo_37: "*Una settimana* F | *di festa* A"

3.1   Experiment 1

The data have been extracted from the "Bonvino" corpus (2005). It consists of 47 utterances selected from 3 out of 12 conversations by speakers from Rome, homogeneous in social level, age, level of education and geographical origin. A reference transcription has been manually added to the extracted waveform to mark the syllabic nuclei needed for prominence identification.

3.2   Experiment 2

The data have been selected from the spoken dialogue sub-corpus of *CLIPS*, stratified through diatopic and diaphasic dimensions (Albano Leoni, 2003). The choice fell on the labeled texts from Rome, to replicate the first experiment using a different data set, Florence and Naples,

so far particularly studied in the AM phonology approach. 184 utterances have been selected: 64 for Rome, 59 for Florence and 61 for Naples.

The results of both experiments, depicted in Table 1, show relevant regularities considering the position of the Main Prominence in relation to the kind of IS. First of all, considering each specific IS, there are no relevant differences between the Italian varieties: the distribution of the Main Prominences seems to follow similar patterns in the different Variety-Corpus pairs. Moreover, the position of the Main Prominence tend to be placed at the border between the two IS components for the TOPIC | FOCUS and the FOCUS | APPENDIX IS, while, in case of BROAD FOCUS utterances, the overall picture seems to be less clear, even if a slight tendency of the Main Prominence to be at the end of the utterance can be found. Figure 1 outlines these regularities for three example utterances from the Bonvino corpus.

It is worth to note that a relevant number of the Main Prominences considered here (e.g. 14 samples out of the 47 from the "Bonvino" corpus) are supported mainly, or uniquely, by force-accents, as shown by the utterance Colosseo_37 in Figure 1, meaning that no intonational phenomena contributed to support them.

These regularities showed to be highly relevant also when testing them by the Fisher exact test.

4.   Demarcation rather than culmination

Table 1 shows that the majority of *Topic-Focus* utterances have the Main Prominence at the Right end of the Topic, while a minority seems not to distinguish between the two units, with comparable Prominences. *Left, Narrow Focus* is always marked by a Main Prominence located at the Right of the Focus itself. About half of *Broad Focus* utterances have the Main Prominence at the Right. The other half show several equivalent Prominences.

In other words, where the Main Prominence is regularly associated is the Right end of constituents located at the Left of the utterance. This suggests that its primary function may be *demarcation*, rather than culmination. There would be *a specific function of the Main Prominence bare presence and position*, whose first effect may be to draw a boundary between two information units, rather than "describing" one of them. For the recognition of *which kind* of units they are, it is sufficient that the contour of the one located to the right signals if this is a Focus or an Appendix.

This may explain that Topics are marked more strongly than both Broad Focuses and Right Focuses after a Topic, though the communicative import of Focuses is greater: because Topics are followed by another major Information Unit, so that the boundary between the two needs to be signaled. Narrow Focuses (at the Left) are also strongly marked, in that they are followed by another information unit within the utterance.

The explanation we propose, based only on the *presence and position*, not on the quality of Prominence and intonation contours, is

A topologic hypothesis on main prominence:

"What is marked through the Main Prominence is the boundary between Information Units within the utterance."

Structurally, the only *qualitative* difference strictly needed in order to recognize the IS of an utterance is that between the marking of a Topic and the marking of a Left (Narrow) Focus, because both are followed by another unit. They can be kept apart either by the different intonation contours of the following units (respectively a Right Focus or an Appendix), and (with some redundancy) by the specific intonational contours of the Topic and the Left Focus themselves. The absence of a Main Prominence, or its being located on the last stressed syllable of the utterance, both signal a Broad Focus (not preceded by a Topic), whose boundaries match those of the whole utterance and don't need to be signaled.

Scheme 1 summarizes the minimal steps by which the addressee can "compute" the IS of an utterance.
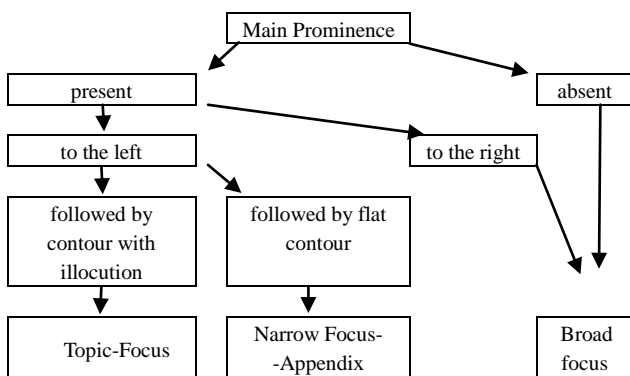
Scheme 1: Steps to the recognition of IS units

If this is true, speakers consistently obey to the *law of least effort*. The only "devices" afforded are (i) one Main Prominence per utterance, and (ii) the difference between a Focus contour and the contour of an Appendix, devoid of illocution. Now, since the different Focus contours are independently needed to express the different linguistic acts, the specific cost required for expressing Information Structure is very low. Culminating each information unit with a specialized Prominence would cost more effort, because distinguishing Topic from Focus would require two different Prominences (one for each) instead of just one at the boundary; and distinguishing Broad Focus from Narrow Focus would require two recognizably different Prominences. As it also happens elsewhere, language prefers to behave economically, marking only the marked element (i.e. Narrow Focus).

5. A continuum

As shown in Table 1, some of the utterances in the corpus that are perceived as Topic-Focus have no Main Prominence. And some of the utterances evaluated as Broad Focuses have an internal Main Prominence, in a position similar to that of Topic-Focus structures.

That utterances acoustically measurable as Broad Focuses can be perceived as Topic-Focus and vice versa, depends on Topic-Focus and Broad Focus being not separate and reciprocally exclusive structures, rather the extremes of a continuum whose center is occupied by utterances with no neat boundary between two units, where the distinction between the two possible ISs remains underspecified. *The speaker is not bound to decide*, at least not prosodically, between Topic-Focus and Broad Focus (possible disambiguation being effected by contextual factors).

In discourse, any content can be focused at different degrees (Daneš, 1974; Firbas, 1989; Sgall *et al.*, 1973), or even remain underspecified from this respect. One should always expect for some utterances to have intermediate status between Topic-Focus and Broad Focus, and to contain information, typically "in the middle", with uncertain information status. In sum, Topic vs. Focus seems not to be a black & white story, rather one in a grey scale.

This is the case for the utterances in Figure 2. Topic-Focus and Broad Focus structures do not always need to be clearly distinguished because they are often possible in the same contexts, and compatible with the same development of discourse.

If we add all utterances underspecified between Topic-Focus and Broad Focus to the patterns explained above within the topologic working of Prominence (summarized in Scheme 1), the matchings between previous perceptive evaluations and the results of measurement all belong in one of the following patterns:

Evaluated IS	Measured position of MP
Topic-Focus	MP at Right end of Topic
Focus-Appendix	MP at Right end of Focus
Broad Focus	MP at Right end, or no MP
Topic-Focus or Broad Focus	No evident MP

The cases that fit this model are almost 90% of the total in the corpus, as shown in Table 2.

	corresponding to the description	not corresponding to the description
Rome – Bonvino	43 (91.49%)	4 (8.51%)
Rome – Clips	55 (85.94%)	9 (14.06%)
Florence – Clips	53 (89.83%)	6 (10.17%)
Naples – Clips	53 (86.89%)	8 (13.11%)
TOTAL	170 (87.88%)	28 (12.12%)

Table 2: confirmation of the analysis by acoustic realizations of IS

6. Conclusions

1. The mere location of Prominence may suffice to signal the demarcation between IS units, allowing speakers to interpret the IS of utterances in discourse. From this respect, the specific intonational contours of the different Information Units may represent a certain amount of redundancy.

2. Acoustically, many utterances remain underspecified for the distinction between Topic-Focus and Broad Focus, with no serious effects on subsequent discourse.

3. Such results seem confirmed by the law of least effort, while the used algorithm receives validation by the very high percent of matching between perceptual evaluations and automatic measurement.
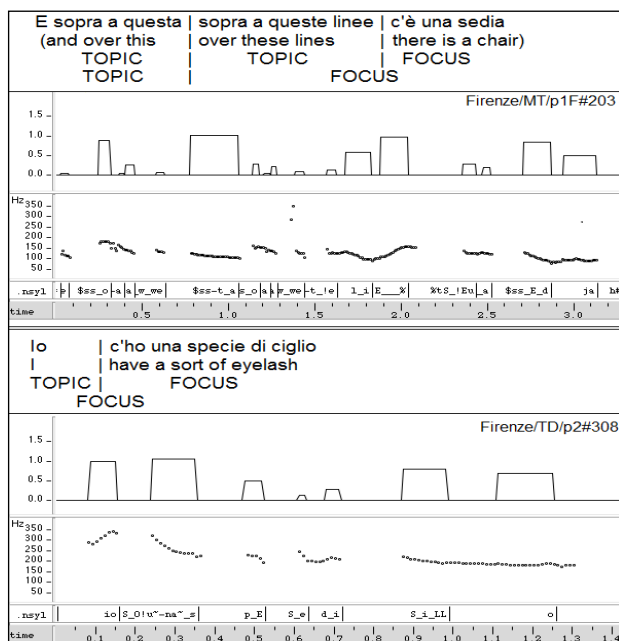
Figure 2: Utterances underspecified between Topic-Focus and Broad Focus

7. References

Albano Leoni, F. (2003). Tre progetti per l'italiano parlato. In *Atti del XXXIV Congr. SLI*, Firenze, pp. 675--683.

Avesani, C., Vayra, M. (2004). Focus ristretto e focus contrastivo in italiano. In F. Albano Leoni, F. Cutugno, M. Pettorino, R. Savy (Eds.), *Il Parlato Italiano. Atti del Convegno Nazionale*, Napoli, pp. 1--20.

Beckman, M.E., Hirshberg, J. and Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S. Jun (Ed.), *Prosodic models and transcription: Towards prosodic typology*, Oxford University Press, pp. 9--54.

Bonvino, E. (2005). *Le sujet postverbal. Une étude sur l'italien parlè*. Paris, Ophrys.

Breen, M., Fedorenko, E., Wagner, M. and Gibson, E. (2010). Acoustic correlates of information structure. In *Language and Cognitive Processes*, 25 (7/8/9): pp. 1044--1098.

Chafe, W. (1987). Cognitive Constraints on Information Flow. In R.S. Tomlin (Ed.), *Coherence and Grounding in Discourse*, Benjamins, pp. 21--51.

Chafe, W. (1992). Information Flow in Speaking and Writing. In P. Downing, S.D. Lima and M. Noonan (Eds.), *The Linguistics of Literacy*, Benjamins, pp. 17--29.

Cresti, E. (2000). *Corpus di italiano parlato*. Firenze, Accademia della Crusca.

Daneš, F. (1974). Functional Sentence Perspective and the Organization of the Text. In F. Daneš (Ed.), *Papers on Functional Sentence Persepctive*. Prague: Academia /The Hague: Mouton, pp. 106--128.

D'Imperio, M. (2002a). Language-specific and universal constraints on tonal alignment: the nature of targets and anchors. In *Proc. Speech Prosody 2002*, pp. 101--106.

D'Imperio, M. (2002b). Italian Intonation: An overview and some questions. In *Probus*, 14(1), pp. 37--69.

Firbas, J. (1989). Degrees of communicative dynamism and degrees of prosodic prominence (weight). In *Brno Studies In English*, 18, pp. 21--66.

Frascarelli, M. (2004). L'interpretazione del Focus e la portata degli operatori sintattici. In F. Albano Leoni, F. Cutugno, M. Pettorino and R. Savy (Eds.), *Il Parlato Italiano. Atti del Convegno Nazionale*, B06, Napoli.

Gili Fivela, B. (2006). Tonal alignment in two Pisa Italian peak accents. In *Proc. of Speech Prosody 2002*, pp. 339--342.

Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in swedish. In *Journal of Phonetics*, 31, pp. 39--62.

Kohler, K.J. (2005). Form and Function of Non-Pitch Accents. In *Prosodic Patterns of German Spontaneous Speech*, AIPUK, 35a: pp. 97--123.

Ladd, D.R. (1996). *Intonational Phonology*. Cambridge University Press.

Lee, Y., Xu, Y. (2010). Phonetic Realization of Contrastive Focus in Korean. In *Proc. of Speech Prosody 2010*, Chicago.

Lombardi Vallauri, E. (2009). *La struttura informativa. Forma e funzione negli enunciati linguistici*. Carocci.

Marotta, G. (2008). Phonology or non phonology? That is the question (in intonation). In *Est. de Fonética Exper.*, Univ. Autònoma de Barcelona, XVII, pp. 177--206.

Mertens, P. (1991). Local prominence of acoustic and psychoacoustic functions and perceived stress in french. In *Proc. ICPhS'91*, Aix-en-Provence, pp. 218--221.

Pierrehumbert, J. (1987). *The Phonology and Phonetics of English Intonation*. Indiana Univ. Linguistics Club.

Pitrelli, J.F., Beckman, M.E. and Hirschberg, J. (1994). Evaluation of Prosodic Transcription Labelling Reliability in the ToBI Framework. In *Proc. ICSLP'94*, Yokohama, pp. 123--126.

Sgall, P., Hajicová, E. and Benesová, E. (1973). *Topic, Focus and Generative Semantics*. Kronberg Taunus, Scriptor.

Sluijter, A., van Heuven, V. (1996). Spectral balance as an acoustic correlate of linguistic stress. In *J.Acoustical Society of America*, 100, pp. 2471--2485.

Streefkerk, B. (1996). Prominent accent and pitch movements. *Inst. of Phon. Sciences Proceedings*, University of Amsterdam, 20, pp. 111--119.

Syrdal, A. and McGorg, J. (2000). Inter-transcriber reliability of ToBi prosodic labelling. *In Proc. ICSLP2000*, Bejing, pp. 235--238.

Tamburini, F. (2005). Automatic Prominence Identification and Prosodic Typology. In *Proc. InterSpeech 2005*, Lisbon, pp. 1813--1816.

Tamburini, F. (2006). Reliable Prominence Identification in English Spontaneous Speech. In *Proc. Speech Prosody 2006*, Dresden, PS1-9-19.

Taylor, P.A. (2000). Analysis and Synthesis of Intonation using the Tilt Model. In *J. Acoustical Society of America*, 107, pp. 1697--1714.

Construction of referents in a corpus of French sportscasts: information structure, syntactic and prosodic realisation, the case of first name + last name referential expressions

Catherine MATHON[1], Sandra AUGENDRE[2]

[1]EA4195 TELEM (TELANCO), Université Bordeaux 3; [2]UMR5263 CLLE-ERSS (ERSSàB), Université Bordeaux 3
mathoncatherine@yahoo.fr, augendre.sandra@wanadoo.fr

The work we present is an analysis of references in the descriptive periods of a sportscast. The successive naming of players and their actions is the very matter of the descriptive part of sport comment. By analysing a corpus of descriptive speech (a rugby match), we want to prove that it is not just a string of player names who currently perform the action but a construction of the referential structure, all along the discourse and in every descriptive period, which makes the discourse coherent. For the 120 descriptive periods of the chosen sportscast, we marked the first introduction of every referent and its further references and resolutions. More precisely, we distinguished the referent activation in a given descriptive period from the coreferent expression(s) or reactivation(s) that followed its introduction, and the properties of all these elements (part of speech, prosodic and syntactic realisations) were noted.

Keywords: reference; French; sportscast; information structure; syntax; prosody.

1. Conceptual and methodological frameworks

The work we present consists in analysing references in a particular type of spoken corpus: live sports comment.

This study is based on a corpus of spoken French, a rugby match sportscast, and we are studying the descriptive periods, directly produced in relation to actions taking place on the field, under the speaker's eyes.

We study the referential structure of these descriptive periods: how referents are introduced and reactivated (information structure) in this particular speech situation at syntactic and prosodic levels?

1.1 The corpus

Our corpus main characteristics (see Lortal & Mathon, 2008 for more information) are:

- Sports event: France-Argentine (Rugby World Cup 2007)
- Sportscasting language: French
- Recorded on TV (TF1)
- Number of speakers: 3
- Duration: 108 mn (total record), 55 mn (total speech), 40 mn (speaker 1), 13 mn (speaker 2)

We distinguish two types of periods:

- Descriptive periods (DP) that are in direct relation to actions
- Comments periods not (directly) related to actions (information about strategy, players' career…)

For this study, we looked at the corpus 120 DP and their structures adapted to a particular production context and motivated by various parameters (see Boulakia & Mathon, 2011).

1.2 Corpus referential structure

At referential level, we consider three types of referents: those introduced for the first time in the discourse (ID), those activated for the first time in a given descriptive period (IP) and the coreferent expressions or reactivations of these referents (R).

Following Chafe (1976: 30) or Lambrecht (1988: 144), we distinguish "newly activated" or "unactivated" referents from "(already) activated" ones.

In the entire match comment, we have 40 different named-entities which are activated as:

- 510 cases of "new" referents in the discourse (ID)
- 470 cases of "new" referents in the DPs (IP)
- 346 cases of "given" referents (R)

In terms of part of speech, 76% of the ID & IP are proper names and more than 50% of the R are pronominal forms (relative pronoun, subject clitic...).

The three examples below illustrate the IP/R distinction:

1. IP + R (relative pronoun):
   les français sont debout avec la balle au fond dans les bras de **Rémy Martin** ouais ah **qui** perce au coeur

2. IP + 2 R (relative & strong pronouns):
   David Skrela qui j(oue) **lui** aussi joue dans les airs

3. IP + Intermediate referents (4) + 2 R (subject clitic and first & last name):
   avec **Rémy Martin** oh les jeux avec <u>Pieter de Villiers</u> avec <u>Heymans</u> maintenant c'est la grande relance française jusqu'à <u>Rougerie</u> <u>Mignoni</u> du rythme du rythme **il** est pris **Rémy Martin**

Our goal is to identify and correlate properties of these three types of referential expressions: syntactic category (NP, PP, Pro, etc.), position, utterance structure, prosodic realization...

1.3   Application to First name (FN) + Last Name (LN) referential expressions

All the corpus referential expressions were annotated, but we began our research by analysing specifically the ones that have the form "first name + last name" and the associate coreferent expressions.

Our test corpus contains 47 ID/IP referential expressions of the form [First Name + Last Name] corresponding to 26 R for these ID/IP.

Our study aims to realize a parallel analysis of these ID/IPs and R and identifies the syntactic and prosodic properties of each kind of referential expressions.

Three levels are taken into consideration:

- A pragmatic level, with three types of referential expressions: ID-IP, IP and R
- A syntactic level, with the identification of syntactic category, function, position, utterance structure, autonomy/dependency…
- A prosodic level, with measures of F0 patterns, F0 register (Low, Medium-Low, Medium-High, High) and F0 range (Delta F0 Max. F0 Min.).

2.   Pragmatic and syntactic structure

In the DPs, discourse organisation is in part related to information progression but also to the iconicity of the production situation:

- Referential expressions' properties depend on the kind of referent: ID/IP are typically introduced by definite descriptions or names, followed by one or more coreferent pronouns...
- Referential expressions properties depend on the sportscaster's relation to the action on the field: speaker's implication and action intensity weigh on the speaker's syntactic and prosodic productions (Boulakia & Mathon, 2009).

At syntactic level, we analysed the referential expressions in terms of syntactic category, function, position, structure and autonomy/dependency. Our goal, here, is to evaluate some tendencies in our corpus such as:

- Postverbal expressions (subject or complement) are typically "new" information and have a particular prosodic realisation,
- Structures like cleft-sentences or dislocations are used to transmit specific information and associated with specific prosodic structures,
- Prepositional phrases and verbless sentences, very present in sportscasts, can be realised as independent groups. To check this last point, we used macrosyntactic categories (Blanche-Benveniste, 1997), to distinguish referential expressions that constitute a unit by

themselves (nucleus, prefix, suffix, postfix) from referential expressions that are embedded in a macrosentence.

The tables below present first the results concerning FN+LN sequences and then those concerning the resolutions of these sequences:

First Name + Last Name	47
Referential category	11 ID-IP & 36 IP
Phrasal part of speech	27 PP & 20 NP
Macro-syntactic category	39/47 embedded 8/47 whole macrophrase
Position within the macrophrases	25/47 at the end 9/47 at the beginning 8/47 whole macrophrase 5/47 in the middle
Function	14 Verb complement 11 Noun complement 8 'utterances' 6 NP's heads 4 subjects 1 ATS 3 juxtapositions

Table 1: Pragmatic and syntactic properties for FN + LN sequences

Let's take some examples of FN+LN expressions:

4. FN+LN at the end of a unit
    à la lutte <u>avec</u> **Nani Corletto**

5. FN+LN at the beginning of a unit
    Felipe Contepomi pour ce drop

6. FN+LN that constitute a unit
    <u>avec</u> **Damien Traille**

7. FN+LN in the middle of a unit
    qui a décalé **Lucas Borges** face à....

As presented in Table 1, referents are mostly not placed at the beginning of a syntactic and prosodic unit. They are mostly not introduced as topics (Player + Action, 9/47), but rather as (part of) focus (Action + Player, 38/47)

For the 8 FN+LN sequences realised as independent units, the principal formal characteristic is that 6 of these expressions are PP, introduced by *avec* ('with') and *pour* ('for').

Table 2 shows corpus resolutions' properties:

Resolutions 26 (17 R1, 8 R2 & 1 R3)	Number of occurrences	Average of intermediate referents
Relative subject	9	0
Last name or First name + Last name	10	2-3
Clitic subject	3	0
NP	3	0
Strong pronoun	1	0

Table 2: Pragmatic and syntactic properties for R

As presented in this second table, there is no intermediate referent between a IP (FN+LN) and a (pro)nominal coreferent expression, whereas there is an average of 2-3 intermediate referents between a IP (FN+LN) and a direct anaphora (FN+LN or LN).

This data indicates that we have to distinguish two types of coreferent expressions: simple anaphora (pronominal forms) without intermediate referents and referents' reactivations ((FN)+LN) with intermediate referents.

Example 8 is a case of reactivation of a referent with FN + LN, since there are intermediate referents:

8. FN+LN(1) + Intermediate Referents + FN+LN(1)
Hernandez le drop avec le pied gauche qui va mourir sous les poteaux où se trouve **Cédric Heymans** avec un arrêt de volée accordé par <u>Monsieur Spreadbury</u> tentative qui a échoué d'un rien hein de la part de <u>Juan Martin Hernandez</u> et le pied gauche de **Cédric Heymans**

In this first example, a referent (Cédric Heymans) is introduced by a FN+LN sequence. After this, two other referents (Monsieur Spreadbury and Juan Martin Hernandez) are mentioned. In order to reintroduce the first referent, the speaker uses a FN+LN sequence, not a pronominal expression that would have been ambiguous.

9. Particular case FN+LN(1) + Ø + FN+LN(1)
ah c'est bien **Rémy Martin** (<u>spk1</u>)
Rémy Martin qui l'a chipé c'est bien (spk2)

This second example is quite particular as there is no intermediate referent between two coreferent FN+LN expressions. There is only this case of direct anaphora [FN+LN] + [FN+LN] in the corpus: the juxtaposition can be explained by the simultaneity of two speaker turns. Speaker 1 introduces the referent and at the same time Speaker 2 (re)introduces the referent for his own production.

3. Prosodic structure

At prosodic level, our study aimed to show that prosody is an indicator of referent activation status, by analysing referential expressions in terms of melodic patterns, F0 range, F0 registers.

Furthermore, prosodic structure helps in identifying corpus macrosentences by distinguishing prosodically autonomous phrases dependent ones.

Our primary interest at this level was to measure the efficiency of prosody as an indicator of the activation status of referents. We focused on melodic variation, especially F0 Range, and F0 registers (Low, Medium-Low, Medium-High, High).
Figure 1 shows F0 range values depending on referents type (ID, IP, R).
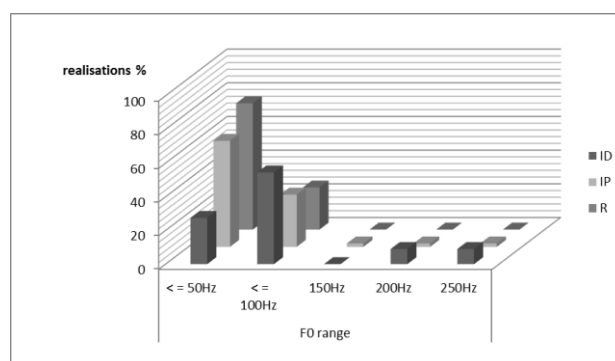
Figure 1: F0 range values depending on referents type (ID, IP, R)

70% of Rs show a very narrow range of variation (50Hz or less).

50% of the IPs present a narrow range of variation (50Hz or less), while 35% show a wider range (100Hz or less).

IDs are more often realised with a wider range of variation (100Hz or less), and it's the only category to show in some cases (20%) a wide range of variation (from 200Hz to 250Hz).

Figure 2 shows the repartition of referent realisations depending on F0 registers and referents type (ID, IP, R).
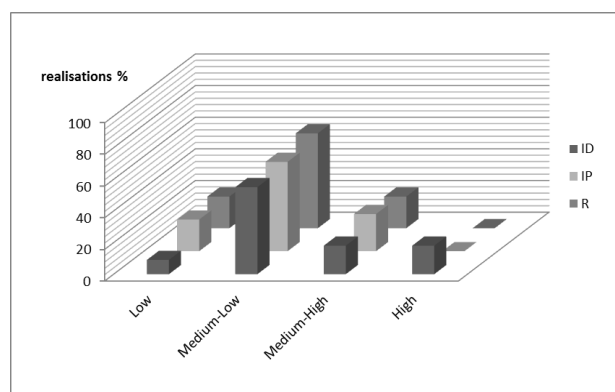
Figure 2: F0 registers depending on referents type (ID, IP, R)

Most of the referents are realised in a Medium-Low register independently from the referential category.

ID referential category is the only one to be realised in some cases in High register, and it is the less to be realised in a Low register.

There is no significant difference between IP and R concerning the F0 registers.

Results showed that there is no statistical evidence for a correlation between F0 range and activation status of referent for one hand, and between F0 registers and activation status of referent for the other. We just note a tendency for High register and wide F0 range being correlated with ID.

4. Prosodic realisations: some examples

We selected some examples of referents prosodic realisations, in order to understand what could be the reason of melodic variations, if not the referent activation status.

We selected utterances presenting the following two referential structures:

- ID-IP + R1 (Clitic subject) + R2 (First Name + Last Name)
- IP1(First Name + Last Name) + IP2 (Nominal Phrase) + R1(PN) + R2 (Relative Pronoun)

We also present a case of prosodic realisation motivated by action on the field.

4.1 ID-IP

Our first assumption was that the referent's degree of activation, at both discourse and descriptive period levels, could be prosodically patterned. The analysis shows that a referent's new introduction in the discourse is not systematically characterized by melodic prominence (see 4.3.).

Figure 3, for example, shows the melodic variations for the ID-IP *Nani Corletto* and its reactivation as a proper name.
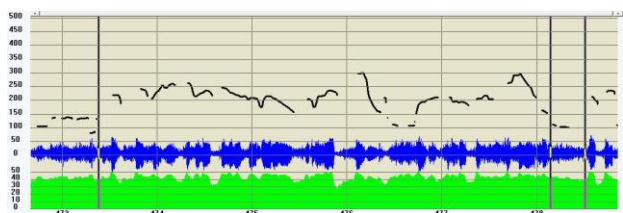
Figure 3: Melodic variations for the utterance
forcer Nani Corletto (IP) à une relance / il (R) y excelle / Nani Corletto (R) une chandelle

The ID-IP Nani Corletto is introduced in a ML Register in an ascending-descending melodic pattern. The referent is first referenced by a clitic subject, which is not accented, then reactivated in a ML Register as well, and in the same pattern, but 50 Hertz lower. This reactivation is a postfix, i.e. an addition of information, given afterward, as a right dislocation. The macrosyntactic function could explain the non-prominent melodic realization of the reactivation (FN + LN).

4.2 IP1 + IP2 + R1 (LN)

Figures 4 and 5 show melodic variations for the utterance

dans le dos de Damien Traille (IP) c'est une touche trouvée par l'Argentine (IP) avec Traille (R) qui (R) saigne hein

which is composed of the following referential structure:

IP1 + IP2 + R1 (PN) + R2
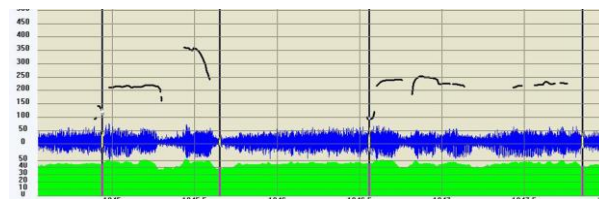
Figure 4: Melodic variations for the utterance
dans le dos de Damien Traille (IP)
[c'est une touche trouvée par l'Argentine (IP) avec Traille (R) qui (R) saigne hein]

On Figure 4, the pitch prominence is on the word *dos* (*back*), while the new referent in the period is activated in a ML register in a plane pattern.
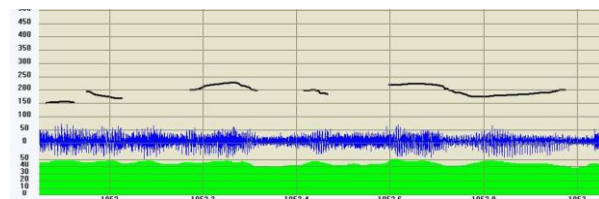
Figure 5: Melodic variations for the utterance
avec Traille (R) qui (R) saigne hein

A new referent *l'Argentine* is introduced. Then the referent *Damien Traille* is reactivated as a Proper Name followed by a relative pronoun, in a type of parenthesis, in a ML-L register and in a plane pattern.

Prosodic realisation does not seem to depend on the degree of activation of the referent, but rather on the impact of the situation on the ground.

4.3 Influence of action

Figure 6 shows melodic variations for IP *avec Pieter de Villiers* in the utterance:

les voilà les meilleurs ballons à jouer avec Jauzion / avec Rémy Martin / {oh les jeux}] / avec Pieter de Villiers (IP) / avec Heymans maintenant c'est la grande relance / française jusqu'à Rougerie
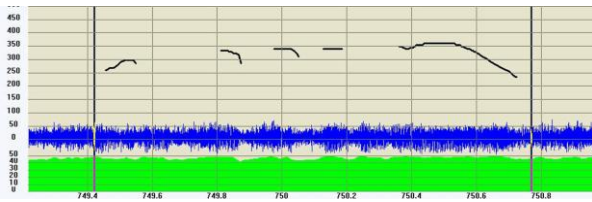
Figure 6: Melodic variations for the utterance
*[les voilà les meilleurs ballons à jouer avec Jauzion /
avec Rémy Martin / {oh les jeux}] / avec **Pieter de
Villiers** (**IP**) / [avec Heymans maintenant c'est la grande
relance / française jusqu'à Rougerie]*

This descriptive period corresponds to a specific offensive action from the French team, whom the speaker supports. He enumerates the players who are part of the offensive action. Each new referent is introduced as an independent unit, a nucleus. The speaker's excitation, his enthusiasm is conveyed by by the use of the MH register, and the F0 range of 100Hz.

In this situation, the iconic function of prosody strongly influences the prosodic realisation of the referent.

5. Conclusion

Our work aimed at describing sport comment's referential structure and more particularly the case of FN+LN sequences and their coreferent expressions.

First of all, the analysis led us to distinguish three types of referential expressions:

- a referent's first introduction by a proper name or a nominal expression,
- a resolution by a pronominal form when the anaphora is direct,
- a reactivation by a proper name or a nominal expression. A referent's reactivation is required as soon as some intermediate referents are introduced between the first introduction of a referent in a descriptive period and its resolution.

Concerning the role of iconicity and sport comment's syntactic and macrosyntactic properties, we can conclude that:

- discourse structure is highly dependent on game actions (iconicity), rather than on information progression principles,
- referents' introduction typically follows action's introduction (81% of the cases) and players-agents are presented as new discourse information,
- specific syntactic structures as *preposition + Proper name* tend to be realized as independent macrosyntactic units

At prosodic level, we saw that there is no prosodic difference between the first introduction of a referent at discourse level, its reactivation(s) in the descriptive periods, and its resolution(s) within the periods. In fact, at descriptive period level, the referential structure is highly related to the action on the field and the iconicity has a more important impact on prosodic realisation than the degree of activation of referents in this kind of discourse.

6. References

Blanche-Benveniste, C. (1997). Macro-Syntaxe (chap.5), In C. Blanche-Benveniste. In A*pproches de la langue parlée en français*, Paris/Gap: Ophrys, pp. 111-123.

Boulakia, G., Mathon, C. (2009). Mise en évidence de l'iconicité prosodique dans le commentaire sportif, Communication orale, ProsIco, 9-10 Avril 2009, Rouen.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C.N. Li (Ed.). *Subject and topic*. New-York: Academic press, pp. 27-55.

Deulofeu, J. (1998). Les commentaires sportifs constituent-ils un "genre", au sens linguistique du terme ?. In *Actes du Colloque Questions de méthode dans la linguistique sur corpus*, Perpignan, 9-11 Mai 98. Paris: Champion.

Lambrecht, K. (1988). Presentational cleft constructions in spoken French. In J. Haiman & S.A. Thompson (eds.), *Clause combining in Grammar and Discourse*. Amsterdam / Philadelphia: John Benjamins, 135-179.

Lortal, G., Mathon, C. (2008). Motion and Emotion or how to align emotional cues with game actions. In *Proceedings of EMOT Worshop*, LREC 2008, Marrakech.

Martin, P. (2009). *Intonation du français*, Paris: Armand Colin.

Mathon, C., Boulakia, G. (2011). Le commentaire sportif en direct: une combinatoire de différentes fonctions de la prosodie, In H.-Y. Yoo, E. Delais-Roussarie (Eds.), *Proceedings from IDP2009*, Interface Discours & Prosodie, Paris, 9-11 Septembre 2009, ISSN 2114-7612, pp. 287--301. Available at:
<http://makino.linguist.jussieu.fr/idp09/actes_en.html>.

Prosodic features of the topic information unit in BP and EP: a corpus based study

Maryualê MITTMANN, Bruno ROCHA

LEEL Laboratory – UFMG
Av. Antônio Carlos, 6627 – Belo Horizonte
maryuale@gmail.com, bbruno791@gmail.com

Abstract

The aim of this study is to describe the topic intonational forms for European and Brazilian Portuguese (EP and BP). The dataset comes from two comparable spontaneous speech corpora: C-ORAL-ROM (EP) and C-ORAL-BRASIL (BP). The theoretical framework is the Language into Act Theory (Cresti, 2000), according to which the utterance corresponds to the shortest linguistic unit that can be pragmatically interpreted. The speech flow is segmented into utterances and in its internal units by prosody. One of the most important informational units it the "topic", which function is to identify the domain of relevance for the illocution conveyed by the utterance. Corpus based studies (Firenzuoli & Signorini, 2003) identified three different intonational forms for the topic unit (types 1, 2 and 3) in Italian. This work shows that BP and EP present the three intonational forms found before and also a fourth one. Prosodic parameters of types 1 and 2 are highly similar in all three languages. Type 3 is the least common form in BP and type 1 is rarely used in EP. Type 4 topics seem to differ in BP and EP regarding the F0 values of the first tonic syllable onset.

Keywords: topic; informational structure; spontaneous speech; corpus; C-ORAL-BRASIL.

1. Introduction

The aim of this study is to describe the topic intonational forms for European and Brazilian Portuguese. The dataset comes from two comparable spontaneous speech corpora: C-ORAL-ROM (European Portuguese – EP) and C-ORAL-BRASIL (Brazilian Portuguese – BP). The theoretical framework is the Language into Act Theory (Cresti, 2000).

2. C-ORAL-BRASIL and C-ORAL-ROM corpora

The C-ORAL-ROM (Cresti & Moneglia, 2005) is a multilingual corpus for Italian, European Portuguese, French and Spanish. It was compiled by a consortium co-ordinated by the University of Florence. The EP session of the C-ORAL-ROM consists of 152 recordings and 317.916 words. The transcriptions are segmented in prosodic/pragmatics units, in order to provide the MEANS ADEQUADO for pragmatic studies. The corpus is integrated with Win Pitch software (Martin, 2004) files, which allows the simultaneous exploitation of the transcription and the acoustic data. The C-ORAL-ROM architecture is designed to cover a large amount of different recording situations in order to document a great variety of speech acts present on spontaneous speech.

The C-ORAL-BRASIL corpus (Raso & Mello, 2012) presents the same architecture of the C-ORAL-ROM and is completely comparable to it. C-ORAL-BRASIL gives a special emphasis on the diaphasic variation of the recordings and it has a very small number or interviews and chats.

3. Language into Act Theory (LAcT)

According to LAcT, the linguistic behaviour is accomplished through speech acts (Austin, 1962). A speech act is understood as the simultaneous performance of three acts: locutionary, illocutionary and perlocutionary. The locutionary act corresponds to the utterance, defined as "the linguistic entity accomplished by the speech act". The utterance is considered the reference unit for the analysis of spoken language and is the shortest linguistic unit that can be pragmatically interpreted (Cresti, 2000). According to LAcT, there isn't a necessary correlation between utterances and propositions and corpus based studies have shown that a high percentage of linguistic units that are pragmatically autonomous don't express a proposition (Crest, 2005).

In this framework, prosody works as an interface between the locutionary and illocutionary acts and it has three important functions: (i) to delimit the utterances in within speech flow; (ii) to assign the illocution conveyed in the utterance; and (iii) to organize information within the utterance.

As for the first function, the utterance is delimited by prosodic breaks perceived by the hearer as conclusive (terminal breaks) and can be parsed into smaller prosodic units (tone units), delimited by prosodic breaks perceived as non-conclusive (non-terminal breaks). The difference between terminal and non-terminal breaks can be seen in example (1), in which "/" marks a non-terminal break and "//" marks a terminal break.

(1) *BAL: as recarregáveis / tão aqui // 🔊

When hearing the sequence until the non-terminal break - 🔊 -, the fluent speaker doesn't perceive the tonal unit as a conclusive and autonomous sequence. It happens only when the speaker hears the whole sequence until the terminal break.

When a utterance is composed by a single prosodic unit it's considered a simple utterance. If the utterance is parsed into two or more prosodic units is, then it is considered complex. The example (1) shows a complex utterance and example (2) shows a simple one.

(2) *PAU: e cê acha que vai gastar mais um // 🔊

The possible internal units are associated with informational functions, through which information is patterned within the utterance. According to LAcT, each prosodic unit corresponds, in principle, to an informational unit. The core of the utterance corresponds to the prosodic unit that bears the utterance's illocutionary force. This unit is called "comment", and it is necessary and sufficient to form an utterance.

Other prosodic units correlate with different information functions, that can be textual (i.e. units that either compose or act on the text), or dialogic (i.e. units that are directed to the addressee and regulate the communicative channel) (Cresti, 2000; Cresti & Moneglia, 2010).

3.1 Topic information unit

Among the textual information functions, the most important and frequent (about 50% of textual units in a BP sample) is the "topic". The topic information function identifies the domain of relevance for the illocutionary act, allowing for the illocution to be distanced from the direct situational context of speech production. The topic provides a linguistic context for the illocution carried by the Comment, when the situational context is not sufficient for the proper interpretation of the speech act (Signorini, 2005). The following example illustrates this.

(3) *CLA: come lei va via la sera /=TOP= nell' ascensore 'un c'è più luce // 🔊

Corpus based studies (Firenzuoli & Signorini, 2003) identified three different intonational forms for the topic unit (types 1, 2 and 3). An intonational form is defined as a set of prosodic features that occurs consistently within an information unit and correlates with its informational function: pitch contour, timing, duration and F0 values. An intonational form is constituted by three distinct tonal portions: preparation, nucleus and coda. The nucleus carries the perceptual prominence associated with the informational function and is, therefore, mandatory. If the syllabic material is greater than what is necessary to accomplish the nucleus, it is distributed in the preparation and/or the coda, which doesn't play any functional role in the topic.

3.1.1. Type 1 form of topic

Type 1 topic is characterized by a rising-falling F0 movement on the nucleus. The rising movement is on the last tonic syllable, and the falling is on the post-tonic syllable(s); tonic and the post-tonic(s) syllables are lengthened.
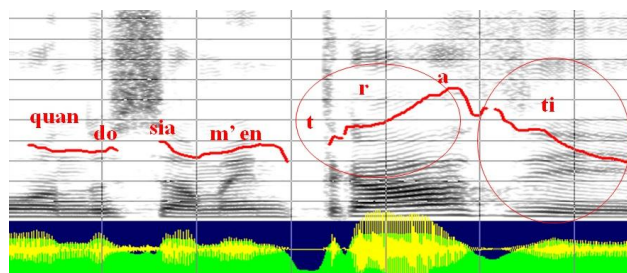
Figure 1: type 1 form of topic 🔊

3.1.2 Type 2 form of topic

Type 2 has a rising intonation profile that begins in the last tonic syllable and continues in any potential post-tonic syllables. Tonic and post-tonic(s) syllables are lengthened.
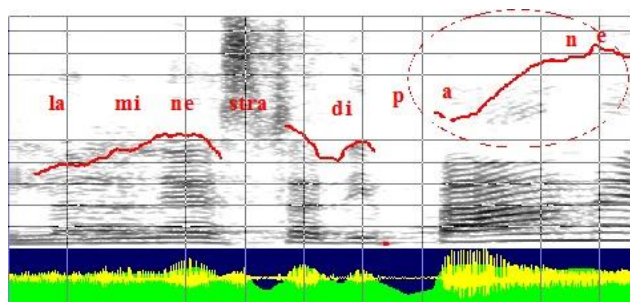
Figure 2: type 2 form of topic 🔊

3.1.3. Type 3 form of topic

Type 3 can be considered holistic, since the nucleus is distributed in two semi-nuclei, together building the topic functional focus. The first semi-nucleus has a falling profile while the second has a rising one, always corresponding to the last syllable of the topic, whether tonic or post-tonic. This syllable is also lengthened.
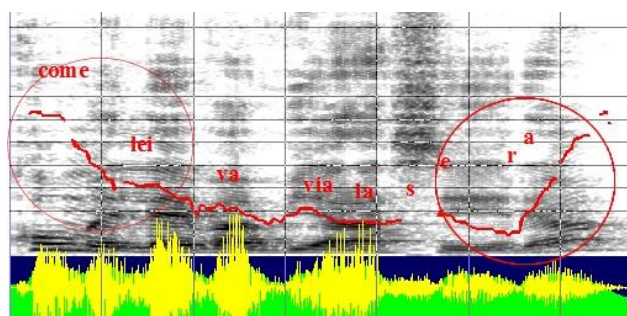
Figure 3: type 3 form of topic 🔊

4.   Methods

In this study, the intonational analysis of topics was carried out in two samples (BP and PE) of speech corpora that were previously prosodically segmented into utterances (simple and complex) according to the methodology developed from the framework of the LAcT (Moneglia & Cresti, 1997). The utterances containing at least one topic unit were extracted from the sample and then it was formed a second sample containing 110 utterances for BP and 72 utterances form EP (EP numbers

are smaller because this corpus wasn't previously informationally tagged). Then we proceeded the analysis through Praat software (Boersma & Weeninck, 2011) through the following steps: (i) identification of the nucleus of the topics; (ii) extraction of F0 values, intensity and duration (syllabic and vocalic) of the nucleus; (iii) stylization of F0 contour of the nucleus; (iv) grouping the topics according to their prosodic parameters; (v) manipulation of prosodic parameters through speech resynthesis. The aim of the manipulations was to identify the most relevant prosodic parameters to the nucleus/semi-nuclei of each intonational form of topic.

4.1 Demonstration of manipulations

In order to clarify the manipulations we did on this research, we present a manipulation of the course of F0 and a manipulation of the duration of the topic of example (4).

(4) *MAR: **e estes espaços** /=i-TOP= por exemplo / **em autores como Camilo Castelo Branco** /=TOP= ou Garrett / são determinantes para a interpretação / &d / dos acontecimentos // 🔊

4.1.1. Manipulation of the course of F0 🔊

With this manipulation we wanted to verify if the rising F0 movement on the topic's second semi nucleus was related to the perception of the topic's function. The original rising F0 movement goes from 182.4Hz to 349Hz. It was manipulated in order to become a flat movement of 182.4 Hz. Hearing both audio files a fluent speaker can notice that the informational unit function is preserved even if the difference between the two audio files can be easily perceived. This manipulation shows that, in this case, the F0 contour of the second semi-nucleus isn't related to the topic's function.

4.1.2. Manipulation of the syllabic duration 🔊

On the other hand, the manipulation of the last tonic's and post-tonic's duration has severe impacts on the unit's function. The original length of these syllables was 0.278s and 0.212s. When they both were reduced to 0.158s (the same duration of another tonic syllable not in the nucleus), the unit is no more recognized as a topic information unit. It means that, for this topic, the duration is a relevant parameter to the perception of the unit's function.

5. Results

PB and PE presents the three intonational forms described for Italian (Firenzuoli & Signorini, 2003), but the EP type 3 form of topic is slightly different from the forms found in BP and Italian: in EP, the rising movement and the lengthening of the second semi-nucleus starts on the last tonic syllable of the topic, and not on the last syllable of the topic. This property can be noticed on example (5).

(5) *MAR: e / **o aspecto dito claramente durativo** /=TOP= é aquele / que / &eh / refere / a relação entre / o discurso do narrador / e / a história // 🔊
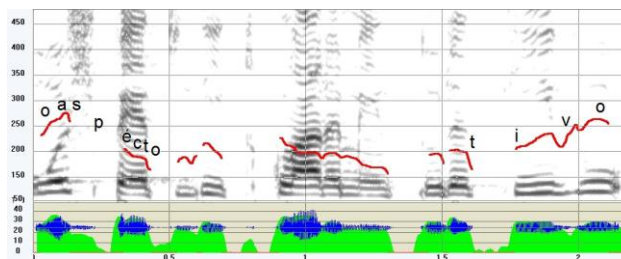
Figure 4: European Portuguese type 3 of topic

Parameters	Syllables					
	o	as	pé(c)	to	ti	vo
syllabic duration (ms)	70	149	200	181	305	258
vowel duration (ms)	70	60	101	75	156	170
F0 peak (Hz)	260	279	251	217	238	265
F0 min (Hz)	234	260	182	178	210	211

Table 1: Example (5) measurements for topic acoustic parameters

However, only one type 3 topic of our BP sample has the stress on the last syllable, which means that further data are needed in order to confirm that BP type 3's topic is similar to Italian's one.

A fourth intonational form of topic (type 4) was found in both Portuguese varieties. Type 4 topic presents two semi-nuclei. The first one is characterized by an extra high onset on the first tonic syllable, with high duration and sometimes intensity as well, followed by a quick pitch fall. The second semi nucleus presents a lengthening and increase of intensity on the last tonic syllable. The F0 contour on the final portion can be either flat, or slightly falling or rising. Example (6) shows a BP type 4 topic with a rising F0 movement on the last tonic syllable. Example (7) shows an EP type 4 topic ending with a falling F0 movement.

(6) *BAL: porque / &he / **de certa forma** /=TOP= a bancada evangélica / eles tão / muito contra / essa coisa / né // 🔊
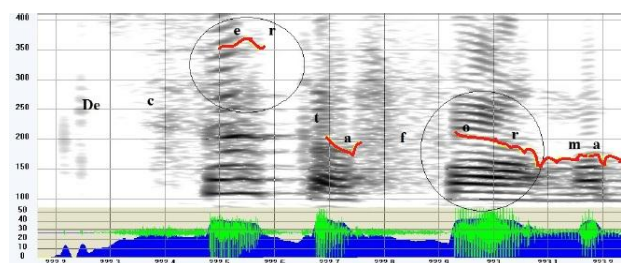
Figure 5: Brazilian Portuguese type 4 of topic

Parameters	Syllables				
	(de)	cer	ta	for	ma
syllabic duration (ms)	-	324	158	361	113
vowel duration (ms)	-	95	57	138	41
F0 peak (Hz)	-	356	189	199	167
F0 min (Hz)	-	340	173	177	149

Table 2: Example (6) measurements for topic acoustic parameters

(7) *MAR: designa-se / **na narratologia** /=TOP= no estudo da narrativa / pausa // 🔊
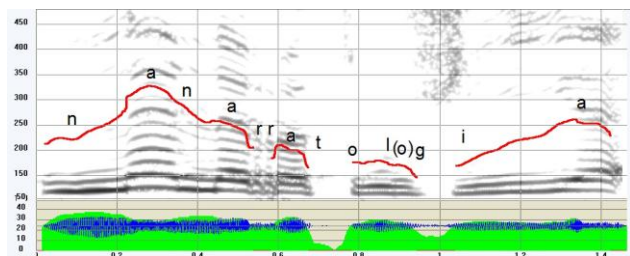
Figure 6: European Portuguese type 4 of topic

Parameters	Syllables					
	na	na	rra	to l(o)	gi	a
syllabic duration (ms)	348	178	142	287	357	149
vowel duration (ms)	121	78	72	172	266	149
F0 peak (Hz)	328	297	234	190	258	262
F0 min (Hz)	218	234	190	155	168	241

Table 3: Example (7) measurements for topic acoustic parameters

Manipulations have shown that when type 4 topics ends with rising movements, the F0 movement tends to be more relevant than the duration for the second semi nucleus. When it ends with a flat or falling F0 movement, duration tends to be more relevant than F0 movements. However, these parameters seems to be in constant interaction and it isn't possible to establish what is the most relevant at all.

The frequency of the intonational forms of topic varies from language to language, as shown by Table 4, on the next page.

Type 4 form of topic seems to be the most commonly used in EP, while it is marginally used in BP. Besides this, European type 4 topics are more complex than Brazilians ones, since they usually have a preparation

between the first and the second semi-nuclei.

Finally, the distribution of the intonational forms of topic regarding the types 1, 2 and 3 for BP is more similar to the distribution found in Italian than in EP.

Lan-guages	Intonational forms of topic				
	TOP1	TOP2	TOP3	TOP4	Total
Italian	58 52%	25 23,8%	22 21%	0	156
BP	39 35%	52 47%	6 5%	13 14%	110
EP	1 1,4%	28 38,9%	4 5,5%	39 54,2%	72

Table 4: distribution of topic intonational forms

6. Conclusions

In summary, BP and EP present the three intonational forms found in Italian and also a fourth one that is not possible in this language. Prosodic parameters of types 1 and 2 are highly similar in all three languages, although it was found only one type 1 topic in EP. Type 3 is the least common form in BP and EP and further data are needed in order to provide a more accurate description. Type 4 topics seem to differ in BP and EP regarding the possibility of a preparation between the semi-nuclei and the F0 values of the first tonic syllable onset (with the Brazilian variety presenting higher values than the European).

Finally, this work raises some interesting questions. Are there any functional differences between the four intonational forms of topic or are they just "intonational allomorphs" (Signorini, 2005)? Why does the distribution of BP intonational forms of topic regarding types 1, 2 and 3 resembles the Italian distribution and not the EP one?

7. Acknowledgements

8. References

Austin, J.L. (1962). How to do things with words. Oxford: Oxford University Press.

Boersma, P., Weenink, D. (2011). Praat: doing phonetics by computer. Available at: <http://www.praat.org/>.

Cresti, E. (2000). Corpus di Italiano parlato. v. 1. Firenze: Accademia della Crusca.

Cresti, E. (2005a) Enunciato e frase: teoria e verifiche empiriche. In M. Biffi, O, Calabrese and L. Salibra (Eds.), *Italia Linguistica: discorsi di scritto e di parlato*. Scritti in onore di Giovanni Nencioni. Siena: Prolagon, pp. 249--260.

Cresti, E., Moneglia, M. (Eds.). (2005). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Cresti, E., Moneglia, M. (2010) Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In A. Panunzi (Ed.), *Bootstrapping Information from Corpora in a Cross-Linguistic*

Perspective. Firenze: FUP.

Firenzuoli, V. (2003). Forme Intonative di Valore Illocutivo dell'Italiano Parlato: Analisi Sperimentale di un Corpus di Parlato Spontaneo. PHD Thesis – Università degli Studi di Firenze.

Firenzuoli, V., Signorini, S. (2003). L'unitá informativa di topic: correlati intonativi. In *Atti delle Giornate del Gruppo di Fonetica Sperimentale - XIII*, Pisa, Novembre 2002 ETS, Pisa, pp. 177--184. Available at: <http://lablita.dit.unifi.it/preprint/preprint-02coll11.pdf>.

Martin, P. (2004). WinPitch Corpus: A text to Speech Alignment Tool for Multimodal Corpora. Lisbon: LREC. Available at:
<http://lablita.dit.unifi.it/coralrom/papers/index.html>.

Moneglia, M., Cresti, E. (1993). Formazione dell'atto linguisitico complesso e intonazione: l'ontogenesi delle relazioni informative in italiano. In E. Cresti, M. Moneglia (Eds.), *Ricerche Sull'acquisizione Dell'italiano*. Roma: Bulzoni editore, pp. 63--114.

Moneglia, M., Cresti, E. (1997). L iéntonazione e i criteri di trascrizione del parlato adulto e infantile. In U. Bortolini, E. Pizzuto (Eds.), *Il Progetto CHILDES Italia.* Pisa: Del Cerro, p.. 57--90.

Raso, T. (2012). O C-ORAL-BRASIL e a Teoria da Língua em Ato. In T. Raso, H. Mello (Eds.), *C-ORAL-BRASIL I. Corpus de referência da fala espontânea informal do português do Brasil*. Belo Horizonte: UFMG..

Raso, T., Mello, H. (Eds.). (2012). *C-ORAL-BRASIL I: Corpus de referência do Português Brasileiro falado informal*. Belo Horizonte: Editora UFMG.

Signorini, S. (2005). Topic e soggetto in corpora di italiano parlato spontaneo. Phd. Thesis, Dottorato di Ricerca in linguistica Italiana, Università degli Studi di Firenze, Departamento di Linguistica, Firenze.

Information patterning strategies in spontaneous speech: a cross-linguistic study

Maryualê M. MITTMANN[1]**, Alessandro PANUNZI**[2]**, Emanuela CRESTI**[2]**, Massimo MONEGLIA**[2]**, Heliana MELLO**[1]**, Tommaso RASO**[1]

[1]Universidade Federal de Minas Gerais, Faculdade de Letras, Belo Horizonte, Brazil; [2]Università di Firenze, Dipartimento di Italianistica, Florence, Italy

maryuale@gmail.com, alessandro.panunzi@unifi.it, elicresti@unifi.it, moneglia@unifi.it, heliana.mello@gmail.com, tommaso.raso@gmail.com

Abstract

We present a cross-linguistic study on information patterning strategies in two romance languages: Italian and Brazilian Portuguese. The language sample comes from two comparable corpora of spontaneous speech: C-ORAL-ROM (Italian section) and C-ORAL-BRASIL. We investigate the occurrence of information units in Italian (IT) and Brazilian Portuguese (BP), thus identifying differences and similarities in the way each language organizes information. Both speech samples are annotated at the informational level according to the textual and dialogic information units established by Language into Act Theory. Results show a prevalence of compound utterances in Italian in comparison with Brazilian, and also an overall tendency in Italian to pattern information at the textual level, while Brazilian presents a more frequent use of dialogic units. These differences could be a result of cultural influences in language use.

Keywords: spontaneous speech; C-ORAL-BRASIL; C-ORAL-ROM; Language into Act Theory.

1. Introduction

In this paper we develop a cross-linguistic study on information patterning strategies in two romance languages: Italian and Brazilian Portuguese. The language sample comes from two comparable corpora of spontaneous speech: Italian section of C-ORAL-ROM (Cresti & Moneglia, 2005) and C-ORAL-BRASIL (Raso & Mello, 2012).

Our aim is to investigate the frequency of occurrence of information units in Italian (IT) and Brazilian Portuguese (BP) and to determine the most frequent information patterns in both languages. We carry out a comparison in the use and distribution of information units according to the type of interaction – monologues, dialogues and conversations (multi-dialogues) – in order to identify the differences and similarities in the way each language organizes information.

2. Theoretical framework

Language into Act Theory (Cresti, 2000) was developed for the analysis of spontaneous speech data. It states a link between prosody, the accomplishment of speech acts (Austin, 1962) the organization of information. The referring unit for the analysis of the spoken language is the utterance, defined as the linguistic counterpart of a speech act. The utterance is the shortest linguistic unit that can be pragmatically interpreted and is delimited in the speech flow by prosodic breaks that bear a conclusive value. Mostly, a prosodically terminated sequence corresponds to the performing of a single speech act. Prosody plays an essential role in the identification of utterances, since through prosody the hearer can perceive the linguistic sequences these pragmatically and prosodically autonomous sequences: the utterances.

The utterance may be prosodically parsed into two or more units, creating a prosodic pattern. The units of the prosodic pattern are associated with informational functions, through which information is patterned in the utterance. Informational Patterning Hypothesis proposes that there is a systematic correspondence between the prosodic pattern and the information pattern of an utterance (Scarano, 2009; Cresti & Moneglia, 2010).

The relation between the prosodic pattern (Hart; Cohen; Collier, 1990) and the information pattern is established by the expression of different information functions with different prosodic profiles. Each prosodic unit corresponds, in principle, to an information unit (IU). The core of the utterance corresponds to the unit that bears the utterance's illocutionary force. It corresponds to the Comment IU.

The comment is the necessary and sufficient unit to form an utterance. Other prosodic units correlate with different information functions, that can be either textual or dialogic. Textual IU participate to the construction of the semantic content of the utterance. Dialogic IU are devoted to the successful pragmatic performance of the utterance (e.g. to regulate the relationship between speakers).

The set of textual information units (and its correspondent tags) is the follow:

a) Comment – COM: accomplishes the utterance's illocutionary force;
b) Topic – TOP: identifies the domain of application for the illocution;
c) Appendix of comment – APC: integrates the text of the comment;
d) Appendix of topic – APT: integration of the information given in the topic;
e) Parenthesis – PAR: adds information with metalinguistic value;
f) Locutive introducer – INT: signals a change of point of view on the subsequent locution.

The dialogic functions are:

a) Incipit – INP: opens the communicative channel while signals a contrastive value with the previous utterance;
b) Conative – CNT: pushes the listener to take part in an adequate way in the dialogue;
c) Phatic – PHA: ensures the maintenance of the communicative channel;
d) Allocutive – ALL: specifies to whom the message is directed, also signaling social cohesion;
e) Expressive – EXP: emotional support of the utterance;
f) Discourse Connector – DCT: signals the continuity of the discourse while establishes a relation between the previous and following units.

There are two cases when one terminated sequence does not correspond to a single illocutionary value: Multiple Comments and "Stanzas".

Multiple Comments – CMM – are a chain of Comments forming an illocutionary pattern. It is an actional model that patterns two or more illocutionary acts for the performance of one conventional rhetoric effect.

A "Stanza" (Cresti, 2009) is a terminated sequence that does not correspond to only one speech act, but to a global linguistic activity, as a result of the intention of performing an oral text, such as narratives and argumentations. It corresponds to a sequence of Bound Comments – COB – with homogeneous illocutionary forces. A "Stanza" may contain other information units forming sub-patterns.

In Language into Act Theory, the information patterning is not explained in terms of given and new information, but rather as the patterning between what conveys illocution and what carries different functions.

3.   C-ORAL: spontaneous speech corpora

The main goal of both the C-ORAL-ROM and the C-ORAL-BRASIL corpora is the documentation of the diaphasic variation, which is needed to represent spontaneous speech. Therefore, besides the variation between private/familiar and public contexts and among the three interactional typologies (monologues, dialogues and conversations), the corpora belonging to the C-ORAL Projects try to document the largest variation in terms of different interaction situations, so allowing a great variation of activity and, as a consequence, of different speech acts and information structures.

As in C-ORAL-ROM corpora, C-ORAL-BRASIL transcriptions incorporate the annotation of prosodic boundaries proposed by Moneglia and Cresti (1997). The annotation scheme segments the speech flow in two distinct levels. The first level deals with the demarcation of the fundamental entity in spontaneous spoken communication (utterances). The second level refers to the internal structure of the utterance, that can be built by

one single tone unit (simple utterance) or by several tone units (compound utterance) (Moneglia & Cresti, 1997; 2006).

In order to study the information structure, the corpus should be tagged regarding information functions. Unlike the tagging of part-of-speech, for which there are already many automatic tools, the tagging of information units is done manually. The samples from IT and BP analysed in this study received informational tagging, using the set of informational units proposed by the Language into Act Theory and the Informational Patterning Hypothesis

4.   Methods

The samples come from the informal sections of C-ORAL-ROM Italian and C-ORAL-BRASIL corpora, selected for a strict comparison with each other. Each sample is detailed in the sections below.

Data were extracted through IPIC, a theoretically-bound XML Database designed for the study of linear relation among Informative Units in spoken language corpora (Panunzi & Gregori, 2012 and in this volume). The database is available for online research and can be accessed at http://lablita.dit.unifi.it/ipic/.

4.1   Brazilian Sample

The selection of texts for the Brazilian Portuguese sample followed a set of criteria adopted to ensure a high quality database to perform information structure studies. At the same time, the same basic structure of the entire C-ORAL-BRASIL informal corpus was preserved (Raso & Mello, 2009).

The BP sample, presented in Table 1 below, has 31318 words, 5483 terminated sequences and 9825 prosodic/information units.

Type	Male/Female	Words	Utterances
Conversations	15/9	9774	2039
Dialogues	6/8	11331	2450
Monologues	7/10	10213	994
Total	28/27	31318	5483

Table 1: Features of BP sample

4.2   Italian Sample

In order to be as much comparable with the BP sample as possible, the Italian sample maintains the same proportion between dialogic and monologic typologies. The texts chosen present a large variety of activities performed by the speakers during the recording sessions.

The Italian sample contains 29414 words, 5276 terminated sequences and 11517 prosodic/information units, as showed in Table 2 below.

Type	Male/Female	Words	Utterances
Conversations	9/11	10141	1986
Dialogues	5/13	12435	1939
Monologues	9/7	11632	1351
Total	23/31	34208	5276

Table 2: Features of IT sample

For more detailed information about the construction of the Brazilian mini-corpus and its comparable Italian counterpart see Mittmann; Raso (2012).

5.   Results

The first important difference to point out regards the distribution of simples versus compound utterances in both samples. In Brazilian shows 71.4% of simple utterance in conversation, 73.6% in dialogue and 55.5% monologue, in Italian these measurements are, respectively, 66.6%, 68.2% and 39.1%.

The prevalence of compound utterances in Italian in comparison with Brazilian is statistically significant (chi-square=52,848 − p<0.0001). Furthermore, in Italian information is more likely to be patterned at the textual level, with high occurrence of compound Utterances with only textual IU (44% of all compound Utterances).

This hypothesis is strengthened by the fact that the number of textual compound utterances is also higher in Italian. While Brazilian shows a percentage of 11.00%, 9.2% and 31.8% of textual compound utterances respectively for conversations, dialogues and monologues, Italian presents 20.0%, 16.00% and 58.9%.

The distribution of illocutionary units shows that the greatest part of illocutionary units for conversations and, specially, dialogues is the Comment unit. In monologues, Bound Comments have a more important role, which is expected, since monologues give rise to more complex and more "textual" discourse, while conversations and dialogues are interactions more action grounded, and therefore present more a greater number and variety of speech acts and dialogic units.

Brazilian shows a relevant use of illocutionary patterns, represented in the Graphic by the Multiple Comments (CMM). Graphic 1 shows the distribution of illocutionary units in both samples.

In Italian there is a strong tendency to organize information in Topic-Comment structures, much more often than in Brazilian Portuguese. Distribution of textual units is showed in Graphic 2.

The only textual unit more frequent in BP is the Locutive Introducer (INT). In Italian, the distribution of INT does not present much variation between conversations, dialogues and monologues, while in the Brazilian the number of INT in monologues is much higher than in the other typologies. This indicates the higher use of reported speech in BP monologues, since reported speech is almost always introduced by an INT unit.
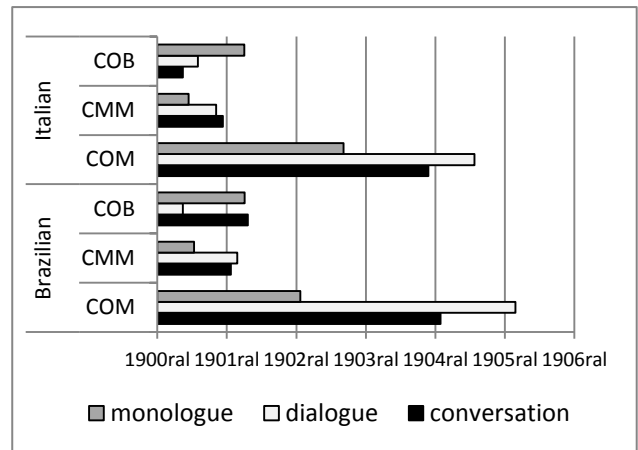
Graphic 1: Distribution of illocutionary units in IT and BP
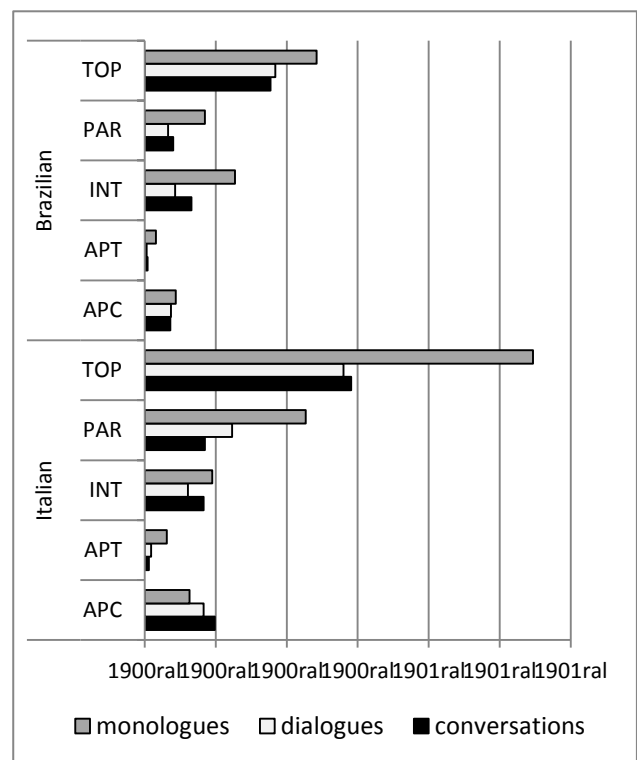
Graphic 2: Distribution of textual units in IT and BP

Graphic 3 below shows the distribution of dialogic units in BP and IT.

Comparing Brazilian and Italian with respect to all the dialogic units, we note that Brazilian uses much more Expressives and Allocutives, while Italian uses much more Incpits and Conatives. When we look at the distribution of dialogic units regarding its position inside the utterance, we notice that the Expressives are very often employed to open the utterance and/or to take the turn. In Italian, those functions are mostly performed by Incipits.

Allocutives and Expressives are signs of social cohesion in discourse, while Incipits signal the speaker's opposition with respect to the previous utterance. It is likely that in Brazilian culture the Incipit is perceived as

an aggressive way to take the turn or begin the utterance. For this reason, Brazilian tends to prefer Expressives to play this role.
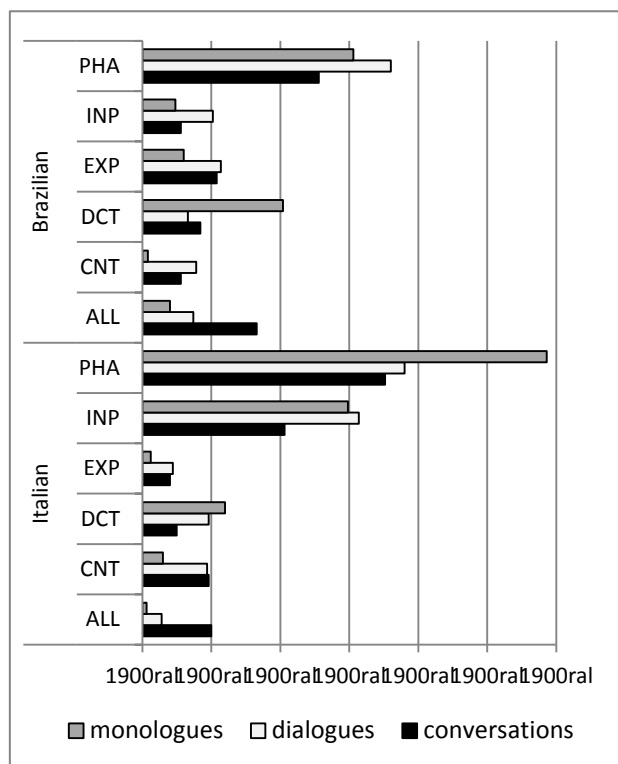
Graphic 3: Distribution of dialogic units in IT and BP

It is important to emphasize that dialogic information units function governing the interaction. Dialogic information units are strongly linked to the interaction (and not the semantic content of the utterance). Therefore, they are sensitive to cultural nuances and, for this reason, they are a good way to investigate how linguistic features can be affected by cultural idiosyncrasies.

6.   Conclusion

The differences observed in the data suggest cultural influences in language use, especially if we consider the distribution of dialogic units. These differences could be interpreted as a result of cultural influences in language use, since dialogic IU like Allocutives and Expressives are signs of social cohesion in discourse. However, a more qualitative look into the data is needed, in order to assure that such differences do not derive from sampling incompatibilities or problems in the information annotation.

Cross-linguistic studies are very valuable, in the sense that through the analysis of different languages we can observe which features are intrinsic to speech as a universal communicative medium and which are specific of each language. Individualizing what is specific to each language is necessary to develop and implement appropriate teaching strategies. The presence of comparable corpora and the study of the information

structure in a contrastive perspective provide many useful elements for L2 teaching. The pragmatic perspective, often invoked in education, still lacks appropriate tools of research. Corpora such as C-ORAL-ROM and the C-ORAL-BRASIL and a theoretical perspective as Language into Act Theory can provide excellent tools to repair this deficiency.

7.   Acknowledgements

8.   References

Austin, J.L. (1962). *How to do things with words*. Oxford: Oxford University Press.

Cresti, E., Moneglia, M. (Eds.). (2005). *C-ORAL-ROM: integrated reference corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins.

Cresti, E., Moneglia, M. (2010). Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In M. Moneglia, A. Panunzi (Eds.), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Firenze: Firenze University Press.

Cresti, E. (2009). Unità di analisi testuale e caratteri costruttivi nell'italiano parlato (spontaneo) e scritto (letterario). Ricerche corpus-based. In A. Ferrari (Ed.), *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione, giustapposizione*. Vol II. Firenze: Cesati.

Cresti, E. (2000). *Corpus di Italiano Parlato*. Firenze: Accademia della Crusca.

Hart, J., Collier, R. and Cohen, A. (1990) *A perceptual study on intonation: An experimental approach to speech melody*. Cambridge: Cambridge University Press.

Mittmann, M.M., Raso, T. (2012). The C-ORAL-BRASIL informationally tagged mini-corpus. In H.R. Mello, A. Panunzi, T. Raso (Eds.), *Pragmatics and Prosody: Illocution, Modality, Attitude, Information Patterning and Speech Annotation*. Firenze: Firenze University Press, pp. 151--183.

Moneglia, M., Cresti, E. (1997). L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In U. Bortolini, E. Pizzuto. *Il Progetto CHILDES Italia*. Pisa:

Del Cerro, pp. 57--90.

Moneglia, M., Cresti, E. (2006). C-ORAL-ROM. Prosodic boundaries for spontaneous speech analysis. In Y. Kawaguchi, S. Zaima and T. Takagaki (Eds.), *Spoken Language Corpus and Linguistics Informatics*. Amsterdam: John Benjamins, pp. 89--112.

Panunzi, A., Gregori, L. (2012). IPIC: an XML data base for the study of Informational Patterning. In *7th LABLITA International Workshop in Corpus Linguistics*, Firenze. Available at : <http://lablita.dit.unifi.it/lablita_workshop/ws7>.

Raso, T., Mello, H. (2009). Parâmetros de compilação de um corpus oral: o caso do C-ORAL-BRASIL. In *Veredas*, pp. 20--35.

Raso, T., Mello, H. (2012). *C-ORAL-BRASIL I. Corpus de referência para a fala espontânea informal do português do Brasil*. Belo Horizonte: UFMG.

Scarano, A. (2009). The prosodic annotation of C-ORAL-ROM and the structure of information in spoken. In L. Mereu (Ed.), *Information structures and its interfaces*. Berlin and New York: Mouton de Gruyter, pp. 51--74.

O Apêndice de Comentário no Português do Brasil: uma análise baseada em *corpus*

Cássia Jacqueline Fernandes OLIVEIRA

Universidade Federal de Minas Gerais

Av. Antônio Carlos, 6.627 - 31270-901 - Belo Horizonte - Brasil

cassiajfo@gmail.com

Resumo

Esse estudo consiste em uma análise baseada em corpora da unidade informacional de Apêndice de Comentário (APC) e teve como objetivo estabelecer uma análise contrastiva dessa unidade no PB e no italiano. Essa pesquisa é sustentada pela Teoria da Língua em Ato (Cresti, 2000), segundo a qual um enunciado é definido como sendo a menor unidade possível de interpretabilidade pragmática. As fronteiras entre enunciados e suas unidades internas são delimitadas pela entonação. Uma das possíveis unidades internas é o APC, que estabelece uma relação de integração com a unidade ilocucionária. Essa investigação foi conduzida em um subcorpus de PB com 20 textos do C-ORAL-BRASIL e em um subcorpus de Italiano com 20 textos do C-ORAL-ROM. A pesquisa demonstrou que no italiano há uma presença maior de unidades terminadas complexas, enquanto o PB apresenta mais enunciados simples. A porcentagem de APC em Italiano nos enunciados simples é quase 50% superior a do PB. Nas estrofes dos monólogos, os APC do PB superam os do italiano. Do ponto de vista entonacional, não aparecem diferenças entre as duas línguas. Informacionalmente, as proporções entre as várias funções nas duas línguas são perfeitamente comparáveis.

Keywords: estrutura informacional; atos de fala; Apêndice de Comentário.

1. Introdução

Ao longo da história, vários pesquisadores se propuseram a estudar a linguagem, principalmente a escrita, deixando em segundo plano a fala. Ainda que semelhante em alguns aspectos, sabe-se que cada uma delas possui as suas especificidades e que analisar a fala pela lente da escrita é um equívoco.

Sem dúvida, uma questão muito discutida pela linguística atual é compreender como o falante organiza a informação na fala, isto é, como se organiza a sua estrutura informacional. A Teoria da Língua em Ato (Cresti, 2000), que serviu de arcabouço teórico para esse estudo, foi desenvolvida para lidar com essas questões, inserindo o estudo da estrutura informacional dentro daquele dos atos de fala (Austin, 1962).

2. A Teoria da Língua em Ato

A Teoria da Língua em Ato (Cresti, 2000) fundamenta-se em um estudo empírico da fala espontânea realizado pelo LABLITA (Laboratorio Linguistico del Dipartimento di Italianistica dell'Università di Firenze).

A fala espontânea, de acordo com essa perspectiva teórica, é considerada como toda a produção linguística sonora dialogada ou monologada em situação natural, realizada livremente, em contextos e situações comunicativas naturais, formais ou informais.

> "A imposição de um molde de segmentação do texto escrito sobre o discurso falado leva o pesquisador a tratar os dados de fala de forma problemática, enviesando especialmente a análise das relações sintáticas no discurso falado. Apesar disso, poucos pesquisadores atentam a este fato e se dão conta da relevância de preservar os aspectos entonacionais da fala em suas transcrições"

(Cresti apud Mittmann, 2012).

Na escrita, todavia, segundo Moneglia (2011) é clara a identificação de unidades linguísticas maiores do que a palavra (unidades da estrutura argumental, sentenças, orações, termos nucleares e dependentes), pois a língua escrita pode ser tranquilamente segmentada de acordo com critérios sintático. Na fala, ao contrario, não é possível utilizar estes mesmos critérios para identificar unidades de referência. Evidências de corpora orais têm mostrado que aproximadamente 30% dos enunciados não apresentam um verbo e não podem ser analisadas conforme parâmetros sintáticos empregados facilmente na escrita.

Em princípio, a unidade linguística que se percebe de maneira mais natural é o turno dialógico. Entretanto, segundo Cresti (2000), o turno dialógico não pode ser considerado como unidade fundamental de referência do discurso falado, porque os turnos apresentam uma ampla variação, podendo ser compostos de apenas uma palavra ou interjeição, ou mesmo de uma longa exposição. O conceito de turno é resultado de uma interpretação antes cognitiva do que linguística.

A Teoria da Língua em Ato parte do princípio de que a unidade linguística da fala deve corresponder à unidade fundamental da atividade comunicativa, já que é esta atividade que 'sustenta' a fala. A unidade linguística fundamental da fala deve corresponder à unidade fundamental da atividade comunicativa: o ato de fala (Austin, 1962). Partindo do princípio de que a fala espontânea consiste na execução de ações, delimitar a unidade de referência da fala deve corresponder a identificar, no fluxo da fala, as sequências linguísticas que se apresentam como suficientes e autônomas do ponto de vista pragmático, isto é, as entidades linguísticas que veiculam ações. Estas unidades são identificadas com o componente linguístico, o ato locutório, do ato de fala, conforme a

perspectiva de Austin (1962). Assim, o enunciado deve ser considerado como a unidade linguística básica da fala, pois corresponde ao componente linguístico de um ato de fala (Cresti, 2000, 2009a; Moneglia, 2000, 2011; Moneglia & Cresti, 1993, 2006).

Esta afirmação fundamenta-se na hipótese de que seja possível estabelecer uma equivalência entre unidades do domínio das ações humanos (atos) e unidades linguísticas (enunciados). Assim, o enunciado é tido como a "contraparte linguística da ação"; isto é, o ato locutório é a contraparte linguística do ato ilocutório, e é interpretável pragmaticamente em autonomia. Isso significa, entre outras coisas, que um enunciado não precisa necessariamente possuir um verbo, e pode, inclusive, ser composto por uma única interjeição, desde que, entoado de maneira a cumprir uma ilocução. Dessa forma, a identificação dos enunciados se realiza através de uma quebra entonacional percebida como conclusiva. Isso significa que uma unidade de enunciado (ou a única se o enunciado for simples) deve ser uma unidade de raiz (o comentário) capaz de veicular autonomia pragmática. Esse princípio baseia-se na teoria perceptiva da entonação ('t Hart, Collier & Cohen, 1990), acarretando uma relação biunívoca entre enunciado e ilocução. A cada enunciado, ou seja, a cada unidade mínima de significado pragmático, corresponde-se uma única ilocução, uma intencionalidade do falante.

Para a Teoria da Língua em Ato, as unidades informacionais são identificadas no enunciado através de três critérios distintos: o critério funcional (função exercida pela unidade no enunciado), o critério entonacional (perfil entonacional característico de cada unidade) e o critério distribucional (posição da unidade no enunciado). Dessa forma, a junção desses três critérios possibilita a identificação das unidades informacionais da fala.

Segundo a Teoria da Língua em Ato há dois tipos de unidades informacionais: as unidades textuais e as unidades dialógicas. As unidades informacionais textuais são aquelas que compõem o texto do enunciado propriamente dito. Dentre elas encontramos as unidades de Comentário (COM), de Tópico (TOP), de Apêndice de Comentário (APC), de Parentético (PAR), de Introdutor Locutivo (INT), de Apêndice de Tópico (APT) e a Unidade de escansão (SCA). As unidades informacionais dialógicas ou não textuais, por sua vez, são aquelas que não contribuem para a constituição semântica de um enunciado, mas dedicam-se ao cumprimento pragmático desse enunciado sendo dirigidas ao interlocutor. São elas: Incipitário (INP), Conativo (CNT), Conector Dialógico (DCT), Fático (PHA), Alocutivo (ALL) e Expressivo (EXP).

A Unidade de Comentário é a mais importante de todas as unidades, pois é a única necessária e suficiente para execução de um enunciado. Sua função é a de realização da força ilocucionária, ou seja, a de cumprir um ato de fala. Entonacionalmente é tida como uma unidade prosódica de raiz que varia conforme o valor ilocucionário; isto é, é interpretável pragmaticamente em autonomia e possui sempre um núcleo, o qual carrega o valor funcional da ilocução. Distribucionalmente pode estar em qualquer posição no enunciado e é com relação a ela que é definida a posição das outras unidades.

A unidade de Tópico é a unidade textual cuja função é especificar no texto do enunciado o domínio de relevância ao qual a força ilocucionária se refere; isto é, o campo de aplicação da força ilocucionária do comentário. Ela tem caráter opcional e é subordinada melodicamente ao comentário, não sendo interpretável autonomamente.

2.1 A unidade de Apêndice de Comentário (APC)

A unidade de APC é por definição uma unidade de integração textual. A maior parte das expressões que são usadas funcionalmente como unidade de APC corresponde a um conteúdo "vazio" ou a um conteúdo genérico do ponto de vista semântico. Funcionalmente, o Apêndice integra textualmente as unidades de Comentário (COM), Comentário Ligado (COB), Comentários Múltiplos (CMM). Entonacionalmente é uma unidade tonal sem foco, com uma F0 sempre mais baixa do que a unidade da qual é apêndice, sempre com perfil nivelado ou descendente e intensidade baixa (Cresti 2000; Ulisses 2008; Oliveira 2009, 2009b, 2010). Distribucionalmente deve suceder a unidade informacional de Comentário. É tida como uma unidade de sufixo.

Ex: *REG: omitir /=COM= só //=APC

2.1.1. A definição de APC e seus critérios de ocorrência

No *subcorpus* de PB foram localizadas 112 unidades informacionais de APC, enquanto que no italiano o número foi de 243 ocorrências. A análise prosódica dos APC encontrados em ambas as línguas revelou a existência de apenas uma forma entonacional (Cresti, 2000; Firenzuoli, 2003) como sendo a de um perfil nivelado e descendente para essa unidade textual.

Observou-se que embora seja uma unidade textual (e ocupe a quarta posição entre as demais unidades textuais no PB, com 10% de ocorrências, sendo mais frequente nos monólogos), a unidade de APC tem como função apenas integrar a unidade de COM, já que não serve de âmbito para aplicação da força ilocucionária como a unidade de TOP, não possui função metalinguística como o PAR, nem introduz uma metailocução como o INT, tampouco possui autonomia pragmática como a unidade de COM.

Funcionalmente deve estar posicionada após a unidade da qual faz a integração, a unidade de COM (CMM e COB) e pode exercer a função de informação tardia, repetição, retomada textual ou preenchimento (Tucci, 2006).

Na tipologia conversação, a informação tardia foi a classificação informacional que mais se destacou com 37,5% de ocorrências. Em seguida encontramos os preenchimentos (34,4%), as retomadas textuais (15,6%) e, por último, as repetições (12,5%). Nos diálogos, há a predominância dos preenchimentos (47,2%), as informações tardias (36,1%), repetições (11,1%) e retomadas (5,5%). Os monólogos procedem de maneira semelhante aos diálogos. Os preenchimentos destacam-se com 45,2% de ocorrências, as informações tardias com 35,7%, as repetições com 16,6% e as retomadas textuais com 2,4%.

Em âmbito geral, notou-se que as unidades de APC no PB desempenham mais a função de preenchimento, com 43% de ocorrências sobre o total de APC, seguido pela informação tardia com 36%, depois as repetições com 14% e, por último, as retomadas textuais incidindo em apenas 7% sobre o total de APC na amostra.

Entonacionalmente, são características da unidade de APC possuir as médias de F0 e intensidade inferiores às médias da unidade de COM e abaixamento do tom de voz. Nessa unidade, nenhum movimento funcional é encontrado, uma vez que ela visa à integração de novas estruturas linguísticas de modo a realizar uma expressão semântica, uma correção ou mesmo uma reestruturação de um enunciado. Do total de 112 APC localizados na amostra, 87% possuem curva analisável e apenas 13% não-analisável.

Distribucionalmente, a unidade de APC localiza-se após a unidade de COM, tendo como função informacional mais comum, nas **conversações,** a informação tardia, (37,5%), os preenchimentos (25%) e as repetições e retomadas textuais juntas (18,8%). Quanto aos **padrões ilocucionários** (CMM), a função mais recorrente é a de retomada textual (6%), repetições e preenchimentos totalizam (6%). Nas **estrofes** (COB), a única função informacional encontrada foi a de preenchimento, com (6%) de ocorrências. Quanto aos **diálogos,** após a unidade de COM, os preenchimentos são mais recorrentes (41,7%), depois aparecem as informações tardias (33,3%), as repetições (8,3%) e as retomadas textuais (2,7%). Nos **padrões ilocucionários,** os preenchimentos se destacam com (5,5%), em seguida com o mesmo valor estão as repetições (2,7%) e informações tardias (2,7%). Nas **estrofes,** a única função exercida pelo APC é de retomada textual (2,7%). Quanto aos **monólogos,** após a unidade de COM, as informações tardias e preenchimentos possuem o mesmo valor percentual (26,2%) cada uma. Em seguida estão as repetições com (16,7%) e as retomadas textuais com apenas (2,4%). Nos **padrões ilocucionários** (CMM), a maior função desempenhada pelo APC é de preenchimento (7,1%), depois de informação tardia (2,4%). Nas **estrofes** (COB), a categoria informacional mais saliente é a de preenchimento (11,9%), seguida apenas pela

informação tardia (7,1%). Distribucionalmente, ainda nos atentamos para o fato de que entre a unidade de Comentário e a unidade de APC podem aparecer intercaladas as seguintes unidades: alocutivos, fáticos, conativos, expressivos e parentéticos; todavia, em nosso estudo, só foram verificadas as presenças das unidades de alocutivo, parentético e fático.

Observou-se, ainda, que há determinados contextos em que duas outras unidades podem ocupar a mesma posição do APC, com um perfil entonacional às vezes parecido, e serem confundidas com o APC. Tratam-se das unidades de PAR e COB. Para distinguir uma unidade de PAR de uma de APC, primeiramente, deve-se observar que se se retirar a unidade de APC percebe-se, frequentemente (do ponto de vista prosódico), a falta de algo para a realização completa do enunciado; o mesmo não ocorre quando da eliminação da unidade de PAR. Segundo, a unidade de PAR possui sempre valor modal ou, pelo menos, constitui uma intervenção metalinguística, cujo ponto de vista é externo àquele do resto do enunciado.

Quanto aos COB em posição de um possível APC, as principais pistas para decidir se tornam o valor cognitivo de 'novo' e as medidas de F0 e intensidade.

Outra situação que merece ser mencionada é a da coda. A coda ocorre quando há uma unidade tonal que parece ter todas as características prosódicas de um APC, mas com características informacionais distintas. Isso pode acontecer quando temos um COM cujo foco funcional é à esquerda e cujo conteúdo locutivo se estenda por várias sílabas de coda. Essa situação faz com que seja impossível ou pelo menos não natural a realização do COM, sem que se produza uma quebra entre o foco e o restante do conteúdo silábico; assim sendo, não há como afirmar que depois da quebra haja um APC, já que essa quebra é praticamente obrigatória; trata-se, portanto, de uma coda que produz uma unidade escansionada (SCA) à direita e não, como é frequente, à esquerda do foco funcional.

Morfossintaticamente, 67% do total de APC analisados nesse estudo são construções sintáticas e 33% são expressões, sendo os ADV os que mais se sobressaem com função de APC em qualquer uma das tipologias. Esse resultado já era esperado, porque é sabido que dentre as várias funções do ADV está a de determinar um fato, ampliando a informação nele contida, função essa desempenhada pelo APC. Outras categorias são bem mais raras, em PB.

Interessante destacar, ainda, a distinção estabelecida entre uma sequência de dois apêndices e um apêndice escansionado. Enquanto o primeiro apresenta um perfil prosódico concluído , o segundo, não.

3.   Análise Contrastiva: APC no PB versus APC no italiano

A fim de estudar a estrutura informacional em uma perspectiva interlinguística, buscou-se analisar o

comportamento da unidade de APC no PB e no italiano. Como o subcorpus brasileiro é altamente acional, para se estabelecer um parâmetro de comparação foi necessário manter a mesma proporção entre as tipologias diálogo e monólogo, e maximinizar o número de atividades realizadas pelo falante no momento da interação.

Observamos que no PB há 82,7% de enunciados nas conversações, 85,6% nos diálogos e 66,6% nos monólogos. No italiano o percentual é parecido, à exceção dos monólogos. Nas conversações há 83,7% de enunciados, nos diálogos 83,4% e nos monólogos, 70,5%. Esses resultados nos permitem afirmar que no italiano os textos dialógicos se comportam de maneira similar aos textos do PB, havendo proporcionalidade de enunciados quando da comparação. Os textos monológicos, entretanto, apresentam medidas muito diferentes, permitindo-nos aventar a hipótese de que as pequenas diferenças encontradas entre as duas tipologias no subcorpus do PB são devidas à presença de duas conversações em que os falantes não realizam qualquer atividade levando, portanto, algumas medidas na direção dos monólogos.

Quanto aos monólogos no italiano, a diferença fundamenta-se no fato de haver nessa tipologia tanto menos estrofes e padrões ilocucionários, quanto mais enunciados.

É interessante notar que a diferença mais significativa entre uma língua e outra está no fato de que em termos percentuais o italiano tem menos enunciados simples que o PB. Enquanto no PB há 71,4% de enunciados simples nas conversações, 73,6% nos diálogos e 55% nos monólogos, no italiano essas medidas são, respectivamente, 66,6%, 68,2% e 39,1%, o que nos conduz ao fato de que no italiano há mais enunciados complexos do que no PB. Esta hipótese é reforçada pelo fato de que as unidades textuais também são superiores no italiano.

Quanto aos enunciados complexos com unidades textuais, enquanto na amostra brasileira o percentual é de 11%, 9,2% e 31% de enunciados complexos com unidades textuais para conversação, diálogo e monólogo, no italiano, respectivamente, encontramos 20%, 16% e 58,9%. O mesmo acontece com os padrões ilocucionários. No italiano, os padrões ilocucionários são mais comuns, enquanto os padrões ilocucionários simples são mais encontrados no PB. A diferença quanto as estrofes não parece muito significativa.

Outra diferença interessante entre os dois subcorpus diz respeito à inversão da distribuição dos APC nas três tipologias. Enquanto no PB há mais APC nos monólogos e menos nas conversações, no italiano os APC são mais comuns nas conversações e menos nos monólogos.

Enquanto no italiano há 243 ocorrências de APC sobre 1018 unidades terminadas com unidades textuais, no PB encontramos apenas 112 APC sobre 1012; isto é, a unidade de APC é muito mais recorrente no italiano,

pois apresenta mais que o dobro (58%) dos APC encontrados no PB.

Em relação aos enunciados complexos com unidades textuais, o que se verifica é que há 3,1% a mais de ocorrência de APC, no italiano, na tipologia diálogo (conversação e diálogo) e nos monólogos apenas 0,8%. Nos padrões ilocucionários há mais 2,8% de ocorrências de APC no italiano do que no PB nos diálogos (conversações e diálogos), e 4%, nos monólogos. E em relação às estrofes, no italiano há mais 4,2% de ocorrências de APC do que o PB nos diálogos (conversações e diálogos), e 0,7%, nos monólogos.

Em síntese, o italiano mostra uma presença muito maior de unidades terminadas complexas, enquanto o PB apresenta muito mais enunciados simples. Esse é um aspecto muito interessante para a comparação entre as duas línguas. Mesmo considerando como *baseline* somente os enunciados complexos, a porcentagem de APC em italiano é quase 50% a mais do que em PB. O maior número de APC se deve principalmente ao padrões ilocucionários, mas também aos enunciados. Não há diferença quanto às estrofes. É interessante notar que nas estrofes dos monólogos, quando a fala necessariamente se complexifica, os APC do PB superam os do italiano nos monólogos. Do ponto de vista entonacional, não aparecem diferenças entre as duas línguas. Do ponto de vista informacional, as proporções entre as várias funções nas duas línguas são perfeitamente comparáveis. Do ponto de vista morfossintático, a comparabilidade entre as duas línguas é estrita.

Assim sendo, a pesquisa confirma a análise da fala espontânea do PB, com base em uma teoria elaborada a partir do italiano; comprova em detalhes as características informacionais, entonacionais e morfossintáticas da unidade de APC; observa que essa unidade é menos presente no PB do que em italiano, e que essa menor presença se justifica pelo fato de, em geral, a estrutura da fala em italiano aparecer mais rica de unidades textuais do que a fala em PB (à exceção do INT).

4. Referências

Austin, J. (1962). How to do things with words. Oxford: Clarendon.

Conte, A.G. (1977). Aspetti dela semântica del linguaggio deontico. In Bernardo, G. di (Ed.), *Logica deontica e semantica*, Bologna, Il Mulino, pp. 147--165.

Cresti, E. (2010). The definition of Focus in Language into Act Theory. In *V° LABLITA workshop &II° Brazilian Seminar on Pragmatics and Prosody: Illocution, modality, attitude, information patterning and speech annotation*. Anais... . Belo Horizonte.

Cresti, E. (2009). La stanza: un´unità di costruzione testuale del parlato. In *Atti del X Congresso SILFI: sintassi storica e sincronica dell´italiano.*

Subordinazione, coordinazione, giustapposizione. Firenze: Cesati.

Cresti, E. (2005a). Enunciato e frase: teoria e verifiche empiriche. In M. Biffi, O. Calabrese and L. Salibra (Eds.). *Italia Linguistica: discorsi di scritto e di parlato. Scritti in onore di Giovanni Nencioni*. Siena: Prolagon, pp. 249--260.

Cresti, E. (2000). Corpus di italiano parlato. Firenze: Accademia della Crusca, 2 voll.

Cresti, E., Moneglia, M. (2010). Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In A. Panunzi (Ed.), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Firenze: FUP.

Cresti, E., Moneglia, M. (2007). C-ORAL-ROM. Comparing Romance Languages in Spontaneous Speech Corpora. In T.C. Silva, H.R. Mello (Eds.), *Conferências do V Congresso Internacional da Associação Brasileira de Lingüística*. Belo Horizonte: UFMG.

't Hart, J., Cohen, A. and Collier, R. (1990). *A perceptual study on intonation: an experimental approach to speech Mellody*. Cambridge: Cambridge University Press.

Marcuschi, L.A. (2001). *Da fala para a escrita: atividades de retextualização*. São Paulo: Cortez.

Mittmann, M. (2012) O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o Tópico no Português Brasileiro. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte.

Moneglia, M. (2011). Spoken corpora and pragmaticis. In *Anais do congresso RBLA*, Belo Horizonte, v. 11, n.2, pp.479--519.

Moneglia, M. (2000). Specifications on the C-ORAL-ROM Corpus. Disponível em: <http://lablita.dit.unifi.it/coralrom/papers/Specifications-CORALROM.pdf>.

Moneglia, M., Cresti, E. (2006). C-ORAL-ROM. Prosodic boundaries for spontaneous speech analysis. In Y. Kawaguchi, S. Zaima, T. Takagaki (Eds.), *Spoken Language Corpus and Linguistics Informatics*. Amsterdam: John Benjamins.

Oliveira, C. (2009). Apêndice de Comentário e Comentários Ligados: uma distinção à luz da Teoria da Língua em Ato. In *Revista Eletrônica de Divulgação Científica em Língua Portuguesa, Linguística e Literatura Letra Magna*. Ano 05, n.10,1° semestre. Disponível em: <www.letramagna.com>.

Oliveira, C. (2009b). Distinção entre Apêndice de Comentário, Comentários Ligados e Inciso, três unidades informacionais, em final de enunciado, à luz da Teoria da Língua em Ato. In *Revista e_Hum* V2. n.1.

Oliveira, C. (2010). O Apêndice de Comentário: classificação informacional e morfossintática à luz da Teoria da Língua em Ato. In *Revista Eletrônica de Divulgação Científica em Língua Portuguesa, Linguística e Literatura Letra Magna*. Ano 06, n.13, 2° semestre. Disponível em: <www.letramagna.com>.

Pottier, B. (2000). *Représentations mentales et catégorisations linguistiques*. Louvain e Paris, Peeters.

Raso, T., Mello, H. (Eds.) (2012). *C-ORAL-BRASIL I. Corpus de referência da fala espontânea informal do português do Brasil*. Belo Horizonte: UFMG.

Tucci, E. (2005/2006). L'unità di appendice in un corpus di italiano parlato (C-ORAL-ROM): caracteristiche intonative, semantiche e morfo-sintattiche. Tesi di laurea triennale in italianistica. Universitá degli studi di Firenze, Facoltà di lettere e filosofia, anno academico.

Ulisses, A.J. (2008). A unidade de apêndice no português do Brasil. Faculdade de Letras da UFMG. (Dissertação de Mestrado).

Intonation and information packaging: evidence from Brazilian Portuguese

Lílian Teixeira de SOUSA

Unicamp

Campus Universitario Zeferino Vaz SN, Cidade Universitaria – 13083-970 – Campinas, Brasil

liliantsousa@gmail.com

Abstract

Intonation is generally understood to be related to information structure. Many scholars claim that new information is characterized by high tone (H), whereas given information is characterized by low tone (L). For Kohler (2004) and Baumann (2008), the NEW/ GIVEN distinction do not explain entirely the relation between intonation and information packaging in German. According to them, it is also relevant to distinguish degrees of givenness. Taking into account the relevance of prominence relations for information packaging, my aim in this paper is to investigate which type of accentuation is used to indicate given/ new information in Brazilian Portuguese (BP) as well as to determine whether degrees of givenness play a central role in intonation for the language in question. Since the contrast-emphasis paradigm is important to the Focus-Background dimension, an additional goal is to compare the interaction between intonation and contrast-emphasis. The analyses have shown relevant relations, on the one hand between falling contours (HL, >HL, LHL) and contrast and, on the other hand, between rising contours (LH, >LH, HLH) and emphasis. As for the degree of givenness, I have not observed a one-to-one relationship between degree of givenness and intonation.

Keywords: information structure; degree of givenness; intonation.

1.   Introduction

Intonation is said to play a role in information packaging in many different languages. English and German, for instance, exhibit similar intonation patterns: high tones (H) indicate new referents in discourse whereas low tones (L) indicate given referents. Some scholars (Kohler, 2004; Baumann, 2008) have distinguished degrees of givenness in their experiments, gathering interesting data. In order to establish the relevance of prominence relations for information packaging, many experimental studies have been conducted in attempt to find similar patterns in other languages. Since studies on intonation and information packaging in Brazilian Portuguese (BP) are scarce, my aim in this paper is to investigate which type of accentuation is used to indicate given/ new information in this language and also to determine whether degrees of givenness play a role in intonation in BP.

Since the contrast-emphasis paradigm is important to the Focus-Background dimension, an additional goal is to investigate the interaction between intonation and contrast-emphasis.

This paper is organized as follows: in Section 2, I summary the literature on intonation and information packaging. Section 3 describes the methodology applied in this study. In Section 4, I examine the relationship between intonation and information packaging in BP. In section 5, I conclude the paper.

2.   Intonation and Information Packaging

Understood as being pitch variation in the course of an utterance, intonation is therefore related to information structure. Many scholars agree that new information is marked by high tone (H) whereas given information is marked by low tone (L). However, Kohler (2004) and Baumann (2008) have argued that the NEW/ GIVEN distinction do not explain entirely the relation between intonation and information packaging in German. In this language, it is also important to distinguish degrees of givenness. Baumann (2008) discusses different aspects of information structure which have been confined in the literature to the concept of givenness (New). According to him, between many issues and terminology found in the literature, three basic dimensions of information structure are often mentioned: 1) The division between what the utterance is about and what comments on it – Theme-Rheme; 2) the division of an utterance into an informative part and an uninformative (or newsworthy) part – Background-Focus; and 3) the cognitive representation of the referent or proposition in the interlocutor's mind – Given-New. The first two dimensions are relational in nature and apply to the level of sentence or utterance, whereas the third dimension is non-relational and applies to the level of discourse (ex. (42) Baumann, 2008):

> (1) A: What about John?
>   B1: {My sister and me} {are going to visit him}
>   Theme   Rheme
>   B2: {My sister and me are going to visit}{him}
>   Focus   Background
>   B3: {My sister} and {me} {are going to visit}{him}
>   New   Given   New   Given

Considering the pragmatic partitioning of an utterance into a focus-background structure, Baumann argues that new information always occurs in the focus part, whereas given or accessible information can occur in the focus or background. Baumann also observes that accentuation does not depend only on the degree of activation. Baumann (2008: 99) says "if a speaker wishes to present a constituent as particularly newsworthy, s/he can highlight this constituent irrespective of its activation status". This happens in contrastive utterances, in which given information may be focused using a particularly salient accent involving a pitch higher than the speaker's

topline.

Authors like Pierrehumbert & Hirschberg (1990) and Kohler (2004) establish some contour patterns related to information packaging for English and for German, respectively. These patterns are summarized in the following tables:

H*	New
L+H*	Addition of a new value
!H*	Accessible
H+!H*	
L*+H	Modification of Given
L*	Given
No accent	

Table 1: Contour patterns related to information packaging according to Pierrehumbert & Hirschberg (1990)

L+H*/ L*+H (Late Peak)	Emphasis (new information)
H* (Medial Peak)	New
H+L*/ H+!H* (Early Peak)	Accessible or Given

Table 2: Contour patterns related to information packaging according to Kohler (2004)

For Pierrehumbert & Hirschberg (1990), an accent on a referring expression contributes to the perceived information status of the referent. The H* pitch accent is said to convey new information. The L+H* has contrast as its central meaning. The H+!H* predicates what is mutually accessible to speaker and listener.

In the same line, Kohler (2004) establishes a relation between meaning and categorical change from early to medial peak and between meaning and a more gradual change from medial to late peak. Early peaks tend to denote established facts or end of an argument. Medial peaks usually indicate a newly introduced fact or the beginning of a new argument. Late peaks add a paralinguistic value to the information expressed, e.g. surprise or incredulity.

Intonational studies on BP have been not referring to degree of givenness, but rather to the focus/ background relationship. Fernandes (2007) claims that focused elements may have the same pitch accent which they generally receive in a neutral context (H*+L versus L*+H) or they may have the same tonal combination which they would receive in a neutral context (L*+H).

3.　Methodology

3.1　On recording

For this study, I have recorded four native male speakers of BP, aged 18 to 30 years old, in an interactional context. To record the speakers, I used a game in which two speakers had to indicate people suspected of a crime, taking into account information available in a set of statements. Each speaker had a different set of statements containing distinct information, e.g., suspect 1 claiming to have been with suspect 2 in the library at the time the crime, and suspect 2 claiming to have been alone in the living room.

3.2　Degree of Givenness

I have considered three degrees of givenness, based on their cognitive status:
1) new or inactive: mentioned for the first time;
2) newsworthy or semi-active;
3) given or active.

The way the referents in the statements were restated by the speaker was used to classify the degree of givenness. If the referent was repeated, the information was considered given; if the speaker used a pronoun or synonym, the information was considered newsworthy. Referents not found in the statements were the only information considered new. Although the referents in this game were controlled, the speakers were able to produce spontaneous sentences.

At the end of the experiment, I selected 59 declarative sentences, which were later analysed with the Praat software (Boersma & Weenink, 2010, version 5.2.11). In those sentences, I distinguished 34 given referents, 44 newsworthy referents and 10 new referents. Since the number of utterances containing new information was significantly lower than both the ones containing given and newsworthy information, it was not possible to proceed to a detailed statistical analysis.

Since the concepts of newness and givenness of information are generally related to the Focus/ Background terminology, I have analysed such interactions in the light of the concepts of contrast and emphasis. I considered contrastive those referents which were used (by the speaker) to correct something in the previous speech and emphasis all highlighted referents with no corrections.

3.3　Dato System

ToBI	DaTo
Pitch Accent	*Level Contours*
L*	L
H*	H
	Dynamic Contours
L+H*	LH
L*+H	>LH
H+!H*	LHL
	HLH
	HL
	>HL
Phrasal Accent	
L-	
H-	
Boundary tones	*Boundary tones*
L%	L%
H%	H%

Table 3: ToBI contours and DaTo contours

The DaTo intonational annotation system (Dynamical Tones of Brazilian Portuguese) was used for my analysis. This system, devised by Lucente (2008), describes focus in intonation taking into account the notion of dynamical contour. The alignment in this approach was formulated according to a synchrony between phonation and articulation. The relation between ToBI contours and DaTo contours are in the Table 3.

4. Brazilian Portuguese (BP) information packaging

On the F0 analyse, I observed the contour type and pragmatic function of the aligned sentential elements. At the end of the data analysis, I distinguished the referents as either new, accessible or given. The number and percentage of referents in relation to degree of givenness are shown in Table 4.

	New		Accessible		Given	
	No.	*%*	*No.*	*%*	*No.*	*%*
LH	6	17	2	5	6	66
>LH	4	11	14	32	-	-
HLH	4	11	-	-	-	-
HL	8	26	18	42	1	1
>HL	1	3	5	12	3	33
H	11	32	2	5	-	-
LHL	-	-	3	6	-	-
Total	34	100	44	100	10	100

Table 4: Number and percentage of new, accessible and given referents

I observed in the percentage data that LH is more frequently used to indicate given information, HL is more often used for accessible information and H for new information. The low frequency of new information in this corpus (only 10 utterances) does not allow for a statistical analysis. However, the number of occurrence indicates that falling contours (HL, >HL, LHL) are somewhat associated with information structure, because they tend to connected to accessible information. HL and H contours are related to new information, in line with many other studies (Kohler, 2004; Yule, 1980; Pierrehumbert & Hirschberg, 1990).

Figure 1 shows the items 'menino' (boy) and 'Rodrigo' (proper name) as accessible and given information, respectively. Both items correspond to accessible/ given information updated in the utterance context. Since these items indicated no contrast or correction, I have considered them to be emphatic.

Figure 2 exhibits HL contour on the contrastive given referent "biblioteca" (library) and >LH on the proper names "Rodrigo" and "Alaíde", which were considered emphatic given referents. These data indicate the use of >LH contour to reintroduce a given referent in the discourse.
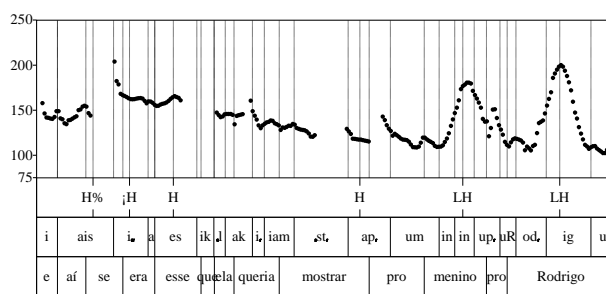
Figure 1: Contours H, !H and LH
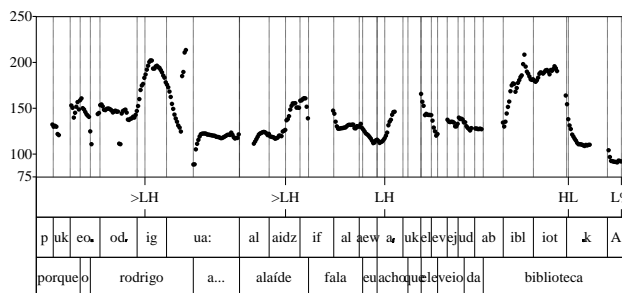
Figure 2: Contours >LH, LH and HL

In Figure 3, there is a >LH contour once again on the given referent "corpo" (body), which is reintroduced in the discourse. Since the expression "suite principal" (master bedroom) corresponds to a correction, we have HL contour again on the contrastive referent.
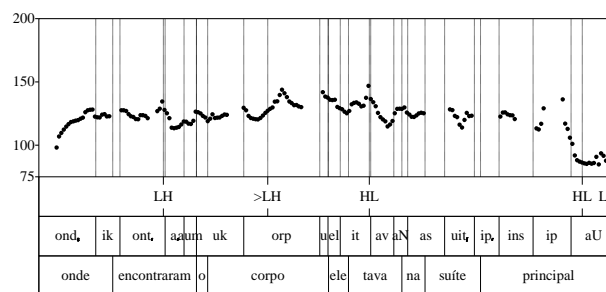
Figure 3: - >LH and HL

Figures 2 and 3 also illustrate a LH contour on the verb, which was in fact another pattern found in the data.
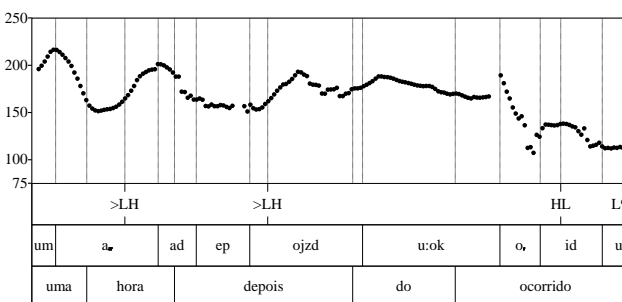
Figure 4: Contours >LH and HL

HL was the most frequent contour applied to express contrast. Moreover, there were some data which exhibit LH contour with contrastive meaning. Figure 4 is an example of LH contour with contrastive meaning.

As previously mentioned, the distinction between emphatic and contrastive referents was made with specific accent types. The number and percentage of the referents in relation to emphasis versus contrast are in Table 5.

	Emphasis		Contrast		Ambiguous	
	No.	*%*	*No.*	*%*	*No.*	*%*
LH	14	26	-	-	-	-
>LH	15	28	2	6	1	25
HLH	4	8	-	-	-	-
HL	5	9	20	65	2	50
>HL	3	6	5	16	1	25
H	12	23	1	3	-	-
LHL	-	-	3	10	-	-
Total	53	100	31	100	4	100

Table 5: Number and percentage of emphatic, contrastive and ambiguous referents

I observed in the percentage data that rising contours (LH, >LH, HLH) are strongly associated with emphasis, since LH and HLH were only used on emphatic referents. The data also show a more frequent percentage of falling contours on contrastive referents – HL (65%), >HL (16%), LHL (10%).

5.   Final Remarks

In this paper, I analysed the interaction between degrees of givenness and intonation in Brazilian Portuguese. Despite the impossibility of a detailed statistical analysis, the results have revealed important relations, on the one hand between falling contours (HL, >HL, LHL) and contrast and, on the other hand, between rising contours (LH, >LH, HLH) and emphasis. Regarding the degree of givenness, the results have not indicated a strong relationship between degree of givenness and intonation. However the results confirmed studies, which have related high tone to new information. Here, new information is also indicated with high tone. In sum, it is possible to state that LH contour is more frequently used to express given information, while HL is more frequently used to convey newsworthy information.

In closing, I would like to highlight that this study was realized using a spontaneous speech corpus, which is an important feature of a first analysis. Nonetheless, further research, including that using different methods, is crucial to confirm or refute the findings of this analysis.

6.   References

Allen, W.S. (1973). *Accent and Rhythm*. Cambridge: Cambridge University Press.

Barbosa, P. A. (2006). *Incursões em torno do ritmo da fala*. Pontes Editores – Campinas, Brasil e Fapesp: São Paulo, Brasil.

Baumann. (2008). The intonation of Giveness. Evidence from German. Tese de doutorado. Saarbrücken: Universität des Saalandes.

Boersma, P.E., Weenink, D. (2010). *Praat: doing phonetics by computer* (Version 5.2.11). Available at: <http://www.fon.hum.uva.nl/paul/praat.html>.

Brown, G. (1983). Prosodic Structure and Given/New Distinction. In A. Cutler, D.R. Ladd (Eds.), *Prosody: Models and Measurements*. Berlin: Springer, pp. 67--77.

Di Cristo, A., Hirst, D. (1998). *Intonation Systems. A Survey of Twenty Languages*. Cambridge University Press, Cambridge.

Fry, D.B. (1958). Experiments in the perception of stress. In *Language and Speech 1*, pp. 126--152.

Kohler, K. (2004). Prosody Revised: FUNCTION, TIME and the LISTENER in Intonational Phonology. In *Proceedings Speech prosody*, Nara, pp. 171--174.

Ladd, D.R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Lambrecht, K. (1994). *Informational Structure and Sentence Form*. Cambrigde: Cambridge University Press.

Lucente, L. (2008). DaTo: Um sistema de notação entoacional do Português Brasileiro baseado em princípios dinâmicos. Ênfase no foco e na fala espontânea. Dissertação de mestrado. Campinas: Unicamp-iel.

Pierrehumbert, J., Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In: P.R. Cohen, J. Morgan and M.E. Pollack (Eds.), *Intentions in Communication*. Cambridge: MIT Press. pp. 271--311.

Yule, G. (1980). Intonation and Giveness in Spoken Discourse. Studies in Language 4, pp. 271--286.

PHONETIC STUDIES

The dynamics of perception and production of VOT patterns in English by Brazilian learners

Ubiratã Kickhöfel ALVES, Márcia Cristina ZIMMER

Universidade Federal do Rio Grande do Sul; Universidade Católica de Pelotas

Brazil

ukalves@gmail.com, marcia.zimmer@gmail.com

Abstract

In this paper, we present the results of an experimental study on the perception and production of English Voice Onset Time (VOT) patterns by Brazilian learners. Twenty-four participants from Southern Brazil took part in the study. All learners sat for both a discrimination and an oral production tasks. The discrimination test consisted of an AxB task, in which we contrasted the three VOT patterns produced by native speakers of English: pre-voicing, short VOT and long VOT. For this test, productions of voiceless plosives were also manipulated on Praat, so that we could obtain artificial short VOT plosives. In the production test, learners were asked to read word-initial /b/, /d/, /g/ aloud. The preliminary results obtained from this experimental study suggest that the acquisition of voicing distinctions, both in terms of perception and production, may be characterized by a multitude of acoustic cues employed by learners, who, in their L2 developmental process, have to learn how to tune in to those cues which are most relevant in the language system to be acquired.

Keywords: VOT; L2 perception; L2 production.

1. Introduction

Learning L2 phonology can be characterized as a non-linear and dynamic process. Variables that are part of this complex system are fully interconnected, systems tend to stabilize for some time in attractor states and language development over time can grow or decline in a nonlinear fashion (Port & Van Gelder, 1995) Therefore, a multitude of variables, which operate at different levels, play a crucial role in second language learning (De Bot *et al*., 2007).

Departing from this dynamic conception of language acquisition, we present the results of an experimental study on the perception and production of Voice Onset Time (VOT) patterns by Southern Brazilian learners of English. The production of English word-initial stops tends to be difficult for Brazilian learners of English. In Brazilian Portuguese, voiced plosives are produced with pre-voicing (i.e, negative VOT), and voiceless plosives are produced with a short VOT pattern (also known as "Zero VOT"). This is different from what can be found in English, in which voiced stops are produced with either some pre-voicing or with Zero VOT, whereas voiceless initial plosives are produced with a long VOT pattern (aspirated). Given the fact that the Zero VOT pattern (short) is used in voiceless stops in BP but in voiced stops in English, Brazilian learners tend to show some problems in discriminating, identifying and producing the distinction between word-initial voiceless and voiced plosive consonants in English.

Our main goals in this article are: (i) to assess whether learners in three different proficient levels are able to distinguish among the production of different VOT patterns of English stop consonants; (ii) to investigate if these students produce VOT values which become gradually similar, according to their proficiency level, to those patterns found in American English; (iii) to study the relation between perception and production in L2 learning. The preliminary results shown in this paper are discussed mainly regarding the dynamics and nonlinearity between the processes of discrimination and production of the L2 VOT patterns.

2. Method

Twenty-four Southern Brazilian learners of English took part in the study. All of them were taking their undergraduate majors in English in one of the institutions of the two authors. After having taken the Oxford Placement Test (Allan, 2004), learners were organized as belonging to three different proficiency groups: proficient (6 participants), intermediate (7 participants) and basic (11 participants). All learners took part in both a perception (discrimination) and a production test.

The discrimination test consisted of an AxB task. In this task, the stimulus presented to learners consisted of triads. In a test booklet, participants were provided with multiple choice questions and were asked to indicate if the initial consonant of the second word was similar either to the 1st word (e.g. beer – beer – peer), or to the 3rd word (e.g. beer – peer – peer), or whether the three words began with the same consonant (e.g. peer – peer – peer). In order to build the stimuli, we invited two speakers of North American English, who had been living in Southern Brazil for less than 6 months, to record their production of the stimuli in a professional studio. These two speakers read a set of pre-selected words, all of which starting with a high vowel (cf. Yavas & Wildermuth, 2006; Yavas, 2008). In building the stimuli, we contrasted the three VOT patterns produced by these native speakers of English: pre-voicing (found in some productions of initial voiced consonants), short VOT (also found in their productions of voiced stops) and long VOT (found in their production of voiceless plosives). For this test, productions of voiceless plosives were also manipulated on Praat (Boersma & Hayes, 2011), so that we could obtain artificial short VOT plosives: as the VOT of the plosives was reduced, the resulting manipulated consonant would have the same VOT duration as a voiced

segment. These artificial voiced stops were contrasted with the three natural VOT patterns in the AxB task. Therefore, four kinds of contrasts were tested in the AxB task: natural zero VOT vs. negative VOT (6 questions), Artificial zero VOT vs. Negative VOT (6), Natural zero VOT vs. artificial zero VOT (6) and positive VOT vs. negative VOT (6). In Figure 1, the overall design of the AxB experiment is presented:

	Number of questions (per learner)	Basic (11 learners)	Intermediate (7 learners)	Proficient (6 learners)	Total tokens per condition
Negative x Nat. Zero	06	66	42	36	144
Negative x. Art. Zero	06	66	42	36	144
Nat. Zero x Art. Zero	06	66	42	36	144
Negative x Positive	06	66	42	36	144
Total tokens per group	24	264	168	144	576

Figure 1: AxB Task Design

Our purpose in testing learners on a manipulated VOT pattern is to assess whether VOT was the only acoustic cue used in their distinction between voiced or voiceless plosives. Should VOT be the only acoustic at play, learners would not be able to discriminate between those plosives starting with a natural Zero VOT and those ones which had their VOT reduced.

In the production test, learners were asked to read words starting with the consonants /p/, /t/ and /k/ aloud. These target words, which were repeated twice, were presented isolated, in a powerpoint presentation shown on a laptop computer. In Figure 2, the overall design of the production test is presented:

	Number of stimuli	Number of repetitions	Basic (11 learners)	Intermediate (7 learners)	Proficient (6 learners)
/p/	3	2	66	42	36
/t/	3	2	66	42	36
/k/	3	2	66	42	36
Total tokens per group			198	126	108

Figure 2: Production Task Design

The design described above allowed us to investigate our three hypotheses for this study. The first three hypotheses concern the results obtained from the AxB task, whereas the fourth hypothesis approaches the results of the production task.

1) There will be no significant differences among the three groups for the contrast between Negative and (natural) Zero VOT.
2) There will be no significant differences among the three proficiency groups in the discrimination of the contrast between Negative and (artificial) Zero VOT.
3) There will be no significant differences among the three groups for the contrast between Natural Zero VOT and Artificial Zero VOT.
4) There will be a significant difference among the three groups in the VOT values for each one of

the consonants (/p/, /t/ and /k/), as the three groups are going to produce native-like VOT values.

As for the first hypothesis, we don't expect Brazilian learners to discriminate between Negative and (natural) Zero VOT, as we hypothesize these learners consider these two patterns to be instances of the same category of voiced stops. With regard to (2), we hypothesized that all learners are able to perceive the difference between negative and manipulated VOT stimuli, so there will be no differences among groups. In (3), we hypothesized all learners are able to perceive the difference between natural and manipulated VOT stimuli, so there should be no differences among groups either. Finally, in (4), significant differences were predicted according to the participants' proficiency level.

The experimental results and the discussion of these hypotheses are presented in what follows.

3.   Results

Table 1 shows the results obtained from the AxB task.

Contrasted VOT	Accuracy (%)			Similarity (%)		
	Prof.	Int.	Bas.	Prof.	Int.	Bas.
Negative x Nat. Zero	2,77	9,52	7,57	91,66	90,47	87,87
Negative x. Art. Zero	77,77	71,42	45,45	16,66	21,42	33,33
Nat. Zero x Art. Zero	47,22	66,66	50,00	30,55	23,80	34,84
Negative x Positive	94,44	85,71	90,90	0,00	2,38	1,51

Table 1: AxB Task Results

Table 1 presents two main labels for its columns: "accuracy", which indicates that learners were able to efficiently distinguish between the two patterns, and "similarity", which presents the frequency rates with which learners chose the "all consonants equal" choice.

As we observe the data in Table 1, we notice that learners proved able to distinguish between negative and positive VOT patterns. These results are in accordance with a previous study carried out by Alves *et al.* (2011), which showed that these same participants reached ceiling effects in a task in which they were asked to identify the voicing of word-initial plosives. In other words, participants are already able to distinguish between voiceless and voiced stops in English. The results of this private study, corroborated by the discrimination findings between Negative and Positive VOT found in Table 3, motivated the present investigation on the discrimination of natural and manipulated stimuli, which gave rise to the three hypotheses guiding this study.

As for the first hypothesis, we concluded that, regardless of the participants' proficiency level, they do not discriminate between Negative and (natural) Zero VOT patterns. Kruskal Wallis tests were run in order to check whether there was a significant difference among the three proficiency groups, but no significant differences were found (Accuracy: n.s. ($X2(2) = 1,196$, p = ,550); (Similarity : n.s. ($X2 (2) = 1,228$, p = ,541).

Hypothesis 1 was thus corroborated. In other words, learners in all proficiency groups tend to accurately judge both the negative and zero VOT patterns as corresponding to instances of voiced stops in English.

The second hypothesis investigated the discrimination between the Negative and the (artificial) Zero VOT patterns. Table 1 shows that learners in the three proficiency groups tended to discriminate these two patterns, as indicating them as referring to two consonants of different voicing categories. However, the results obtained from Kruskal-Wallis tests indicated a significant difference among the three groups with regard to their Accuracy rates (Accuracy: s. (X2(2) = 7,916, p = ,019); Mann-Whitney (Prof and Basic) Similarity: n.s. (X2 (2) 2,353, p = ,308). This is mainly explained by the lower rates found in the answers provided by learners in the Basic Proficiency group, which seemed to be more doubtful about discriminating these two patterns. Given these findings, hypothesis 2 was not corroborated.

The last hypothesis on the perception task investigated the discrimination between Natural and Artificial Zero VOT. This comparison is of great importance to the present study, as it may be indicative of whether VOT is the single acoustic cue Brazilian learners of English make use of when distinguishing between voiceless and voiced consonants. The results in Table 1 show high rates of discrimination between these two VOT patterns, regardless of the learners' proficiency level. This was also confirmed by the results obtained from the Kruskal Wallis test, which showed there were no significant differences among the three groups (n.s. (X2(2) 1,968, p = ,374); Similarity : n.s. (X2(2) =, ,392 p = ,822), as the three of them tended to discriminate artificial and natural zero VOT patterns. Our third hypothesis was, therefore, corroborated.

Still regarding the results of the AxB task, as we pursued the perception of the artificial Zero VOT pattern further, we ran post-hoc pairwise comparison in which we contrasted the accuracy levels of all students in two different contrasts: Natural Zero vs. Negative and Artificial Zero vs. Negative. The results obtained from this Paired T-Test indicated a significant difference (t(23)=-9,364, p = 0,000) between the rates given for each of these constrasts. This result may be understood as we consider the fact that learners do not to discriminate between the Natural Zero and Negative VOT patterns, but do differentiate Artificial Zero and Negative VOT. Once again, this is indicative that learners tend to treat the Natural and Artificial Zero VOT patterns differently.

As for the production tests, the results, organized according to place of constriction, are shown in Table 2.

Consonant	Proficient (6)		Intermediate (7)		Basic (11)	
	Tokens	Mean (SD)	Tokens	Mean (SD)	Tokens	Mean (SD)
/p/	35	43,94 (31,17)	41	22,73 (13,03)	63	26,30 (15,06)
/t/	36	61,67 (22,36)	40	61,55 (21,93)	65	57,69 (24,02)
/k/	35	87,09 (31,31)	42	75,79 (24, 68)	65	86,08 (19,38)

Table 2: Production Task results (mean VOT in ms)

The results shown in Table 2 suggest that, regardless of the learners' proficiency group, nativelike VOT values for /p/ and /t/ were not yet produced. With regard to the velar consonant, the three proficiency groups tended to present native-like values in their VOT production. This will be discussed further in the following section.

As VOT values tend to decrease the more fronted the place of constriction of the consonant is, we investigated our fourth hypothesis in each one of the consonants (/p/ , /t/ and /k/) taken separately. As for the labial consonant, a One-Way Anova indicated a significant difference among groups (F (2) = 3,493, p = 0,049). This can be explained as the mean VOT value produced by the Proficient Group is much higher than those presented by the Intermediate and Basic Learners. Even though not even the proficient participants were able to produce near-native VOT values for /p/ (around 60ms, cf. Ladefoged and Cho, 2004), Hypothesis 4 was confirmed for this consonant.

The fourth hypothesis, however, was not confirmed for /t/ or /k/, due to two different reasons. Another One-Way Anova showed no significant differences – (F (2) = 0,102, p = 0, 903) among groups in their mean VOT values for /t/, as the three groups of learners presented similar VOT values, which did not resemble the nativelike ones (about 75ms, cf. Cho & Ladefoged, 1999). Although significant differences could not be found in the mean VOT values found for /k/ (F (2) = 0,904 , p = 0,420) either, it is important to point out that, unlike what was shown in the values for the other two consonants, the main VOT values for this stop seem to be produced in a nativelike fashion by learners in the three proficiency groups. In other words, even though our fourth hypothesis was confirmed only for /p/, the mean VOT values found in each one of the three consonants tend to show a different behavior. This will be explained further in the section that follows.

4.   Discussion

As already mentioned, the present investigation was motivated by a previous study developed by Alves et al. (2011), in which the same participants of this study presented high accuracy levels in an identification test. By considering the fact that these learners were already able to identify voicing patterns in English, but still seemed to show several problems concerning the production of aspirated (positive VOT) consonants, we inquired whether other factors, besides VOT, might have an influence in the discrimination and production of voiced and voiceless plosives in Brazilian Portuguese-English interlanguage.

Our discrimination task results confirmed our hypothesis that learners could not discriminate between productions of Negative and natural Zero VOT, since both patterns would be considered to be indicators of voiced stops, as can be found in the production of native speakers of English. Our hypothesis that participants in all proficiency levels discriminate Natural from Artificial VOT patterns was also confirmed, as simply reducing the VOT length of an aspirated stop was not enough to

prevent learners from distinguishing them from voiced stops which presented short VOT values.

The conclusion discussed above may be of great relevance for future investigations on the perception and production of VOT patterns. Should VOT length be the only acoustic pattern taken into consideration by Brazilian learners in their distinction between voiceless and voiced stops in English, discrimination rates concerning the distinction between the Natural and Artificial Zero VOT patterns would be low. This might suggest that Brazilian learners make use of other acoustic cues, besides VOT, in order to distinguish voiceless from voiced stops.

Speech sounds are categorized by a multitude of acoustic cues that do not act in isolation. This considered, learning to perceive (and consequently produce) the sounds of a second language implies having learners tune in to those cues which play a more decisive role in this new sound system. This might imply giving importance to some cues whose role was not imperative in their first language system.

This may seem to be the case of the participants in this study. Although VOT patterns are regarded as the most important acoustic cue among native speakers of English (cf. Lisker & Abramson, 1964), this does not seem to be the single or most important aspect considered by our learners. Further studies need to investigate which other aspects might be playing a role, among which burst intensity, might have an effect on the perception of these voicing patterns.

The possible role of burst intensity should also be highlighted as we consider the production data. Significant differences were found only for the production of /p/, even though none of the three proficiency levels were able to achieve the target VOT values for this consonant. As for the velar consonant, no significant differences were found, as the three groups seemed to have achieved the target VOT values. Finally, no significant differences were found among groups for /t/, even though learners seem to be closer to achieve the target VOT values for this consonant than they are with regard to /p/. These results seem to be very interesting as the role of burst intensity is taken into consideration. If we consider that the cue of burst intensity is stronger for /p/, a possibility might be that, in other to distinguish between /p/ and /b/, learners might be making use of this cue more regularly than they attend to VOT values. In other words, it might be the case that acoustic cues vary not only in terms of the learners' proficiency level, but also in terms of the place of constriction of the target consonant, as the acoustic correlates of VOT length and burst intensity may vary between /p/, /t/ and /k/. Additional statistical tests. which take each one of the consonants separately in the AxB task, may be indicative of a possible connection between perception, production and the use of different acoustic cues according to the place of constriction of the target consonant.

The possibilities discussed above deserve further investigation, as future studies should provide more detailed knowledge into what other acoustic cues are used not only by Brazilian learners of English, but also by learners of English from different first language systems. As to our future directions, it seems to us that data on the production of Brazilian Portuguese /p/, /t/ and /k/ must also be measured, so that we can investigate whether different acoustic cues are also at play in the production of these learners' L1 stop consonants. Furthermore, a control group with American participants also proves necessary, as it is imperative to confirm whether VOT is really the main cue which allows native speakers of English to distinguish between voiceless and voiced stops.

We believe that the results to be obtained from these future studies on the role of different acoustic cues, according to the learners' L1 system, involved both in perception and production, can provide further insight into the view of language as a complex adaptative system (Beckner *et al.*, 2009), according to which learning a second language is heavily attached to a complexity of variables in interaction (Herdina & Jessner, 2002, DeBot *et al.*, 2007). This seems to be the case of learning an L2/L3 sound system, which cannot be confined to an "all or nothing" issue, since a variety of acoustic correlates might be playing a role in this complex process.

5. References

Allan, D. (2004) *Oxford Placement Test 1*. Oxford University Press.

Alves, U., Schwartzhaupt B.M. and Baratz, A.H. (2011). Percepção e produção dos padrões de VOT do inglês (L2) por aprendizes brasileiros. In G. Ferreira-Gonçalves, M. Brum-de-Paula, M. Keske-Soares (Eds.), *Estudos em Aquisição Fonológica - Volume 4*. Pelotas-RS: Editora e Gráfica Universitária - UFPel, pp. 179--190.

Boersma, P., Weenink, D. (2011). *Praat: Doing Phonetics by Computer* (version 5.2.19). Available at: <www.praat.org>.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M.H., Croft, W., Ellis, N., Holland, J., Jinyun, K.E. Larsen-Freeman, D. and Schoenemann, T. (2009). Language is a complex adaptive system: position paper. Language Learning, 59, supp. 1, pp. 1--26.

Cho, T., Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. In *Journal of Phonetics,* 27, pp. 207--229.

De Bot, K., Lowie, W. and Verspoor, M.A. (2007). Dynamic Systems Theory approach to second language acquisition. In *Bilingualism: Language and Cognition*, 10 (1), pp. 7--21.

Herdina, P., Jessner, U. (2002). A dynamic model of multilingualism: Perspectives of change in psycholinguistics. Clevedon: Multilingual Matters.

Lisker, L., Abramson, A. (1964). A Cross-language study of voicing in initial stops: Acoustical measurements. *Word*, v.3, n.20, pp. 384--422.

Port, R.F., van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

Yavas, M. (2008). Factors influencing the VOT of English long lag stops and interlanguage phonology. In A. Rauber, M. Watkins and B. Baptista (Eds.). *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech.* Florianópolis: Universidade Federal de Santa Catarina, pp. 492--498.

Yavas, M., Wildermuth, R. (2006). The effects of place of articulation and vowel height in the acquisition of English aspirated stops by Spanish speakers. In *IRAL*, v. 44, pp. 251--263.

Acoustic description of the English and Brazilian Portuguese front vowel systems of Brazilian EFL teachers

Clerton Luiz Félix BARBOZA, Wilson Júnior de Araújo CARVALHO, Katiene Rozy Santos do NASCIMENTO

UERN-UFC, UECE, UECE

clertonluiz@gmail.com, wiljrac@gmail.com, kattiene@hotmail.com

Abstract

The present research has as its main objective the description and comparison of the front-vowel systems of both Brazilian Portuguese (BP) and English as a foreign language (EFL) as realized by English teachers in western Rio Grande do Norte-Brazil. We focus on an usage-based analysis of the phonetic details, such as duration, Euclidian distance, F1 and F2. Our methodology made use of a set of four experiments used to elicit BP and EFL vowels in a CVC or CVCV context. Two experiments were used to collect data from each language. The first made use of reading carrier sentences and the second used a street map as the main cue for eliciting data. Results regarding spectral data show overlap was found as regards the high-front vowel system of both languages. On the other hand, low-front vowels did not show acoustic overlap. As for duration, it seems to be used as the main acoustic cue to distinguish the exemplars of both languages, as EFL vowels are significantly longer than BP ones.

Keywords: front-vowels; EFL; BP.

1. Introduction

Traditional phonological theories assume the mental representation of the phonological level is simple, free of the details and redundancies found on the phonetic level. Much effort is made to try to find a set of rules, processes or restrictions capable of explaining satisfactorily the mapping from this simple mental representation to the phonetic level, which is complex. On the other hand, phonological theories based on use defend a mental representation capable of retrieving the phonetic details considered redundant by traditional theories. Once it assumes the mental representation to be complex, mapping from this representation becomes simple, as it is not necessary to use a set of rules, processes or restrictions which aim at simplifying or normalizing the phonetic realization. The view above is in consonance with the one defended by Bybee (2001) and Johnson and Mullenix (1997), respectively seminal texts as regards Phonology of Use and the Exemplar Model.

Having in mind the usage-based approaches commented above, the present research has as its main aim the description and comparison of the front-vowel systems of both Brazilian Portuguese (BP) and English as a foreign language (EFL) as realized by English teachers in western Rio Grande do Norte state, in north-eastern Brazil. Our specific focus lies on the analysis of spectral and duration cues of front-vowels as produced by EFL teachers in the aforementioned region. Our main hypothesis states EFL exemplars are markedly influenced by BP as regards their spectral and duration phonetic details.

On the following pages we present a brief overview of previous research on BP EFL vowel production and perception, our research methodology, our main results, as well as our conclusions.

2. Literature overview

Studies involving English vowels production and perception have been carried out for quite a long time. However, only quite recently usage-based approaches started to be used in this field of research. A glimpse of an exemplar approach to phonology is observed in the most seminal research on the field, presented by Peterson and Barney (1952). In this study, it is made clear the enormous amount of variation vowels are subject to. Such variation, however, is not enough to impede a correct perception by listeners most of the time. Figures presented by the authors are similar to the exemplar clouds reported on the present study, as a series of vowel realizations are more or less associated with a central, most robust exemplar.

Similar usage-based inferences are allowed in the research of a multitude of scholars and their work on the production of English as a second or foreign language (Baker & Trofimovich, 2005; Flege, Schirru, & MacKay, 2003; Cebrian, 2006), on perception only (Højen & Flege, 2006; Flege & MacKay, 2005), or on both skills (Jia *et al.*, 2006).

Having in mind this study focus on BP EFL subjects, we discuss below only the results presented by Baptista (2000), Rauber *et al.* (2005), Bion *et al.* (2006), Rauber (2006), and Nobre-Oliveira (2007). Such studies focus on BP speakers of EFL and are therefore worth reviewing.

Baptista (2000) is a longitudinal production study which describes the acquisition of English vowels of BP speakers living in the US. Results indicate a holistic approach to vowel acquisition. For example, in acquiring [ɪ] some subjects lowered the production of the first sound of the diphthong [eɪ]. Other changes in the system are also mentioned, as the need to make the front-vowel space longer, once English has more vowels than BP.

Rauber *et al.* (2005) investigated the relationship between English vowel perception and production by advanced EFL learners in Brazil. Perception data indicate a good accuracy for distinguishing the [i, ɪ] pair, but a poor perception of the [ɛ, æ] pair. The same results were obtained in production, with the former pair being well produced, and the latter, poorly realized.

Bion *et al.* (2006) also involved production and perception of natural stimuli, but added synthesized vowels with fixed duration but variable spectral quality. Once again natural data indicated the pair [i, ɪ] was better

perceived and produced than the [ɛ, æ] pair. Synthesized stimuli revealed even with controlled duration, the former pair is easier to perceive than the latter. This results indicates duration is not a primary cue for distinguishing [i, ɪ], whilst it is important for improving [ɛ, æ] perception.

Rauber (2006) again used synthesized vowels to study perception. Results indicated BP subjects use duration as their primary cue to distinguish both [i, ɪ] and [ɛ, æ] vowel pairs. Once again, the former pair was easier to perceive than the former. As regards production, the same problem arises, with [ɛ, æ] showing grater vowel overlap than [i, ɪ]. Duration was found to be important as well, once duration of the constituents in both pairs were significantly different.

Finally, Nobre-Oliveira (2007) carried out a study whose main focus was on perceptual training using both natural and synthesized stimuli. Production research was also carried out. The group which used synthesized stimuli had better production and perception results than the one which used natural stimuli.

Even though all the aforementioned studies are not grounded on a usage-based view of phonology, their results fit perfectly on the frame. Phonetic details associated with the realization of the BP front vowel system are seen influencing EFL throughout all the studies. BP influence seems to be concentrated on the spectral level, once most results indicate some degree of overlap especially of the [ɛ, æ] pair. Results involving duration, on the other hand, show a greater independence between the BP and EFL vowel systems, indicating training on this acoustic cue is important for our learners. It can be stated, thus, that BP vowel exemplars characteristics are more strongly linked to their EFL correlates, probably through a network, specially the low-front pair [ɛ, æ].

Next section deals with the research methodology used on this study.

3.   Methodology

Our subjects were a group of 20 male English teachers. All but one had university level. None had ever been abroad. Four experiments were carried out. Two experiments involved the reading of CVC (EFL) and CVCV (BP) words in carrier sentences. Two involved role-playing location information over a small city map in both EFL and BP.

Exemplars of the BP front-vowels [i, e, eɪ, ɛ] were collected using the carrier sentence "X. Diga Y alto". X and Y were words containing the same vowel, but only Y was acoustically analysed. Each sentence was repeated 3 times. 720 BP vowel exemplars were thus collected on this experiment, from now on called L1-1.

The second BP experiment involved the use of a small city map in which street names were used as cues to elicit the same vowel exemplars. Subjects were asked about how to go from one place to another. Each word was recorded 5 times. We analysed, thus, 400 vowel exemplars in this experiment, which we called L1-2.

First EFL experiment was similar to L1-1. [i, ɪ, eɪ, ɛ, æ] vowel exemplars were collected in the carrier-sentence "X. Say Y again." Once again, X and Y were words with the same vowel exemplar, but only the one in Y position was acoustically analyzed. 1500 vowel exemplars were collected on this experiment, called L2-1.

The second EFL experiment also used a small city map. Procedures were identical to experiment L1-2, but given the bigger number of vowel exemplars analyzed in EFL data collection, the total number of tokens reached 500. This experiment was called L2-2.

Overall number of tokens reached the total of 3,120, with 8,680 values of F1, F2 and duration being analyzed. Statistical test used were mainly paired-samples t-test and repeated measures ANOVAS. SPSS was used to carry out all the statistical treatment of the data.

Acoustic analyses were carried out using Praat, version 4.6.21. Formant analyses were carried out in a point in the middle of the vowel, except for the diphthong [eɪ] in both languages, which had only the middle of the first vowel analyzed. Duration measures excluded VOT when applicable and included only the pressure peaks of the exemplar vowels with visible formants on the spectrogram. No duration was measured for the diphthongs.

Recordings were made in a quiet, but not acoustically treated, room. We used a Shure SM-58 unidirectional dynamic microphone and a digital Microtrack 24/96 recording WAVE 16-bit, 44Khz files.

Next section presents our data analyses results & discussion.

4.   Results & Discussion

For the sake of brevity, we chose not to present a huge amount of tables with exact spectral and duration measurements. Instead, we will focus on presenting informative figures as much as we can. In case exact data is needed, we invite our readers to send us an email.

Paired-sample t-tests revealed significant (p < ,001) differences between the [i, ɪ] in all tests. We can observe in Figure 1 there is no overlap between the high-front EFL exemplars in experiment L2-1. Experiment L2-2 showed a very similar picture and is not reproduced
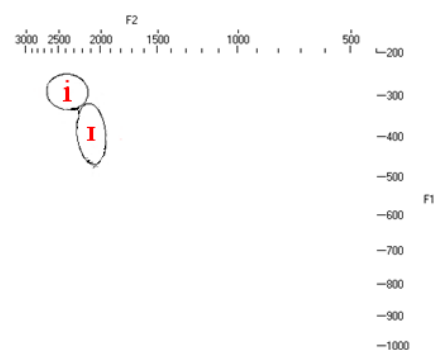
Figure 1: EFL exemplars [i, ɪ] in experiment L2-1

Further evidence of the motor control our subjects have in dealing with the aforementioned pair is found

when Euclidian distance is analyzed, once paired-sample t-tests indicate non-significant differences (p.= .693) between L2-1 and L2-2 values.

Comparisons of the English high-front exemplars [i] and [ɪ] with Brazilian Portuguese [i] and [e] in a spectral level were also carried out. Repeated-measures ANOVAs involving the L2-1 exemplars [i] and [ɪ] and L1-1 [i] found non-significant differences in both F1 and F2 between L2-1 [i] and L1-1 [i] (both p. > .05). However, the same test found significant differences between L2-1 [ɪ] and L1-1 [i] (both p. < .05). Figure 2 indicates graphically this high degree of exemplar overlap between L2-1 [i] and L1-1 [i].
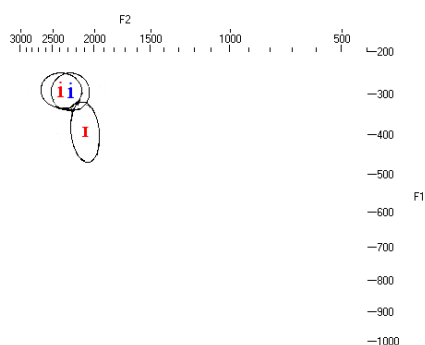
Figure 2: L2-1 exemplars (red) and L1-1 (blue)

Results of repeated-measures ANOVAs involving L2-2 exemplars [i, ɪ] and L1-2 [i], found similar results, except for a significant difference in F2 for L2-2 exemplar [i] and L1-2 exemplar [i] (p. < .05). The resulting figure, however, was very similar to the one presented above and was therefore not presented. We focus on the comparison between EFL [ɪ] and BP [e].
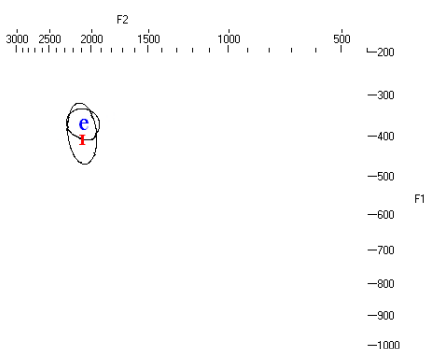
Figure 3: L2-1 exemplar (red) and L1-1 (blue)

Figure 3 shows a high degree of exemplar overlap between EFL [ɪ] and BP [e], indicating a degree of gesture influence as big as the one found between EFL [i] and BP [i]. Paired-sample t-tests involving L2-1 [ɪ] and L1-1 [e] found significant differences only for F1 (p.= .006). A comparison between L2-2 [ɪ] and L2-2 [e] reached non-significant levels for both F1 (p.= .06) and F2 (p.=.469), indicating an even bigger degree of exemplar overlap.

As regards duration of the EFL and BP high-front vowel exemplars, results indicate all duration values were

substantially different across experiments. Paired-sample t-tests found significant differences (p. < .001) for all comparisons, with both L2-1 and L1-1 exemplars having a longer duration than the values found in L2-2 and L1-2 experiments. Focusing on the EFL results, Figure 4 indicates the exemplar [i] is significantly longer in both L2-1 (p.= .003) and L2-2 (p. < .001).
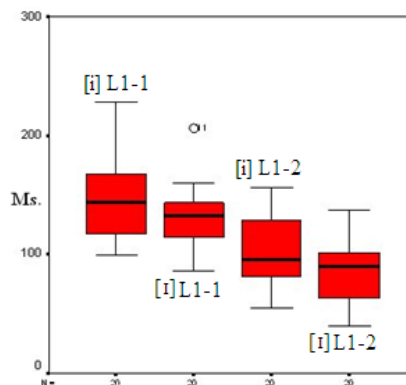
Figure 4: duration for the exemplars [i, ɪ] in experiments L2-1 and L2-2

Repeated-measure ANOVAs found a significant difference in duration between EFL L2-1 [i, ɪ] and BP [i] (p. < .05). A paired-sample t-test also found a significant difference in duration between L2-1 [ɪ] and L1-1 [e] (p.= .026). Even though new ANOVAs involving EFL L2-2 exemplars [i, ɪ] and BP L1-2 [i] showed similar results (p. < .05), a t-test for L2-2 exemplar [ɪ] and L1-2 [e] achieved only non-significant levels (p.= .824). This non-significant result reinforces the shared exemplar feature of EFL [ɪ] and BP [e] on the duration level as well as the spectral.

We now turn to the spectral characteristics of the first element of the diphthong [eɪ] in both EFL and BP. We remind our reader no duration measurements were made for these two exemplars. A high degree of vowel overlap is again observed in Figure 5, involving L2-1/L1-1 [eɪ]. Paired-sample tests indicate, however, a significant difference for F2 (p. < .001), but not for F1 (p.= .232). These results are opposed to the ones found in the comparison of L2-2/L1-2 [eɪ] (F2 (p.= .258); F1 (p. < .001). We present only one figure owing to lack of room and to the high degree of similarity between them.
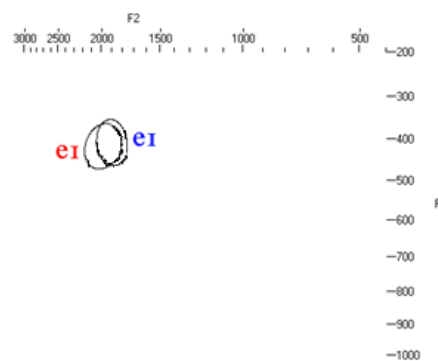
Figure 5: L2-1 (red)/ L1-1 (blue) [eɪ]

We now change our focus from high-front to low-front vowel exemplars in both BP and EFL. A superficial analysis of the spectral data reveals a more stable exemplar [ɛ] than [æ], as indicated by the standard deviation ellipses seen on Figure 6. We can also observe a huge amount of exemplar overlap of the two vowels. It indicates a good number of our subjects treat the EFL pair [ɛ, æ] as a single sound. Even though this overlap is easily observed, paired-sample t-tests revealed a significant difference between the exemplar pair in experiment L2-1 (F1 (p. < .001); F2 (p.= .238)).
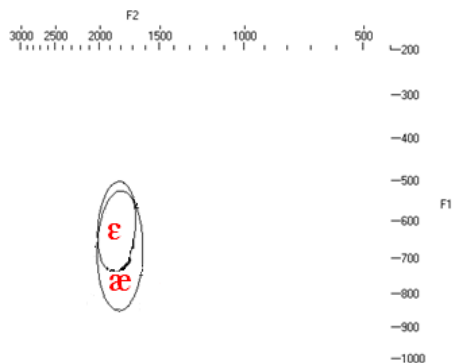
Figure 6: L2-1 exemplars [ɛ, æ]
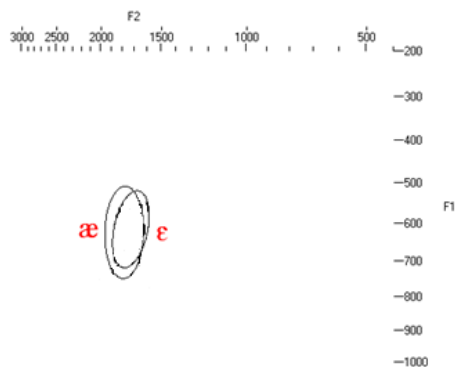
Figure 7: L2-2 exemplars [ɛ, æ]

The view that most of our subjects treat the EFL exemplar pair [ɛ, æ] as a single exemplar is reinforce by the data presented in Figure 7, regarding experiments L2-2. The exemplar overlap is even higher than the one found in Figure 6. However, this time a significant difference was found for F2 (p.= .014) but not for F1 (p.= .425) by the paired-sample t-tests.

Final piece of evidence BP speakers treat EFL exemplar pair [ɛ, æ] mostly as a single exemplar was found by the analysis of the Euclidean distance between these vowels across experiments. A paired-sample t-test confirmed the non-significant difference (p.= .443) between experiments L2-1 and L2-2.

A comparison between EFL [ɛ, æ] and BP [ɛ], presented in Figure 8, shows the BP exemplar is significantly higher (p. < .05) than its EFL counterparts [ɛ, æ]. However, significant difference was found for F2

(p.= .475) when comparing L2-1 and L1-1 exemplars. The same results were found for L2-2 and L1-2.
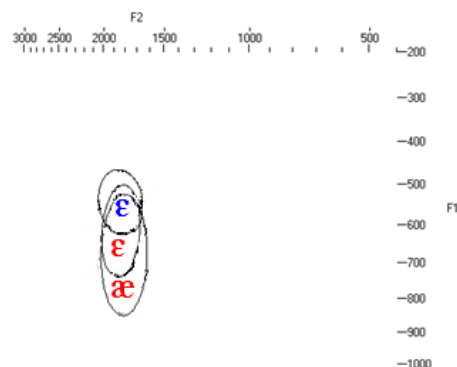
Figure 8: L2-1 (red)/ L1-1 (blue) exemplars.

Such results indicate that in as regards spectral cues, advanced BP speakers of EFL tend to create a new exemplar which is associated with the English pair [ɛ, æ].

Finally, as regards duration measurements Figure 9 presents a boxplot of the data regarding L2-1 [ɛ, æ] and L1-1 [ɛ] vowel exemplars. Both EFL exemplars were realized with a longer duration than the BP one as revealed by a repeated-measures ANOVA which reached significant results (p. < .05). As regard duration differences between the EFL vowel exemplars [ɛ, æ], the same test failed to show a significant difference (p. > 0,05), indicating our subjects do not realize the exemplars differently.
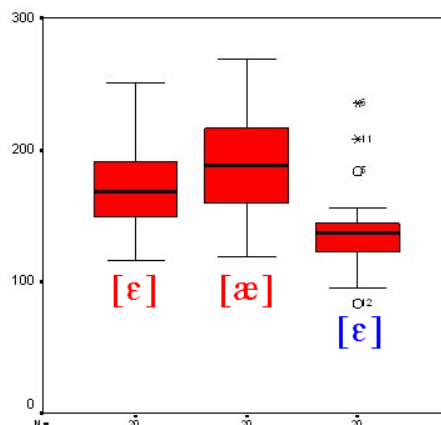
Figure 9: L2-1 (red)/ L1-1 (blue) durations

L2-2 and L1-2 data is very similar to the figure presented above. We therefore do not present the bloxplot regarding this set of data. Once again PB vowel exemplar [ɛ] was significant shorter in duration when compared to EFL [ɛ, æ] (p. < .05). And non-significant results were obtained between the EFL exemplar pair [ɛ, æ] (p. > .05).

5. Conclusion

A long tradition of interlanguage studies emphasizes the transfer of mother-tongue phonetic-phonological characteristics in the acquisition of a given foreign/second language. This tradition has created a perception that once

a positive or negative interference across languages has been noticed, all learners will face the same problems. This idea, however, is not completely true, once a huge amount of variables can influence positively or negatively a language acquisition construction course. The present research has concluded that BP dialectal variation is also responsible for EFL variation, once some of our results do not match previous research, made with subjects who speak other BP southern dialects, especially as regards duration as an important cue for producing the low-front vowel exemplars [ɛ, æ]. Other researches involving BP dialectal variation and its influence on EFL production are therefore necessary to achieve a more detailed view of EFL acquisition by BP learners.

As regards our own data, we are able to state our informants rely heavily on their BP vowel exemplars in order to produce EFL vowel categories. This could be observed mostly in the realization of the high-mid EFL vowels [i, ɪ, eɪ] which overlapped significantly with BP [i, e, eɪ] in this study. This overlap was not found in the same degree with the EFL low-front exemplars [ɛ, æ], as these exemplars were realized significantly lower than BP [ɛ] exemplar. Duration results, by their turn, indicated EFL exemplars to be different than BP ones. Such acoustic cue seems to be important for the production of BP speakers of EFL, even though most fail to realize significant differences between the low-front [ɛ, æ] EFL pair.

Pedagogical implications for the teaching of EFL to BP speakers involve the early association of BP and EFL [i] exemplars, as well as BP [e] and EFL [ɪ] vowel exemplars. Such early association would avoid the production of English [i, ɪ] as similar to BP vowel exemplar [i]. This was precisely what Baptista (2000) observed in her research. Another important implication is related to the EFL low-front exemplars [ɛ, æ]. Despite the fact they constituted a new vowel exemplar separate from BP [ɛ] exemplar, this creation of a single vowel category for two EFL categories indicates the high degree of training BP EFL speakers need in order to make their front-low vowel space longer so that to accommodate the English vowel space.

Finally, as a limitation of the present study we emphasize the lack of treatment frequency effects had in our research. Aware of the importance of frequency to the exemplar model as well as to usage-based phonology, a next logical step or our research is to include this variable in our future studies, alongside with the phonetic detail analyzed in the present research.

6.   Acknowledgements

7.   References

Baker, W., Trofimovich, P. (2005). Interaction of native- and second-language vowels system(s) in early and late bilinguals. *Language and Speech,* 48 (1), pp. 1--27.

Baptista, B. (2000). *The acquisition of English vowels by Brazilian-Portuguese speakers.* Florianópolis: Universidade Federal de Santa Catarina.

Bion, R., Escudero, P., Rauber, A. S., & Baptista, B. (2006). Category formation and the role of spectral quality in the perception and production of English front vowels. In *Proceedings of Interspeech 2006.* Pittsburgh, United States: Interspeech. pp. 1363--1366.

Bybee, J. (2001). *Phonology and language use.* Cambridge, UK: CUP.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics,* 34, pp. 372--387.

Flege, J.E., Schirru, C., & MacKay, I. (2003). Interaction between the native and second language phonetic subsystem. *Speech Communication,* 40, pp. 467--491.

Flege, J.E., MacKay, I. (2005). Perceiving vowels in a second language. *Studies in Second Language Acquisition,* 26, pp. 1--34.

Højen, A., Flege, J.E. (2006). Early learners' discrimination of second-language vowels. *Journal of the Acoustical Society of America,* 119 (5), pp. 3072--3084.

Jia, G., Strange, W., Wu, Y., Collado, J. and Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America,* 119 (2), pp. 1118--1130.

Johnson, K., Mullennix, J.D. (1997). Complex representations used in speech processing: overview of the book. In K. Johnson, J.W. Mullenix (Eds.), *Talker variability in speech perception.* San Diego: Academic Press, pp. 1--9.

Nobre-Oliveira, D. (2007). *The effect of perceptual training on the learning of English vowels by Brazilian Portuguese speakers.* Unpublished doctoral dissertation. Florianópolis: Universidade Federal de Santa Catarina.

Peterson, G.E., Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America,* 24 (2), pp. 175--184.

Rauber, A. (2006). *Perception and production of English vowels by Brazilian EFL speakers.* Unpublished doctoral dissertation. Florianópolis: Universidade Federal de Santa Catarina.

Rauber, A.S., Escudero, P., Bion, R. and Baptista, B. (2005). The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. In *Proceedings of Interspeech 2005.* Lisbon, Portugal, pp. 2913-2916

Flutuação do acento em palavras produzidas por falantes nativos do Português Brasileiro

Aline de Lima BENEVIDES

Universidade de São Paulo (USP)

aline.benevides@usp.br

Resumo

O objetivo desta pesquisa é investigar o estímulo da flutuação do acento em algumas palavras produzidas por falantes nativos do português brasileiro. O corpus foi composto a partir das gramáticas de Cunha & Cintra (2001), Lima (2002) e Bechara (1976; 2005) e de alguns testes de produção realizados com falantes nativos do PB. O processo estudado é tratado pelos gramáticos como silabada. Na primeira fase da pesquisa, as principais teorias do acento regular foram consultadas (Bisol, 1992 *apud* Collischon, 2010; Camara Jr., 2001; Lee, 1995). A partir dessas revisões, verificou-se que elas não conseguem explicar os padrões acentuais do PB sem fazer uso de exceções. Na segunda fase da pesquisa, constatou-se que os pressupostos da Fonologia de Uso (Bybee, 2001) podem ajudar a justificar as oscilações acentuais. A análise dos dados indica que as oscilações encontradas são decorrentes da baixa frequência de uso, que se utiliza de associações fonológicas com palavras de maior frequência de uso. Contudo, constata-se ainda a necessidade da realização de alguns testes de produção, os quais serão realizados no desenvolvimento da pesquisa e de possíveis interfaces com outras teorias.

Palavras-Chaves: Fonologia; Acento; Português Brasileiro.

1. Objetivo

A finalidade desta pesquisa é verificar o condicionamento e/ ou o estímulo da flutuação do acento em palavras produzidas por alguns falantes nativos do Português Brasileiro – doravante PB.

2. Composição do corpus

O corpus é constituído por um conjunto de palavras cuja pronúncia apresenta variação sendo considerada pelos gramáticos como fora da dita "norma culta", por exemplo: [ɡɾaˈtujtʊ] > [ɡɾatuˈitʊ] e [noˈbɛw] > [ˈnɔbɛw]. O processo estudado é tratado como silabada, que, segundo os gramáticos, é "o erro de prosódia que consiste na deslocação do acento tônico de uma palavra" (Bechara, 2005: 90).

As etapas de formação do corpus foram: i) a partir das gramáticas de Cunha & Cintra (2001), Lima (2002) e Bechara (1976; 2005), listamos as palavras consideradas "mais usuais", totalizando 79 vocábulos; ii) apresentamos essa lista de verbetes para 12 colegas, os quais deveriam apontar em quais palavras já escutaram a pronúncia oscilante, resultando em 54 marcados e 2 palavras sugeridas para acréscimo; e, iii) realizamos um teste de produção com os 25 vocábulos não marcados, além deles, 11 verbetes que apresentaram poucas marcações também foram inseridos no teste, finalizando com 36 palavras.

A metodologia utilizada foi formar frases e solicitar para 14 falantes nativos as lerem. Ressalta-se que não houve rigor com os critérios sociolinguísticos, de modo que a faixa etária, o sexo, o grau de escolaridade e a naturalidade são diferentes, contudo, não há uma distribuição equivalente. Vale lembrar também que a marcação da tonicidade por meio do acento gráfico pode direcionar a leitura para uma ou outra forma, da mesma forma que a ausência dessa marcação também a direciona, uma vez que a neutralidade não é possível, optamos por seguir a ortografia oficial da língua portuguesa.

Das 36 palavras testadas, 25 foram produzidas com oscilação e, surpreendemente, 2 palavras que não estavam sendo testadas também sofreram flutuação por 2 falantes. Assim, conclui-se a formação do corpus em 72 palavras.

3. O acento regular em Português Brasileiro

Entre as diversas teorias que tentam explicar a realização do acento regular em PB, que se contradizem em partes, fundamentalmente todos os teóricos admitem a realização nas três últimas sílabas a partir da margem direita da palavra. Diferente de outras línguas, como o francês, em que a tonicidade se dá sempre na última sílaba, o acento em português não é totalmente previsível.

Na literatura do PB, encontramos inúmeras teorias que buscam esclarecer o acento regular, diante da impossibilidade de explorar todas elas, resolvermos apresentar as três principais hipóteses para a atribuição do acento regular, segundo Ferreira-Netto (2007), que são:

- ✓ Hipótese do Acento Livre - previamente definido no léxico (Camara Jr., 2001);
- ✓ Hipótese do Molde Trocaico - definido pela característica rítmica padrão (Bisol, 1992 *apud* Collischon, 2010);
- ✓ Hipótese do Acento Morfológico - definido pela qualidade do morfema portador (Lee, 1995).

A primeira proposta prediz que o acento é livre, assim não há uma regra para a atribuição acentual, o que pode ocorrer é uma maior tendência a uma dada terminação. Seguindo esta hipótese, teríamos o acento sendo atribuído no léxico. A lacuna encontrada aqui diz respeito em como ocorre à organização desses vocábulos no léxico, que não é prevista pela teoria.

A segunda hipótese propõe que o peso silábico e o pé métrico são os mecanismos responsáveis pela atribuição acentual. As sílabas finais pesadas atraem o acento, caso não sejam pesadas, o acento cai na penúltima sílaba. Todos os casos que fujam a tais regras são inseridos na extrametricidade.

A terceira proposta se vale de regras diferentes para verbos e não verbos. Nesse, o acento cai na última vogal do radical derivacional. Assim, as paroxítonas com sílabas finais pesadas e as proparoxítonas, que não se enquadram nesse padrão, são consideradas casos marcados lexicalmente.

Como vemos, na segunda e na terceira propostas, centenas de palavras são englobadas na excepcionalidade, que incluem todas as proparoxítonas, algumas paroxítonas e oxítonas. Elas são denominadas pelos autores de casos extramétricos ou marcados, respectivamente. Ora, se um padrão acentual inteiro, o esdrúxulo, mais alguns casos dos outros padrões são considerados desvios às regras acentuais, podemos realmente considerá-los exceções?

O estudo realizado por Araújo *et al.* (2007) refuta os principais argumentos empregados pelos teóricos que inserem as proparoxítonas na excepcionalidade. O autor demonstra que o padrão esdrúxulo não deve ser considerado excepcional, posto que apresenta a mesma regularidade de entrada na língua que os demais padrões; os processos que reduziriam as proparoxítonas em paroxítonas, como a síncope ou a apócope, não podem afetar todas as palavras por gerar palavras agramaticais, como *[ˈmɛdkʊ] e *[ˈbɛbdʊ]; e, por fim, que sua frequência de ocorrência está diretamente relacionada com o número de sílabas, assim as palavras trissílabas proparoxítonas possuem uma frequência semelhante a dos demais padrões.

Tendo isso em vista, observamos que tais teorias parecem não esclarecer a regularidade do acento primário no PB, uma vez que nesse corpus, considerando apenas a pronúncia regular, aproximadamente 50% dos verbetes seriam compreendidos na extrametricidade ou na marcação lexical.

4. Os dados

Como vimos, o peso silábico é um dos fatores comumente tratados como influente para a atribuição acentual. As propostas afirmam que as sílabas com coda silábica, i. é., as sílabas pesadas, atraem o acento. Além disso, o acento paroxítono por ser o padrão mais produtivo é considerado o padrão acentual do PB, de modo que esta seria a tonicidade atribuída às novas entradas lexicais. Se essas características são realmente importantes para o acento, presume-se que a tonicidade oscilante ocorre em direção a elas.

Entretanto, tal fato não é verificado com uma percentagem significativa no corpus. Observa-se que dentre 34 vocábulos[1] que poderiam ter a oscilação

motivada pelo peso silábico, apenas 23,5% deles saem de uma sílaba leve em direção a uma sílaba pesada, contrapondo-se a 35,3% que sai de sílaba pesada em direção a uma sílaba leve. As demais percentagens referem-se a: 20,6% saindo de sílaba pesada para outra sílaba pesada e 20,6% saindo de sílaba leve em direção a uma sílaba leve, tendo como opção uma sílaba pesada.

Embora possamos considerar que oscilar de uma sílaba pesada para outra sílaba pesada não seja uma violação a sensibilidade da língua ao peso silábico. Além disso, que em casos como *ínterim* a flutuação vai de encontro à preferência por sílabas pesadas finais (Collischon, 2010), é possível encontrar casos, como *condor*, em que a oscilação se opõe a essa predileção.

Ao considerar o padrão acentual, observamos que as palavras proparoxítonas possuem, aproximadamente, 93% de oscilações a favor do padrão, sendo que os 7% restantes flutuam para as sílabas finais pesadas. Se tivéssemos apenas esses dados, pressuporíamos que as duas características apontadas são essenciais para a língua. Logo, a proposta de Bisol (1992 *apud* Collischon, 2010) seria a melhor hipótese para descrever o acento regular em PB. Contudo, o corpus também apresenta vocábulos paroxítonos, os quais oscilam para outros padrões.

Dentre 36 verbetes com acento na penúltima sílaba, 41,7% flutuam para a antepenúltima, 13,9% para a última sílaba e 44,4% mantêm a penúltima sílaba tônica. Essas são formadas por ditongos mediais que se tornam hiatos ou encontro vocálicos finais, que quando ditongos tornam-se hiatos e quando hiatos tornam-se ditongos. Esses casos ainda não serão considerados, pois receberão tratamento diferenciado ao longo da pesquisa, a saber, serão submetidos a testes.

Atente-se ao fato de que uma percentagem considerável de paroxítonas oscila em direção ao acento proparoxítono, que além de ser considerado um desvio, é apontado como um caso a ser evitado.

Com relação às palavras oxítonas, não há dados suficientes para qualquer afirmação, uma vez que todas elas são vocábulos dissílabos, desse modo não possuem outra opção para a flutuação.

Essa pequena apresentação dos dados foi apenas uma tentativa de esboçar reflexões que devem ser exploradas no desenvolvimento deste trabalho. Até o momento, não é possível tecer qualquer afirmação valendo-se apenas desses dados. Mas alguns questionamentos se instauram: ora, se há uma preferência pelo acento paroxítono, não esperaríamos que a oscilação partisse dele; se o PB evitasse o acento esdrúxulo, não encontraríamos uma percentagem alta de flutuações em sua direção; se há sensibilidade da língua ao peso silábico, elas não só segurariam o acento, mas também os atrairiam.

Com a finalidade de observar se essas podem ser características pertinentes para a colocação acentual, faremos um teste de produção com alguns falantes nativos do PB. Nele, iremos formar um texto com palavras inventadas, as quais possuirão os principais

---

[1] Ressalta-se que não incluímos as sílabas que possuem ditongos finais ou mediais, posto que daremos um tratamento diferenciado a eles no desenvolvimento da pesquisa.

padrões silábicos permitidos na língua. Com isso, se o PB for sensível ao peso silábico ou tender ao acento paroxítono, haverá um grande número de atribuição de tonicidade a esse padrão; ou, se esses aspectos não forem relevantes para a língua, não verificaremos uma percentagem significativa de tonicidade nesse padrão acentual.

Diferentemente do teste de composição do corpus, esse irá valer-se de algumas variáveis sociolinguísticas, a saber, faixa etária, sexo e grau de escolaridade. Para a faixa etária, faremos três divisões, que são: 1ª) a partir de 20 anos até 34 anos; 2ª) de 35 anos a 59 anos; e, 3ª) mais de 60 anos. Para o grau de escolaridade, distinguiremos os indivíduos que possuem ensino superior (cursando ou completo) dos demais. Ao cruzar esses critérios, teremos 12 perfis distintos.

Para que possamos realizar uma análise mais sólida, contaremos com 3 informantes para cada perfil, totalizando assim em 36 gravações. O resultado deste experimento será apresentado em trabalhos posteriores.

5.   A Fonologia de Uso

A Fonologia de Uso está compreendida nos modelos de língua baseados no uso, os quais consideram o uso como o principal fator para a formação da gramática dos falantes. Dentro desta teoria, a frequência é a responsável pela organização das representações mentais e pelos processos fonológicos, morfológicos e semânticos que ocorrem na língua.

Este modelo não exclui da análise ou trata de modo diferenciado os padrões menos produtivos. Dentro dele, todos os vocábulos podem ser analisados de modo semelhante. O estudo realizado por Greenberg (1966 *apud* Bybee, 2001) mostra que os membros não marcados são os mais frequentes. Esse fato demonstra que, provavelmente, a frequência é o mais básico fator dessa relação de marcação, de modo que a distinção feita por alguns autores, entre verbetes marcados ou não marcados, tem como premissa a frequência de ocorrência.

Os modelos teóricos anteriormente esboçados, parecem se valer direta ou indiretamente da frequência de uso dos vocábulos, uma vez que os verbetes que são inseridos na excepcionalidade são os menos comuns, os padrões menos produtivos, ou ainda, os desvios às regras do acento. Se a frequência é um dos mecanismos utilizados por eles e ao considerarmos que o corpus, coincidentemente, é constituído de palavras de baixa frequência, por que não partir dela para analisar os dados?

Foi partindo desse questionamento que nos adentramos nessa teoria. Os estudos ainda estão em andamento, entretanto, apresentaremos alguns princípios essenciais da teoria e como eles parecem explicar as oscilações encontradas no PB.

5.1   Uma análise à luz da Fonologia de Uso

A Fonologia de Uso é um modelo de língua que tem

como premissa o uso da língua e a frequência de uso. Nela, a gramática é formada fundamentalmente pelo uso da língua, de modo que as representações mentais estão em constantes modificações e reestruturações, ou seja, o uso constrói e modifica as representações mentais. Tais estruturas são construídas por associações fonológicas ou semânticas, ou ainda, quando ambas compõem a estrutura, temos associações morfológicas.

Os itens lexicais não são apenas usos concretos, mas servem de gatilho para as novas entradas lexicais e para os usos menos frequentes. Os padrões produtivos por serem acessados mais rapidamente tornam-se o gatilho da língua, i.e., os lexemas mais frequentes fornecem seus padrões aos menos frequentes.

É a alta frequência que é a responsável pela facilidade no acesso das palavras, pela produtividade e pela extensão dos padrões da língua. Além disso, os itens com alta frequência de ocorrência têm força lexical, por isso, possuem resistência morfológica e são menos suscetíveis a mudanças por analogia. Em contrapartida, são mais propícios a sofrer processos fonológicos, como reduções e apagamentos.

Bybee ainda demonstra que a frequência influencia a aquisição de determinadas formas. Em um estudo do Antigo Inglês, a partir de Phillips (*apud* Bybee, 2001), mostra que a aquisição do ditongo <eo> se dá de modo diferenciado, conforme a frequência dos vocábulos. Os mais frequentes são adquiridos corretamente, enquanto que, os menos frequentes sofrem simplificações para /æ:/ e /æ/, e, posteriormente, a vogal arredondada anterior perde o arredondamento, tornando-se /e:/ e /e/.

Se tal processo pode ocorrer na aquisição dos fonemas sendo influenciado pela frequência das palavras, acredita-se que processo semelhante ocorra com a tonicidade dos vocábulos, posto que os verbetes do corpus possuem baixa frequência de uso e o processo se daria por difusão lexical. Assim, hipotetiza-se que as palavras de baixa frequência se associam com palavras de alta frequência por similaridades fonológicas, resultando na extensão da tonicidade de um vocábulo a outro. A Figura 1 é um exemplo de como as teias de conexões fonológicas são formadas:
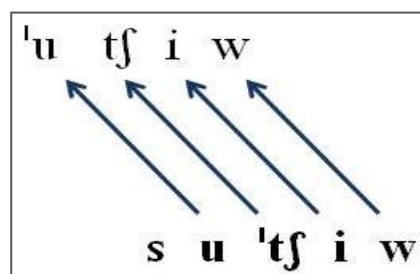
Figura 1: Conexões fonológicas por ['utʃiw] entre *útil* e *sutil*

No exemplo apresentado, *útil*, com frequência de

3.514, seria o atrator de *sutil*, com frequência de 11.125[2], e forneceria seu padrão acentual a ele, resultando na forma com tonicidade oscilante.

Essas teias associativas organizam os vocábulos na representação mental do falante e a frequência reforça essas conexões. Com o tempo, o uso da forma oscilante em detrimento da forma regular conduz a modificações desse vocábulo na estrutura mental. A Linguística Cognitiva denomina essas estruturas frequentes, que são formadas por meio da experiência de uso, de estruturas arraigadas.

Possíveis evidências de que algumas das formas oscilantes estão sendo arraigadas pode ser verificada no teste de composição do corpus. Observamos que alguns verbetes que não foram apontados pelos colegas como sujeitos a flutuação acentual são produzidos pela maioria deles com o acento oscilante, por exemplo: *cateter* e *dúplex*.

Um indício de que esse seja um processo de mudança começa a ser encontrado em alguns dicionários (Aulete & Valente, 2012; Ferreira, 2004), os quais apresentam duas entradas lexicais para alguns vocábulos, sendo eles: a forma regular e a forma oscilante. Em geral, eles fazem referência um ao outro e há, em alguns deles, observações, no caso o Dicionário Aulete (2012), indicando a pronúncia do acento tônico, como se observar a seguir:

> **Autópsia**
> [Var. pros. de autopsia.] – 1. Exame de si mesmo. 2. Med. Impr. Necropsia.
> **Autopsia**
> [Do gr. autopsía.] – s.f. 1. Autópsia (q.v.). [Cf. autopsia, do v. autopsiar.]
> **Observação:**
> [Nota: A 1ª ed. deste Dicionário marcou a pronúncia com o acento tônico no i, de acordo com o étimo. Porém; o uso português consagrou a forma esdrúxula autópsia, pelo que se adota esta acentuação.] (grifo nosso).

Com a finalidade de verificar em que medida as gramáticas também estão sendo afetadas, comparei duas edições da gramática de Bechara, uma editada em 1976 e a outra editada em 2005, que é uma edição revisada. Vale lembrar que os gramáticos costumam apontar dentro do processo de silabada, algumas palavras que admitem dupla prosódia, a forma regular e a oscilante são aceitáveis pela norma culta. Ora, se as gramáticas não são afetadas pelo uso, não encontraremos mudanças entre as duas edições. Mas se houver modificação, pressupõe-se que o uso também pode afetá-las. Para tanto, examinamos o que é admitido como dupla prosódia em ambas as edições, como resultado tem-se o acréscimo de 6 palavras que passam a ser aceitas como verbetes com dupla prosódia (ver Figura 2). Dessa maneira, pressupõe-se que o uso já está afetando as gramáticas.

Tendo isso em vista, postula-se que, em um primeiro momento, o uso modifica as representações mentais, em um segundo, começa a afetar os dicionários, que já registram algumas das formas oscilantes, e, em um terceiro momento, afeta as gramáticas, as quais passarão a aceitar as duas formas como prosódias possíveis. Claro que, as modificações nas gramáticas são mais lentas devido à resistência normativa.

Figura 2: Palavras com dupla prosódia

6.   Conclusões parciais

Como vimos, encontramos na literatura do PB diversas teorias que objetivam explicar o acento regular, que vão desde teorias métricas até teorias que consideram aspectos morfológicos. De modo geral, elas se utilizam de um grande número de exceções, que na maioria dos casos incluem um padrão acentual inteiro, o esdrúxulo. Em contrapartida, encontramos estudos que defendem o acento proparoxítono através de dados quantitativos.

O objetivo desta pesquisa é buscar um modo de análise que inclua todos os padrões acentuais, posto que admitimos que o acento antepenúltimo não é um caso excepcional. Ao longo das revisões, observou-se que a frequência é o principal mecanismo que inclui ou exclui determinados padrões acentuais das análises. Ora, se ela é realmente um fator relevante por que não iniciarmos por ela? Foi essa pergunta que nos fizemos e é por ela que adentramos nossos estudos na Fonologia de Uso.

Até o momento, as pesquisas indicam que o uso da língua e a frequência de ocorrência são os principais motivadores das oscilações. Acredita-se que as representações mentais são modificadas a cada interação e que a cada nova ocorrência, da forma regular ou da oscilante, as estruturas se fortificam, conduzindo ao fortalecimento de uma das formas. No entanto, quando as duas formas são produzidas pelo mesmo indivíduo, o que parece ocorrer, em alguns casos, é uma especificação de uso.

Como viemos salientando, muito ainda se tem a dizer e explicar a respeito da flutuação acentual. Este é apenas um estudo piloto que apresenta questões que devem ser exploradas em novos trabalhos. Os próximos passos desta pesquisa serão baseados em testes de produção, os quais têm como finalidade investigar a sensibilidade da língua ao peso silábico, se há tendência

---

[2] Segundo o índice de frequência do Projeto Aspa (Avaliação Sonora do Português Atual).

do acento cair no padrão acentual da língua (o paroxítono), o que ocorre nos encontros vocálicos [ia] e [uj] que os fazem oscilar em determinadas posições, se as oscilações têm algum correlato com as variantes sociolinguísticas, tais como grau de escolaridade e faixa etária? Em alguns verbetes como *recorde*, a tonicidade da língua inglesa parece influenciar na pronúncia desse vocábulo, assim, questiona-se: em que medida a origem etimológica influencia as flutuações? Em palavras como *Nobel*, o uso das formas regular e oscilante parece ser motivado pelo contexto. A hipótese é que quando o falante se refere à livraria produzirá [noˈbɛw] e quando faz referência ao prêmio utilizará [ˈnɔbɛw]. Tendo isso em vista, será que há especificação de uso entre as duas formas nesse e em outros vocábulos?

Muitas respostas ainda devem ser dadas para se concluir as motivações da flutuação acentual no PB. Para tanto, buscaremos no decorre desta pesquisa, além de realizar testes, novos apoios teóricos na tentativa de comprovar a motivação do deslocamento acentual. A princípio, as teorias que estão sendo estudadas são: a Fonologia de Uso, a Teoria dos Exemplares e a Linguística Cognitiva, contudo, se no decorrer da pesquisa novos aparatos teóricos surgirem, eles também serão incorporados ao estudo.

7. Agradecimentos

8. Referências

Aulete, F.J.C., Valente, A.L. dos S. (2012). *Dicionário Digital Aulete*. Lexicon Editora Digital Ltda. Disponível em: <http://aulete.uol.com.br/site.php?mdl=aulete_digital&op=creditos#ixzz1nJEUxOpA>.

Araújo, G.A. de. *et al.* (2007). As proparoxítonas e o sistema acentual do português. In G.A. Araújo (Ed.), *O acento em português: abordagens fonológicas*. São Paulo: Parábola. pp. 37--60.

Aspa. *Projeto ASPA: Avaliação Sonora do Português Atual*. Disponível em: <http://www.projetoaspa.org/buscador2/xlogin.php>.

Bechara, E. (1976). *Moderna Gramática Portuguesa*. 20. ed. São Paulo: Companhia Editora Nacional.

Bechara, E. (2005). *Moderna Gramática Portuguesa*. 37. ed. rev. e ampl. 15. reimp. Rio de Janeiro: Lucerna.

Bechara, E. (2009). *Moderna Gramática Portuguesa*. 37. ed. rev., ampl. e atual. Rio de Janeiro: Nova Fronteira.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Camara Jr., J.M. (2001 [1970]). *Estrutura da Língua Portuguesa*. 34. ed. Petrópolis: Editora Vozes.

Collischon, G. (2010 [1996]). O Acento em Português. In L. Bisol (Org), *Introdução a estudos de fonologia do português brasileiro*. Porto Alegre: Edipucrs. 5 ed., rev. pp. 132--165.

Cunha, C., Cintra, L. (2001). *Nova Gramática do Português Contemporâneo*. 3. ed. Rio de Janeiro: Nova Fronteira.

Ferreira, A.B. de H. (2004). *Novo Dicionário Eletrônico Aurélio versão 5.0*. 3. ed. Editora Positivo.

Ferreira-Netto, W. (2007). O Acento na Língua Portuguesa. In G.A. Araújo (Ed.), *O acento em português: abordagens fonológicas*. São Paulo: Parábola. pp. 21--36.

Lima, R. (2002). *Gramática Normativa da Língua Portuguesa*. 42. ed. Rio de Janeiro: José Olympio.

Lee, S. (1995). Morfologia e Fonologia Lexical do Português do Brasil. Tese (Doutorado em Linguística). Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas.

Revisiting the acoustic and auditory approach to speech analysis

Luiz Carlos CAGLIARI, Mariane CARVALHO, Mariana MORETTO Gementi

Capes/CNPq (Proc. N. 303440/2009-0); Capes; Fapesp; UNESP Universidade Estadual Paulista "Júlio de Mesquita Filho"

lccagliari@gmail.com, mazicarvalho@bol.com.br, mariana_moretto@hotmail.com

Abstract

The difficulty to combine articulatory interpretation of speech with acoustic analysis has produced an epistemological conflict recently. To overcome this uneasy situation it is necessary to improve the phoneticians' ear-training and performance skills. A good acoustic analysis can be interpreted in audible based frameworks in the same way as an auditory analysis can be analyzed in acoustic based frameworks. Today it is important to ask how phoneticians carry on their scientific work at phonetics labs and in fieldworks. A short review of the history of phonetics shows that the conflict between acoustic and auditory approaches to phonetics in recent years is new and it has particularities not found in old times. The target is not to make a criticism on what is being produced recently in phonetics. However, the engineering tendency with all modern and technical facilities might not lose the main target of phonetics that is to produce a linguistic relevant analysis of speech.

Keywords: auditory analysis; acoustic analysis; ear-training; phonetic skills.

1.   Revisiting the question

In this congress the linguistic *corpora* is the focus. This is an important and today topic aiming the development of the linguistic science, in particular for the description of languages, in search of universal phonetic and phonological principles and for the definition of particular languages parameters. However to achieve such goals with real scientific results there are required good theories with appropriated approach to the investigating object.

2.   A historical conflict

It is important to ask how phoneticians carry on their scientific work at phonetic labs and in fieldwork. This question is not irrelevant, since it is expected that all scientist masters the approaches to the science they do. When technology is used in human science, the question often rises conflicts. Ladefoged (1973) and Yi Xu (2010), in different span of time put the question to phoneticians. Particularly, they showed the conflict between auditory and acoustic approaches to describe phonetic entities.

This conflict has created an epistemological situation in which the areas of phonetics and phonology took different ways, not rarely, producing contradictory results. In this way, the acoustic approach to speech restrict itself to the physics of speech, declaring the real science product. On the other hand, the auditory work describing the sounds of speech were treated as unscientific, idiosyncratic, highly individualized and without scientific value. Only with the support of an acoustic evaluation, the speech could be analyzed and describe properly. With all the recent facilities to carry on acoustic analyses of speech (PRAAT, WinPitch, SFS, ProTools, etc.), more people found convenient to produce acoustic works. If a paper has no acoustic printouts, statistic tables, graphics most certainly the paper will not be accept for publication and even for presentation. This is an awkward situation inside phonetics. The auditory description of speech has been used for centuries, has sophisticated the methodology and produced very nice, original and consistent pieces of linguistic description.

3.   A false conflict

The scenario presented above is typical of some groups of researchers and cannot be extended to the phoneticians in general. Congresses and periodical still accept papers based entirely on auditory researches.

A short review of the history of phonetics shows that the conflict between acoustic and auditory approaches to phonetics in recent years is new and it has particularities not found in old times. As a matter of fact, since the time when the technology to study the acoustics of sounds were presented to phoneticians (beginning of XX century), they started to look at the speech differently. The introduction of such technology and the set up of phonetics laboratories obliged the researchers, trained to do auditory analysis, to sophisticate their work, introducing in parallel acoustic analysis. The good company were welcomed because it helped linguistics to be seen as a science in modern terms. Besides, phonology were the linguistic area that brought more significant contributions to this idea at that time. So it seemed obvious that speech should be treated acoustically to be more scientific and audibly to be able to produce good phonological analysis..

A good example of the marriage between acoustic and auditory data to produce linguistic analyses is the MIT Report *Preliminaries to Speech Analysis: the distinctive features and their correlates* by Jakobson, Fant and Halle (1951). The reason by which the old phoneticians work with the two approaches is the fact that they used to do good ear-training and performance courses when students (Cagliari, 2007: 51-65, 130-131). It was unthinkable to work in phonetics without such training. On the other hand, the phoneticians found in the acoustic analysis an indispensable tool to check their auditory analyses. The two approaches were complementary.

Fant (1960) set up a definitive acoustic theory of speech, but he acknowledged the importance of auditory based analyses to achieve good acoustic based data. He said:

"The rules relating speech waves to speech production are in general complex since one articulatory parameter, e.g., tongue height, affects several of the parameters of the spectrogram. Conversely, each of the parameters of the spectrogram is generally influenced by several articulatory variables. However, to establish and learn these analytical ties is by no means a hopeless undertaking. Some elementary knowledge in acoustics is valuable, but the main requirement is a sound knowledge of articulatory phonetics" (Fant, 1967: 95).

Gordon Peterson recognized the difficulty in doing phonetics: "... it is clear that phonetics is a discipline of substantial complexity requiring much further experimental and theoretical research" (Peterson, 1968: 171; see also Fry, 1973, 1979: 4). Ladefoged was conscious of the necessity to work with an auditory and an acoustic approach to describe adequately the sounds of speech. He said: "Understanding speech is, in essence, a process of obtaining information from an auditory stimulus. This process involves discriminating between some sounds and considering other sounds to be similar". (Ladefoged, 1967: 143).

And, in another place, he comments: "Furthermore, although we could (with difficulty) characterize all possible systematic phonetic contrasts entirely in physiological terms, it would be ridiculous to overlook the fact that some phonological rules obviously work in terms of acoustic properties of sounds" (Ladefoged, 1971: 4).

Many other phoneticians share the same scientific point of view. As a matter of fact it should be obvious to think in that way. But things have never been smoothly in scientific agreement. The conflict started when people stopped receiving good ear-training and performance courses when students, mostly because these phoneticians came from engineering areas, like telephony and communication, or even from linguistics, but getting a different phonetics education. Obviously, when an engineering or a phonetician look at a spectrogram or other kind o printout they need to listen to the sound recorded to proceed any type of analyses. It means, in other words, that they do use auditory analysis to carry on any kind of interpretation for any kind of acoustic parameters. So, the point we make in this paper needs to be better understood.

As we know, the speaker's intuition is an essential tool to check linguistic value of data and language rules, following the generative (Chomsky, 1965) and the functional (Halliday, 1970a) approach to linguistic analysis. The speaker's consciousness of language works differently in different levels of linguistic analysis. A person recognizes that the word *horse* does not apply to the object *pencil*, and so on. An English speaker knows that it is wrong to say: *ball the kicked boy backyard the in.* The correct is: *the boy kicked the ball in the backyard.* The

intuition about the language structure works better in the area of semantics and syntax. It works rather well in relation to the phonological system of the language, but the same cannot be said when the intuition assess phonetic data. Without a good training in recognizing and producing the speech sounds according to the phonetic categories linguistically determined (cf. Catford, 1968: 309-333; IPA phonetic transcription symbols), the naive speaker may fall in many strange and erroneous conclusions about the sound he is inquired to explain. For instance, it is difficult for a person without specific training to categorize the vowels of a language, even when it is his native language. Ladefoged (1973) carried out a famous experiment in this respect and showed that phoneticians trained in the cardinal vowels system could agree in the identification of vowels quality. But phoneticians without such training committed many inaccuracies and mistakes. It is hard to convince at first sight a Portuguese speaker that he pronounces differently the "a" in words like *mais* and *maus* (a front and a back low vowel), because the language treats them as belonging to the same phoneme /a/. A phonetician without the appropriated training may describe these vowels acoustically as being unique. With this kind of analysis it is impossible to interpret the language sounds in appropriated terms. The criticism must be extended to all phonetic parameters. This is the reason by which some acoustic analysis does not reflect the linguistic rules of the language. Statistics cannot save a wrong basic phonetic interpretation.

Another aspect of the question (conflict) is crash between traditional phonetic theories based on linguistic approaches to language sounds and new acoustic theories proliferated recently. The discrepancy between the old and the new has being seen as a motif to introduce a new theory if when the results of the analysis and interpretation of the data are in clear contradiction with the linguistic analysis. In this respect, for instance, some papers show an interpretation of pitch variations that mischaracterizes the stress system and the rhythm of the language, since the oscillation between peaks and valley are interpreted differently from the way the speakers of the language do. Obviously any linguistic analysis must always convince the native speaker that the analysis refers to his own language. The most notorious example, however, is the acoustic interpretation of the typology of rhythm for languages. What sounds reasonable to the speakers ears that the rhythm may not change when the cadence varies has being interpreted acoustically as a chaos. It is hard to believe how some phoneticians look only to statistic data and not to the music structure of speech. As a consequence of such awkward interpretation of the rhythm, other levels of phonetic and phonological analysis has generated awkward categories of data and rules for the language. It is absolutely naive to believe that an acoustic analysis is performed without an auditory analysis, based on specific training. On the other hand, if we can account for a good acoustic analysis of speech, why not use them?

4. Doing articulatory and acoustic analysis

It is perfectly possible to interpret in acoustic terms analysis that has being made originally in an auditory framework. On the other hand it is equally possible to transform an acoustic interpretation into an auditory analysis. If the job has being done adequately this kind of interchanging approaches permuting acoustic and auditory analysis might be carried out easily. If not, it means that there is something wrong. However there is cases in which one approach does not match exactly to another, for the specification lack of some essential parameter or for unacceptable procedures producing unacceptable results. In spite of that, a good acoustic analysis can be converted into an auditory interpretation and vice-versa. A good example is the analysis of intonation produced according to Pierrehumbert (1980) and Halliday (1970) theoretical models. The first is inserted into the generative grammar and the second into the functional approach to grammar. Pierrehumbert's model is essentially acoustical and Halliday's is an audible based model. In both cases we have a record of the utterance that can be interpret in one or in another model. We converted Pierrehumbert examples into Halliday's analysis, taking as point of departure the location o focus. What comes before it constitutes the pretonic component in Halliday terms, and from the focus up to the end of the utterance there occurs the tonic component, and the definition of the tones. In the other way round we took Halliday's analysis and converted them into a sequence of High and Low pitch tones following Pierrehumbert theoretical framework. Some results of this job are presented as follow.
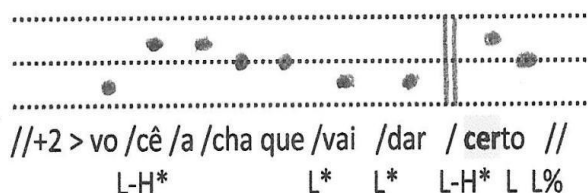
Figure 1: An utterance analyzed according to Halliday's theory and compared with the interpretation in Pierrehumbert's framework[1]

Figure 2: PRAAT printout showing the intonational analysis of the utterance *Você acha que vai dar certo*?
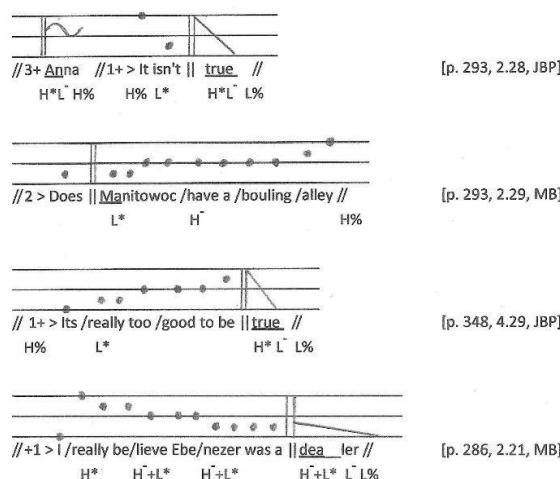
Figure 3: Examples from Pierrehumbert (1980) interpreted according to Halliday's (1970) framework

5. Conclusion

The constitution of linguistic corpora is as important as the theories which motivate and give them a scientific support. However it is useless to have a good corpus if there is no well trained phonetician to study it. Moreover, it is not enough to gather the required data in a good statistic program. A solid phonetic theory, compromised with the linguistic description of a language, is fundamental to produce nice pieces of work. It is generally admitted that phonetics science needs to take into account either auditory description and acoustic interpretation of speech. Other instrumental techniques are also complementary. Behind the action of viewing, hearing and interpreting speech data, there must always be the phoneticians' mind and the phonetic skill, acquired through specific ear-training and performance training with somebody who knows how to conciliate auditory and acoustic analysis. This kind of training cannot be achieved exclusively by reading textbooks or practicing individually. In this regard, doing phonetics is very similar of doing music.

6. References

Cagliari, L.C. (2007). Elementos de fonética do português brasileiro. São Paulo: Editora Paulistana.

Catford, J.C. (1968). *The articulatory possibilities of man*. In: *Manual of Phonetics*. B. Malmberg (Ed.). Amsterdam:

[1] In the example, the tone values in Hz are: Mid-High: 160.55; 162.54; 150.94. Mid:112.60; 117.56; Mid-Low: 110.42; 101.69; 91.84.

North-Holland Publishing Co. pp. 309--333.

Chomsky, N.A. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Fant, C.G..M. (1960). *Acoustic theory of speech production.* The Hague: Mouton and Co.

Fant, C.G..M. (1960). Descriptive analysis of the acoustic aspects of speech. In I. Lehiste (Ed.), *Reading in acoustic phonetics*. Cambridge: The M.I.T. Press, pp. 93--107.

Fry, D.B. (1973). Linguistic theory and experimental research. In W.E. Jones, J. Laver (Eds.), *Phonetics in linguistics: a book of readings.* London: Longman. pp. 66--86.

Fry, D.B. (1979). *The physics of speech*. Cambridge: Cambridge University Press.

Jakobson, R., Fant, G. and Halle, M. (1951). *Preliminaries to Speech Analysis: the distinctive features and their correlates*. MIT Report. Cambridge: The MIT Press.

Jones, W.E.; Laver, J. (Eds.). (1973). *Phonetics in linguistics: a book of readings.* London: Longman.

Halliday, M.A.K. (1970). *A course in spoken English: intonation*. London: Oxford University Press, 1970.

Halliday, M.A.K. (1970a). Language Structure and Language Function. In J. Lyons (Ed.) *New horizons in linguistics*. London: Penguin Books, pp. 140--165.

Ladefoged, P. (1967) P. Units in the perception and production of speech. In *Three areas of experimental phonetics*. London: Oxford University Press, pp. 143--172.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics.* Chicago: The University of Chicago Press.

Ladefoged, P. (1973). The value of phonetic statements. In W.E. Jones, J. Laver (Eds.), *Phonetics in linguistics: a book of readings.* London: Longman. pp. 218--228.

Lehiste, I (Ed.). (1967). *Reading in acoustic phonetics*. Cambridge: The M.I.T. Press.

Malmberg, B. (Ed.). (1968). *Manual of Phonetics*. Amsterdam: North-Holland Publishing Co.

Peterson, G.E. (1968). The speech communication process. In B. Malmberg (Ed.), *Manual of Phonetics*. Amsterdam: North-Holland Publishing Co., pp. 155--172.

Pierrehumbert, J.B. (1980) *The Phonology and Phonetics of English Intonation*. PhD. thesis, MIT, Published by Indiana University Linguistics Club.

Xu, Y. (2010). In defense of lab speech. In *Journal of phonetics.* London: London W C1N, pp. 329--336.

O alinhamento do pico da F0 na questão total da região sudeste: um estudo preliminar

Joelma CASTELO

Universidade Federal do Rio de Janeiro (UFRJ), Brazil

Av. Horácio Macedo, 2151 – Cidade Universitária, Rio de Janeiro, RJ – 21941-917

joelmacastelo@gmail.com

Resumo

O presente estudo objetiva analisar a localização do pico da F0 na sílaba nuclear das questões totais das capitais do sudeste brasileiro, utilizando o *corpus* do projeto ALiB.Os resultados apontam para uma diferenciações regional que opõe Rio de Janeiro e São Paulo a Vitória e Belo Horizonte.

Palavras-chave: entoação; alinhamento; prosódia regional.

1. Objetivo

O objetivo do presente trabalho é descrever o fenômeno entoacional do alinhamento em enunciados do tipo questão total, produzidos por informantes cultos dos quatro estados do sudeste brasileiro - Belo Horizonte, Vitória, Rio de Janeiro e São Paulo. Essas questões foram recolhidas do *corpus* do projeto ALiB. Considerando análise feita com amostras de fala de informantes não-cultos retirados da mesma matriz, postula-se que o referido fenômeno pode ser objeto de uma diferenciação regional.

Essa descrição ajuda a enriquecer o conhecimento a respeito da diversidade de traços que caracteriza a questão total do português brasileiro encontrada por Silva (2011).

2. Pressupostos teóricos

O alinhamento do pico localizado na última sílaba tônica do enunciado interrogativo está sendo estudado sob perspectivas fonológicas e fonéticas. Do ponto de vista fonológico, esse comportamento prosódico mostra-se como peça-chave para distinguir a pergunta do pedido no PB (Moraes & Colamarco, 2007). A oposição fonológica entre esses dois atos ilocutórios se realiza através de um movimento ascendente na curva da F0, quando se produz uma questão total neutra, e de um movimento descendente na curva da F0, quando se produz um pedido. Silva, Couto e Pinto constatam que os falantes nativos do PB transferem essa marca quando falam uma língua estrangeira. No espanhol, língua investigada pelas autoras, a pergunta é realizada por meio de uma curva ascendente, já os brasileiros usam o contorno circunflexo com pico alinhado à direita para produzi-la. O mesmo ocorre com o pedido; ao passo que no espanhol ele é realizado por meio de uma curva descendente, o brasileiro falante de espanhol como L2, produz essa diretiva através de um contorno circunflexo com pico alinhado à esquerda.

Além das diferenças fonológicas, o alinhamento do pico na sílaba tônica também revela diferenças diatópicas entre as línguas. Segundo Ladd (1999: 128), "two languages or dialects may have the same tonal sequence used the same way, but align the tonal targets differently with respect to the stressed syllable". O autor cita o fenômeno do alinhamento do pico como sendo uma variante linguística encontrada em dialetos do Sueco e do dinamarquês. A respeito do Português do Brasil, Antunes (2011) faz um estudo comparativo preliminar entre a entoação de enunciados interrogativos e assertivos neutros de duas cidades de Minas Gerais: Belo Horizonte e Mariana, com base no *corpus* do projeto AMPER. A autora constata que, enquanto no falar de Belo Horizonte o alinhamento do pico ocorre à esquerda da sílaba tônica em enunciados interrogativos, em Mariana essa localização é simetricamente oposta, ou seja, à direita da sílaba tônica final.

Silva (2011), ao comparar as variedades faladas nas capitais brasileiras, descreve para região sudeste uma representativa quantidade de enunciados interrogativos cujo pico está alinhado à esquerda da última sílaba tônica. Observa-se no gráfico abaixo a proporção em que ocorre o alinhamento do pico à direita (vermelho), padrão mais comum, em comparação à realização do pico à esquerda da sílaba (azul). Observa-se em relação a este tipo de comportamento o seguinte contraste: 37% no sudeste e menos de 20% nas demais regiões. Chama-se a atenção para os dados de duas capitais: Vitória, onde o movimento descendente na tônica é o comportamento predominante; e Belo Horizonte, onde esse mesmo contorno se realiza em cerca de 50% dos dados.
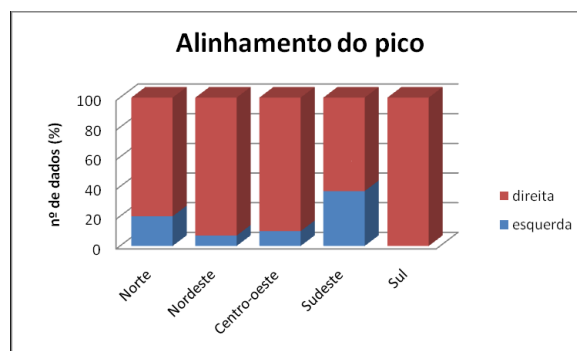
Gráfico 1: Valores percentuais do alinhamento do pico nuclear na fala dos não- cultos

3.   Metodologia

3.1   Dados

A amostragem é composta por 19 dados de fala semi-espontânea retirados do questionário de prosódia do ALiB, cujos enunciados são apresentados a seguir. Em cada tópico, aparece primeiro a questão total que se espera como resposta do informante e, em seguida, a diretiva que o inquiridor formula para obtê-la.

- Você vai sair hoje?

  Se você / o (a) senhor (a) quer saber se alguém vai sair hoje, como é que você / o (a) senhor(a) pergunta?

- Eu vou sair hoje, doutor?

  Uma pessoa está internada em um hospital e quer saber do médico se vai sair naquele dia. Como é que pergunta?

3.2   Perfil sociolinguístico do informante

Quatro informantes cultos e naturais de cada localidade equitativamente divididos em duas faixas etárias , 18 a 30 anos e 50 a 65 anos , e entre os dois gêneros.

4.   Análise

Dividiu-se a duração da última vogal acentuada em três partes iguais, denominadas de esquerda, meio e direita. Aferiram-se os valores da F0 nesses três pontos a fim de verificar o comportamento da entoação neles. Entende-se que, ao localizar o ponto máximo, pode-se descrever, de forma mais detalhada, os movimentos ascendente e descendentes nessas sílabas, isto é, conhecer se o seu pico está alinhado no início, no meio ou no final da vogal.

5.   Resultados

Os resultados da presente pesquisa confirmam o falto de o padrão fonológico L+<H*L%, descrito por Moraes, não ser o mais expressivo em termos percentuais na região sudeste. A realização do pico ocorreu com mais frequência no meio da última vogal tônica, 47 % dos dados, ficando o restante dos dados divididos entre o alinhamento à direita da vogal, 42% dos dados e o alinhamento à esquerda da vogal, 11% dos dados.

Em termos relativos, Belo Horizonte foi a capital em que o pico alinhado no meio da tônica esteve mais presente. Ao contrário do que foi encontrado para os falantes não-cultos, não foi verificado nessa capital nenhum enunciado em que o pico tenha ocorrido mais próximo à fronteira esquerda do constituinte. Já em Vitória, 10% dos dados apresentaram o pico alinhado à esquerda e 15% dos dados, o pico alinhado à direita. O alinhamento do pico ao meio da sílaba também se mostrou predominante em Vitória, somando 75% dos

dados. No Rio de Janeiro, o padrão ascendente da F0 descrito por Moraes foi realizado na maioria dos dados, totalizando 70% dos dados, embora tenha-se encontrado também 15% de dados com pico alinhado ao meio e 15 % de dados com pico alinhado à esquerda da última vogal tônica. Em São Paulo, por fim, os dados estão equitativamente divididos entre alinhamento do pico no meio e alinhamento do pico à direita, não sendo encontrado para essa capital um movimento descendente na vogal tônica.

		Medidas na última vogal tônica nuclear			
		Duração (s.)	Frequência (Hz)		
			E	M	D
Belo Horizonte	j138.6.2	0.25	204	329	291
	j138.6.3	0.26	189	277	240
	j138.7.1	0.19	121	135	98
	j138.8.1	0.22	196	227	312
Vitória	j190.5.1	0.17	131	214	251
	j190.5.2	0.19	125	158	125
	j190.6.1	0.24	209	175	167
	j190.7.1	0.17	99	112	99
	j190.7.2	0.14	94	105	91
	j190.8.1	0.20	131	155	166
	j190.8.2	0.19	137	164	129
Rio de Janeiro	j202.5.1	0.16	137	169	170
	j202.5.2	0.22	137	156	162
	j202.2.1	0.19	191	268	280
	j202.7.2	0.16	80	123	133
	j202.7.3	0.18	66	83	80
	j202.8.2	0.18	211	184	162
São Paulo	j179.5.1	0.15	104	123	114
	j179.7.1	0.18	93	121	130

Legenda	
E	Esquerda
M	Meio
D	Direita

Tabela 2: Valores de duração e F0 dos picos na última vogal tônica nuclear
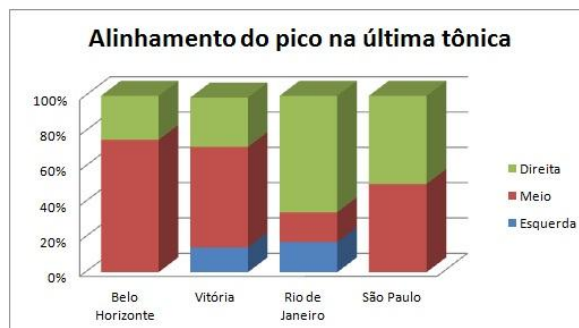
Gráfico 3: Valores percentuais do alinhamento do pico nuclear na fala dos culto

5.1   Algumas ilustrações

a.   Alinhamento no meio da última tônica
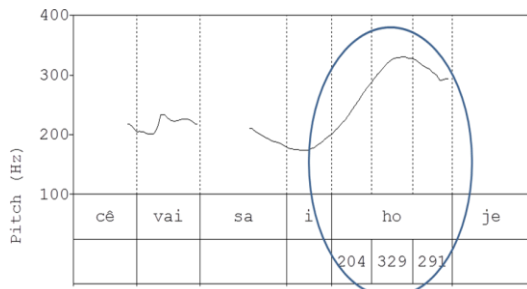
Figura1: Frase *Cê vai sair hoje?, dita pela informante M1 de Belo Horizonte*

b.   Alinhamento à esquerda da última tônica
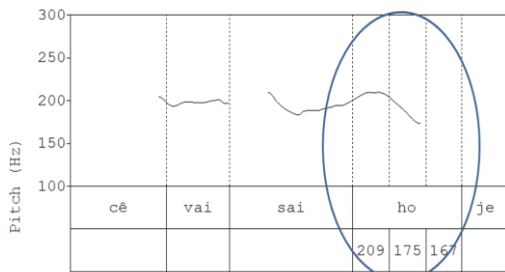
Figura 2: Frase *Cê vai sair hoje?, dita pela informante M1 de Vitória*

c.   Alinhamento à direita da última tônica
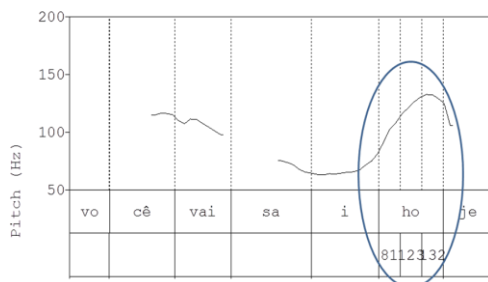
Figura 3: Frase *Você vai sair hoje?, dita pela informante H1 do Rio de Janeiro*

6.   Considerações Finais

O presente estudo corrobora o fato de os falares do sudeste brasileiro apresentarem particularidades prosódicas no domínio intrassílabico da questão total neutra. Nas demais regiões, Silva (2011) constata que, na fala dos não-cultos, o padrão descendente ocorre em menos de 20% dos dados, ao passo que na região sudeste esse quantitativo cresce para quase 40%. Para fala dos cultos, os resultados supracitados mostram que o alinhamento do pico no meio da vogal é predominante nas capitais de Belo Horizonte e Vitória, capitais essas que também apresentaram comportamentos semelhantes na fala dos não-cultos. Observou-se ainda que, no falar carioca, ocorre o predomínio do alinhamento à direita e que, no falar paulistano, os movimentos intrassílabicos ascendente e descendente dividem o nº de ocorrências.

7.   Referências

Antunes, L. (2011). Análise prosódica de sentenças declarativas e interrogativas do dialeto mineiro (Brasil) com diferentes Sintagmas Nominais (SN's) na posição de sujeito. In *Revista Internacional de Linguística Iberoamericana* (RILI) La prosodia en lenguas y variedades del ámbito iberorrománico. v. IX, n. 17. Sevilha: Vervuert.

Moraes, J.A., Colamarco, M. (2007). Você está pedindo ou perguntando? Uma análise entonacional de pedidos e perguntas na fala carioca. In *Revista de estudos da linguagem*, v. 15, pp. 113--126.

Ladd, R. (1996). *Intonational phonology*. Cambrige: Cambridge University Press.

Silva, J. (2011). Caracterização Prosódica dos falares brasileiros: as orações interrogativas totais. Dissertação de mestrado em língua portuguesa. Rio de Janeiro, Faculdade de Letras, UFRJ.

Silva, C., Couto, L and Pinto, M. (2011). Pedidos de Informação e Pedidos de Ação em Português do Brasil, fala carioca e em Espanhol Europeu, fala madrilena: variantes ou padrões entonacionais distintos? In *Anais do III Colóquio Brasileiro de Prosódia da Fala*, realizado de 6 a 8 de junho de 2011. Belo Horizonte, UFMG..

Rhythm variation in spontaneous and induced speech

Amedeo DE DOMINICIS

University of Tuscia

Via del Paradiso, 47 - 01100 - Viterbo - Province of Viterbo - Italy

dedomini@unitus.it

Abstract

This paper investigates to what extent the metrics of speech can be induced by the pragmatic conditions of communication. We elicit two Italian corpora: the first was elicited by means of an experimental collaborative task; the other is a natural polemical interaction with overlapping turn-taking. The analysis of the former shows that the trend towards a syllable-timed or a stress-timed rhythm can be experimentally induced and it is an effect of the communicative interaction. The analysis of the polemical corpus demonstrates that rhythmical patterns vary according to the conversational goals of the speakers. The qualitative and statistic results confirm that no stable rhythmic pattern exists. Furthermore, the metric trend of each turn changes according to the conversational purposes: in particular, the speaker may borrow his interlocutor rhythm – or the opposite one – in order to collaborate with or to dominate him by cutting or easing the antagonist rhythm.

Keywords: rhythm; conversation; Italian.

1.   Introduction

Two main approaches to linguistic rhythm exist in the literature: the hypothesis of rhythmic discrete types and the assumption of rhythm as a variable property which does not belong to the linguistic system, but to conversational interaction. In this latter approach the rhythm function is to handle cooperation and conflict among the speakers. Therefore it is not stable, but varies according to its conversational functions.

1.1   Rhythm-property of the system

The hypothesis of rhythmic types goes back to the forties (Lloyd James, 1940; Pike, 1945; Abercrombie, 1967; Faure, Hirst & Chafcouloff, 1980; Dauer, 1983). It mainly consists of a binary classification (syllable-timed/stress-timed languages). But it has not yet been clearly experimentally validated (e.g. Shen & Peterson, 1962; Bolinger, 1965; O'Connor, 1965; Uldall, 1971; Lea, 1974; Lehiste, 1977; Donovan & Darwin, 1979; Roach, 1982; Wenk & Wiolland, 1982; Borzone de Manrique & Signorini, 1983; Dauer, 1983; Drake & Palmer, 1993). According to a weaker hypothesis, rhythm is a perceptual impression arising from the convergence of some clusters of phonological properties typical of a given language (e.g. Dasher & Bolinger, 1982; Nespor & Vogel, 1986; Dauer, 1987; Bertinetto, 1981, 1989; Nespor, 1990; Ramus, Nespor & Mehler, 1999). The linguistic typology (syllable/stress-timed) is not discrete and different systems are spread out over a *continuum*[1].

1.2   Rhythm-variable property of conversation

This hypothesis derives from conversational analysis studies, and represents rhythmic features in Gestalt terms. Recently, a new impetus has been given by the so-called phonetic-details studies (cf. Sacks, Schegloff & Jefferson, 1974; Erickson, 1982; Erickson & Shultz, 1982; Cutler, 1991; Couper-Kulhen, 1989, 1990, 1993, 2001; Buder

1986, 1991, 1996; Auer *et al.*, 1999; Buder & Eriksson, 1997, 1999; Local, 2003; Fon, 2006; House, 2007; Russo & Barry, 2008; Arvaniti, 2009; Reed, 2010). In this paradigm, during interaction, rhythm may vary due to the conversational tasks, it is not a property of the system, but a tactical resource of the speaker[2].

2.   Experimental analysis

We ran an experimental test in order to verify to what extent the speech metrics can be induced by some pragmatic conditions. Two corpora have been elicited. The first corpus was obtained by an experimental collaborative task in which the subjects were asked to synchronize their speech with a recorded one. The second is a natural corpus in which two speakers are engaged in a polemical interaction (the so-called quarrel between Vittorio Sgarbi and Mike Bongiorno during the TV show *Telemike* in 1991). In each of the two corpora we took two measurements: the interstress intervals (henceforth "Acc" = the temporal distance between the stressed syllables); the syllabic intervals (henceforth "Syl" = the duration of stressed and unstressed syllables). These measurements were used to check the metrical typology (stress/syllable-timing) and its variation along the corpus.

2.1   Experiment on the collaborative corpus

We recorded the sentence *Il capostazione ha spento la luce* ('The station master has turned the lights out'). Then we manipulated the signal in order to build new ones with constant Syl or Acc. On these signals we built a "Listen & Repeat" test. The working hypothesis was that listening to these signals will induce the listener to a syllable- or a stress-timing rhythm, according to the manipulated signals. To the same purpose, before the original signal,

---

[1] Also according to the PVI hypothesis (Low & Grabe, 1995; Low, Grabe & Nolan, 2000; Grabe & Low, 2002; Patel & Daniele, 2003), rhythm is an intrinsic property of the system.

[2] To this paradigm three more branches belong: the studies on the metrical feet variability (*heterometry*) (Brown & Weishaar, 2010); the studies on rhythm as an entrainment phenomenon (Cummins & Port, 1998; Port, 2003; Cummins, 2009); the studies of rhythm as an Adaptive Oscillator (Port, Cummins & Gasser, 1996).

we inserted three beeps[3], 150 ms apart (equal to the mean Syl in the original signal) and 496 ms apart (equal to the mean Acc in the original signal). The signals for this listening test (*passive corpus*) are listed in tab. 1. The *active corpus* is composed of the sentences that the subjects recorded after listening to the passive corpus.

Five university students took part in the experiment: S1 (man, age 52, born in Rome where he lives); S2 (woman, age 24, born in Terni where she lives); S3 (woman, age 51, born in Orune-Nu, but living in Tuscania-Vt); S4 (woman, age 19, born in Alatri-Fr where she lives), S5 (woman, age 19, born in Civitavecchia where she lives). They listen to signal A, and are asked to utter the same sentence into the microphone (signal 1 of the active corpus). Then they listen to the further 4 signals and repeat them into the microphone, trying to imitate them and keep as close to the same timing of the signal they listened with headphones. Thus, the *active corpus* contains 25 signals (5 per subject), as shown in tab. 1.

2.2   Experimental expectations and results

Compared to signal 1 (recorded at the beginning of the session), signals 4 and 3 should show equalized Acc durations and close to 496 ms (but the absolute value depends on the speaking rate). Signals 5 and 2, should show equalized Syl durations and close to 150 ms (but the absolute value depends on the speaking rate). If these expectations are confirmed, then the syllable-timed or stress-timed rhythm is an effect of the communicative interaction. As we are able to induce it, it is not a property of the linguistic system. The results validate these expectations. Tab. 2 shows the Syl and Acc durations in the active corpus, and their standard deviation (σ). Signals 2 and 5 systematically approach the reference value (150 ms) as compared to signal 1; likewise, 3 and 4 systematically approach the reference value (496 ms). Therefore the σ decreases.

2.3   Experiment on the polemical *corpus*

As for the polemical corpus it is the quarrel between two Italian TV showmen: Mike Bongiorno and Vittorio Sgarbi (*Telemike* in 1991) - 🔊 . It was downloaded from *YouTube*. Its low audio quality creates no problem with the duration measurements. It is a communicative situation where the speakers do not collaborate, but manage to hinder and sabotage each other.

2.4   Experimental expectations and results

In the polemical *corpus* we expect a minimal degree of rhythmic integration: i.e. anisochrony. The results confirm these expectations. Indeed, no stable rhythmic pattern exists. Furthermore, the metric of each turn changes according to the conversational purposes; in particular, the speaker may borrow his interlocutor rhythm – or the opposite one – in order to dominate by cutting or easing the antagonist rhythm. The chaining and the syntagmatic

succession of metrical types in each speech turn is a function of the conversational strategy of the speaker to create dominance. As is seen in tab. 3, at the beginning both speakers alternate different metrics, in a sort of "skirmish" (opposed rhythm speech turns). A "truce" follows: Sgarbi tries to disrupt the metrical strategy of Mike, using an asynchronous rhythm. Then, both speakers resume their confrontation, but change their tactics: there is the first instance of speech turns overlapping ("mimetic") where their rhythm is common: stress-timed and synchronized. Next is a second overlapping ("rolling") where, again, the metrics of the quarrellers are completely asynchronous, but dynamically tuned. Then, there are two "interval" speech chains: a stress-timed trend chain by Mike and a following one by Sgarbi, showing an opposed syllable-timed trend. Then, the "mimetic" tactic resumes, but the rhythmic features are reversed, as compared to the previous one: there is a third turns overlapping with a common syllable-timed and synchronized metric. In the "end", three chains by Mike show an alternate rhythmical trend: stress- and syllable-timed. Four examples of these turns are given below[4].

1. Truce-rhythm (tab. 3: chains 14-15; fig. 1-2). Sgarbi shows no rhythmic isochrony: an extreme case of polemical strategy (maybe in order to sabotage Mike's rhythm). His rhythm is anisochronous: the σ values are very high both for the Acc and the Syl mean durations.

2. Mimetic-rhythm (tab. 3: chain 16; fig. 3). It is a first overlapping where both speakers tend to have a common stress-timed rhythm, and synchronized interstress boundaries: the difference between their mean Acc duration is not significantly different, as assessed by Student's and Anova tests.

3. Rolling-rhythm (tab. 3: chain 17; fig. 4), a second overlapping. The turns are anisochrounous (the difference between their mean Acc duration is significantly different, as assessed by Student's and Anova tests), but with a peculiarity: both speakers undertake a sort of rhythmic "rolling relay", where the metrics of the both quarrellers is *dynamically tuned*: i.e. each turn takes up the Acc durational trend towards increasing or decreasing of the previous one, uttered by the interlocutor. As you see in fig. 5, Sgarbi produces a sequence of three increasing intervals (62-190-388 ms), followed by a reply by Mike with three equally increasing intervals (336-322-417 ms); then, Sgarbi reverses the trend, realizing a 170 ms interval and Mike pursues the decreasing trend with a 291 ms interval. Finally, Sgarbi reverses again the trend and produces a 236

---

[3] At least three evenly spaced beats are required in order to establish an isochronous chain (Couper-Kuhlen, 1990: 16).

[4] The signals are annotated by means of 6 *Praat Tiers*, as follows: (1) orthographic transcription: Mike, (2) orthographic transcription: Sgarbi, (3) Mike's Syl: IPA and boundaries, (4) Sgarbi's Syl: IPA and boundaries, (5) Mike's Acc: IPA and boundaries, (6) Sgarbi's Acc: IPA and boundaries.

ms interval, and Mike replies with the same increasing trend (365 ms).

4. Mimetic-rhythm (tab. 3: chain 20; fig. 6). In this third overlapping, both speakers tend to have a common syllable-timed rhythm and they even tend to overlap, to make isotopic (synchronized) their syllabic boundaries: the difference between their mean Syl is not significantly different, as assessed by Student's and Anova tests.

3. References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Arvaniti, A. (2009). Rhythm, timing, and the timing of rhythm. *Phonetica*, 66 (1-2), pp. 46-63.

Auer, P. et al. (1999). *Language in Time*. New York: Oxford University Press.

Bertinetto, P.M. (1981). *Strutture prosodiche dell' italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*. Firenze: Accademia della Crusca.

Bertinetto, P. M. (1989). Reflections on the dichotomy 'stress' vs. 'syllable-timing'. *Revue de Phonétique Appliquée*, 91/93, pp. 99--130.

Bolinger, D. (1965). Pitch accent and sentence rhythm, *Forms of English: Accent, morpheme, order*. Cambridge, MA: Harvard University Press.

Borzone de Manrique, A.M., Signorini, A. (1983). Segmental durations and the rhythm in Spanish. *Journal of Phonetics*, 11, pp. 117--128.

Brown, S., Weishaar, K. (2010). Speech is "heterometric": The changing rhythms of speech. In M. Hasegawa-Johnson *et al.* (Eds.), *Proceedings of the Fifth International Conference on Speech Prosody*, paper 074. Chicago: ISCA.

Buder, E.H. (1986). *Coherence of speech rhythms in conversations: Autocorrelation analysis of fundamental voice frequency*. Toronto: Toronto Semiotic Circle.

Buder, E.H. (1991). *Vocal synchrony in conversations: spectral analysis of fundamental voice frequency*. PhD diss. Madison, WI: University of Wisconsin.

Buder, E.H. (1996). Dynamics of speech processes in dyadic interaction. In Watt, J., Van Lear, C. A. (Eds.), *Dynamic Patterns in Communication Processes*, pp. 301-325. Thousand Oaks, CA: Sage.

Buder, E.H., Eriksson, A. (1997). Prosodic cycles and interpersonal synchrony in American English and Swedish. In G. Kokkinakis, N. Fakotakis, E. Dermatas. (Eds.), *Eurospeech '97 Proceedings*, Grenoble: European Speech Comm. Association, vol. 1, pp. 235--238.

Buder, E.H., Eriksson, A. (1999). Time-series analysis of conversational prosody for the identification of rhythmic units. In A.C. Baile (Ed.), *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco: Univ. of California, vol. 2, pp. 1071-1074.

Couper-Kulhen, E. (1989). Speech rhythm at turn transitions: its functioning in everyday conversation. Part I, *KontRI* Working Paper No. 5, University of Konstance, Fachgruppe Sprachwissenschaft.

Couper-Kulhen, E. (1990). Discovering rhythm in conversational English: Perceptual and acoustic approaches to the analysis of isochrony, *KontRI* Working Paper No. 13, University of Konstance, Fachgruppe Sprachwissenschaft.

Couper-Kulhen, E. (1993). *English Speech Rhythm: Form and Function in Everyday Verbal Interaction*. Amsterdam: Benjamins.

Couper-Kulhen, E. (2001). Intonation and Discourse: Current Views from Within. In Schiffrin, D., Tannen, D., Hamilton, H. E. (Eds.), *The Handbook of Discourse Analysis*, Oxford: Blackwell, pp. 13--34.

Cummins, F. (2009). Rhythm as Entrainment: The Case of Synchronous Speech. *Journal of Phonetics*, 37(1), pp. 16--28.

Cummins, F., Port, R.F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26 (2), pp. 145--171.

Cutler, A. (1991). Prosody in situations of communication: salience and segmentation. In *Proceedings of the 12th International Congress of Phonetic Sciences*, 1, Aix-en-Provence, pp. 264--270.

Dasher, R., Bolinger, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association*, 12, pp. 58--69.

Dauer, R.M. (1983). Stress-timing and Syllable-timing Reanalyzed. *Journal of Phonetics*, 11, pp. 51--62.

Dauer, R.M. (1987). Phonetic and Phonological Components of Language Rhythm. In *Proceedings of the 11th International Congress of Phonetic Sciences*, 5, Tallinn: Academy of Sciences, pp. 447--450.

Donovan, A., Darwin, C.J. (1979). The perceived rhythm of speech. In *Proceedings of the 9th International Congress of Phonetic Sciences*, 2, Copenhagen: Institute of Phonetics, pp. 268--274.

Drake, C., Palmer, C. (1993). Accent structure in music performance. *Music Perception*, 10(3), pp. 343--378.

Erickson, F. (1982). Money tree, lasagna bush, salt and pepper: Social construction of topical cohesion in a conversation among Italian Americans. In D. Tannen (Ed.), *Analyzing discourse: Text and talk*, Washington, D.C.: Georgetown University Press, pp. 43--70.

Erickson, F., Shultz, J. (1982). *The Counselor as Gate Keeper: Social Interaction in Interviews*. New York: Academic Press.

Faure, G., Hirst, D.J. and Chafcouloff, M. (1980). Rhythm in English: Isochronism, pitch and perceived stress. In L.R. Waugh, C.H. van Schoolneveld (Eds.), *The Melody of Language*, Baltimore: University Press Park, pp. 71--79.

Fon, J. (2006). Cross-dialectal turn exchange rhythm in English interviews. In R. Hoffman, H. Mixdorff (Eds.), *Proceedings of the 3rd Speech Prosody Conference*, PS4, Dresden: TUDpress, pp. 16--179.

Grabe, E., Low, E.L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven C, W. Warner (Eds.), *Papers in Laboratory Phonology 7*. The Hague: Mouton de

Gruyter. Available at: <http://www.phon.ox.ac.uk/~esther/ivyweb/Grabe_Low.doc>.

House, J. (2007). The role of prosody in constraining context selection: a procedural approach. *Nouveaux cahiers de linguistique française*, 28, pp. 369--383.

Lea, W.A. (1974). *Prosodic Aids to Speech Recognition: IV. A General Strategy for Prosodically-Guided Speech Understanding*. St. Paul, Minnesota: Univac Report PX10791, Sperry Univac, DSD.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5, pp. 253--263.

Lloyd James, A. (1940). *Speech Signals in Telephony*. London. Pitman.

Local, J. (2003). Phonetics and talk-in-interaction. In M. Solé, D. Recasens, J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona: Casual, pp. 115--118.

Low, E. L., Grabe, E. (1995). Prosodic patterns in Singapore English. In K. Elenius, P. Branderud (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences*, 3, Stockholm, pp. 636--639.

Low, E.L., Grabe, E. and Nolan, F. (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language & Speech*, 43 (4), pp. 377--401.

Nespor, M. (1990). On the rhythm parameter in phonology. In I.M. Roca (Ed.), *Logical issues in language acquisition*, Dordrecht: Foris, pp. 157—175.

Nespor, M., Vogel, I. (1986). *Prosodic phonology.* Dordecht: Foris.

O'Connor, J. D. (1965). *The perception of time intervals* (Progress Report 2) Phonetics Laboratory, London: University College.

Patel, A.D., Daniele, J.R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, 87 (1), B35-B45.

Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Port, R. (2003). Meter and speech. *Journal of Phonetics*, 31, pp. 599--611.

Port, R., Cummins, F. and Gasser, M. (1996). A Dynamic Approach to Rhythm in Language: Toward a Temporal Phonology. In B. Luka, B. Need (Eds.), *Proceedings of the Chicago Linguistics Society*, Dept. of Linguistics, University of Chicago, pp. 375--397.

Ramus, F., Nespor, M. and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 72, pp. 1--28.

Reed, B.S. (2010). Speech rhythm across turn transitions in cross-cultural talk-in-interaction. In *Journal of Pragmatics*, 42 (4), pp. 1037--1059.

Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (Ed.), *Linguistic controversies: essays in linguistic*

theory and practice in honour of F.R. Palmer, London: Edward Arnold, pp. 73--79.

Russo, M., Barry, W. (2008). Isochrony reconsidered: Objectifying relations between rhythm measures and speech tempo. In P.A. Barbosa, S. Madureira and C. Reis (Eds.), *Proceedings of the Speech Prosody 2008 Conference*, Campinas, pp. 419--422.

Sacks, H., Schegloff, E. and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, pp. 696--735.

Shen, Y., Peterson, G.G. (1962). Isochronism in English. *University of Buffalo Studies in Linguistics - Occasional papers*, 9, pp. 1--36.

Uldall, E.T. (1971). Isochronous stresses in RP. In Hammerich, E. (Ed.), *Form and Substance*, Copenhagen: Akademisk Forlag, pp. 205--201.

Wenk, B.J., Wiolland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, 10, pp. 193--216.

Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Port, R. (2003). Meter and speech. *Journal of Phonetics*, 31, pp. 599--611.

Port, R., Cummins, F. and Gasser, M. (1996). A Dynamic Approach to Rhythm in Language: Toward a Temporal Phonology. In Luka, B., Need, B. (Eds.) *Proceedings of the Chicago Linguistics Society*, Dept. of Linguistics, University of Chicago, pp. 375--397.

Ramus, F., Nespor, M. and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 72, pp. 1--28.

Reed, B.S. (2010). Speech rhythm across turn transitions in cross-cultural talk-in-interaction. *Journal of Pragmatics*, 42 (4), pp. 1037--1059.

Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal. (Ed.), *Linguistic controversies: essays in linguistic theory and practice in honour of F.R. Palmer*, London: Edward Arnold, pp. 73--79.

Russo, M., Barry, W. (2008). Isochrony reconsidered: Objectifying relations between rhythm measures and speech tempo. In P.A. Barbosa, S. Madureira, and C. Reis (Eds.), *Proceedings of the Speech Prosody 2008 Conference*, Campinas, pp. 419--422.

Sacks, H., Schegloff, E., Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, pp. 696--735.

Shen, Y., Peterson, G. G. (1962). Isochronism in English. *University of Buffalo Studies in Linguistics - Occasional papers*, 9, pp. 1--36.

Uldall, E. T. (1971). Isochronous stresses in RP. In Hammerich, E. (ed.) *Form and Substance*, pp. 205--201. Copenhagen: Akademisk Forlag.

Wenk, B. J., Wiolland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, 10, pp. 193--216.

4.   Appendix

Figure 1: Sgarbi: *no non puoi dirle perché dici delle cazz* ['no, you can't tattle as you talk bullshit'] ◀))

Figure 2: Sgarbi: *questo è il concetto* ['this is the idea'] ◀))

Figure 3: Mike: *Dici tu. Va bene. Io non dico nessuna caz.* ['You say. Ok. I don't talk bullshit']; Sgarbi: *Sì d'accordo, le diciamo insieme, siamo insieme a dirlo..io* ['Yes, ok, we talk at once. We talk at once, ..I'] ◀))

Figure 4: Mike: *Adesso parlo io. Sì adesso mettila giù* ['No it's my turn. Yes now put it down']; Sgarbi: *Ora litighiamo. Vuoi fare a pugni con me? No puoi parlare. No* ['Now we are going to have a quarrel. Do you want to box with me? No you can talk. No'] ◀))
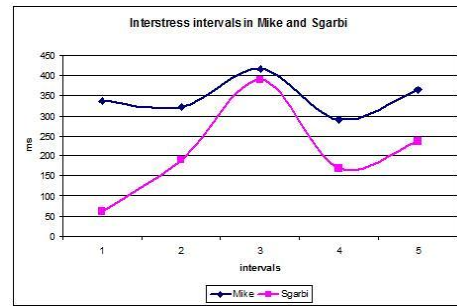
Figure 5: Rolling-rhythm. Variation of interstress intervals duration by Mike and Sgarbi, arranged in sequence

Figure 6: Mike: *Io non ho ancora.. Io non ho* ['I have not yet .. I have not..']; Sgarbi: *Quelli che parlano a vanvera. È andato* ['Those who blether. It's gone'] ◀))

Signal		Passive corpus
A	◀))	the natural original signal (writer's voice)
B	◀))	three beeps (150 ms apart) + signal A
C	◀))	three beeps (496 ms apart) + signal A
D	◀))	signal A: equalized Acc durations: 496 ms
E	◀))	signal A: equalized Syl durations: 150 ms
		Active corpus
1		the natural original signal (subject's voice)
2		signal recorded after listening to the signal B
3		signal recorded after listening to the signal C
4		signal recorded after listening to the signal D
5		signal recorded after listening to the signal E

Table 1: Passive and active corpora

		Signal 1		Signal 2		Signal 5	
		ms	σ	ms	σ	ms	σ
Syllables	S1	141.2	43.78	159.8	47.44	149.0	36.38
	S2	201.7	82.06	197.5	47.07	190.3	46.42
	S3	176.3	74.82	156.4	57.47	156.5	50.17
	S4	177.3	55.12	163.9	35.78	152.1	35.70
	S5	144.8	44.82	152.7	36.56	149.2	23.50
		Signal 1		**Signal 3**		**Signal 4**	
		ms	σ	ms	σ	ms	σ
Interstress	S1	402.3	78.61	476.0	76.21	474.0	53.69
	S2	620.0	253.74	585.3	110.88	550.3	152.16
	S3	560.6	232.02	493.0	124.38	500.3	87.06
	S4	510.6	115.86	444.0	46.16	477.0	18.68
	S5	418.3	65.54	477.3	29.29	479.6	23.09

Table 2: Mean Syl/Acc durations (ms) & their std. dv. (σ).

Strategy	Chain	Mike	Sgarbi	Sylld	σ (Sylld)	Accd	σ (Accd)	T	ANO
Skirmish	1	A		151.5	52.15	266.1	33.20		
	2		S	110.6	22.03	272	107.25		
	3		S	121.5	17.67	440.5	228.39		
	4		A	100.8	44.25	254.8	20.53		
	5		A	108.9	52.79	225.4	15.80		
	6		S	164	4.24	244	74.95		
	7		S	319	2.82	478.5	144.95		
	8		S	392	11.31	467	224.86		
	9		A	150.1	85.41	538.5	10.60		
	10	S		117.3	6.02	1 stress	--		
	11		A	187.8	106.39	540	14.14		
	12		A	77	29.17	172.2	16.91		
	13	S		120.6	8.96	297.3	35.83		
Truce	14		*	114.6	53.01	190.6	94.11		
	15		*	156.2	47.78	249.5	149.19		
Mimetic	16	Ac	Ac	Mike: 180.4 Sgarbi: 144.3	Mike: 43.8 Sgarbi: 45.7	Mike: 376 Sgarbi: 345.4	Mike: 14.4 Sgarbi: 27.2	0.07	0.08
Rolling	17	*	*	Mike: 126.8 Sgarbi: 128.3	Mike: 56.5 Sgarbi: 80.7	Mike: 346.2 Sgarbi: 217.6	Mike: 47.7 Sgarbi: 102.5	0.01	0.02
Interval	18	A		168.2	63.38	348.2	11.67		
	19		S	144	13.1	302.7	166.44		
Mimetic	20	Sc	Sc	Mike: 181.7 Sgarbi: 166.5	Mike: 22.09 Sgarbi: 36.33	Mike: 566 Sgarbi: 494.3	Mike: 90.50 Sgarbi: 217.19	0.50	0.50
End	21	A		168.1	73.82	401	59.99		
	22	A		195.4	62.51	359	41.76		
	23	S		263.6	27.5	362	76.36		

Table 3: Sequence of turns and rhythm trend. A = stress-timed; S = syllable-timed; * = asynchronous rhythm; c = synchronized durations; Sylld = syllables mean duration; Accd = interstress intervals mean duration; σ = Std Dv; T = two sample Student's t-test, assuming an unequal variance & a level of confidence α = 0.05; ANO = one-way, or single-factor ANOVA (α = 0.05). Non-significant values are on a grey background

Some more applications of the tonal grid annotation

Amedeo DE DOMINICIS

Università degli Studi della Tuscia

Via del Paradiso, 47 - 01100 - Viterbo - Province of Viterbo - Italy

dedomini@unitus.it

Abstract

In this paper we give some more detailed evidence concerning the operating mechanism of the tonal grid annotation. Using some chunks of speech in different languages, we will show how a grid works to find the unexpected perturbations, their tonal shape, and the relations the grid establishes among the macroIP constituents which are their functors or bearing units. Then we will argue that the tonal grid is not only a mere annotation technique, but also (and above all) a new theoretical approach to understand the constituency of the Intonational Phrase (IP). Particularly, the architecture of the grid helps to find the relation between the F0 prominences (material prominences) and the prominences that result from the metrical or syntactic hierarchies (metalinguistic prominences) within the same IP or across IPs (i.e. within what we name a macroIP). The relation between two or more material and metalinguistic prominences identifies what we call a Nucleus.

Keywords: Tonal Grid; Nucleus; Intonation; Annotation.

1. Introduction

This paper aims at providing a theoretical setting for the tonal grid annotation system (De Dominicis, 2010a, 2010b). The tonal grid (henceforth TG) is a graphic device which represents linguistic phenomena such as (grammatically) unexpected and recurrent tonal and segmental perturbations, syntactic and lexical discontinuities, and pragmatic functions (e.g. the focalizations). Moreover, the TG reconstructs the relations among these phenomena: i.e. the possible tonal 'rhyme(s)' between two (or more) tonal perturbations at close/remote range, or the phoric relationship between constituents within the clause.

The TG is an upgraded version of the syntactic grid (Blanche-Benveniste, 1990, 1997; Blanche-Benveniste *et al*., 1979; Blanche-Benveniste *et al*., 1990; Bilger, 1982; Bilger *et al*., 1997), which is specially suited to highlight the disfluencies and the fragmentary nature of speech (false starts, hesitations, repetitions), and how they contribute to build meaning and grammatical functions. A syntactic grid consists of two main dimensions: the horizontal axis represents the sequence of the syntagmatic positions (or constituents); the vertical axis shows the possible different paradigmatic occurrences lying on the same position. By adding a syntagmatic construction to its paradigmatic fragments one gets a *discursive configuration*. It may recur at regular intervals, like a rhyme, and so give the discourse a peculiar architecture.

The TG supplements the syntactic one by adding tonal (or intonational) features: it highlights the recurrence of the same tonal pattern on different syntagmatic positions, or on the whole paradigmatic set of constituents belonging to the same syntactic position. In both cases, if a given tonal perturbation recurs, then each instance is an occurrence of a tonal pattern *rhyme*. Moreover, the TG brings up the recurrence of some special segmental perturbations: so these ones establish another kind of *rhyme* (or phoric relation). Inside a TG some places are provided, where several kinds of *rhyme* interface. These relations – specially if they link two or more F0 prominent positions (due to a focalization or to the metrical hierarchy) – identify what we call an *intonational Nucleus*: a new (relational) definition of the Nucleus in a *macro-Intonational Phrase* (macroIP).

2. Interfaces

The theoretical starting point of our approach relies on a multidimensional conceit of the intonation theory (and of the Nucleus). Firstly, the reference units should be defined at the interface among F0 contour, metrical hierarchy, syntactic and pragmatic functions. Some special syntactic functions of the oral production (short and juxtaposed clauses, often without nominal constituents) should also be added to those interfacing functors.

Secondly, as for the intonation analysis, strictly speaking, we will consider tonal rhymes (i.e. non automatic, intentional tonal perturbations), and some (likewise intentional) segmental perturbations.

On the whole, the intonation is considered as the main mean for achieving the textual cohesion (Couper-Kuhlen & Selting, 1996; Ladd, 1996; Selkirk, 2000; Truckenbrodt, 2007; De Dominicis, 2009, 2010a, 2010b).

3. Case studies

The data concern Italian and English. They come from CLIPS *corpus* (dialogue "DGmtA01T"-Turin archive) and Human Communication Research Centre (HCRC) *corpus*.

The data have been analyzed by *Praat*. For each speaker the F0 has been extracted. Then both corpora have been syntactically annotated. As for the intonational annotation, the INTSINT system (Hirst & Di Cristo, 1998; Hirst, Di Cristo & Espesser, 2000) has been adopted, according to its automatic version (mono_momel-intsint.praat15). Using the typical Praat *Tiers & TextGrid* annotation system, the two labelling levels (syntactic & intonational) have been aligned to the audio signal. On this basis, the TGs have been built. They are shown in § 6.

The INTSINT system encodes the intonation curve as a sequence of tonal targets whose succession represents

the F0 contour in a stylised format. The symbols are as follows: "H" (Higher) and "L" (Lower); "S" (Same), "D" (Downstepping) and "U" (Upstepping). Three more symbols refer to the speaker tonal range: "T" (Top), "B" (Bottom), "M" (Mid).

Other syntactic symbols are: "V" (verb), "Prep" (preposition), "NP" (noun phrase), "VP" (verb phrase), "PrepP" (prepositional phrase), "PRO" (pronoun), "Adv" (adverb), "NEG" (negation), "Art" (article), "N" (noun), "Conj" (conjunction), "Modif" (modifier).

The data are organized into the TGs as follows. The first line is the orthographic transcription. Each grid represents a clause (C) or an Intonational Phrase (IP). PROM means a point of tonal prominence. The speakers are labelled "P1" and "P2" (in Italian), or "G" and "F" (in English). The constituents that are marked by an intonational prominence are underlined.

The first case study is an Italian conversation and consists of three speech turns:

P2: *passando sopra gli sci ?* ('passing over the ski?')
P1: *no gli sci io non ce li ho* ('no the ski, I don't have ski')
P2: *io passo sopra gli sci o no ?* ('do I pass over the ski or not?') ◀))

The figures 1-2 show their F0 contours and three annotations tiers (from the top to the bottom: orthographic transcription, syntactic segmentation and INTSINT tonal annotation). In the figures 1 and 2 the system has not detected the final F0 rising (ranging respectively 25 and 38 Hz).

The tables 1-3 represent the three TGs that correspond to the three speech turns. Each speech turn corresponds to an intonational phrase (IP1-3). For each IP the grid shows the speaker (P1 and P2). In the grids the syntactic functions (for instance V, Prep, NP) and the tones on each constituent are annotated.

In these three IPs (and speech turns) we note two parallel phenomena referring to the promotion of some constituents along the hierarchical prosodic structure as a consequence of increasing their tonal prominence.

IP1 (by P2) contains the entry "sci" ('ski') in a PrepP syntactic position and with a B tone. In IP2 (by P1) it has a different syntactic function (in a VP), but it has a higher tone (H) and a PROM function. Finally, in IP3 (by P2) "sci" ('ski') is again in a PrepP syntactic position (like in IP1) and it is also PROM (like in IP2) with a UU tone. Similarly, the same destiny applies to another entry. It is "io" ('I'), the personal pronoun that occurs in IP2 (with a B tone) and it is repeated in IP3 (with a M tone): in IP2 it is not PROM, but it becomes so in IP3 where it also shows an additional increasing of tonal prominence. Therefore, two entries ("sci", "io") have the same parallel destiny: passing from one IP to another, and from one speaker to another, they increase both their degree of tonal prominence and their hierarchical rank in the prosodic structure. In the last IP (IP3 by P2) both attain the PROM position. So, in the same IP two PROMs

cohabit.

According to the theoretical hypothesis of the present paper, only these two last PROMs ("io" and "sci" in IP3) are the real intonational Nuclei. The construction of their hierarchical position is the result of two cooperative strategies of the interlocutors: both a syntactic-lexical cross-reference architecture and a growing enhancement of the prominences. This macrostructure may be interpreted in a functionalist way, as a discursive cohesion mechanism, which results from an intra-speaker/inter-speaker involvement. A further remark concerning the linguistic theory of the intonational Nucleus may be added to this functional interpretation. The data show that the two Nuclei are established not on the basis of the simple physical prominence of F0 curve, but by means of a linguistic mechanism relying on the tonal relationship among constituents. This outcome is interesting and complex because the nuclear splitting (or doubling) is not supported by the predictions of the intonational phrasing theory. The two Nuclei govern a single macroIP which is composed of three IPs

The second case study consists of two pieces of an English conversation:

G: *(erm) have you got a collapsed shelter*
F: *yes I do*
G: *right*
G: *you've to go up north and then round the collapsed shelter* ◀))

G: *(just is) there a site of the plane crash*
F: *uh-huh I've got that I've got a site of plane crash*
G: *well it's just below there*
F: *just below that* ◀))

The figures 4 and 5 show their F0 contours and three annotation tiers (from the top to the bottom: orthographic transcription of G and F, syntactic segmentation and INTSINT tonal annotation).

In the first four IPs (Tables 4-7) we remark a promotion of some constituents along the hierarchical prosodic structure as a consequence of increasing their tonal prominence. IP1 and IP4 (by G) contain the phrase «collapsed shelter» in the same syntactic position (NP). The repetition by the same speaker is complemented by an increasing tonal prominence: from the initial M to the final S (= H). Therefore, both are good candidates to be the functors of the relation constituting the intonational Nucleus of the macroIP, even if they are not the most prominent part of the F0 contour, in a physical meaning.

In the following four IPs (Tables 8-11) we remark two intonational prominences: the first on «plane crash» (NP in IP1 by G) and the second on «that» (PrepP in IP4 by F). Their tonal complements are M and H.

The increase of their tonal prominence results from the cooperation between the two speakers.

Therefore, their relation constitutes the real Nucleus of the macroIP.

4. Effects of the manipulation

In order to experimentally verify our hypothesis concerning the intonational Nucleus of the macroIP, we have manipulated the corresponding signals by erasing the tonal prominences that were originally associated with the nuclear constituents (figures 6-7). The tables 12-15 and 16-19 represent the corresponding four TGs.

The first HCRC conversation is an information exchange ("G: *have you got a collapsed shelter? -* F: *yes, I do -* g: *right -* G: *you've to go up -* G: *north -* G: *and then round -* G: *the collapsed shelter*".). On the contrary, after the manipulation, it becomes something different: every relation between the "collapsed shelter" and "to go up north and then round" has disappeared. The conversation may be interpreted: "G: *have you got a collapsed shelter? -* F: *yes, I do -* G: *right -* G: *you've to go up -* G: *north -* G: *and then round. The collapsed shelter...*" - 🔊).

The second HCRC conversation is an information exchange too ("G: *There a site of the plane crash -* F: *uh-huh -* F: *I've got that, I've got a site of plane crash -* G: *well, it's just -* G: *below there -* F: *just below that*"). On the contrary, after the manipulation, it becomes something different: F seems to be in need to know, whereas G does not collaborate. The conversation may be interpreted: "G: *There a site of the plane crash... -* F: *uh-huh -* F: *I've got that? I've got a site of plane crash? -* G: *well, it's just -* G: *below there! -* F: *just below that?*" - 🔊).

5. Conclusions

In this paper we gave some more detailed pieces of evidence concerning the operating mechanism of the TG annotation. Using some chunks of speech, in different types of language, we have demonstrated how a grid works to find the unexpected perturbations, their tonal shape, and the relations the grid allows to establish among the macroIP constituents that are their functors or bearing units.

Then we argued that the TG is not only a mere annotation technique, but also (and above all) a new theoretical approach to understand the constituency of the IP. Particularly, the architecture of the grid helps to find the relation between the F0 prominences (*material prominences*) and the prominences that result from the metrical or syntactic hierarchies (*metalinguistic prominences*) within the same IP or across IPs (that is, within what we name a macroIP). The relation between two or more material and metalinguistic prominences identifies what we call *Nucleus* of the macroIP.

The theory claims that the Nucleus must be one (per IP) and obligatory. So, in order to verify this outcome, we simply predict that by removing a single prominence (no matter what kind, material or metalinguistic, it is) there should not be a change of the phonological type of the IP (or macroIP), whereas by erasing all the prominences that enter into a relation to form a Nucleus, a categorical shift of the IP or macroIP (e.g. a change in sentence modality, or syntactic interpretation) would be triggered.

6. References

Bilger, M. (1982). Contribution à l'analyse en grille. *Recherches sur le français parlé*, 4, pp. 195--215.

Bilger, M., Blasco, M., Cappeau, P., Pallaud, B., Sabio, F. and Savelli, M., (1997). Transcription de l'oral et interprétation; illustration de quelques difficultés. *Recherches sur le français parlé*, 14, pp. 57--86.

Blanche-Benveniste, C. (1990). Un modèle d'analyse syntaxique «en grille» pour les productions orales. *Annuario de Psicologia*, 47, pp. 11--28.

Blanche-Benveniste, C. (1997). *Approches de la langue parlée en français*. Paris: Ophrys.

Blanche-Benveniste, C., Bilger, M., Rouget, Ch., van den Eynde, K. (1990). *Le français parlé. Etudes grammaticales*. Paris: Editions du Centre National de la Recherche Scientifique.

Blanche-Benveniste, C., Borel, B., Deulofeu, J., Durand, J., Giacomi, A., Loufrani, C., Meziane, B., Pazery, N. (1979). Des grilles pour le français parlé. *Recherches sur le français parlé*, 2, pp. 163--205.

Couper-Kuhlen, E., Selting, M. (eds.) (1996). *Prosody in conversation*. Cambridge: Cambridge University Press.

De Dominicis, A. (2009). Tones and intonation in Gizey. *Studi Magrebini*, 7, pp. 145--166.

De Dominicis, A. (2010a). Griglie tonali. In M. Pettorino, A. Giannini, F.M. Dovetto (Eds.), *La comunicazione parlata 3*, vol. II, pp. 45-69. Napoli: Università degli studi di Napoli L'Orientale. Available at: <http://opar.unior.it/327/>.

De Dominicis, A. (2010b). *Intonazione. Una teoria della costituenza delle unità intonative*. Roma: Carocci.

Hirst, D.J., Di Cristo, A. (1998). A survey of intonation systems. In D.J. Hirst, A. Di Cristo (Eds.), *Intonation systems. A survey of twenty languages*. Cambridge: Cambridge University Press, pp. 1--44.

Hirst, D.J., Di Cristo, A. and Espesser, R. (2000). Levels of representation and levels of analysis for intonation. In M. Horne (Ed.), *Prosody: Theory and Experiment*. Dordrecht: Kluwer Academic Publishers, pp. 51--87.

Ladd, D.R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Selkirk, E. (2000). *The interaction of constraints on prosodic phrasing*. In M. Horne (Ed.), *Prosody: theory and experiment*. Dordrecht: Kluwer Academic Publishers, pp. 231--262.

Truckenbrodt, H. (2007). *The syntax-phonology interface*. In P. de Lacy (Ed.), *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press, pp. 435--456.

7. Appendix
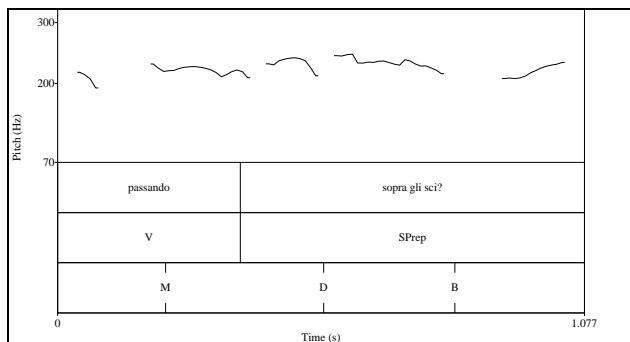
Figure 1: First CLIPS conversation turn (P2). F0 contour, transcription, syntactic annotation, INTSINT annotation
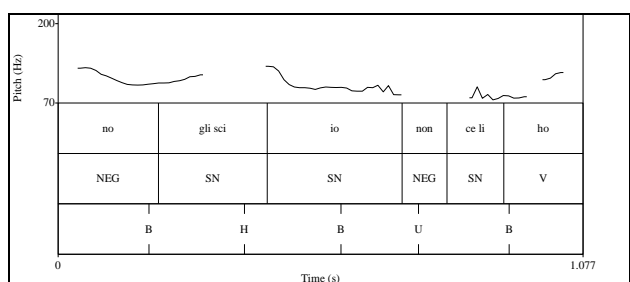
Figure 2: Second CLIPS conversation turn (P1). F0 contour, transcription, syntactic, and INTSINT annot
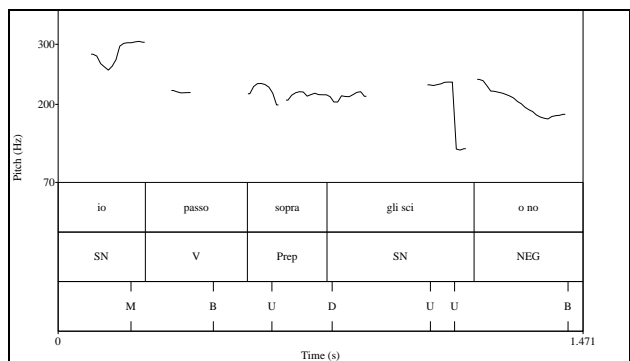
Figure 3: Third CLIPS conversation turn (P2). F0 contour, transcription, syntactic, and INTSINT annotations
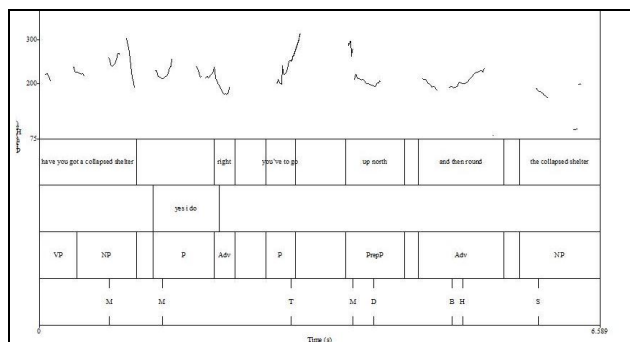
Figure 4: First HCRC conversation. F0 contour, 'g' and 'f' transcription, syntactic, and INTSINT annotations
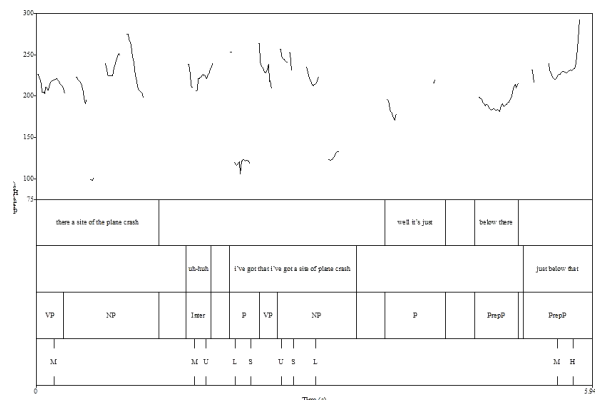
Figure 5: Second HCRC conversation. F0 contour, 'g' and 'f' transcription, syntactic, and INTSINT annotations
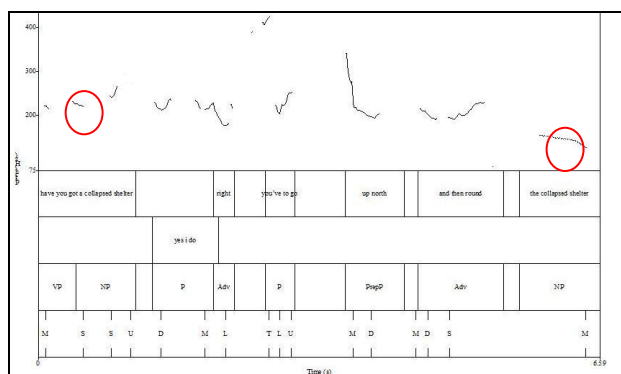
Figure 6: First HCRC conversation after manipulation
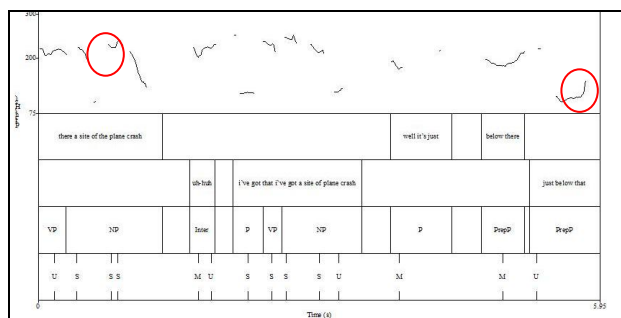
Figure 7: Second HCRC conversation after manipulation

IP1	*passando sopra gli sci ?*		
	P₂	passando	sopra	gli sci
		V	Prep	NP
		PROM: M	D	B

Table 1: TG of the IP1: 'passing over the ski?'

IP2	*no gli sci io non ce li ho*						
	P₁	no	gli sci	io	non	ce	li	ho
		NEG	NP	PRO	NEG	Adv	PRO	V
		B	PROM: H	B	U		B	

Table 2: TG of the IP2: 'no the ski, I don't have ski'

IP3	**P₂**	*io passo sopra gli <u>sci</u> o no ?*						
		io	passo	sopra	gli	sci	o	no
		PRO	V	Prep	Art	N	Conj	NEG
		PROM: M	B	U	D	PROM: UU	B	
		Nucleus				Nucleus		

Table 3: TG of the IP3: 'do I pass over the ski or not?'

IP1	**G**	*have you got a <u>collapsed shelter</u> ?*	
		have you got	a collapsed shelter
		VP	NP
		M	
			Nucleus

Table 4: TG of the IP1.

IP2	**F**	*yes I do*
		yes I do
		P
		M

Table 5: TG of the IP2

IP3	**G**	*right*
		right
		Adv

Table 6: TG of the IP3

IP4	**G**	*you've to go up north and then round the <u>collapsed shelter</u>*			
		you've to go	up north	and then round	the collapsed shelter
		P	PrepP	Adv	NP
		T	M D	B H	S
					Nucleus

Table 7: TG of the IP4

IP1	**G**	*there a site of the <u>plane crash</u>*	
		there	a site of the plane crash
		VP	NP
		PROM: M	
			Nucleus

Table 8: TG of the IP1

IP2	**F**	*uh-huh I've got that I've got a site of plane crash*			
		uh-huh	I've got that	I've got	a site of plane crash
		Inter	P	VP	NP
		M U	L S		U S L

Table 9: TG of the IP2

IP3	**G**	*well it's just below there*	
		well it's just	below there
		P	PrepP

Table 10: TG of the IP3

IP4	**F**	*just below <u>that</u>*
		just below that
		PrepP
		M H
		Nucleus

Table 11: TG of the IP4.

IP1	**G**	*have you got a collapsed shelter ?*	
		have you got	a collapsed shelter
		VP	NP
		M	S S U

Table 12: TG of the IP1

IP2	**F**	*yes I do*
		yes I do
		P
		D M

Table 13: TG of the IP2

IP3	**G**	*right*
		right
		Adv
		L

Table 14: TG of the IP3

IP4	**G**	*<u>you've to go</u> up north and then round the collapsed shelter*			
		you've to go	up north	and then round	the collapsed shelter
		P	PrepP	Adv	NP
		T L U	M D	M D S	M
		Nucleus			

Table 15: TG of the IP4

IP1	**G**	*there a site of the plane crash*	
		there	a site of the plane crash
		VP	NP
		U	S S S

Table 16: TG of the IP1

IP2	**F**	*uh-huh I've got that I've got <u>a site of plane crash</u>*			
		uh-huh	I've got that	I've got	a site of plane crash
		Inter	P	VP	NP
		M U	S	S	S S U
					Nucleus

Table 17: TG of the IP2

IP3	**G**	*well it's just below there*	
		well it's just	below there
		P	PrepP
		M	M

Table 18: TG of the IP3

IP4	**F**	*just below that*
		just below that
		PrepP
		U

Table 19: TG of the IP4

A prosódia das interrogativas absolutas na fala carioca: leitura versus fala espontânea

Vivian PAIXÃO, Dinah CALLOU

UFRJ/Capes/CNPq

vivian_bp@live.com, dcallou@gmail.com

Abstract

This paper analyzes absolute/total questions in Rio de Janeiro dialect, taking into consideration the prosodic differences between isolated read sentences and the ones produced in spontaneous colloquial speech. Another topic brought into discussionis the kind of *corpora* for the study of spontaneous speech prosody, since, in general, non-controlled speech *corpora* have insufficient recording quality. Besides, it is difficult to find comparable samples, due to the variety of attitudes and emotions whichdetermine speech prosody. In the first stage of the study, interrogative sentences obtained in the UFRJ Acoustics Phonetics Laboratory were analyzed, as well as affirmative sentences were manipulated in order to make its melodic contour similar to the interrogatives. These manipulated sentences were submitted to perception tests with native listeners, who confirmed its authenticity as interrogatives in Portuguese, without significant differences in relation to the original questions. In a second moment, we studied spontaneous speech data, taken from the *corpus* of NURC / RJ project, and we compared the results of both *corpora*: the same patterns were found, however some differences were detected at the micro-melodic level.

Keywords: prosody; total questions; spontaneous speech *versus* reading.

1. Introdução

1.1 Objetivos

O objetivo desta pesquisa é analisar as chamadas interrogativas absolutas – segundo a terminologia sugerida por Font-Rotchés & Mateo-Ruiz (2011) – na fala carioca, observando as diferenças prosódicas entre as sentenças lidas de forma isolada e aquelas produzidas espontaneamente dentro do contexto de fala.Utilizou-se, para tanto, um *corpus* de fala lida, obtido em laboratório, e uma amostra de fala espontânea.

O trabalho visa também a discutir a questão dos *corpora* no estudo da prosódia da fala espontânea, uma vez que, nesta área de pesquisa, esbarra-se frequentemente com uma dificuldade no que se refere à obtenção de dados: em geral, *corpora* de fala não controlada apresentam uma qualidade técnica de gravação muito inferior, o que dificulta e chega a impedir, em determinados casos, a análise acústica de alguns trechos. Além disso, ao lidar com um *corpus* de fala espontânea, o pesquisador depara-se ainda com outro problema: a dificuldade de encontrar amostras comparáveis, devido à grande variedade de atitudes e emoções que influenciam na prosódia da fala. Trabalhos como o de Moraes (2006) mostram que, mesmo entre as interrogativas globais (yes/no questions), há uma variedade de contornos melódicos. A depender do tipo de pergunta que representam (neutra, com desconfiança ou confirmativa, por exemplo), sentenças estruturalmente idênticas têm entonações distintas.

1.2 Trabalhos anteriores

O português brasileiro, e, mais especificamente, o português falado no Rio de Janeiro carece de estudos relativos à prosódia que utilizem dados de fala espontânea.

Os trabalhos de Moraes (1996, 2006, 2008), elaborados a partir de dados obtidos em laboratório (fala controlada) são algumas das maiores referências para o estudo da entonação modal a partir da análise acústica.

O recente trabalho de Silva (2011) traz um estudo comparativo entre todas as capitais do Brasil, mas utiliza um *corpus* de fala semiespontânea do projeto Atlas Linguístico Brasileiro (ALiB).

No que se refere especificamente às interrogativas na fala espontânea do Rio de Janeiro, além dos trabalhos aos quais este dá continuidade (Paixão, 2011a, 2011b), encontramos apenas o trabalho de Souza (1995), que utiliza o *corpus* do projeto NURC para analisar três tipos de perguntas (totais, disjuntivas e parciais). Quanto às perguntas totais, os resultados obtidos apontam diferenças em relação àquelas produzidas em contexto de leitura, analisadas por Moraes (1984). A pesquisa de Souza, assim como o presente trabalho, mostra que, no contexto de fala espontânea, as interrogativas não seguem um padrão melódico tão regular quanto as produzidas em contexto de leitura, e dá margem a uma investigação mais aprofundada acerca de quais seriam os fatores determinantes dessas diferenças.

2. *Corpora*

Os dados utilizados para este trabalho foram retirados de duas amostras distintas: uma de leitura de sentenças isoladas e outra de fala espontânea.

2.1 *Corpus* de fala controlada (leitura)

A primeira amostra, de fala controlada, é composta de treze sentenças interrogativas e outras treze afirmativas, que se diferenciam das perguntas apenas pela entonação. Todas as sentenças, gravadas no Laboratório de Fonética Acústica da UFRJ por uma informante do sexo feminino, tinham estrutura sintática similar, seguindo a ordem prototípica do português brasileiro sujeito-verbo. Procurou-se ainda controlar a tonicidade da última palavra das sentenças: cinco delas eram terminadas em

palavras proparoxítonas, quatro em paroxítonas e quatro em oxítonas – houve uma sentença a mais com proparoxítona para garantir que houvesse dados com esse tipo de palavra, uma vez que elas sofrem frequentemente processos de sincope e consequente ressilabação. O tamanho das frases também é similar: todas têm entre sete e dez sílabas.

A lista foi lida três vezes pela informante em diferentes ordenações, tendo sido a primeira e a última gravações descartadas.

2.2 *Corpus* de fala espontânea

O *corpus* de fala espontânea é composto deoutras dez sentenças retiradas de inquéritos do tipo diálogo entre dois informantes (D2) do projeto NURC-RJ.

O número de sentenças analisadas é restrito devido a várias dificuldades encontradas na busca pelos dados e sua análise. As interrogativas globais são escassas nos *corpora* obtidos através de entrevistas: raramente ocorrem perguntas nos diálogos entre informante e documentador (DID) ou nas elocuções formais (EF: aulas e conferências). Quando isso acontece, em geral, consistem nas chamadas *tag questions*, pedidos de confirmação terminados pelo marcador "né" – esse tipo de pergunta, muito frequente no português brasileiro, tem uma conformação prosódica diferenciada, conforme observa Serra (2009) em capitulo dedicado exclusivamente a essas sentenças.

Nos inquéritos do tipo D2 (diálogos entre dois informantes), as interrogativas são um pouco mais frequentes, mas ainda escassas. Além disso, por se tratar de fala espontânea, muitas vezes não se respeitam os turnos e as falas são sobrepostas, inviabilizando a análise dos enunciados.

3.   Metodologia

A metodologia utilizada na análise e na ressíntese dos dados foi a do Método de Análise Melódica da Fala de Cantero&Font-Rotchés (2009). Esse método permite comparar vozes de diferentes indivíduos (inclusive homens e mulheres), uma vez que os gráficos são plotados com base em números estandardizados, e não nos valores brutos de frequência da voz.

3.1   Análise acústica

A análise dos dados, feita através do programa Praat, deu-se da seguinte maneira: primeiramente, os enunciados foram segmentados em sílabas. Aferiu-se, então, a medida da frequência fundamental (F0) de um ponto central da vogal de cada sílaba. Nos casos em que havia uma sílaba mais prolongada, que apresentasse oscilação de mais de 10% na medida de F0 de uma mesma vogal (o que equivale a um semitom musical, aproximadamente), consideraram-se dois pontos na mesma sílaba.

Uma vez obtidos os valores absolutos em Hertz, fez-se a estandardização desses valores, isto é, mediu-se a distância tonal, em porcentagens, entre uma vogal e a vogal subsequente, para poder criar a curva melódica

representada em um gráfico gerado pelo Microsoft Excel, como se vê no exemplo a seguir (Gráfico 1).

	A	Pris	ci	l(a)u	l(a)u*	sa	ó	ó*	cu	los?
Hz	207	210	250	244	292	220	193	247	250	205
Perc.	100,	1,4%	19,0	-2,4	19,7	-24,	-12,	28,0	1,2%	-18,
●– C. Est.	100	101	121	118	141	106	93	119	121	99
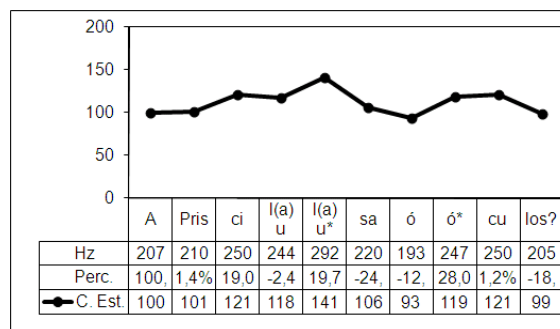
Gráfico 1: Curva estandardizada da sentença "A Priscila usa óculos?" (fala controlada) 🔊

3.2   Manipulação acústica (ressíntese)

As sentenças afirmativas gravadas pela informante foram submetidas à manipulação acústica, também no programa Praat, conforme descrito a seguir: segmentaram-se as sentenças em sílabas da mesma forma como havia sido feito com as interrogativas, e marcou-se um ponto na curva de F0 na vogal de cada uma das sílabas – com exceção daquelas em que havia uma diferença de mais de 10%, em que foram marcados dois pontos. Em seguida, cada um dos pontos foi deslocado para o mesmo valor em Hertz da sílaba correspondente na gravação original (interrogativa).

Dessa forma, a diferença na curva de F0 das perguntas originais e manipuladas foi, principalmente, a regularidade: nas interrogativas "verdadeiras", temos uma curva mais sinuosa, com mais oscilações da frequência entre os pontos marcados para medição, enquanto nas ressintetizadas a curva mostra-se mais regular (cf. figura 1).
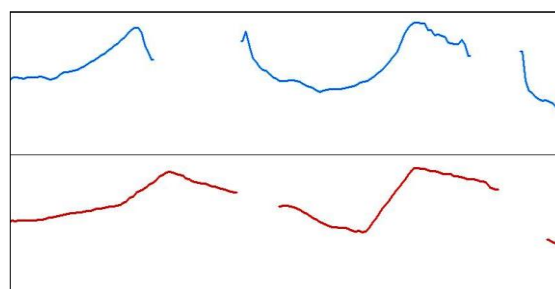
Figura 1: Comparação entre as curvas de F0 da sentença "A Luísa estuda música?" original (acima) - 🔊 - e ressintetizada (abaixo) - 🔊 -

4.   Resultados

4.1   Resultados com dados de fala controlada

Nesta primeira etapa, verificou-se um padrão melódico para as interrogativas condizente com aquele descrito pela literatura: há um primeiro pico seguido de uma descida e uma inflexão final ascendente, no caso de sentenças terminadas em palavras oxítonas; ou circunflexa, quando

a sentença é terminada em palavras paroxítonas ou proparoxítonas. Tanto no primeiro quanto no segundo picos, a subida representou um aumento de, em média, 30% em relação ao tom inicial da sentença.

Mesmo em se tratando de dados fornecidos por uma mesma informante, a similaridade entre os gráficos das diferentes sentenças é impressionante. Ao sobreporem-se os desenhos das curvas estandardizadas de duas sentenças com tamanho similar, percebe-se que elas são praticamente iguais. As diferenças, aparentemente, devem-se apenas à localização das sílabas acentuadas em cada uma delas (cf. Gráfico 2).
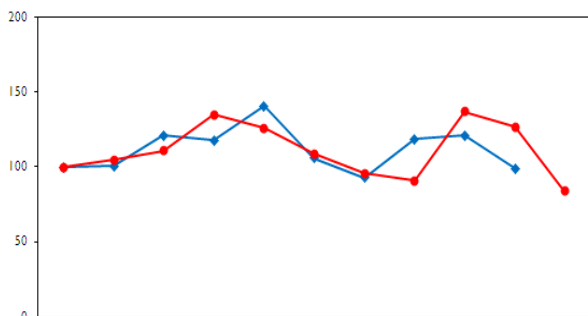
Gráfico 2: Comparação entre as curvas estandardizadas das sentenças "A Priscila usa óculos?" (em azul) e "A Luísa estuda música?" (em vermelho)

4.1.1. Manipulação acústica e testes de percepção

Os enunciados foram submetidos a testes de percepção com ouvintes nativos cariocas. Foi solicitado aos voluntários que ouvissem duas gravações: a primeira delas continha as perguntas originais. Em seguida, escutaram as sentenças ressintetizadas (afirmativas transformadas em interrogativas).

Os voluntários reconheceram todas as sentenças como sendo, indubitavelmente, interrogativas autênticas da língua portuguesa. Quando indagados quanto a possíveis diferenças entre as sentenças, os participantes da pesquisa disseram apenas ter achado aquelas pertencentes à segunda gravação (ressintetizadas) "menos enfáticas" ou "ditas sem muita vontade".

Curiosamente, uma das informantes declarou que as frases do "grupo 1" e do "grupo 2" haviam sido gravadas por pessoas de sexos diferentes, como se pode inferir a partir da resposta transcrita a seguir:

O grupo 2 tem uma voz mais intensa, transmitindo mais firmeza. Aparentemente, na frase 6, trocaram os grupos, sendo assim, quem fala a frase do grupo 1 é o integrante do grupo 2, e no grupo 2 quem fala a frase é a integrante do grupo 1.

Esta ouvinte, em particular, atribuiu a diferença entre as sentenças originais e ressintetizadas a uma diferença de qualidade vocal, e, mais do que isso, ao sexo do falante – a voz ressintetizada foi identificada como voz masculina, e a voz natural, como feminina. Embora essa opinião tenha sido emitida por apenas um dos voluntários,

levanta um interessante questionamento sobre uma possível tendência da fala masculina a uma maior regularidade no nível micromelódico, isto é, a apresentar menos oscilações na frequência fundamental nos pontos da frase em que esta variação não é um parâmetro decisivo para a caracterização modal da sentença.

4.2 Resultados com dados de fala espontânea

Os resultados obtidos com os dados de fala espontânea, mesmo com todos os empecilhos que envolvem o trabalho com esse tipo de *corpus*, confirmam, em certa medida, aqueles da fala controlada.

Os gráficos de alguns enunciados não apresentam os picos tão destacados quanto nos dados de fala lida, mas ainda assim pode-se perceber uma primeira e uma segunda subida, de acordo com o esperado. O Gráfico 3, a seguir, ilustra a configuração um pouco mais planificada dos dados de fala espontânea, com um primeiro pico muito sutil, enquanto o Gráfico 4 assemelha-se mais aos de leitura, uma vez que tem os dois picos bem destacados.

	(Vo)cês	cos	tu	mam	(sa)ir	em	gru(pos)?
Hz	220	198	200	189	173	151	180
Perc.	00,0%	10,0%	1,0%	-5,5%	-8,5%	-12,7%	19,2%
C. Est.	100	90	91	86	79	69	82
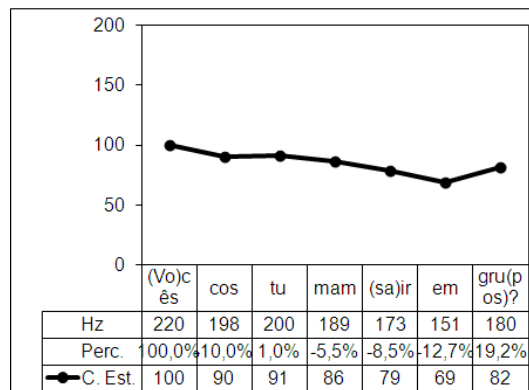
Gráfico 3: Curva estandardizada da sentença" Vocês costumam sair em grupos?" (fala espontânea) 🔊

	Vo	cê (é)	de	fa	mí	lia	de	ls	ra(e)	li(tas)?
Hz	178	223	201	204	196	186	171	166	176	234
Perc.	100,	25,3	-9,9	1,5%	-3,9	-5,1	-8,1	-2,9	6,0%	33,0
C. Est.	100	125	113	115	110	104	96	93	99	131
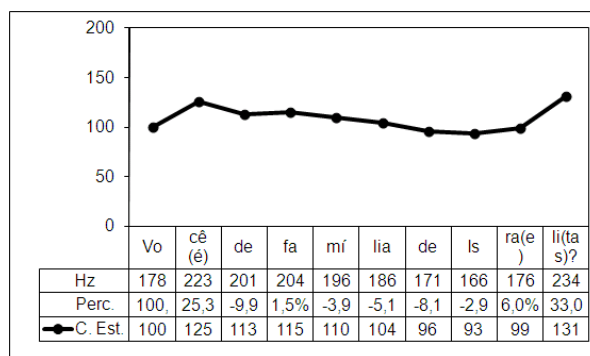
Gráfico 4: Curva estandardizada da sentença "Você é de família de israelitas?" (fala espontânea) 🔊

5. Conclusões

5.1 Fala controlada (leitura)

Os resultados obtidos com os dados de fala controlada confirmam aquilo que a literatura afirma a respeito das interrogativas absolutas no português: a presença de um pico de F0 na tônica final é, certamente, a característica

mais marcante desse tipo de sentença. O pico inicial, no entanto, também se faz presente, embora bem menos destacado do que o final em alguns dados.

5.2   Fala ressintetizada

Com relação às sentenças afirmativas ressintetizadas, foi possível perceber que, mesmo realizando-se a alteração da medida de F0 em cada silaba, as diferenças no nível micromelódico (regularidade na oscilação de F0 entre os pontos demarcados na curva) foram suficientes para que ouvintes nativos as interpretassem como tendo expressões de atitudes distintas relacionadas à ênfase.

5.3   Fala espontânea

Os resultados com dados de fala espontânea, apesar de se assemelharem em certa medida aos de leitura, mostram menor oscilação da F0, o que pode se dever à velocidade de fala e à tendência a uma demarcação mais nítida, na leitura, dos contornos prosódicos que determinam a modalidade da sentença.

5.4   Questões levantadas

A resposta de uma das informantes do teste de percepção realizado com as sentenças ressintetizadas permitiu levantar a hipótese de que a fala masculina apresente, no dialeto carioca, menor oscilação de F0.

Além disso, a observação dos dados, em especial daqueles de fala espontânea, nos leva a pensar na questão da neutralidade de atitudes e supor uma possível influência de fatores de ordem morfossintática na melodia da sentença. É sabido que a ênfase em um dos elementos da frase é fator determinante de alterações em diferentes parâmetros acústicos (frequência, intensidade, duração). A hipótese que se pode aventar é a de que determinados itens lexicais ou classes de palavras possam conter uma ênfase inerente. Em outras palavras, levanta-se a possibilidade de que, ainda que o falante procure dizer a sentença de forma neutra, sem focalizar nenhum elemento, determinados itens tenham a tendência a receber maior ênfase, seja por sua carga semântica ou papel sintático, e o que interferiria no contorno melódico da sentença.

6.   Ideias para a expansão da pesquisa

Conforme já se viu, a necessidade de realização de trabalhos referentes à prosódia da fala espontânea no português brasileiro esbarra na dificuldade de se trabalhar com os grandes *corpora* de fala disponíveis, como o do projeto NURC. Em se tratando de sentenças interrogativas, a dificuldade é ainda maior, uma vez que esse tipo de sentença não é recorrente nas entrevistas.

Apesar disso, é possível elaborar um *corpus* de fala que concilie alto grau de qualidade técnica e espontaneidade de fala Pretende-se dar continuidade ao estudo das interrogativas com um *corpus* elaborado especificamente para isso, que se utiliza de estratégias para induzir os informantes a produzir perguntas. Tem-se, por exemplo, o HCRC MapTask (Anderson *et alii*, 1991), utilizado por pesquisadores de diversas línguas, que

consiste em uma tarefa a ser realizada em duplas e que, para tanto, exige que um participante faça perguntas ao outro. O *corpus* utilizado por Pinto (2009) na análise de transferências prosódicas também é uma ideia interessante: consiste em um "jogo da verdade" entre pessoas conhecidas.

A realização de um trabalho sobre as interrogativas no português falado no Rio de Janeiro a partir de um *corpus* específico é, portanto, necessária para se confirmar resultados aqui apresentados e expandi-los, na medida em que se podem controlar melhor a possível interferência de fatores de ordem morfossintática e lexical.

7.   Referências

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S. and Weinert, R. (1991).The HCRC Map Task Corpus. In *Language and Speech*, 34, pp. 351--366.

Boersma, P., Weenink, P. (1992-2008): PRAAT. Doing phonetics by computer.Institute of Phonetic Sciences, Univerty of Amsterdam. Available at: <http://www.praat.org>.

Cantero, F.J., Font-Rotchés, D. (2009). Protocolo para elanálisis melódico delhabla. In *Estudios de Fonética Experimental*, n. XVIII, pp. 17--32.

Font-Rotchés, D., Mateo-Ruiz, M. (2011). Absolute interrogatives in Spanish, a new melodic pattern. In *Anais do VII Congresso Internacional da Abralin*. Curitiba.

Moraes, J.A. (1984). Recherches sur l'intonation modale du portugais brèsilien parlé à Rio de Janeiro. Tese de Doutorado, Université de la Sorbonne Nouvelle. Paris.

Moraes, J.A. (2006). "Melodic contours of yes/no questions in Brazilian Portuguese", In A. Botinis (Ed.) *Proceedings of ISCA Tutorial andResearch Workshop on ExperimentalLinguistics*, Athens: University of Athens, pp. 117--120.

Moraes, J.A. (2008). The Pitch Accents inBrazilian Portuguese: analysis by synthesis. In P. Barbosa, S. Madureira, S. and C. Reis (Eds.), *Proceedings of the Speech Prosody 2008: Fourth Conference on Speech Prosody*, Campinas, pp. 389--397.

Paixão, V. (2011a). A entonação das interrogativas absolutas neutras no português do Rio de Janeiro. In *Anais do Colóquio Brasileiro de Prosódia da Fala*, v. 1, pp. 1257.

Paixão, V. (2011b). Padrões entonacionais da interrogativa na fala carioca. Apresentação de trabalho na XXXIII Jornada de Iniciação Científica, Artística e Cultural Giulio Massarani. Universidade Federal do Rio de Janeiro. Rio de Janeiro, out 2011.

Pinto, M.S. (2009).Transferências prosódicas do português do Brasil/LM na aprendizagem do Espanhol/LE: enunciados assertivos e interrogativos totais. Tese de Doutorado, Programa de Pós-graduação em Letras Neolatinas da Universidade Federal do Rio de Janeiro. Rio de Janeiro.

Serra, C.R. (2009). Realização e percepção de fronteiras prosódicas no Português do Brasil: fala espontânea e

leitura. Tese de Doutorado. Programa de Pós-graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro. Rio de Janeiro, pp. 130--137.

Silva, J.C.B. (2011). Caracterização prosódica dos falares brasileiros: as orações interrogativas totais. Dissertação de Mestrado. Programa de Pós-graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro. Rio de Janeiro.

Souza, M.L.M. (1995). Padrões Entoacionais da Interrogação na Fala Espontânea Carioca. Trabalho apresentado no II Encontro de Língua Falada e Escrita. Universidade Federal do Alagoas, 20-24 set.

The perception of foreign accented speech. Segmental and suprasegmental features affecting the degree of foreign accent in L2 Italian

Elisa PELLEGRINO

University of Naples "L'Orientale"

Via Duomo, 219. 80138 Napoli

epellegrino@unior.it

Abstract

The factors affecting the degree of foreign accent have been a matter of debate for years. This study is intended to investigate the role of suprasegmental and segmental features that make L2 Italian speech perceptively deviating from Italian native speech. Fifty-six Italian listeners listened to the excerpts of read speech produced by 8 Chinese learners and 2 Italian native speakers and rated them for degree of accentedness. The L1 and L2 corpora were spectro-acoustically analyzed. At suprasegmental level, we calculated the following rhythmic and prosodic parameters: articulation and speech rate, fluency, tonal range, percentage of silence and mean duration of empty pauses. At segmental level, we measured the duration of stressed and unstressed vowels and syllables and the length of stressed open and closed syllable. We also considered the syllable composition and pronunciation mistakes. The comparison between the results of the perception test and the data from the spectro-acoustic analyses have shown that both some suprasegmental parameters (i.e. fluency, tonal range and composition of speech time) and some segmental ones (i.e. duration of stressed and unstressed vowels and syllables, percentage of mispronunciations mistake) are relevant features differentiating the strength of perceived foreign accent of Chinese speakers.

Keywords: foreign accent; segmental and suprasegmental cues; Chinese-accented Italian.

1. Introduction

Researchers have generally come to a consensus that the age at which the acquisition of a second language begins may greatly affect the outcomes of the process itself. Late second language acquisition indeed has often been considered one of the primary factors preventing the attainment of a native-like proficiency especially at the level of L2 speech perception and production (Birdsong 2006; Matsuoka & Smith 2008). Nevertheless, there is no widespread agreement among researchers on the role played by segmental and suprasegmental cues in foreign accent detection.

Over the years, the bulk of studies on the perception and production of non-native speech have focused mainly on the segmental features deviating from the native speakers' pronunciation. (Flege, Bohn & Jang, 1997; Flege, Munro & MacKay, 1995; Walley & Flege, 1998). Similarly, most recent theoretical models accounting for L2 speech production and perception (Flege, 2003; Best, 1995; Major, 2001) have examined above all the production and perception of segments and have investigated the phonetic transfers from L1 to L2.

As a consequence, for years the role played by the suprasegmental features of speech in the perception of foreign accent has been relegated to a subordinate position (Piske, MacKay & Flege, 2001). Nevertheless, in recent decades the trend has changed. Recent studies on second language acquisition, (De Meo & Pettorino, 2011, 2012; Horgues, 2010), on the perception of foreign accent (Boula de Mareüil *et al.*, 2004; Boula de Mareüil & Vieru-Dimulescu, 2006; Marotta & Boula de Mareüil, 2010), research undertaken on speech synthesis (Magen, 1998; Munro, 1995; Ramus & Mehler, 1999), on automatic approaches for foreign accent identification (Piat, Fohr & Illina, 2008) argue for a major role of prosody in the perception of non-native speech and in the

recognition of the foreign speaker' L1. On the same wavelength there is the research on the relationship between foreign accent, communicative effectiveness, credibility, reliability and persuasiveness in L2 Italian (Pettorino *et al.*, 2011; De Meo, Pettorino & Vitale, 2012). Such work based on a pragmatic and acquisitional approach assessed both qualitatively and quantitatively the role played by the single rhythmic and prosodic parameters in carrying out effective communication.

2. The study

2.1 Objectives

Since it was shown that the suprasegmental features of speech are as essential as segmental ones both in the perception of foreign accent and in the detection of non-native speakers' mother tongue, in the present study we considered both levels of analysis.

In order to facilitate the reading of the study results, we have divided this article into two main sections. One is devoted to the identification of the suprasegmental correlates of foreign accentedness. The other is dedicated to figuring out the contribution of phonemic deviations to the strength of perceived foreign accent in Chinese-accented Italian.

2.2 Participants

In the study we recruited two groups of participants with distinct roles: 10 speakers and 56 native listeners.

2.2.1. The speakers

The group was composed of 8 non-native speakers (NNS) from China and 2 Italian native speakers (NS) from Campania Region.

The Chinese students ranged in age from 20 to 22 years of age. They had already studied Italian in China for three years and had attained an intermediate level of

linguistic competence (B1- CEFR). On their arrival in Naples, they were enrolled in a course of Italian specifically designed to help them improve Listening and Speaking skills. At the time of the test, they were following a study curriculum in Italian and Linguistic Studies at the University of Naples "L'Orientale".

The Italian native speakers, aged between 23 and 26, were students of Foreign Languages and Literatures at the same University. They constituted the control group.

2.2.2.   The listeners

The group of listeners was composed of 56 native speakers of Italian, from Campania region, ranging in age from 20 to 50 years old. Since the competence in the foreign speaker' L1 and the familiarity with a specific foreign accent were proved to mitigate the strength of perceived foreign accent and to facilitate the recognition of the interlocutors' provenience (Marotta & Boula de Mareüil, 2010), none of the listeners were competent in Chinese or familiar with Chinese-accented Italian.

2.3   Materials and Methods

In order to prevent the topic, the word length and the syntactic sentence structure from affecting the results of the study, the Italian native and non-native speakers were involved in a read speech task. The subjects were instructed to read a 50-word text on jet-lag symptom drawn from an Inflight Magazine but with simplified lexicon and syntax. The recordings were taken in single sessions with every speaker in an anechoic chamber, at 44.100 Hz sampling rate.

The Italian listeners listened to the single excerpts of read speech produced by the 10 speakers in a randomized order. Each speech sample was rated through an accent degree rating test, based on a four-point scale: 0= native speaker (N); 1= mild accent (M); 2= strong accent (S); 3= very strong accent (VS).

After the perception test, the corpora of L1 and L2 Italian were spectro-acoustically analyzed for single speech chains, that is the part of a spoken utterance comprised within two silent pauses (Pettorino & Giannini, 2010). For each speech chain, we measured the number of syllables really uttered, their duration, the lowest and highest f0 values, the occurrence of disfluencies, the length of silent pauses between the speech chains. On the basis of these measures, we calculated articulation and speech rate, fluency, tonal range, percentage of phonation time, silence time and disfluency time, mean duration of silent pauses.

At segmental level, we carried out analyses on the duration of stressed and unstressed vowels and syllables and the length of stressed open and closed syllables[1]. We also considered syllable composition and pronunciation mistakes. We used the open-source softwares Wavesurfer 1.8.8 and Praat (v. 4.1) for speech analysis.

2.3.1.   Brief description of the analyzed prosodic features[2]

- Articulation Rate (AR) was calculated as the ratio between the number of syllables really uttered and phonation time (syl/s). It is considered as a qualitative index because it measures the level of accuracy of the articulatory gesture and gives indications on the spatiotemporal organization of speech. With high values of AR the perfect achievement of the articulatory targets is compromised. Lower values of AR, on the contrary, allow the articulators to perfectly reach the expected acoustic targets. It is a rather stable parameter because its variations are limited by the anatomical and physiological constraints of the phonatory organs.

- Speech Rate (SR) was calculated as the ratio between the number of syllables and total time of the utterance, including both silent pauses and disfluencies (syl/s). Unlike AR, SR is considered as a quantitative parameter that measures the productivity level of speech. Its variations depend on the number and length of silent and non silent pauses. The more and longer empty and filled pauses, the lower the speech rate.

- Fluency (F) was calculated as the ratio between the number of syllables and the number of speech chains (syll/SC). It measures the frequency of silence, indeed, the higher the fluency, the fewer the empty pauses. Like SR, fluency is not a stable index. Its variations can be ascribed to many sender-dependent factors such as his/her socio-cultural background, emotional state, degree of control of the communicative event, ability to organize discourse on line and intention to give emphasis to his/her own speech.

- Tonal range. This corresponds to the interval between the f0 minum and maximum and it was measured in semitones (st). Low values of tonal range signal a monotone and flat speech. On the contrary, a wide interval between f0 minimum and maximum shows a more varied and dynamic speech.

- Percentage of phonation time, silence time and disfluency time. Phonation time is constituted by the syllables actually uttered in the speech. The silence time, instead, includes empty pauses, that is respiratory and emphatic pauses. The latter are commonly used to give the speech more emphasis in order to attract the listeners' attention on a specific portion of the discourse.

---

[1] In Italian the stressed vowels and syllable are longer than unstressed ones. See a.o. Avesany et al (2006).

[2] See a.o. Barr, 2001; Giannini, 2010; Giannini & Pettorino, 2010; Pettorino, 2003; Pettorino & Giannini, 2005, Savastano, Giannini & Pettorino, 1995).

Unlike silence time, disfluency time comprises filled pauses such as false start, vocalizations, nasalizations, lengthening, repetitions, and corrections. These occur above all in spontaneous speech together with the words that speakers plan and utter. They may signal to the listeners when the speaker is uncertain, or when he/she has to make choice and the speech planning process is delayed. They may also inform about the speaker's confidence in what he/she is saying .

3. Results

3.1 Perception test

The results of perception test have showed that 96% of listeners rated the two Italian participants as NS. The remaining 4% did not answer. As for the rates given to foreign speakers (Figure 1), 7 Chinese students out of 8 were unanimously recognized as NNS. Only the speaker no. 7 was rated as NS by 4% of listeners.
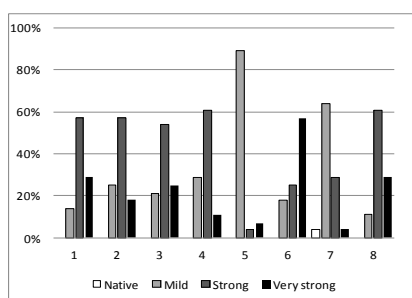
Fig 1: Accent rate for foreign speaker

If we consider the rates for degree of accentedness:

- Speakers no. 5 and 7 were rated "mildly accented" respectively by 89% and 64% of listeners.

- Speaker no. 6 was perceived with a very strong foreign accent by 57% of listeners. Only 25% rated her accent "strong" and 18% rated her speech as "mildly accented".

- Speakers no. 1-4 and 8 were rated with "strong foreign accent" by more than 50% of listeners, with very high difference in percentage from those who rated their accent as "very strong" (about 30%) and "mild" (30-40%).

In order to determine the acoustic correlates of degree of foreign accent, the results of the perception test were then compared to the acoustic data.

3.2 Suprasegmental features

The comparison between the results of perception test and the data from spectro-acoustic analysis have showed some evident differences between native and non-native speech at suprasegmental level.

First of all, the speech rated as "native" is the one with the highest values of articulation rate, speech rate, tonal range and fluency (Table 1).

	AR (syll/s)	SR (syll/s)	F (syll/SC)	TR (st)
N	6.2	5.1	13.8	9.5
M	4.6	4	13.2	8.2
S	4.1	3.6	10.7	7.7
VS	4	3.3	8.7	5.8

Table 1: Mean values of suprasegmental features per groups of speakers

Secondly, if we shift our attention to the supra-segmental correlates of mild, strong and very strong foreign accent, from Table 1 it is possible to infer that there are both some stable parameters within the three groups of speakers (M, S, VS) and some suprasegmental features that instead differentiate the three degrees of foreign accent.

The most steady parameters are AR and SR; their values indeed do not change meaningfully among the three groups of participants (Table 2).

	AR (syll/s)	SR (syll/s)
M	4.6	4
S	4.1	3.6
VS	4	3.3

Table 2: Mean values of AR and SR per degree of foreign accent

All foreign participants, indeed, speak at an articulation rate of about 4.2 syl/s and at a speech rate ranging from 3.3 syl/s of the VS speaker to the 4 syl/s of M speakers. The stable values of AR and SR can be ascribed to the particular kind of speech, that is read speech uttered by speakers with the same L1 and level of competence in L2 Italian. Generally speaking, indeed, read speech is more uniform among speakers in terms of rate than other kinds of speech, such as spontaneous speech, because it does not involve on-line speech planning processes.

The suprasegmental features that instead differentiate the three degrees of foreign accent are tonal range and fluency. Table 3, on the next page, summarizes the values for these two parameters for degree of foreign accent.

	F(syll/SC)	TR(st)
M	13.2	8.2
S	10.7	7.7
VS	8.7	5.8

Table 3: Mean values of fluency and tonal range per degree of foreign accent

M speakers have the broadest tonal range and are the most fluent readers. As shown in Table 3, they produce speech chains of about 13 syllables, instead S and VS speakers utter speech chains that are respectively 10.7 and 8.7 syllables long. The diverse values of fluency are due a different amount of silent pauses made by the 3 groups of readers (Table 4).

	Silent pauses	
	Number	Mean duration (s)
M	8	0.3
S	10	0.3
VS	12	0.3

Table 4: Number and mean duration of silent pauses

Even though silences do not diverge in their mean duration, M speakers pause less than S and VS speakers do. The reason for the discrepancy in the frequency of silences seems to be imputable to speakers' adoption of diverse pausing strategies while reading the target text.

For example, M speakers pause between complete sentences separated by full stops and coherently with the thematic organization of the sentence. The S and VS speakers, instead, silence when there are sentence boundaries marked by full stops and when there are boundaries of lower syntactical levels, usually marked by commas in the text. VS speaker also produces empty pauses within a sentence, that do not correspond to any syntactic boundaries. These latter rather occur when she produces disfluencies like word repetitions. The utterance of extra-words make the number of syllable to utter increase and consequently the speaker is lead to silence even before a syntactic boundary.

The divergences between the three groups extends also to the composition of speech time (Figure 2). Higher percentage of phonation time to detriment of silence time and disfluency time signals to native listeners a gradual reduction of foreign accent.
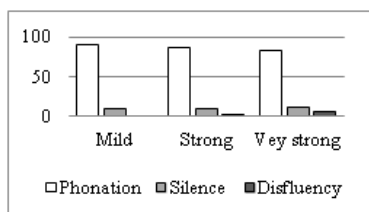
Figure 2: Composition of speech time per degree of foreign accent

3.3 Segmental features

At segmental level, we carried out contrastive analyses on the length of stressed and unstressed vowels and syllables[3]. We measured the duration of stressed open and closed syllables uttered both by Italians and by Chinese participants. The syllable composition and segments mispronunciations were considered too. The data concerning mean duration of stressed and unstressed vowels (Figure 3), stressed and unstressed syllable (Figure 4), stressed open and closed syllables (Figure 5) mirror the difference in the mean values of articulation rate between native and non native speakers (Table 2).

The higher the articulation rate, the shorter the vowel and syllable length.
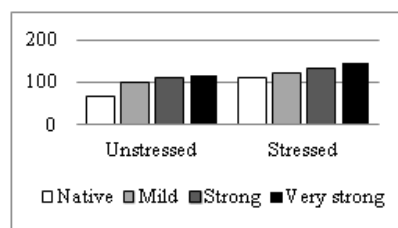
Figure 3: Mean duration of stressed and unstressed vowels (ms)
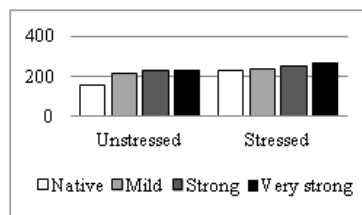
Figure 4: Mean duration of stressed and unstressed syllables (ms)
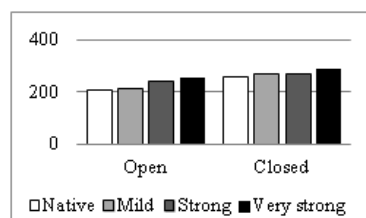
Figure 5: Mean duration of stressed open and closed syllables (ms)

As shown in figures 3-5, there seems to be a direct correlation between the rate of accentedness and vowel and syllabic duration. The speaker with very strong foreign accent always utters the longest vowels and syllables, regardless if they are stressed or not. On the contrary, the milder the rate of foreign accentedness, the lower the difference in vowels and syllabic duration from L1 Italian speakers.

---

[3] In Italian stressed are longer than unstressed vowels. For a review of the acoustic and articulatory differences between stressed and unstressed vowels, see a.o. Avesani et al. (2007).

However, if we consider the duration ratios between vowels and syllable uttered by NSs and NNSs (Table 5), the major differences between L1 and L2 speech lie in the articulation of unstressed vowels and syllable. These latter are much longer than those uttered by the native speakers.

	M	S	VS
Unstressed Vowel	1:1.52	1:1.72	1:1.75
Stressed Vowel	1:1.09	1:1.20	1:1.31
Unstressed Syllable	1:1.34	1:1.45	1:1.48
Stressed Open Syllable	1:1.03	1:1.16	1:1.25
Stressed Closed Syllable	1:1.05	1:1.05	1:1.13

Table 5: Duration ratios between vowels and syllable uttered by the native speakers and foreign speakers per degree of foreign accent

The different values of vowel and syllable length between NSs and NNSs lead to a different syllable composition. On the one hand, we find the native speakers in whose syllables the consonantal component occupy the largest portion (54.4%). On the other, there are S and VS speakers. In their syllables, indeed, the vowel represents the longest sounds (S 48.1%; VS 48.5%). Mildly accented speakers lie in the middle. In their syllables, the consonantal part tends to be equivalent to the vocalic portion (50.6%).

Other factors affecting the performance of the Chinese participants are pronunciation mistakes. Their speech, indeed, is characterized by a great deal of the typical deviations that mark the interlanguage phonology of Chinese L2 Italian learners (Costamagna, 2011; Dal Maso, 2005).

In the corpus, for example, there is evidence of:

- phoneme substitutions (i.e.['ra:pido]>['la:pido]), due to the students' difficulty to articulate the opposition [l]/[r],

- phoneme alterations (i.e. [akkom'paɲ:a]> [*akkom'paɲ:a]) imputable to the Chinese learners' tendency to replace unvoiced stops with their voiced versions and

- phoneme insertion or deletion [dʒeneral'mɛnte]> [*dʒenerala'mɛnte]; ['bordo]>[*'bɔ:do], depending both on the speakers' tendency to simplify the Italian syllabic structure from CVC to CV and on their difficulty to produce syllable with vibrant or liquid codas.

Additionally, other attested phonemic deviations were concerned with accent shift and geminate timing[4]. Some

proparoxytone words like ['sindrɔme], for instance, were treated as paroxytones [*sin'drɔ:me], while some geminates were uttered as singleton (i.e. [ap:e'ti:to] >[*ape'ti:to]).

Nevertheless, among the segmental errors mentioned above, those that most significantly affect the native listeners' rate of accentedness were shifts of accents and misarticulation of long consonantal sounds. The lower the percentage of these two kinds mistakes, indeed, the milder the rate of foreign accent (Table 6).

	M	S	VS
Shift of accent	0.4%	1.4%	2.6%
Geminate timing	4%	6.3%	9.6%

Table 6: Percentage of pronunciation mistakes per degree of foreign accent

Conversely, Italian listeners, though neither competent in Chinese language, nor familiar with Chinese accented Italian, seem to remain neutral when Chinese speakers produce the errors that are consistent with their interlanguage phonology. As shown in Table 7, there is no direct relationship between the strength of perceived foreign accent and the percentage of insertion, substitution, deletion and alteration mistakes.

Kind of mistake	M	S	VS
Insertion	4.4%	1.6%	1.7%
Substitution	2.2%	3.3%	2.6%
Deletion	0.4%		0.9%
Alteration		1.7%	1.2%

Table 7: Percentage of pronunciation mistakes per degree of foreign accent and kind of deviation

However, the total number of pronunciation mistakes seem to affect the listeners' rate. The speaker with very strong foreign accent is the one with the highest percentage of segmental errors (17.6%). The S and M speakers, instead, produce the 15.1% and 11.4% of pronunciation mistakes.

4. Conclusions

The role of segmental and suprasegmental features in the perception of foreign accent has been a very controversial issue. For years research interests have focused on segmental deviations from native pronunciation and on phonetic transfers from L1 to L2.

In the light of the recent increasing attention paid to the contribution of prosody to the perception of foreign accent, the present study was intended to determine both

---

[4] In Italian geminate consonant timing is about twice singleton duration. In addition, the vowel that immediately precedes a

geminate consonant is shorter than the vowel preceding a singleton (Bertinetto, 1981; Zmarich & Gili Fivela, 2005).

the suprasegmental features and the segmental ones leading Italians to formulate the rate of foreign accent when listening to Chinese speakers of L2 Italian.

The comparison between the results of perception test and data from the spectro-acoustic analyses of L1-L2 Italian corpora has revealed that both levels play a role in influencing the rate of perceived foreign accent.

At suprasegmental level, fewer silences, fewer disfluencies, higher fluency and wider tonal range are perceived as signals of mild foreign accent.

At segmental level, instead, the speakers with native-like pronunciation are those whose speech was characterised by shorter duration of stressed and unstressed vowels and syllable, lower percentage of mispronunciation mistakes and errors relating to geminate timing and shift of accent.

5.   References

Avesani, C., Vayra, M., Zmarich, C., Paggiaro, R. and Sperandio, D. (2007). Le basi articolatorie della prominenza accentuale in italiano. In *Atti del Terzo Convegno Nazionale AISV.* CD-ROM. Torriana (RN): EDK editore.

Barr, D.J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In S. Santi, I. Guaïtella, C. Cave and G. Konopczynski (Eds.) *Oralité et gestualité, communication multimodale, interaction.* Paris: L'Harmattan, pp. 597--600.

Bertinetto, P.M. (1981). *Strutture prosodiche dell'italiano.*Firenze: Accademia della Crusca.

Best, C.T. (1995). A direct realistic view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience*. Baltimore (MD): York Press, pp. 171--206.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, 56, pp. 9--49.

Boula de Mareüil, P., Marotta, G. and Adda-Decker, M. (2004). Contribution of prosody to the perception of Spanish/Italian accents. In B. Bel, I. Marlien, (Eds.), *Speech Prosody*, Nara (Japan), pp. 681--684.

Boula de Mareüil, P., Vieru-Dimulescu, B. (2006). The Contribution of Prosody to the Perception of Foreign Accent. In *Phonetica, Int. Journal of Phonetic Science*, 63 (4), December, pp. 247--267.

Costamagna, L. (2011). L'apprendimento della fonologia dell'italiano da parte di studenti sinofoni: criticità e strategie. In *Atti del XV seminario AICLU.* Pavia: Pavia University Press, pp. 49--65.

Dal Maso, S. (2005). Processi di semplificazione della forma delle parole nell'italiano di cinesi in relazione alla struttura e allo statuto della sillaba. *Linguistica e Filologia,* 21, pp. 209--228.

De Meo, A., Pettorino, M. (2011). L'acquisizione della competenza prosodica in italiano L2 da parte di studenti sinofoni. In *Atti del XV seminario AICLU.* Pavia: Pavia University Press, pp. 67--78

De Meo, A., Pettorino, M. (2012). Prosodia e italiano L2: cinesi, giapponesi e vietnamiti a confronto. In R.

Bozzone Costa, L. Fumagalli and A. Valentini (Eds.), *Apprendere l'italiano da lingue lontane: prospettiva linguistica, pragmatica, educativa*. Perugia: Guerra, pp. 59--72.

De Meo, A., Pettorino, M. and Vitale, M. (2012). Non ti credo: i correlati acustici della credibilità in italiano L2. In *Atti dell'XI Congresso dell'Associazione Italiana di Linguistica Applicata.* Perugia: Guerra Edizioni, pp. 229--248.

Flege, J.E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer, N. Schiller (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities.* Berlin: Mouton de Gruyter, pp. 319--355.

Flege, J.E., Bohn, O.S. and Jang, S. (1997). Effects of experience on non-native speaker's production and perception of English vowels. *Journal of Phonetics*, 25, pp. 437--470.

Flege, J.E., Munro, M.J. and MacKay, I.R.A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, 16, pp. 1--26.

Giannini, A. (2010). Uno sguardo al ritmo e alla prosodia. In Mazzei, F., Carioti, P. (Eds.), *Oriente, Occidente e dintorni… Scritti in onore di Adolfo Tamburello*. Napoli: Il Torcoliere, Vol. III, pp. 1227--1239.

Horgues, C. (2010). Prosodie de l'accent français en anglais et perception par des auditeurs Anglophones. Thèse de Doctorat Université Paris-Diderot 7, 2010, publiée en Actes et Thèses de Association Francophone de la Communication Parlée.

Magen, H.S. (1998). The perception of foreign-accented speech. *Journal of Phonetics*, 26, pp. 381--400.

Major, R. C. (2001). *Foreign Accent: The Ontogeny and Philogeny of Second Language Phonology*. Mahwah, NJ: Erlbaum.

Marotta, G., Boula de Mareuil, P. (2010). Persistenza dell'accento straniero. Uno studio percettivo dell'italiano L2. In *Atti della V Convegno Nazionale AISV.* Torriana (RN): EDK Editore, pp. 475--494.

Matsuoka, R., Smith, I. (2008). Age Effects in Second Language Acquisition: Overview. *J Nurs Studies NCNJ,* 7(1), pp. 33--40.

Munro, M.J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition*, 17, pp. 17--34 .

Pettorino, M. (2003). La velocità di articolazione. In *Atti delle XIV Giornate di Studio del Gruppo di Fonetica Sperimentale – GFS.* Viterbo, 2003.

Pettorino, M., De Meo, A., Pellegrino, E., Salvati, L. and Vitale, M. (2011). Accento straniero e affidabilità del messaggio. Un'indagine acustico-percettiva. In *Atti dell'XI Convegno AISV*. Roma: Bulzoni editore.

Pettorino M., Giannini, A. (2005). Analisi delle disfluenze e del ritmo del dialogo romano. In F. Albano Leoni & R. Giordano (Eds.) *Italiano parlato. Analisi di un dialogo*. Napoli: Liguori, pp.89--104.

Pettorino, M., Giannini, A. (2010). Il parlato dei mass

media. In Pettorino, M., Giannini, A., Dovetto, F.M. (Eds.), *La Comunicazione Parlata 3*, vol. 2. OPAR - Università degli Studi di Napoli "L'Orientale", Napoli (2010), pp. 71--83.

Piat, M., Fohr, D. and Illina, I. (2008). Foreign accent identification based on prosodic parameters. *INTERSPEECH* - 2008, pp. 759--762.

Piske, T., MacKay, I.R.A. and Flege, J.E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29, pp. 191--215.

Ramus, F., Mehler, J. (1999). Language Identification With Suprasegmental Cues: A Study Based on Speech Resynthesis. *Journal of the Acoustic Society of America,* 105(1), pp. 512--521.

Savastano, E., Giannini, A. and Pettorino, M. (1995). Aspetti prosodici del parlato dei politici. In *Atti del XXIII Congresso Nazionale dell'AIA*. Bologna: Editore Cocchi, pp. 171--76.

Walley, A.C., Flege, J.E. (1998). Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics*, 27, pp. 307--332.

Zmarich, C., Gili Fivela, B. (2005). Consonanti scempie e geminate in italiano: studio cinematico e percettivo dell'articolazione bilabiale e labiodentale. In *Atti del I° Convegno Nazionale AISV*. Torriana (RN): EDK editore, pp. 429--448.

Perceptual competence and persuasiveness: L1 and L2 compared

Luisa SALVATI[1]**, Anna DE MEO**[2]**, Massimo PETTORINO**[3]

[1]University for Foreigners of Siena; [2,3]University of Naples "L'Orientale"

[1]Centro di Eccellenza – Via Sallustio Bandini 35 – 53100 Siena; [2]C.I.L.A. – Via Nuova Marina 59 – 80133 Napoli; [3]Via Duomo 219 – 80138 Napoli

salvati@unistrasi.it, ademeo@unior.it, mpettorino@unior.it

Abstract

This research aims at analysing the perception of prosodic features in learners of L2 Italian, from a comparative perspective with L1 Italian. In particular, we chose spontaneous argumentative speech, which implies the perlocutionary act of convincing, in order to investigate the relationship between the degree of persuasiveness of a speaker and the prosodic features characterizing her/his speech, in relation to the perceptual competence of non-native learners. A corpus of argumentative speech in L1 and L2 Italian has been collected. For the corpus in L1, 8 Italians, divided into two groups, were asked to take part in a debate and argue for or against a specific topic. The aim was to convince an audience of 19 Italians, who evaluated the persuasiveness of each speaker, judging it as "positive" or "negative". For the corpus in L2 Italian, we carried out the same procedure with 10 Chinese learners of Italian, who argued (5 pros, 5 cons) in front of an audience made up of 8 Chinese people. The data obtained are significant because they show not only that there is a relationship between persuasiveness and prosodic features, but that this relationship is strongly influenced by the perceptual competence of the listeners.

Keywords: prosody; perception; second language acquisition; persuasiveness.

1. Introduction

The prosodic competence in a second language is the result of a complex of variables, such as the quantity and the quality of exposure to the second language, the way of using L1 and L2, the language learning pathways, and the individual differences in terms of motivation, attitude, affective filter and age. The last factor is probably one of the influential: the period in which an individual can develop the same skills of a native speaker is limited to the first years of life. After this phase, it is very difficult that a non-native speaker is able to acquire an L2 prosodic proficiency comparable to that of a native speaker (Birdsong, 1999).

To these variables it must be also added the influence of the L1 prosodic models on the L2 perception. Some studies have recently focused on the influence of the perceptual segmentation and the resulting phonetic and phonological identification of the acoustic elements on the rhythmic organization which characterize the speech production in the various languages (Flege, 1991; Best & Tyler, 2007). In the case of a foreign language acquisition, it seems that a high degree of typological similarity between the languages in contact may cause positive transfer for the learning of morphosyntax, vocabulary and pragmatics, while a negative impact of the L1 or of other known languages may occur with regards to L2 pronunciation. Flege (1987) notices that this influence is also active on the L2 perceptual competence, since learners have real difficulties in discriminating the L2 sounds, particularly if they are similar to those of the native language.

1.1 Perceptual competence of Chinese learners of L2 Italian

The Chinese is a tonal and isolating language, typologically distant from Italian. Therefore, when dealing with the study of the Italian language, Chinese learners spend a lot of time trying to understand a language which is completely different from their L1, unless they have previously learned another foreign language typologically close to Italian. In particular, from the point of view of oral comprehension, Costamagna (2011) states that Chinese learners access to speech understanding with great difficulty, because they are unable to perceive and segment the Italian speech chain. The development in comprehension is also influenced by the Italian morphological organization: in the early stages, Chinese learners try to grab the prominent elements that can facilitate the comprehension, as they perceive the linguistic message in L2 as an indistinct mass of sounds without distinguishing the discriminatory elements. In more advanced levels, they develop a greater awareness of the distance existing between the two languages, above all as regards the prosodic structure. The skill of using variations of intonation for pragmatic purposes can be seen only in advanced levels, since in the early stages, they generally recognize interrogative and exclamative sentences.

Therefore, what characterizes the perceptual competence in Chinese learners of L2 Italian is a little progression from one stage of interlanguage to the other one, as shown by De Meo & Pettorino (2011) in a study on the relationship between language proficiency and prosodic competence. A Chinese can achieve a C1 high level of language competence (C1 level of the Common European Framework of References - CEFR) and, at the same time, not adequately develop the ability to communicate effectively with Italian native speakers using the appropriate prosody and intonation. Oral comprehension is also delayed by the different Chinese and Italian pragmatic-communicative models and this often makes the oral interaction in L2 Italian difficult (Costamagna, 2011).

2.   Material and method

The present research aims at analyzing the perception of rhythmic and prosodic features in Chinese learners of L2 Italian, in a comparative perspective with the Italian native speakers. Using the task of the debating the relationship between the degree of persuasiveness achieved by the speaker and the related rhythmic and prosodic features of her/his speech was investigated.

It should be clarified that the study was carried out with the awareness that the argumentation and a speaker's persuasiveness is the result of a series of elements: the content of the text, the way the speaker expresses her/his opinions, the body language. Given these variables, the prosodic component was isolated to verify its influence on the ability to persuade the audience, not only because, through the voice, the speaker can arouse emotion and, therefore, persuade, but also because the voice may be spectro-acoustically analysed, allowing measurable and comparable results.

2.1   The debating structure

The debating is not simply a discussion where speakers argue about a topic, but it is rather an interactive exchange of ideas, with a strict protocol of rules which imply the alternation of arguments for and against a given topic, imposing a time limit to respect and finally involving the audience judgment called upon to evaluate individual speakers on linguistic, paralinguistic and extralinguistic parameters. When the debating involves also foreigners who argue in L2, it becomes an intercultural interaction between natives and non-natives, who are characterized by different linguistic behaviours and cultural backgrounds. In this perspective, the features normally defining the debating become even more complex because of cognitive factors related to language learning processes, sociolinguistic and sociopragmatic factors. This type of arguing involves intercultural communication skills as well, i.e. proper skills to interact by negotiating meanings, values, symbols, ideas, on a "common ground" (Fetzer & Fischer, 2006) between natives and non-natives.

For this research, a debating was held between a team of Italians and a team of Chinese learners of Italian. The debating took place in two phases. In the first one, chaired by a moderator, members of each team alternatively argued on the topic, having a time limit of two minutes. In the second phase, both groups had a time limit of six minutes to discuss freely, without any moderator, in order to convince the audience.

2.2   The corpora

The corpus in L1 and L2 Italian was audio-recorded using Goldwave 5.58 and videotaped by a Sony handycam HDR-SR8E and then orthographically annotated on the basis of the indications given by the CLIPS project "Lexicons and Corpora of Written and Spoken Italian"

(Albano Leoni & Giordano, 2005). Here we will refer to the spectro-acoustic analysis conducted, using Wavesurfer 1.8.8, on the corpus recorded during the first phase of the debating, where each speaker talked without being interrupted.

For each speaker measures were performed in order to determine the number of speech chains, the number of syllables for each speech chain, the duration of each speech chain, the duration of silent pauses, the duration of non-silent pauses or disfluencies, the maximum and the minimum f_0 value for each speech chain. Furthermore, for each speaker the following calculation were carried out: articulation rate AR, i.e. the ratio between the number of syllables and the speech chain duration (syll/s), speech rate SR, i.e. the ratio between the number of syllables and the utterance time (syll/s), fluency (F), i.e. the ratio between the number of syllables and the number of speech chains (syll/SC), the percentage of silence duration, the mean duration of silent pauses (s), the percentage of disfluencies duration, the tonal range, i.e. the difference between the maximum and the minimum f_0 value in an utterance, measured in semitones (st) in order to compare data relating to different speakers.

2.3   The native and non-native participants

The Italian speakers were three female and one male university students, aged between 20 and 25, all coming from the Campania region (southern Italy). The Chinese participants, two male and two female students of Italian at the University of Tianjiin, aged between 20 and 25, who had been living in Naples for four months, had a language competence of Italian corresponding to B2 level of CEFR. Before the debating, rules were explained to both groups separately and some tips on how to practice for the discussion, both individually and in groups, were given. Moreover, a large part of this introductory phase was dedicated to comment on the parameters the speakers would have been judged on: persuasiveness, voice volume, speech rate, pauses, intonation, posture and gaze, gestures, language use and competence. Afterwards, several debating simulations were held. For the Chinese learners, a textbook aiming at development of the argumentative skills in L2 Italian was used (Barki & Diadori, 1994).

3.   The perception of persuasiveness in the L1 corpus

The corpus in L1 Italian consists of a debating between native speakers (NS) in front of an audience of native listeners (NL) about the following topic: "It is better to live in a big city". The team in favor was made up of one man and three women, while the team against was formed by 4 women. The audience, consisting of 19 NLs, male and female, had to judge the persuasiveness of each speaker in terms of "positive" or "negative."

To investigate the perceptual level, the persuasiveness degree of each speaker was related to the prosodic features of her/his speech, in order not only to verify the existence of a link between persuasiveness and prosody. The most significant relationships were found

between persuasiveness and AR (Figure 1), fluency (Figure 2), mean duration of silent pauses (Figure 3), disfluencies (Figure 4).
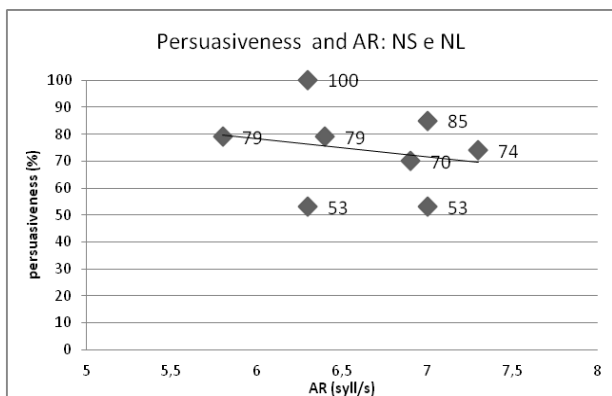
Figure 1: Articulation rate and persuasiveness
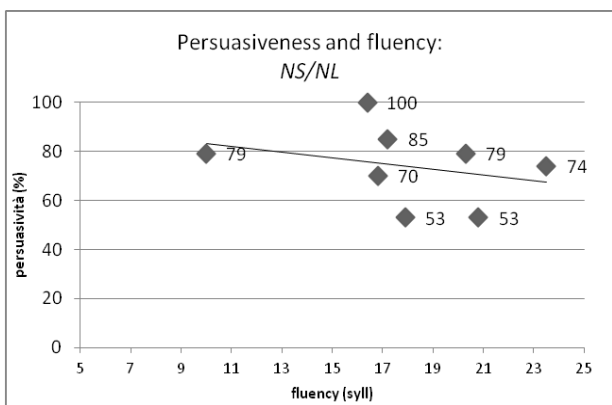
Figure 2: Fluency and persuasiveness
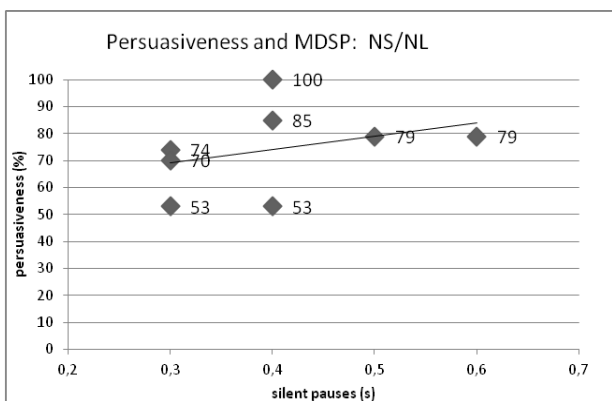
Figure 3: Mean duration of silent pauses and persuasiveness

The graphs show that the Italian listeners tend to accept an argumentation pronounced with a greater articulatory accuracy and many medium-long silent pauses, which may give them time to think about what they have just listened: the more the Italian speaker produces long silences, the more the native listener

perceives him/her as more persuasive. Instead, persuasiveness decreases if disfluencies increase, as if the native listener perceives those silent pauses, which are used to fill the spaces between sentences, as disturbing elements.

In conclusion, from the results it is possible to assume that a native listener tends to perceive an hyper-articulated speech with many long silent pauses and few disfluencies as more persuasive. Furthermore there are not significant relationships among persuasiveness, speech rate and tonal range.
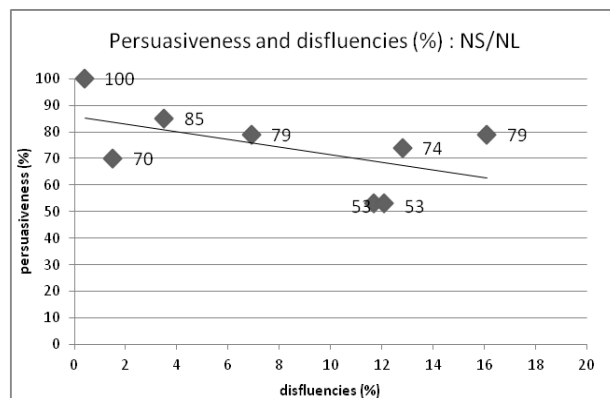
Figure 4: Disfluecies and persuasiveness

4. The perception of persuasiveness in the L2 corpus

The corpus in L2 Italian consists of the debating between two groups of non-native speakers (NNS) in front of non-native listeners (NNL). The team in favour consisted of four female and one male Chinese students, while the team against was made up of three male and two female Chinese students. In order to eliminate, as far as possible, the text variable, the assigned topic was the same as the one used in the previous L1 debating.

In this section, the relationship between the prosodic features characterizing the speech of NNSs and their ability to persuade a non-native audience will be analysed. Data were used to evaluate if the L1 and the L2 debating share the same characteristics, and to determine how NNLs perceive their peers speaking in a foreign language. In literature there are very few studies which deal with these questions and they mainly relate to foreign languages others than Italian.

Results show that the 84% of NNLs judged in a very positive way all the speakers, regardless of prosodic features. However it is worth reflecting upon how the relationship between persuasion and prosody is related to the perceptual ability of the listener. Indeed, a comparison between NLs evaluation (in the L1 debating) and that of the NNLs (in the L2 debating) points out that, while the NSs perceive a clear relationship between persuasiveness and the related prosodic features, the Chinese learners competence does not seem enough to detect a significant connection between prosody and persuasiveness. With regards to this difference, it can be assumed that there are

two co-existing causes. On the one hand, the assignment of a judgment on persuasiveness involves a four-step process: listening to speech, understanding the acoustic message, comparing it with one own opinions, and finally giving the judgment. It seems that the non-native learner pays more attention to single words rather than to the argument as a whole, unlike the NL, who has the tools to reach the next phases of the comprehension process. The speech perception in L2, indeed, is strongly influenced by the mother tongue prosodic structure, which may affect the learner's oral comprehension ability. Chinese learners, who have a native language characterized by rhythmic and intonation structures very different from Italian language, access to speech perception with great difficulty, because they are unable to perceive and segment the speech chain effectively.

On the other hand, there are idiosyncratic sociolinguistic and cultural mechanisms in the NNSs: from this perspective, the Chinese students positively evaluate their peers to reward the effort and the commitment they face dealing with another language. The development of the L2 perceptual competence, therefore, slows down because of the different pragmatic-communicative patterns of the learners.

The combination of these two elements - one cognitive, the other one socio-linguistic - leads the NNSs to identify with difficulty the suprasegmental components and their pragmatic value. This is even more interesting when we consider that the CEFR, with reference to the listening comprehension skills of B2 learners, indicates that s/he is able to understand the main ideas of a complex text on both concrete and abstract topics, including technical discussions in their field of specialization.

Considering the results obtained by this research, it can be added that an L2 Italian learner, with the so-called autonomy level of a language knowledge, is able to perceive and decode complex messages, but s/he is less able to evaluate them in terms of persuasiveness. The data shed new light on the studies regarding perceptual competence from an acquisitional point of view and on the ability of oral understanding. They also reveal a certain lack of attention to the prosodic dimension of L2 communication, both in acquisition and teaching, and finally, in the assessment field, because of the absence of any reference to language suprasegmental aspects in the CEFR descriptors.

5.   Conclusion

The task of the present study was to analyze the perception of rhythmic-prosodic features in Chinese learners of L2 Italian in argumentative speech, from a comparative perspective with L1 Italian. To this purpose,. a relationship between prosody and persuasiveness was outlined: it emerged that Italian listeners find most persuasive a well structured L1 speech, with many long silent pauses and few disfluencies. These data about spontaneous speech confirm the research carried out by De Meo *et al.* (2011) on read speech.

Instead, with regard to the non-native speakers, for spontaneous argumentative speech, there is no significant relationship between persuasiveness and prosody, since the Chinese students have always attributed highly positive evaluations, which do not allow detecting a trendline that can link the above variables. Regarding this issue, this study proposes two exlanations, one cognitive and the other cultural-pragmatic.

Further research could have repercussions in the field of language teaching, since the oral texts administered to learners should be constructed, adapted and chosen not only on the basis of morphosyntactic structures and language functions, but also according to the various levels of perceptual competence that the L2 learners develop. Finally, it is interesting not only to extend the investigation to the relationship between persuasiveness and textual/kinesic variables, but also to study the link existing with the prosody by the technique of low-filtering, in order to eliminate other variables.

6.   References

Albano Leoni, F., Giordano, R. (Eds.) (2005). *Italiano parlato. Analisi di un dialogo*. Napoli: Liguori.

Barki Coricelli, P., Diadori, P. (1997). *Pro e contro. Conversare e argomentare in italiano. Livello intermedio. Guida per l'insegnante*. Roma: Bonacci.

Best C.T., Tyler M.D. (2007). Nonnative and second-language speech perception. In O. Bohn, M.J. Munro. (Eds.), *Language Experience in Secon Language Speech Learning*. In honor of James Emil Flege. Amsterdam/Philadelphia: John Benjamins Publishing, pp. 13 --34.

Costamagna, L. (2011). L'apprendimento della fonologia dell'italiano da parte di studenti sinofoni: criticità e strategie. In E. Bonvino, S. Rastelli (Eds.), *La didattica dell'italiano a studenti cinesi: il programma Marco Polo ed altre esperienze, Proceedings of XV Workshop AICLU*. Pavia: Pavia University Press, pp. 49--66.

De Meo, A., Pettorino, M. (2011). L'acquisizione della competenza prosodica in italiano L2 da parte di studenti sinofoni. In E. Bonvino, S. Rastelli (Eds.), *La didattica dell'italiano a studenti cinesi: il programma Marco Polo ed altre esperienze, Proceedings of XV Workshop AICLU*. Pavia: Pavia University Press, pp. 67--78.

De Meo, A., Pettorino, M. and Vitale, M. (2012). Non ti credo: i correlati acustici della credibilità in italiano L2. In G. Bernini, C. Lavinio, A. Valentini and M. Voghera (Eds.), *Competenze e formazione linguistiche. In memoria di Monica Berretta, Proceedings of XI Congress Associazione Italiana di Linguistica Applicata*. Perugia: Guerra Edizioni, pp. 229 -- 248.

Flege, J.E. (1987). The Production of "New" and "Similar" Phones in a Foreign Language: Evidence for the Effect of Equivalence classification. In *Journal of Phonetics*, 15, pp. 47 -- 65.

Flege, J.E. (1991). Perception and production: the relevance of phonetic input to L2 phonological learning.

In T. Huebner, C. A. Ferguson (Eds.). *Cross currents in Second Language Acquisition and Linguistic Theory*. Amsterdam/Philadelphia: John Benjamins Publishing, pp. 249--290.

Pettorino, M., Pellegrino, E., Salvati, L., Vitale, M. and De Meo, A. (in press). La voce dei media. Trapianti ritmico-intonativi per un'analisi diacronica dell'italiano parlato. In *Coesistenze linguistiche nell'Italia pre- e postunitaria, Proceedings of XLV International Congress of Società di Linguistica Italiana.* Aosta/Bard/Torino, 26-28th September 2011.

Medidas da variação prosódica diatópica no espaço românico

Antonio ROMANO, Lurdes de Castro MOUTINHO, Rosa Lídia COIMBRA, Albert RILLIARD

Laboratorio di Fonetica Sperimentale "A. Genre" - Dip. di Lingue e L.S. e C.M. - Università degli Studi di Torino (Itália);
Centro de Línguas e Culturas - Universidade de Aveiro (Portugal); LIMSI-CNRS - Orsay (França)
antonio.romano@unito.it, lmoutinho@ua.pt, rlcoimbra@ua.pt, albert.rilliard@limsi.fr

Abstract

The aim of this article is to propose an experimental method for automatic assessment of prosodic similarities between dialects within a large linguistic domain (Romano *et al.*, 2011; Moutinho *et al.*, 2011). Data have been collected in the framework of the international *AMPER* project (Atlas Multimédia Prosodique de l'Espace Roman / Atlas Multimédia Prosódico do Espaço Românico) and measurements taken into account for this experiment refer to varieties of European Portuguese and regional Italian varieties. General indexes such as coherence and congruence have been tested and, between different varieties, prosodic similarity is measured on the basis of a weighted correlation formula providing elements for the definition of dialectometric distances. Italo-romance dialects were also considered in some case in order to enlarge the testing to the assessment of prosodic persistence between similar languages spoken by the same speakers (Romano, 1999). Since intonation within the Romance domain may show different patterns, this study is intended to provide useful elements explaining how these patterns could define homogeneous contiguous areas vs. discontinuous dialectal spaces or converging solutions between separate regions.

Keywords: Prosody, Dialectometry, Italo-Romance, European Portuguese.

1. Introdução

A necessidade de uma descrição e comparação dos traços prosódicos das variedades linguísticas do espaço românico está na origem do projeto internacional *AMPER* (Atlas Multimédia Prosódico do Espaço Românico, cf. Moutinho & Coimbra, 2000; Romano, 2001). Fazem parte deste projeto equipas de diversos laboratórios europeus e latino-americanos, adotando todas elas estratégias comuns de constituição, recolha e análise de corpora.

Na verdade, a entoação no espaço linguístico românico apresenta diferentes esquemas, marcados diatopicamente, que podem ser convergentes ou divergentes de uns falares para outros no mesmo domínio linguístico (Contini, 1992; Romano, 1999; Contini, 2008; Contini *et al.*, 2008; Turculeţ, 2008) e o presente estudo propõe um método de avaliação objetivo destas dinâmicas[1].

2. O projeto *AMPER*

A participação no projeto *AMPER* permitiu-nos a obtenção de dados prosódicos comparáveis num conjunto de pontos bastante denso em algumas das regiões linguísticas que fazem parte deste Projeto[2]. Depois de termos trabalhado sobre um conjunto limitado de frases, com comparações escolhidas entre perfis duma seleção de variedades italianas, portuguesas e brasileiras (Romano, 1999; Romano & Moutinho, 2004; Interlandi *et al.,* 2007; Felloni, 2011), pretendemos, neste artigo, discutir resultados de análise relativos a semelhanças e diferenças manifestadas nas configurações entoacionais obtidas para numerosas estruturas frásicas em diversas variedades do Italiano (em *AMPER-ITA*) e do Português europeu continental (em *AMPER-POR*).
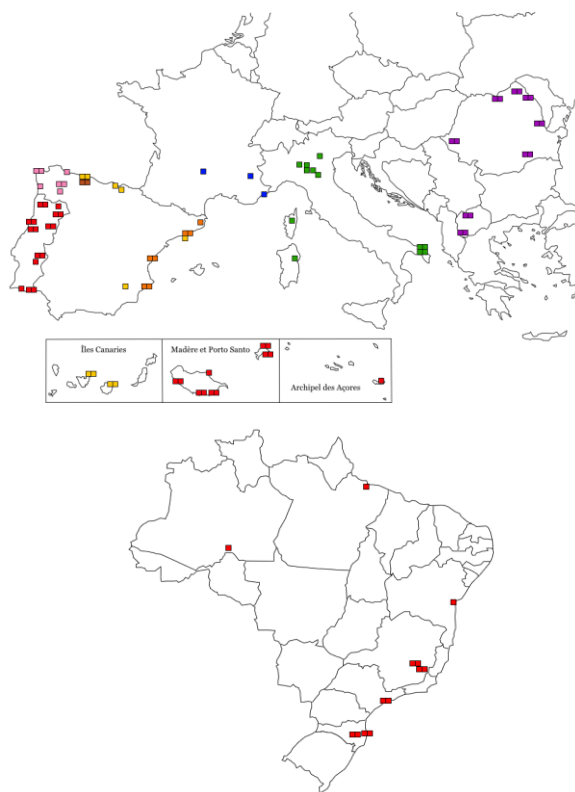
Figura 1: Os dados do *DVD AMPER* 2011 (ed. por P. Mairano): 62 pontos de inquérito e 108 falantes

---

[1] Ainda que baseado num trabalho comum dos 4 autores e numa versão preliminar coletiva, para fins académicos explicitamos que a versão final deste artigo foi redigida por A. Romano (§§2, 4.1 et 5), L. de Castro Moutinho (§3), L. de Castro Moutinho e R.L Coimbra (§1 et 6) e A. Rilliard (§4.2).

[2] Uma parte destes dados foram recentemente publicados (*DVD AMPER* 2011). O tratamento em curso de umas duas dezenas de outros inquéritos deveria melhorar, nos próximos meses, a cobertura do *AMPER-ITA*, nomeadamente a secção do projeto consagrada aos falares italo-romanos e às variedades do italiano regional. Também no âmbito do *AMPER-POR* (português brasileiro e europeu), estão em curso novos inquéritos e análise de dados, que serão objeto de nova publicação.

3.   Dados

O corpus submetido a esta análise é constituído por um total de 28 frases declarativas e 28 interrogativas, com as mesmas estruturas sintáticas e obedecendo a restrições de tipo fonético e sintático[3]. Esta recolha foi efetuada com a colaboração de 36 informantes (18 homens e 18 mulheres), provenientes de diferentes regiões. Da totalidade do corpus gravado, foram selecionadas, para cada informante, 3 repetições de cada uma das estruturas e modalidades, o que perfaz um total de 6048 enunciados analisados para o estudo que aqui apresentamos.

4.   Metodologia

4.1   Medidas de correlação

Com o objetivo de serem estabelecidas semelhanças entre os dados de duas variedades, comparam-se as sequências de valores de frequência fundamental (f_0), duração (D) e energia (I) com base numa variável determinada em Romano (1999, 2008):

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sigma_x \cdot \sigma_y} \tag{1}$$

donde:

$$-1 \leq \rho_{x,y} \leq 1 \text{ (em percentagem } -100\% \leq \rho_{x,y} \leq 100\%)$$

e:

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y) \tag{2}$$

X e Y representam sequências de n valores de f_0, embora pudessem igualmente referir-se a séries de dados relativos à energia e à duração.

O resultado assim obtido precisa de ser validado por referência a um patamar previamente definido, estabelecendo-se o intervalo de oscilação da variável, quando se trata de repetições da mesma frase, na mesma variedade e produzida pelo mesmo falante (v. abaixo).

Apresentamos em seguida um exemplo da utilização do index de correlação de Romano & Miotti (2008).

Na Figura 2 podemos observar os valores de correlação especialmente elevados (0,76÷0,84), quando comparamos enunciados declarativos, representados nos gráficos da esquerda e caraterizados por curvas bastante similares. Os valores positivos de correlação apresentam-se, somente em dois casos, para os locutores escolhidos para este caso específico (gráficos à direita, de baixo para cima): uma fraca correlação (0,26) diz respeito às duas questões nas variáveis 0905 e 0820, que são globalmente dbastante similares, mas com diferenças localizadas bem visíveis, as quais são colocadas em evidência com setas a tracejado, no início dos enunciados. O index baixa para 0,04, na comparação entre 0276 e

0905: as diferenças localizam-se, desta vez também, na parte terminal dos contornos, depois da realização do perfil correspondente ao acento de frase, com evoluções completamente inversas. Na comparação entre 0276 e 0820, as diferenças estendem-se sobre outras porções da curva (afetando também as vogais acentuadas) e a correlação torna-se negativa (-0,23).
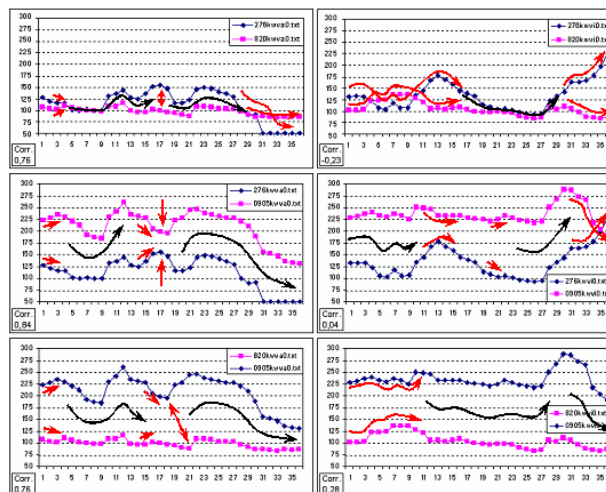
Figura 2: Comparações entre curvas (Romano & Miotti, 2008). Frases interrogativas de variedades de três espaços diferentes (0820, 0276, 0905). Medida de correlação em baixo à esquerda de cada gráfico

No entanto, este index não tem em conta a importância percetiva de certas modificações da curva, em correspondência com as posições acentuais ou de fronteira prosódica. Como já foi referido, mostra-se muito sensível às variações individuais e precisa de uma avaliação prévia sobre as repetições de um mesmo locutor e as realizações de locutores de uma mesma localidade, definindo, respetivamente uma medida de "coerência" e de "congruência" (v. infra e §4).

4.2   Medidas de semelhança

Para evitar estas avaliações prévias, para melhor poder ter em consideração a energia e também para limitar a influência de valores absolutos de f_0, uma nova medida foi proposta em Moutinho *et al.* (2011)[4]. Esta medida constitui uma avaliação objetiva da semelhança percetiva entre duas curvas entoacionais comparáveis. Permite também ignorar as diferenças de registo que poderiam existir entre dois locutores, para se concentrar na proximidade morfológica dos contornos. Esta medida mostrou-se pertinente na avaliação da proximidade percetiva de dois contornos prosódicos e, nesse aspeto, pareceu-nos adaptada a este tipo de situação. A referida medida baseia-se na seguinte fórmula:

---

[3] Estas restrições foram estabelecidas, desde o início, para o projeto *AMPER* (v. *DVD AMPER* 2011).

[4] Esta medida de correlação é obtida graças aos escriptes Matlab™. Outros métodos de avaliação de distâncias são estudadas por A. Rilliard (cf. Rilliard *et al.*, 2011).

$$r_{f_1 f_2} = \frac{\sum_i w(i)(f_1(i) - m_1)(f_2(i) - m_2)}{\sqrt{\sum_i w(i)(f_1 - m_1)^2 \sum_i w(i)(f_2(i) - m_2)^2}} \quad (3)$$

Aqui, *f1* e *f2* representam os valores de f_0 dos dois contornos entoacionais (expressos em semi-tons); *m1* e *m2* os valores médios destes contornos de f_0 para a totalidade da frase, e *w* a ponderação devida à energia do sinal, calculada como a média dos dois valores de energia medidos num ponto dado para as duas frases comparadas , expressos em dB[5].

O índice *i* varia entre 1 e o número de pontos de medida de f_0 para a frase considerada. Os valores de f_0 e de energia extraídos segundo o protocolo *AMPER* são utilizados para a seguinte medida: 3 pontos de f_0 por vogal, ponderados a partir do mesmo valor de energia média da vogal.

A distribuição das medidas de correlação não seguem uma lei normal, considerámos o valor da mediana como indicadora central, preferindo aquele ao valor da sua média.

5. Resultados

Comparando os valores da semelhança entre repetições para um mesmo falante e para falantes dum mesmo dialeto, obtivemos medidas de coerência e congruência.

Nos gráficos da Figura 3, propomos a avaliação da coerência de seis locutores de *AMPER-ITA* (pontos 061, 062, 06g e 06h, v. *DVD AMPER* 2011).

Os dados dos quatro primeiros locutores do primeiro diagrama acima representado (0616, 0621, 0625 e 06g5)[6] evidenciam uma boa coerência (>90%), enquanto que para o locutor 06g6 a dispersão de valores assinala a presença de repetições com curvas bastante diferentes e para o locutor 06h7 uma coerência reduzida (mesmo assim >80% em média).

No segundo diagrama, avalia-se a congruência entre os dados de 0621 e 0625 (da localidade 062) e entre os dados de 06g5 e 06g6 (da localidade 06g): Os dados relativos ao ponto 062 estão associados a valores de congruência bastante elevados (à volta de 94%) e com uma dispersão bem concentrada (superior a 90%), enquanto que os dados de 06g apresentam uma congruência média inferior a 85% (no entanto ainda bastante elevada) e oscilações que poderiam ser consideradas localmente mais importantes[7].

Se, pelo contrário, compararmos, os valores desta medida para falantes de dialetos próximos, obteremos

uma estimativa da (dis)semelhança entre as amostras.
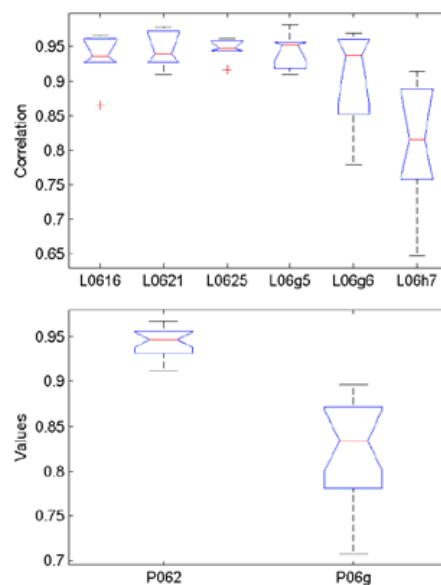
Figura 3: Medidas de coerência intra-falante e de congruência inter-falantes (dados *AMPER-ITA*: Romano, 1999 – 061-062; Felloni, 2011 – 06g-06h)
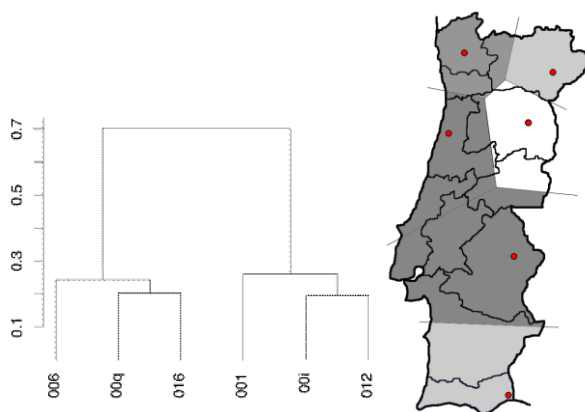
Figura 4: Dendrograma com o agrupamento de *clustering* hierarquizado e mapa dialectométrico da distância prosódica média dos dados das diferentes regiões em relação aos dados do ponto 016 (Trinta, Beira Alta) [Moutinho *et al*. (2011)]. A intensidade do cinzento é uma função linear da distância entre os pontos. Códigos: 006 = Alfândega da Fé (Trás-os-Montes), 00q = Monte Gordo (Algarve), 001 = Prado (Braga, Minho), 00i = Monforte (Alto Alentejo), 012 = Aradas (Beira Litoral)

Uma quantificação dos resultados obtidos encontra-se detalhada em Moutinho *et al*. (2011), com a apresentação de vários casos de congruência reduzida para certas localidades (20-60%), melhor para outras, como é o caso do ponto 016 (Trinta, Beira Alta) e que escolhemos como localidade de referência para uma primeira proposta de avaliação geoprosódica destes dados. Para dar conta das relações existentes entre os dados de outras localidades com os obtidos para estas (somente 25% de semelhança entre 016 e 012 e 75% entre 016 e

---

[5] Esta ponderação foi introduzida a partir de Moutinho *et al*. (2011). Acerca do interesse desta ponderação, v. também discussão em Lai & Rilliard (2008) e Romano & Miotti (2008).
[6] O último algarismo acrescentado ao código da localidade designa o código do locutor.
[7] Isso não significa necessariamente que o inquérito realizado no ponto 062 reproduza uma estimativa da prosódia típica desta localidade melhor do que a obtida para o ponto 06g: uma congruência menos boa poderia ser o sintoma de uma prosódia mais variável localmente entre variedades diastráticas e/ou os idioletos de género.

061), foi adotado o método de análise dialectométrica (cf. Goebl, 1981, 1996) baseada numa avaliação cruzada da distância prosódica média dos dados das diferentes regiões em relação aos dados deste ponto, tendo sido proposto um agrupamento hierarquizado (v. Figura 4)[8].

6. Conclusões

As medidas que nós aplicámos aos dados das variedades presentes na *BD AMPER*, mesmo que não possam substituir a análise do dialectólogo foneticista tradicional, permitem, sem dúvida, colocar em evidência algumas divergências e convergências prosódicas de diferentes falares. Estas constatações proporcionam indicações sobre a distância percetiva que podemos esperar encontrar entre dialetos.

Nas nossas pesquisas, depois de termos discutido as possibilidades e modalidades de aplicação da distância propostas em trabalhos precedentes, fizemos a avaliação da variabilidade prosódica em dados referentes a uma primeira seleção de variedades. Deste modo, este estudo deve ser lido como um esboço de dialectometria prosódica.

É indispensável que estes resultados sejam confirmados através de análises que incidam sobre um número mais vasto de falantes de cada uma das variedades e sobre a base de um conjunto de pontos de inquérito mais densa e completa, para ambas as línguas, mas especialmente no que diz respeito ao *AMPER-ITA*.

7. Agradecimentos

8. Referências

AMPER (2005). Projet *AMPER* - Atlas Multimédia Prosodique de l'Espace Roman. *Géolinguistique*, hors série 3 (J.P. Lai, ed.).

AMPER (2011). Intonations Romanes. *Géolinguistique*, hors série 4 (P. Mairano, ed.).

Contini, M. (1992). Vers une géoprosodie. In *Actas do "Nazioarteko Dialektologia Biltzarra Agiriak"* (Bilbao, Espanha, 1991), Bilbao: Real Academia de la Lengua Vasca, pp. 83--109.

Contini M., Lai J.-P., Romano A., Roullet S., de Castro Moutinho L., Coimbra R.L., Pereira Bendiha, U. and Secca Ruivo, S. (2002). Un projet d'atlas multimédia prosodique de l'espace roman. In *Proc. of the International Conference Speech Prosody 2002* (Aix-en-Provence, França, 2002) (B. Bel & I. Marlien eds.), pp. 227--230.

Contini, M. (2008). Le projet *AMPER*: passé, présent et avenir. In L. de Castro Moutinho, R.L. Coimbra (Eds.), *Actas das I Jornadas Científicas AMPER-POR* (Aveiro, Portugal, 2007), pp. 9--19.

Felloni, M.C. (2011). *Prosodia sociofonetica. L'italiano parlato e percepito a Parma*. Milano: FrancoAngeli.

Fernández Planas, A.M., Roseano P., Martínez Celdrán, E. and Romera Barrios, L. (2011). Aproximación al análisis dialectométrico de la entonación en algunos puntos del dominio lingüístico catalán. In *Estudios de Fonética Experimental*, XX, pp. 33--55.

Goebl, H. (1983). Eléments d'analyse dialectométrique (avec application à l'AIS). *Revue de Linguistique Romane*, 45, pp. 349--420.

Goebl, H. (1996). La convergence entre les fragmentations géo-linguistiques et géo-génétiques de l'Italie du Nord. In *Revue de linguistique romane*, 60, pp. 25--49.

Interlandi, G.M., Romano, A., de Castro Moutinho, L. and Coimbra, R.L. (2007). Cross-linguistic prosodic variation in two Romance languages: Portuguese and Italian. In *Actas do III Congreso Int. de Fonética Experimental* (Santiago de Compostela, Espanha, 2005), pp. 377--387.

Lai, J.P., Rilliard, A. (2008). Outils pour le calcul et la comparaison prosodique dans le cadre du projet *AMPER*- l'exemple des variétés Occitane et Sarde. In A. Turculeţ (Ed.), pp. 217--229.

Moutinho, L. de C., Coimbra, R.L. (2000). Para a construção de um atlas prosódico multimédia das variedades românicas. In *RUA-Letras*, 17, pp. 111--118.

Moutinho, L. de C., Coimbra, R.L., Rilliard, A. and Romano, A. (2011). Mesure de la variation prosodique diatopique en portugais européen. In *Estudios de Fonética Experimental*, XX, pp. 33--55.

Rilliard, A., Lai, J.P. (2008). La Base de Données AMPER et ses interfaces: structure et formats de données, exemple d'utilisation pour une analyse comparative de la prosodie de différents parlers romans. In L. de C. Moutinho, R.L. Coimbra (Eds.), *Actas das I Jornadas Científicas AMPER-POR* (Aveiro, Portugal, 2007), pp. 127--139.

Rilliard A., Allauzen, A. and Boula de Mareüil, Ph. (2011). Using Dynamic Time Warping to compute prosodic similarity measures. In *Proc. of INTERSPEECH 2011* (Florência, Itália, 2011), pp.2021--2024.

Romano, A. (1999). Analyse des structures prosodiques des dialects et de l'italien régional parlés dans le Salento (Italie): approche linguistique et instrumentale. Thèse de Doctorat de l'Université Stendhal de Grenoble (Lille: Presses Univ. du Septentrion, 2001).

Romano, A. (2003). Un projet d'Atlas multimédia prosodique de l'espace roman (AMPER). In F. Sánchez Miret (Ed.), *Actas do XXIII CILFR* (Salamanca, Espanha, 2001), vol. I, Tübingen: Niemeyer, pp. 279--294.

Romano, A. (2004). Indices acoustiques supra-segmentaux dans la caractérisation des langues

---

[8] Uma comparação automática dos dados *AMPER* com outros critérios, tendo também em vista uma análise dialectrométrica, foi também aplicada por Fernández Planas *et al.* (2011).

romanes : identification de variétés linguistiques et description des traits prototypiques. In *Actes du congrès MIDL 2004 "Identification des langues et des variétés dialectales par les humains et par les machines"* Paris: École Nationale Supérieure des Télécommunications, pp. 91--92.

Romano, A. (2008). Éléments théoriques et pratiques de l'analyse multiparamétrique de la prosodie dans le cadre d'AMPER. In L. de C. Moutinho & R.L. Coimbra (eds.), *Actas das I Jornadas Científicas AMPER-POR* (Aveiro, Portugal, 2007), pp. 115--126.

Romano, A., Moutinho L. de C. (2004). Alguns esquemas entoacionais característicos da leitura de um texto por portugueses e brasileiros. In *Revista de Estudos da Linguagem*, 12/2, pp. 93--109.

Romano, A., Miotti, R. (2008). Distancias prosódicas entre variedades románicas. In A. Turculeţ (ed.), pp. 231--249.

Romano, A., Contini M., Lai J.-P. and Rilliard A. (2011). Distancias prosódicas entre variedades románicas en el marco del proyecto AMPER. In *Revista Internacional de Linguística Iberoamericana – RILI*, Vol. IX, No. 1 (17), pp. 17--26.

Turculeţ, A. (Ed.). (2008). *La variation diatopique de l'intonation dans le domaine roumain et roman*. Iaşi: Editura Universităţii Al. I. Cuza.

F2 transition as a cue to place of articulation in Brazilian Portuguese coronal fricatives

Rui ROTHE-NEVES, Fabiana Andrade PENIDO

Laboratório de Fonética da FALE/UFMG

Av. AntonioCarlos, 6627 – 31270-217, Belo Horizonte/MG - Brazil

rothe-neves@ufmg.br, fabianapenido@gmail.com

Abstract

In Portuguese, frication noise has been tested for consonant recognition, but to the best of our knowledge formant frequencies were not investigated yet. We tested whether the second formant (F2) transition said to be useful in English are also a cue to place of articulation for coronal fricatives in Brazilian Portuguese. Subjects performed a rating task in which they had to listen to a syllable and quickly respond whether they heard [(s)a] or [(ʃ)a] (1st. block) and [(s)u] or [(ʃ)u] (2nd. block), and rate their confidence in their answers using a 3-pt scale. Hit and false alarm rates of all response alternatives to [(s)a]-[(ʃ)a] and [(s)u]-[(ʃ)u] were computed. Slope and A_z were estimated by maximum likelihood estimation of ROC. For [a] there was good evidence in our data that F2 transition is an important and sufficient cue to place of articulation in coronal fricatives. Also, F2 transition described for English, once adapted to the formant frequency values reported for BP, are useful to distinguish between [sa]-[ʃa]. However, listeners could not distinguish between [(s)u]-[(ʃ)u] only on the basis of F2 transition. This result points to a possible role of F3 transition, which was said to become an important cue for rounded vowels.

Keywords: speech perception; phonetics; phonology; Brazilian Portuguese.

1. Introduction

Two kinds of cues are shown to be used in the distinction between coronal voiceless fricatives: the spectral shape differences in the frication noise (4-8kHz for [s], 2-4kHz for [ʃ]) and the spectral changes in formant transitions between the noise and the adjacent vowel (Harris, 1958; Heinz & Stevens, 1961; Hughes & Halle, 1956; Dorman, Raphael & Isenberg, 1980 for English; Guerlekian, 1981 for Spanish). In European Portuguese, frication noise has been described to have center frequencies around 5kHz for [s] and 3kHz for [ʃ] in the European variety (Lacerda, 1982; Jesus, 1999), and in Brazilian Portuguese (BP), Haupt (2008) and Santos (1987) described values around those of English: for [s] 4,5-7,4kHz, for [ʃ] 2-4,6kHz. Lacerda (1982) tested those frequencies for consonant recognition, but to the best of our knowledge formant frequencies were not investigated yet. We decided to test whether second formant (F2) transition said to be useful in English is also a cue to place of articulation for coronal fricatives in Brazilian Portuguese.

2. Methods

2.1 Subjects

Twelve female subjects with age varying from 14-28 years with no history of hearing problems participated in the study.

2.2 Stimuli

Four vowels were synthesized withHLSyn (Sensimetrics, Inc.), two tokens for [a] and two for [u]. One token of either vowel was manipulated in F2 transition in the first 50ms: one compatible with a transition from [s], the other from [ʃ] (Table 1). Formant values for the steady-state part were those presented by Escudero *et al.* (2009), and the F2 transition values, from Nittrouer and Miller (1997).

		[(s)a]	[(ʃ)a]	[(s)a]	[(ʃ)a]
F1	Initial	483	483	310	310
	Final	683	683	310	310
F2	Initial	1449	1769	1391	1711
	Final	1329	1329	761	761
F3	Initial	2624	2624	2609	2609
	Final	2324	2324	2309	2309

Table 1: Initial and Final Transition Formant values (Hz)

A 160ms raw frication noise with no filtering was synthesized using the Klatt cascade model implemented in Praat (Boersma & Weenink, 2011). Noise was subsequently single-pole filtered in different formant frequencies with a bandwidth of 230Hz. Noise filter frequencies were taken from a normally distributed, randomly generated 100 number sequence with mean = 4830 and sd = 50. With this procedure, each frication noise was not exactly the same, so that subjects would not get tired responding many times to one and the same stimulus, but at the same time the main effect would normally cluster around halfway between the center frequency of a [s] and that of a [ʃ]. The 100 different frication noises were then concatenated with the four vowel tokens (150ms) to produce 400 synthetic syllables of 310ms duration.

2.3 Procedure

The 200 tokens for [(s)a] and [(ʃ)a] were presented in a block, the 200 tokens for [(s)u] and [(ʃ)u] in another block within the same session. Subjects were allowed to have a break between blocks. A rating task was used (Macmillan & Creelman, 2005), in which subjects were required to listen to a syllable and respond whether they heard [(s)a]

or [(∫)a] in the first block, and [(s)u] or [(∫)u] in the second block, and to rate their confidence in their answers using a 3-point scale.

Data was collected in a quiet room. PercEval (LPL, CNRS/Université de Provence) in the BP version was used for sound presentation with a circumaural headphone and response entry[1].

3.   Results

Hit and false alarm rates of all response alternatives to [sa]-[∫a] and [su]-[∫u] were computed. Instead of d', a more common measure of sensitivity in classification task, we used the area under the ROC curve (A_z) produced by the cumulative d' for each response level in the 3-point scale. Slope and A_z were estimated with ROC-kit (Metz, Herman & Shen, 1998). So, the standard assumption of unit slope in d' is unnecessary. According to the results in the "slope" column in Tabs. 2 and 3, this would have been here a very strong assumption.

As A_z values range from .5 (no sensitivity, or confusion) to 1 (complete sensitivity), for [a] there is good evidence in our data that F2 transition is an important and sufficient cue to place of articulation in coronal fricatives (Table 2). Also, F2 transitions described for English, once adapted to the formant frequency values reported for BP, are useful for listeners to distinguish [sa] from [∫a].

Subject	n(∫|s)	Slope	A_z	St. Dev.
1st run				
ATR	100/99	2,227	0,976	0,011
MTR	97/92	2,888	0,811	0,035
PGF	100/99	0,679	0,965	0,013
RPA	98/99	0,299	0,839	0,043
2nd run				
ACA	100/99	0,274	0,977	0,016
CM	90/62	0,685	0,867	0,034
DOS	100/100	0,869	0,967	0,011
DSF	100/100	2,564	0,984	0,012
FAP	100/99	2,406	0,964	0,018
LM	100/100	1,131	0,644	0,041
SHA	100/99	1,46	0,95	0,016
TPA	100/100	1,04	0,846	0,029

Table 2: Results for the [∫a]-[sa] distinction

For [u], however, things were more confuse (Table 3). In the 1st run, A_z under .5 in 3 out of 4 subjects mean here that subjects made more false alarms than hits. We expected that the results would not be like those for the [a] tokens, since [u] is acoustically less clear. We then re-synthesized the [u] tokens with a longer steady-state part (270ms) and re-run the experiment. Subject PGF was re-tested a month after the 1st run. Then, 8 new subjects

were tested on both blocks. The longer [u] tokens resulted in better classification performance, but with results around .5 most subjects were barely sensitive to a difference between [su] and [∫u]. All subjects in the 2nd run were then pooled in a single set of results: [sa] and [∫a] seemed to be almost 60% as different as [su] and [∫u].

Subject	n(∫|s)	Slope	A_z	St. Dev.
1st run				
ATR	100/100	1,392	0,58	0,041
MTR	100/98	2,512	0,422	0,043
PGF	100/99	0,92	0,387	0,043
RPA	99/99	1,341	0,414	0,053
2nd run				
PGF	98/97	0,909	0,662	0,041
ACA	100/100	0,744	0,665	0,046
CM	69/66	1,11	0,529	0,054
DOS	100/100	0,72	0,559	0,044
DSF	100/100	1,168	0,544	0,044
FAP	99/100	1,04	0,943	0,017
LM	100/100	1,151	0,589	0,041
SHA	100/97	1,614	0,555	0,048
TPA	99/100	1,471	0,571	0,044

Table 3: Results for the [∫u]-[su] distinction

4.   Conclusion

For [a] there is good evidence in our data that F2 transition is an important and sufficient cue to place of articulation in coronal fricatives. However, listeners could not distinguish between [(s)u]-[(∫)u] only on the basis of F2 transition. This result points to a possible role of F3 transition, which were said to become important in rounded vowels. It will be a matter of future studies.

5.   Acknowledgements

6.   References

Boersma, P.,Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.2.40. Available at: <http://www.praat.org/>.

Dorman, M.F., Raphael, L.J. and Isenberg, D. (1980). Acoustic cues for a fricative-affricate contrast in word-final position. *Journal of Phonetics*, 8, pp. 397--405.

Gurlekian, J.A. (1981). Recognition of the Spanish fricatives /s/ and /f/. In *Journal of the Acoustical Society of America*, 70(6), pp. 1624--1627.

Harris, K.S. (1954). Cues for the identification of the fricatives of American English. In *Journal of the Acoustical Society of America*, 26, pp. 952 (A).

Haupt, C. (2008). As fricativas [s], [z], [S] e [Z] do Português Brasileiro. *Letras&Letras*, 24(1), pp. 59--71.

---

[1] The Brazilian version of PercEval (including manual) is available at http://www.letras.ufmg.br/perceval_BR/

Heinz, J.M, Stevens, K.N. (1961).On the properties of voiceless fricative constants. In *Journal of the Acoustical Society of America*, 33(5), pp. 589--596.

Hughes, G.W., Halle, M. (1956).Spectral properties of fricative consonants.*Journal of the Acoustical Society of America*, 28(2), pp. 303--310.

Jesus, L.M.T. (2000). Acoustic phonetics of European Portuguese fricative consonants. In *Technical Report*, Department of Eletronics and Computer Sciences, University of Southampton.

Lacerda, F.P. (1982). Acoustic perceptual study of the Portuguese voiceless fricatives. In Journal of Phonetics,10, pp. 11--22.

Macmillan, N.A.,Creelman, C.D. (2005). *Detection theory: A user's guide*. Erlbaum.

Metz, C.E., Herman, B.A. and Shen, J.H. (1998). Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. In *Statistics in Medicine*, 17(9), pp. 1033--1053.

Transplanting credibility into a foreign voice: an experiment on synthesized L2 Italian

Massimo PETTORINO, Anna DE MEO, Marilisa VITALE

University of Naples "L'Orientale"

Naples, Italy

mpettorino@unior.it, ademeo@unior.it, vitalem@unior.it

Abstract

This study intends to verify through perceptual tests conducted on original and artificially modified speech whether a relationship exists among the degree of comprehensibility of an utterance, the foreign accent and the credibility of the message. Four bizarre-but-true news read in Italian by four non-native speakers were artificially modified with Praat and WaveSurfer. Each piece of news was transplanted, so that segmental and prosodic features of a text read by a native speaker were transferred onto the same text uttered by a non-native speaker. The corpus was administered to 265 native Italian listeners, who were requested to indicate the degree of comprehensibility, the level of foreign accent and the truthfulness of each item. The results point out the existence of a close inverse relationship between comprehensibility and credibility. The presence of foreign accent, providing an impediment to the understanding of the message, tends to create an attitude of distrust in the listener. The most important features for the foreign accent reduction are the suprasegmental ones and, in particular, the durations of the phones and the pitch movement.

Keywords: foreign accent; comprehensibility; credibility; L2 Italian; prosody.

1. Introduction

Our recent study on socio-cultural effects of foreign accent on communication effectiveness (De Meo *et al*., 2012) revealed the relevance of comprehensibility factors - such as disfluency, frequency of silences, pitch range variation, silent pauses, segmental errors - on message credibility. A hundred seventy-five native Italian listeners, after hearing a set of 10 news uttered in Italian by one native speaker of Italian and four non-native speakers of L1 Chinese, Vietnamese, Arabic and Japanese, were asked to assess the comprehensibility, i. e. listener's estimation of difficulty in understanding an utterance (Munro & Derwing, 1999), and the truthfulness of each news item. The four non-native speakers, all late bilinguals with a basic (A2) and a mid (B1) level of competence as laid out in the Common European Framework of Reference, and an average stay time in Italy of 6 months, were chosen for the study after a global foreign accentedness rating test which was administered to 70 male and female native Italian listeners. Listeners rated the degree of foreign accentedness of a short read text on a 4-point scale (0 = native speaker; 3 = strong foreign accent). The results allowed to select four L2 speakers of Italian with a strong foreign accent.

Ten bizarre-but-true news from around the world read by the native speaker and the four non native speakers were presented to native listeners in form of radio news magazines, each combining the four voices reading different news, same news sequences but random voice order, pretending to administer a survey on media reliability, in order to avoid to focus the attention on foreign voices.

Obviously each single piece of news revealed to have its own degree of credibility, in accordance with the textual content of the message. However results showed that, within the same text, ratings were significantly different depending on the auditory comprehensibility level.

The study showed that when there are no comprehensibility problems the assessment of real/false is maintained around 50%, so in a range of randomness. On the contrary, when the level of comprehensibility lowers, due to various acoustic factors (disfluencies, errors, percentage of silence, tonal variation, etc.), the judgments of "false" increase rapidly, reaching 90% when the statement proves to be poorly understandable for the 40% of listeners. Therefore, there seems to be a threshold of comprehension tolerance, i.e. a level of difficulty in understanding an utterance at which the listener's effort to understand the message leads him to believe that what he has just heard is not credible.

Following this line of research, our current study intends to carry out a perceptual test on artificially modified speech, in order to evaluate the role played by both segmental and suprasegmental features in the achievement of an L2 comprehensible communication and find out if there is a relationship between the perceived degree of foreign accent and credibility.

2. Methods and materials

The corpus used for this study, taken from the one used in De Meo *et al*. (2012), consists of 4 news artificially modified with Praat (Boersma & Weenink, 2012). Each single piece of news was manipulated, so that disfluencies and errors were removed, and the prosodic features of the native speaker's utterances were transferred onto the same utterances produced by the non-native speakers (prosodic transplantation technique).

2.1 Corpus and Informants

1) Informants: 5 female voices
- 1 Italian speaker (L1)
- 4 L2 Italian speakers (Chinese, Vietnamese, Japanese, Arabic L1s)

2) Corpus: 18 audio files (bizarre-but-true news)
- 8 original news (4 L1, 4 L2)

- 10 L2 artificially modified (4 with removed disfluencies and cloned pauses, 2 with removed errors, 4 with cloned durations and pitch contour).

2.2   The transplantation technique

The rhythmic-prosodic transplantation technique is based on the algorithm PSOLA (pith-synchonous overlap-add, Moulines & Charpentier, 1990), implemented in Praat and illustrated in Yoon (2007) with regard to the English productions of Korean speakers. The prosodic features that can be transplanted from one voice to another are essentially four: the length of the segments, the pitch contour of the utterance, the intensity contour and the silent pauses.

The procedure of transplantation must follow a well-defined sequence of steps, since each of them is preparatory for the subsequent ones. The five phases are: anomalies treatment (disfluencies removal, pause cloning, errors elimination), segmentation and labelling, transplantation of the duration, intensity transplantation, pitch contour superimposition.

This technique seems to be a rather effective tool for the study of the spoken L2, since the manipulation of an utterance allows to evaluate the role played by individual acoustic parameters at the pragmatic-communicative level.

2.3   Perceptive test

The whole corpus was administered in a randomized order to 265 native listeners (male and female, mean age 21, university students) organized into 5 groups, so that nobody could listen to the same news more than once. As the purpose of the survey was to assess the credibility, the repeated exposure to a same input would have affected the reliability of the test results.

For each utterance, listeners were asked to evaluate the comprehensibility (poor, sufficient, good), assess the degree of perceived foreign accent (native accent, mild foreign accent, strong foreign accent) and judge on its truthfulness (true/false).

3.   Results and discussion

In this section we will examine the results of the abovementioned test, in order to evaluate the relevance of each manipulated factor on the perceptual level. The discussion will be organized into three parts corresponding to the different steps of the synthesis procedure. For the data analysis the One-Way ANOVA was performed.

3.1   First step: Removing disfluencies and cloning native silences

Figures 1, 2, 3 show the average percentage values of the judgements given to the utterances, both original and modified, produced by the native (NS) and the non-native speakers (NNS), with respect to comprehensibility, degree of foreign accent and credibility.
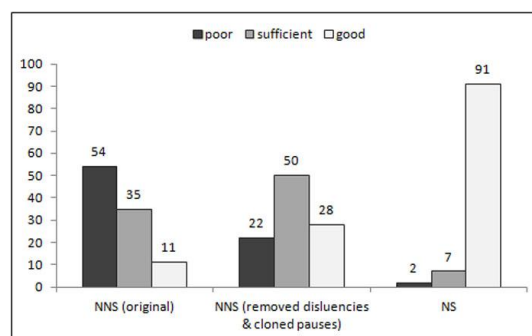
Figure 1: Average comprehensibility values (%) of the NS and the NNS

As for the foreign accent, both the NS and the NNSs were correctly recognized by almost all the listeners (Figure 2). The modifications carried out on NNSs' utterances produced a decrease of about 20% of the judgments of "strong foreign accent" (from 79% to 60%). In addition, it is worth noting that the 5% of the listeners assumed to have heard a native voice. Data are statistically significant (p<0.001).
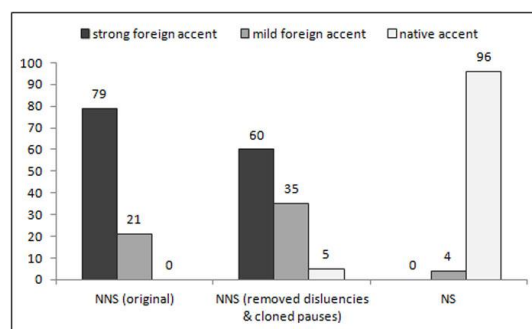
Figure 2: NS and NNSs' average percentage values of the foreign accent ratings
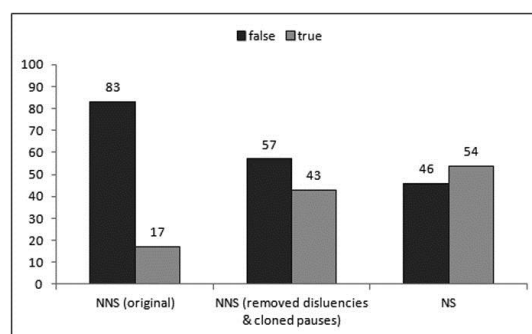
Figure 3: NS and NNSs' average percentage values of credibility

The removal of disfluencies and the repositioning of the silences determine a statistically significant improvement (p<0.001) of the NNSs' utterances comprehensibility (Figure 1). As a result of the manipulation, the majority of the listeners (78%) judged the non-native productions at least sufficiently comprehensible. Obviously, the NS proved to be highly comprehensible.

Removal of disfluencies and changes of silences (Figure 3) determined a significant increase (p<0.005) in the level of news credibility (+26%), taking the NNSs' values to levels very similar to those obtained by the native speaker.

3.2   Second step: Errors removal

For the second step of the study data are limited to the A2 level NNSs, since in the other speakers' productions there were no particular segmental irregularities. Using WaveSurfer, phones perceptually detected as wrong by three native trained phoneticians were artificially modified or substituted through a self-transplantation procedure, i.e. using adequate micro-segments produced by the speaker within the same utterance.

Because of the large variability and unpredictability of the errors, this phase is technically the most problematic. The typology and frequency of errors require operations that may damage the quality of the synthesized audio file and interfere with the perceptive evaluation.

Our data show that the segmental modifications give anyhow rise to a slight but significant improvement in terms of foreign accent assessment (p<0.005; "strong" from 77% to 66% and "mild" from 20% to 31%). No significant variations were observed for the comprehensibility and the credibility (p>0.05).

3.3   Third step: Duration and pitch transplantation

The final step of the transplantation procedure involved the cloning of the duration of each segment and, subsequently, the superimposition of the intonation contour from the NS's utterances to the NNSs' ones. The perceptive test outcomes are generally satisfactory. Figure 4, concerning comprehensibility, shows that, if compared to the first step of the procedure (disfluencies removal and silences cloning), the negative judgments decreased by 10% (from 23% to 12%) in favour of the "sufficient" ratings, while the "good comprehensibility" values did not change (p<0.05). It should be noted that the results of the overall transplantation process, when compared to the original utterances, reveal a remarkable improvement: the "poor comprehensibility" lowers by 42% and the "good comprehensibility" rises by 16% (p<0.001).

The most evident effects of this last step are those related to the degree of foreign accent (Figure 5), with a gain of about 30% for the judgment of "native" and a 60% reduction with regards to the judgment of "strong foreign accent" (p<0.001).

Finally, the values of credibility (Figure 6) do not undergo further significant variations (p>0.05).
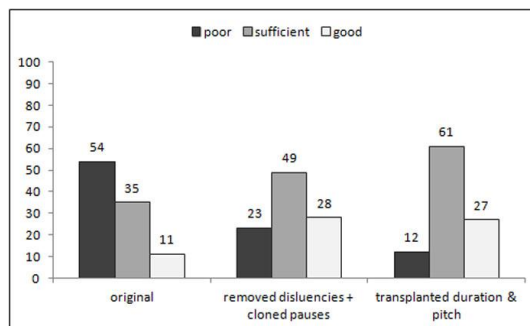
Figure 4: Average comprehensibility values (%) of the NNS
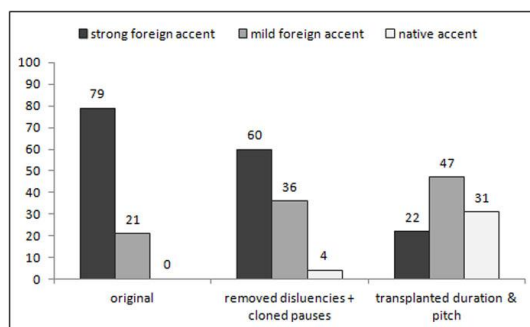
Figure 5: NNSs' average percentage values of the foreign accent ratings
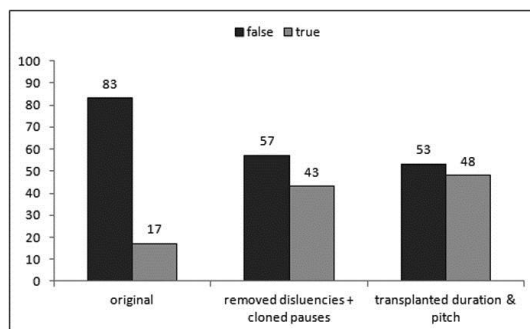
Figure 6: NNSs' average percentage values of credibility

4.   Conclusion

In conclusion, the study confirms the existence of a close relationship between comprehensibility and credibility, both for original and manipulated audio files (p<0.001). The more the utterance is easy to understand the more the listener is led to believe true what he/she has just heard. In this perspective, for our informants, beginners (A2) and low-intermediate (B1) speakers of L2 Italian, whose speech is characterized by disfluencies, anomalous silences, segmental errors, and inappropriate pitch contour, foreign accent provides an impediment to the understanding of the message and, consequently, tends to create an attitude of distrust in the listeners. However, it is worth to emphasize that it is not the "foreignness" as such to cause a lowering of credibility, but it is rather the difficulty of decoding the message determined by the presence of anomalies typical of an early L2 speech.

5.   References

Boersma, P., Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3.23, retrieved 7 August 2012 from <http://www.praat.org/>.

De Meo, A., Pettorino, M. and Vitale, M. (2012). Non ti credo: i correlati acustici della credibilità in italiano L2. In G. Bernini, C. Lavinio, A. Valentini and M. Voghera (Eds.), *Atti dell'XI Congresso dell'Associazione Italiana di Linguistica Applicata. Competenze e formazione linguistiche. In memoria di Monica Berretta.* Guerra Edizioni, Perugia, pp. 229--248.

De Meo, A., Vitale, M., Pettorino, M. and Martin, P. (2011). Acoustic-perceptual credibility correlates of news reading by native and Chinese speakers of Italian. In *Proceedings of the 17th International Congress of Phonetic Sciences.* Hong Kong, PRC, pp. 1366--1369.

Moulines, E., Charpentier, F. (1990). Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones. *SpeechCommunication*, 9, pp. 453--467.

Munro, M.J., Derwing, T.M. (1999). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. In *Language Learning*, 49(s1), pp. 285--310.

WaveSurfer 1.8.8. Available at : <www.spee.kth.se/wavesurfer/>.

Yoon, K. (2007). Imposing native speakers' prosody on non-native speakers' utterances: The technique of cloning prosody, *Journal of the Modern British & American Language & Literature*, 25(4), pp.197--215.

SPEECH PATHOLOGIES

Narrative perseverations in MCI patients

Luciana BRANDI, Stefania LUCCHESINI

DiLCo Laboratory - University of Florence
Via S. Reparata 93-95, 50129 Florence
luciana.brandi@unifi.it, stefania.lucchesini@unifi.it

Abstract

Narrative perseverations, defined as those repetitive verbal behaviours that appear to be intentional attempts at fully propositional utterances and narrative texts within conversations, are examined in six patients affected by Mild Cognitive Impairment (MCI). The role of the connections between the classical language areas is considered so to explain echolalic types of language productions.

Keywords: auto-echolalia; fronto-temporal degeneration; neurolinguistics.

1. Persistence in normal speech

Persistence of activation is a normal feature in the language processing system, and its effects are observable in the domain of speech production: at word level, in word-naming tasks, at phonological level, in speech errors, and at syntactic level (Levelt, 1989; Bock & Loebell, 1990; Dell *et al*., 1997a,b). Recurrent linguistic strings in spontaneous oral stories (Wray & Perkins, 2000) represents the modality to assure linguistic economicity and efficacy of the produced text and hence they constitute a sort of 'recitative speech' which makes communication easier. These perseverative effects in normal speech production may be related to the role and function of formulaic language in communication, seen as a blending of generative and formulaic sequences, each one resulting from the selected choice, for the speaker, of a holistic or an analytic processing strategy at any given moment. The very existence in normal subjects of such perseverative effects is of great consequence for interpreting the verbal perseverations produced by brain damaged patients, as in aphasia, mild cognitive impairment or dementia: the role of the impairment would not be to newly generate a protracted activation of previous utterances but only to disclose and abnormally maximize the shared verbal behaviour through the pathological form of overt perseverations.

2. Persistence in abnormal speech production

In the literature, recurrent perseveration is defined as the inappropriate occurrence of a previous response following the intervening presentation of a new stimulus within the context of a task set (Christman *et al*., 2004). Information processing models account for the phenomenon of recurrent perseveration as for the involuntary reactivation of an old memory trace in the context of a purposive attempt to respond to a new stimulus in a given task. Normally, memory traces retain a certain amount of post activation strength that either decay naturally over time or undergoes active cognitive inhibition. Hence inhibition failure can explain perseverations in the sense that once a response is produced, it is retained in working memory as an active trace that is subsequently available for rehearsal processes, with the effect that it disrupts the registration of new material in working memory and compromise search and retrieval from long term memory (Goldstein, 1948; Sandson & Albert, 1984; Cohen & Dehaene, 1998; Bayles et al. 1985, 2004; McNamara & Albert, 2004).

3. Narrative Perseverations

In the present study, narrative perseverations are defined as those repetitive verbal behaviours that appear to be intentional attempts at fully propositional utterances and narrative texts within conversations in patients with Mild Cognitive Impairment (MCI), a clinical construct that describes individuals with mildly impaired performance on objective neuropsychological tests but relatively intact global cognition and daily functioning (Petersen *et al*., 2001). MCI has been validated as qualitatively different from both normal aging and dementia (Petersen, 2004) and is a risk factor for the development of dementia (Smith *et al*., 2003). The invented recurrent utterances in recurrent texts recall the Verbatim texts in Becker's (1975) basic six category taxonomy of adult native speaker formulas, but their textual dimension and the temporal distance intervening between the recurrent texts make it questionable that the explanation would rest on priming effects and in general on a simple information processing or memory processing hypothesis (Brandi, 2011).

3.1 Methods and materials

The study was conducted in six patients (2 males and 4 females), aged from 70 to 78 years. They fulfilled the criteria for M.C.I. The data were collected as recorded spontaneous speech in familiar conversations (in the period 2009-2011): the corpus is filed at DiLCo Lab. The choice: given the curvilinear relation between severity of dementia, task type and frequency of perseveration, we decided to ecologically examine recurrent perseveration in spontaneous speech so that no task effect would be present.

3.2 Results

The corpus is characterized by the perseveration of quite extended narrative texts, that is extended linguistic sequences through which the patient is telling about an episode of his/her life. We show that perseveration does

not range over words or phrases alone but also over sentences and sequences of sentences, that is texts. Their main features are: the recurrent perseverations of narratives originate from later reiterations of the patient's own previous narrations, where previous means a) within the same conversational unit , hence in temporally near or concomitant stages; b) in different conversational units, temporally at a distance of days or even weeks.

With respect to perseverations in aphasia tests, narrative perseverations are not due to a problem of working memory because of the temporal span involved: infact they occur:

- in the same story text at distance of few minutes;
- in the same story text at distance of days;
- in the same story text at distance of weeks.

Ex:

Patient M. F.

01.06.2011
A. C.: Do you remember your sisters' name ?
M. F.: Nives and Nisarde, strange names, Nives is beautiful, but Nisarde is very ugly. I don't realize how my mother could choose a similar name. Maybe she read it, there isn't any other, in P. and V. and it is an ugly name, while Nives is a beautiful one. Nives, but it is difficult to pronounce because of its final s, and we don't stop at it, we say Nivesse.

09.07.2011
M. F.: Nisarde, I don't know, probably my mother read it, nobody in the valley has the same name, while Nives is difficult to pronounce, we never stop at the s, we say Nivesse.

16.07.2011
A. C.: And your sisters?
M. F.: Nives and Nisarde, a very strange name, I don't realize how my mother could choose it, there isn't any other in whole T., maybe she read it somewhere, Nisarde, is very ugly
S. L.: No, I like it, it has a nice sound
M. F.: No I didn't like it, while Nives is a beautiful name, we have to stop at the s but it's difficult, and many people said Nivesse

Patient P. M.

28.02.2011
[02.00] Of course I go, we were on shift together, but, you you I don't stay for dinner, so they prepare a little pizza for me and one of these boys bring me home.
[11.00] When we were on shift, then they stay for dinner, while I don't stay, no, they prepare a little

pizza for me and I go home, they prepare a little pizza and one of these boys bring me home.
[31.06] I was on shift with these boys, I was on shift with these boys, but they bring me home because I don't stay for dinner, they prepare for me something to eat and bring me home.

Long-term memory can be differently affected in MCI patients: as narrative perseverations show:

- semantic memory is fully spared;
- old episodic memory is spared;
- new episodic memory is partly affected.

The patient R.S. is able to give specific features and details in talking about the birth of his nephew, held eighteen years before, while C.B. may relate only generically on his very recent trip to USA:

24.03.2011
R. S.: It was wonderful when E. was born, we were at the Careggi, at the Mayer, no, my daughter was... no, at the Mayer, and she was under the doctor of the maternity ward, first we saw all the new-born babies over a trolley, then I said to my wife he is the one, and she replied how can you say that? And I: you will see. The others were dark with hairs, he was the only one blonde without hair.
A C.: Was he the one?
R. S.: When we went to see him I said to M.P.: Was he the one? and she: yes, you were right. And even if think back I was right, he was the one.

Patient C. B.

13.10.2011
C. B.: Colorado is beautiful, beautiful, beautiful, beautiful... Go there if you can.... Every kind of animal, there're there're... they pass between the cars, last july my friend from Piombino called me: Do we go to Colorado? Colorado? right, we do go to Colorado. We were fourteen people in a minibus, we saw something.... So, beautiful, natural, I am surprised that Americans left a place in that way, so... natural.

As the process involved is an inability to inhibit the iterate repetition of one's own previous productions, even as external stimuli change, the proper term would be auto-echolalia – i.e. the accurate reproduction of his own/her own previous uttered texts:

A neurolinguistic model is required, linking the observed linguistic behaviour to inferred dysfunctions within distributed neural networks. Narrative perseverations may be explained as changes in functional brain integration due to progressive white matter loss.

The perisylvian network for language involved

mirrors the language territories for echolalic autism. Following the analysis from Catani and ffytche (2005) for the arcuate fasciculus, the connections between the classical language territories, that is Broca's and Wernicke's area, show a more complex structure, adding to the known direct pathway two indirect ones. Specifically, the indirect pathway appears to relate to semantically based language functions (such as auditory comprehension and vocalization of semantic content), whereas the direct pathway relates to phonologically based language functions (such as automatic repetition).

This is not to say that these functions are restricted to perisylvian areas, but merely that within the parallel pathways we describe, the two functions are anatomically dissociable (Brandi, 2005; Lucchesini, 2010).

Given that the evolution/devolution of the blending between echolalic and creative language strongly correlates to the neural processes of connectivity and lateralization involving the arcuate fasciculus, the occurrence of auto-echolalic perseverative language in the speech of M.C.I. patients and of echolalic speech in children with autism could be traced to the same assumptions. If the echolalic speech of the autistic child has to be due to lack of maturational processes in neural connectivity its features could be related to hyperfunction in the direct pathway connecting Broca's and Wernicke's territories (Catani & ffytche, 2005). Perseveration in M.C.I. can be seen as a sort of auto-echolalia equally descending from loss of neural connectivity within the same language territories.

4.   Acknowledgements

5.   References

Bayles, K.A., Tomoeda, C.K., Kaszmak, A.W., Stern, L.Z. and Eagans, K.K. (1985). Verbal perseverations in dementia patients. *Brain and Language*, 25, pp. 106--116.

Bayles, K.A., Tomoeda, C.K., McKnight, P.E., Helm-Eastbrooks, N. and Hawley, J.N. (2004). Verbal perseverations in individual with Alzheimer's disease. *Seminars in Speech and Language*, 25, pp. 335--347.

Becker, J. (1975). The phrasal lexicon. In *Proceedings of the 1975 workshop on Theoretical issues in natural language processing*. Cambridge, MA: Association for Computational Linguistics, pp. 60--63.

Bock, J.K., Loebell, H. (1990). Framing sentences. *Cognition*, 35, pp. 1--39.

Brandi, L. (2005). Linguaggio e comunicazione: dis/giunzioni autistiche. *Quaderni del Dipartimento di Linguistica, Università degli Studi di Firenze*, 15, pp. 169--192

Brandi, L. (2011). Il bisogno di raccontarsi. Perseverazioni narrative nel MCI. *Quaderni del Dipartimento di Linguistica, Università degli Studi di Firenze*, 21, pp. 195--217.

Catani, M., ffytche, D.H. (2005). The rises and falls of disconnection syndrome. *Brain*, 128, pp. 2224--223.

Christman, S.S., Boutsen, F.R. and Buckingam, H.W. (2004). Perseverations and other repetitive verbal behavior: functional dissociations. *Seminars in Speech and language*. 25: pp. 295--308.

Cohen, L., Dehaene, S. (1998). Competition between past and present. Assessment and interpretation of verbal perseverations. *Brain*, 121, pp. 1641--1659.

Dell, G.S., Burger, L.K., Svec, W.R. (1997a). Language production and serial order: a functional analysis and a model. *Psychological Review*, 104, pp. 123--147.

Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M. and Gagnon, D.A. (1997b). Lexical access in aphasic and non-aphasic speakers. *Psychological Review*, 104, pp. 801--838.

Goldstein, K. (1948). *Language and language disturbances*. New York: Grune and Strutton.

Levelt, W.J. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Lucchesini, S. (2010). Ripensare l'ecolalia, nel disturbo autistico e nello sviluppo regolare del linguaggio. Tesi di dottorato in Linguistica. Università degli Studi di Firenze.

McNamara, P., Albert, M.L. (2004). Neuropharmacology of verbal perseverations. *Seminars in Speech and Language*, 25, pp. 309--322.

Petersen, R.C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256, pp. 183--194.

Petersen, R.C, Doody, R., Kurz, A., Mohs, R.C., Morris, J.C., Rabins, P.V., Ritchie, K., Rossor, M., Thal, L. and Winblad, B. (2001). Current Concepts in Mild Cognitive Impairment. *Archives of Neurology*, 58, pp. 1985--1992.

Sandson, J., Albert M.L. (1984). Varieties of perseveration. *Neuropsychologia*, 22(.6), pp. 715--732.

Smith, G. E., Ivnik, R.J. (2003). Normative neuropsychology. In R. C. Petersen (Eds.), *Mild Cognitive Impairment: Aging to Alzheimer's Disease*. New York: Oxford University Press, pp. 63--88.

Wray, A., Perkins, M.R. (2000). The functions of formulaic language: an integrated model, *Language and Communication*, 20, pp. 1--28.

A interação conversacional entre afásicos e não afásicos

Érica de Araújo COUTO, César REIS

Laboratório de Fonética
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6.627 - 31270-901 - Belo Horizonte - Brasil
ecoutobhz@gmail.com, creisufmg@ufmg.com

Abstract

Em uma situação comunicativa o interlocutor não afásico interpreta e dá significado aos segmentos estereotipados do sujeito afásico, a partir da variação da entonação e de outras formas de expressão como os gestos de apontar, mímica facial e a escrita. O interlocutor não afásico utiliza ainda outras estratégias comunicativas como perguntas e afirmações, provocando a concordância ou não do sujeito afásico, tornando possível a comunicação e a interação social. Buscando compreender a interlocução entre afásicos e não afásicos na estereotipia verbal, delineou-se um estudo onde foram realizadas entrevistas semi estruturadas com os familiares não afásicos de 4 indivíduos afásicos que utilizam a estereotipia não lexical e gestos como forma de expressão. O grau de parentesco é cônjuge (3) e irmã (1) e que foram convidados a participarem do estudo considerando o contato diário com o afásico em atividades cotidianas, acompanhamento a médicos e atividades de lazer. Como conclusão observa-se que: o interlocutor não afásico se sente como um tradutor da expressão do afásico; a variação da entonação é importante mas não o suficiente para uma comunicação efetiva; o contexto e familiaridade são essenciais e finalmente relatam uma dificuldade na compreensão de uma informação nova fornecida pelo afásico.

Palavra-chave: afasia; comunicação; gestos; escrita.

1. Introdução

Durante uma interação comunicativa entre sujeitos afásicos e não afásicos observa-se que o interlocutor não afásico interpreta e dá significado aos segmentos estereotipados do sujeito afásico, a partir da variação da entonação e de outras formas de expressão como os gestos de apontar, a mímica facial, a escrita, o desenho. O interlocutor não afásico utiliza ainda outras estratégias comunicativas como perguntas e afirmações provocando a concordância ou não do sujeito afásico tornando possível a comunicação e a interação entre afásicos e não afásicos.

A estereotipia verbal é uma alteração da expressão oral em afásicos caracterizado pela emissão de segmentos sonoros que são automaticamente repetidos todas as vezes que o indivíduo tenta se comunicar. As estereotipias verbais se dividem em *não lexicais*, constituídas de uma sequência de fonemas, palavras sem significado e emissões ininteligíveis; e *lexicais*, constituídos de palavras com significado, frases e partículas sim/não. Muitas vezes as estereotipias verbais não lexicais são compostas de sílabas com estruturas simples como consoante-vogal (CV) ou consoante-vogal-consoante (CVC).

Uma das características mais marcantes da estereotipia é a entonação. As estereotipias parecem interagir com a entonação e com elementos do contexto, possibilitando uma interpretação parcial, senão total, do enunciado em uma situação específica de fala. Na ausência de elementos sintáticos e semânticos significativos e associados a habilidades pragmáticas, a prosódia possibilita a manutenção de habilidades comunicativas como a alternância de papéis na conversação.

Um número de afasiologistas tem expressado a visão de que pacientes com estereotipia podem utilizar sua entonação para transmitir significado: eles habilidosamente modulam sua estereotipia para expressar necessidades, pensamentos, sentimentos (Lebrun, 1993) A observação clínica indica que os indivíduos afásicos produzem uma expressão fluente com variações de entonação, com a intenção de transmitir informação comunicativa. Code (1994) também aponta para o fato de que, na prática clínica, o individuo parece manter habilidades pragmáticas como a alternância de papéis na conversação, o que torna a interação possível, apesar da ausência de elementos sintáticos e semânticos.

Outros estudiosos, no entanto (Pell & Baum, 1997; Bleser & Poeck, 1985), apontam para o fato de que afásicos com estereotipia com alto grau de severidade apresentam um baixo desempenho nas tarefas de compreensão oral, o que não lhes permitiria desenvolver e exercer um controle cognitivo sobre suas emissões. Em uma situação de conversação com um interlocutor não afásico, muito provavelmente este interlocutor irá se adaptar ao baixo nível de informação transmitido e, com o auxílio de certo grau de compreensão verbal e não verbal e estratégias não verbais, interprete a resposta do parceiro afásico utilizando a variação da prosódia como adequada. Os estudos mencionados acima também apontam para a existência de uma troca de turnos conversacionais nesses pacientes, tornando possível a interação conversacional apesar da ausência de elementos semânticos e sintático.

2. Método

Buscando compreender a interlocução entre afásicos e não afásicos na estereotipia verbal, delineou-se um estudo onde foram realizadas entrevistas semi estruturadas com os familiares não afásicos de 4 indivíduos afásicos que tem a estereotipia não lexical e a utilização de gestos como forma de expressão. O grau de parentesco é cônjuge (3) e irmã (1) e que foram convidados a participarem do estudo considerando o contato diário com o afásico, em atividades do cotidiano, acompanhamento a médicos e atividades de lazer. Os familiares eram entrevistados pela pesquisadora e as entrevistas gravadas. Em algumas perguntas eram dadas alternativas caso o entrevistado demonstrasse alguma incerteza ou incompreensão. Uma análise descritiva das respostas dos entrevistados foi realizada e os resultados encontrados foram divididos em

estratégias comunicativas utilizadas pelos afásicos e estratégias comunicativas utilizadas pelos não afásicos.

> 1) Como você se comunica com o seu familiar?
>   (1) fala e faz gesto.
>   (2) fala somente .
>   (3) faz gesto somente.
> 2) Como ele se comunica com você?
>   (1) fala e faz gesto.
>   (2) fala somente.
>   (3) faz gesto somente.
> 3) Quando vocês estão em algum lugar que alguém se aproxima e inicia a conversação:
>   (1) você deixa ele responder mesmo com
>   dificuldade.
>   (2) responde por ele.
>   (3) explica para a pessoa que ele tem
>   dificuldades para falar e então responde
>   por êle.
> 4) Você entende o que o seu familiar fala ou tenta adivinhar ? perguntando ou fazendo algum gesto ou ação?
> 5) Seu familiar afásico:
>   (1) inicia a conversa.
>   (2) espera você iniciar.
>   (3) mantém a conversa mesmo com dificuldade.
>   (4) encerra a conversa se alguém não entende
>   (5) fica nervoso, com raiva, tímido triste.
> 6) Você acha que ele sabe que está falando de forma diferente? Aparenta ter vergonha do jeito que fala?
> 7) Que atividades o seu familiar faz sozinho?

Quadro 1: Roteiro da Entrevista

3.   Resultados

As estratégias comunicativas utilizadas pelos afásicos, descritas pelos interlocutores foram: estereotipias verbais onde a variação da entonação se destaca; gestos de apontar; mímica facial; gestos de ação e da forma do objeto; escrita e desenho. Relatam que os afásicos não iniciam a conversação; que quando a compreensão da expressão do afásico por parte do interlocutor torna-se difícil há o abandono do processo comunicativo; que os afásicos não se utilizam de modalidade (prosódia linguística), mas apresentam a prosódia afetiva. Quanto ao interlocutor não afásico este se utiliza das seguintes estratégias em sua comunicação com o afásico: combinação de diferentes formas de comunicação como gestos, mímica facial e escrita, juntamente com a variação da entonação e apoio do contexto; utilização de estratégia comunicativa do tipo "hint and guess",onde o interlocutor não afásico sugere e ou adivinha o significado do enunciado e o afásico confirma ou não.

4.   Conclusão

Como conclusão que em uma situação comunicativa onde a estereotipia verbal é a forma de expressão oral do afásico, o interlocutor não afásico assume o papel de um tradutor desta forma de expressão. A variação da entonação é importante, mas não o suficiente para uma comunicação efetiva sendo o contexto e familiaridade com o tópico da conversação essenciais para uma boa compreensão do que é expresso através da estereotipia verbal. E finalmente, os interlocutores não afásicos relatam muita dificuldade em compreende uma informação nova dada pelo afásico.

5.   Referências

Bleser, R., Poeck, K. (1985). Analysis of Prosody in the Spontaneous Speech of Patients with CV-Recurring Utterances. *Cortex*, n.21, p.405-416.

Code, C. (1989). Speech Automatisms And Recurring Utterances. In: Code, C. *The Characteristics of Aphasia*. London: Ed. Taylor and Francis.

Basso, A. (2003). Severe Aphasia and Pragmatics. In: Basso, A. *Aphasia and its Therapy*. Ed. Oxford University Press.

Laakso, M., Klippi, A. (1999). A Closer Look at the "hint and guess" Sequences in Aphasic Conversation. In *Aphasiology*, 13, 4, pp. 345--363.

Prosody and communicative function in verbal stereotypy of aphasic individuals

Érica de Araújo COUTO, César REIS

Phonetics Laboratory - Universidade Federal de Minas Gerais, Belo Horizonte, Brasil
Av. Antônio Carlos, 6627 - Belo Horizonte/MG - Brasil
ecoutobhz@gmail.com, creisufmg@ufmg.com

Abstract

Prosodic aspects in aphasic adults were assessed to gain insight into aphasic verbal stereotypy in Portuguese-speaking subjects. We employed language tasks with repetition and naming to test the hypothesis that aphasic individuals, who use stereotypies as a form of expression, appropriately use prosodic features to communicate effectively. Our results suggest that there is a strong individual component in the development of stereotypy at both the segmental and prosodic level. The intonation pattern of studied aphasic individuals did not match the expected intonation pattern of normal speech, and their acoustic parameters showed variability with highly specific characteristics. We suggest the existence of stereotyped prosody in aphasics that results from automatic processing and a lack of cognitive control and communicative intent.

Keywords: aphasia; prosody; verbal stereotypy; communication.

1. Introduction

In speech and language disorders of acquired neurological origin, such as aphasia, there is a variety of changes in oral and written language skills. These can involve both understanding and expression and are due to dysfunction in specific brain regions. One oral expression disorder that has captured the attention of clinicians and researchers is the emission of sound segments that are automatically repeated every time the individual attempts to communicate. These sound segments, also called "recurring utterances," "permanent verbal stereotypies," and "speech automatisms," differ considerably from patient to patient and may occur for days, weeks, months, or even years.

One of the most striking features of verbal stereotypy is intonation. Stereotypies seem to interact with intonation and context elements, enabling the partial, if not total, interpretation of a statement in a specific speech situation. In the absence of syntactic and semantic elements associated with meaningful and pragmatic abilities, prosody enables the maintenance of communication skills, such as alternating roles in conversation.

In an attempt to gain greater insight into this feature of stereotypy, we sought to answer the following questions: which communicative role has the intonation in the speech of aphasic individual? Is it a communication strategy developed by aphasic individuals or is it the product of automated processing as published studies suggest? Is this variation intentional, i.e., is it emitted by aphasic individuals for conveying meaning? Or, is the non-aphasic listener inferring meaning from the prosodic issuance variations by aphasic individuals?

According to Code (1989), verbal stereotypies are divided into non-lexical, consisting of a sequence of phonemes; nonsense words and unintelligible emissions; and lexical, consisting of meaningful words, phrases, and yes/no particles. Often, non-lexical verbal stereotypies are composed of syllables with simple structures, such as consonant-vowel (CV) or consonant-vowel-consonant (CVC). Stereotypies are always pronounced in the same way, but they may have temporary phonetic variations. They are produced easily, smoothly, and without apparent effort for an indefinite period, predominantly as a verbal expression of the individual or, in some cases, as their only form of expression. Each aphasic individual has a limited, individual repertoire of verbal segments with specific variations in frequency, intensity, and pace, but we cannot affirm whether these prosodic features contain any meaning, e.g., rising and falling intonations to distinguish questions and assertions.

A number of aphasiology professionals have expressed the view that patients with stereotypy can use their intonations to convey meaning; they skillfully modulate stereotypy to express their needs, thoughts, and feelings (Lebrun, 1993). Others, however, such as Code (1994) observe that changing intonation is possible for some patients, but this does not follow the intonation pattern proposed for non- pathological speech. We believe that non-stereotypic individuals have unique intonation patterns, and patients that cannot vary intonation make changes at the level of arrangement.

With the aim of increasing understanding of prosodic functioning in aphasic patients, especially those with stereotypies, we sought to confirm the hypothesis that aphasic individual who uses stereotypes as a form of expression make appropriate use of prosodic resources to communicate effectively.

2. Methods

We assessed non-lexical stereotypies with linguistic tasks that enabled the collection of data and subsequent quantitative and qualitative analysis. The tasks chosen were repetition and confrontation naming where the participants were presented with a picture and were asked to name the object or its function. Repetition implies that the subject's processes of encoding and decoding segmental and non-segmental aspects of speech are preserved, and we expected them to reproduce different intonations of utterances. Additionally, repetition allows greater control over utterance duration and the number of pauses, syllables, and accents. During the repetition task,

the content to be repeated was comprised of 6 short and long illocutionary speech acts, assertion (I am very tired/I am very tired because these bags are too heavy), question (Do you want to dance with me?/Do you want to dance with me this last song?), and order (Get out! Out!/Get out of here! Get out; you're making a mess in here). The phrases were chosen based on modality and length.

The naming task was designed to induce a more spontaneous emission and allow us to assess whether the subjects' responses resembled the control word with regard to duration, number of syllables, reproduction of the word's emphasis, and prosodic organization. The stimuli consisted of 11 words with different numbers of syllables and accents.

WORD	CLASSIFICATION
Trem	monosyllable, with accent on the whole word
Vaca	dissyllable, with accent on the penultimate syllable
Boné	dissyllable, with accent on the last syllable
Ônibus	trisyllable, with accent on the antepenultimate syllable
Cadeira	trisyllable, with accent on the penultimate syllable
Macarrão	trisyllable, with accent on the last syllable
Professora	four syllables, with accent on the penultimate syllable
Helicóptero	five syllables, with accent on the antepenultimate syllable
Escorregador	six syllables, with accent on the last syllable
Aspargo	trisyllable, with accent on the penultimate syllable
Ábaco	trisyllable, with accent on the antepenultimate syllable
Esfinge	trisyllable, with accent on the penultimate syllable

Table 1: Words used in Nomination Task

The experimental group (EG) was comprised of 8 aphasic patients (5 males and 3 females) with chronic global aphasia and non-lexical verbal stereotypy who were able to understand the linguistic tasks. The control group (CG) consisted of 4 subjects (3 females and 1 male) without language impairment.

Utterances of both groups were analyzed using the Pratt computer program for acoustic analysis; CG assessment was based on Halliday's Theoretical Model of Intonation (1970), and the aphasic group analysis was based on the findings of Rizzo (1981) and his claim that intonation plays an important role in speech acts.In our study, we found that aphasic individuals, who had control over prosodic parameters, were able to use them appropriately in different communication situations. The utterances of both groups were described in detail and descriptively analyzed at the segmental and prosodic levels.For each utterance an orthographic transcription, an orthographic transcription adapted to the pronunciation, and a phonetic transcription, was made, as demonstrated below:

C – I am very tired

T.O – "Odi ôôôôôri"

T.F – [o'dʒi] [o: ɾi]

Figure 1: Example of utterancy transcription

We also generated a table with the analyzed prosodic parameters and their values. The following parameters were measured, and their experimental values were compared to the control values: duration, number of pauses (in the repetition task), maximum F0, minimum F0, tessitura, initial F0, final F0, and intensity.

S	D	P	FO Max	FO Mn	Tes	F0 In	F0 fin	I
C	1.964	0	210	108	102	90	143	42
G	2.730	1	156	65	91	125	65	72

Table 2: Repetition Task–Assertion[1]

Figure 2: Praat screen following an utterance

We note that the term word was defined as the segmented speech sequence consisting of one or more

---

[1] The meaning of the labels are as follows: S- subject; D- duration; P- pause; Fo Max- maximum fundamental frequency; Fo Min- minimum fundamental frequency; T- tessitura; Fo In- initial fundamental frequency; Fo Fin-final fundamental frequency; I- intensity

syllables that was preceded and followed by a pause. In the case of stereotypies, these show a basic repeated structure, e.g., V, CV, and CCV. Utterances were defined as a process for spoken enunciation between 2 pauses; the term was used to refer to speech production that had a sound sequence between pauses greater than 0.168 seconds. This length was selected because it is the shortest duration between words.

3. Results

For the repetition task the length of utterances of EG were sometimes higher and sometimes lower than the utterance duration of CG, leading to the conclusion that knowledge about the physical size of the utterance is not preserved. We found that the tessitura pattern varies both between individuals and between utterances of different sizes and modalities. Considering the parameter F0, one can say that stereotypies have a standard falling intonation in all utterances, regardless of the modality. The intonation pattern presented is unique to each individual and can be considered stereotypical. With respect to intensity, the subjects studied showed an upward-descending curve, which is considered standard for normal speech. It was difficult to fit the pace of stereotypy within patterns of accentual and syllabic rhythm. Most of the time, what we observed was the production of sequences of syllables that we refer to as syllabic pace.

In general, the nomination task was similar to the repetition task. The intonation pattern was mostly ascending in the first syllables and descending in the last ones. The duration of the utterance remained long, with values much higher than the target words. Word organization, with reference to the number of syllables, was not observed, and there was no correspondence between the target word and utterance.

4. Conclusions

Our hypothesis that aphasic individuals who use stereotypes as a form of expression make appropriate use of prosodic features to communicate effectively was not confirmed. The evidence indicates that stereotypies are strongly influenced by automatic processing, without the interference of a cognitive control suggested by Bleser & Poeck (1985) and Blanken, Wallesh, and Pagano (1990) and that only through the development of this control would it be possible to reverse the stereotypy. The data also suggest that there is a strong individual component in the development of stereotypy at both the segmental and prosodic levels. The intonation pattern presented by the studied aphasic individuals does not match the expected intonation pattern of normal speech, and the acoustic parameters show variability with very specific characteristics. The results point to a stereotyped prosody, i.e., resulting from automatic processing, limited in repertoire, and without the interference of cognitive control and communicative intent. However, we consider that the prosody in stereotypic speech may contribute to the dialogue by providing clues about the information to the non-aphasic listener, who together with other forms of language, such as gestures, facial movements, and discursive resources, interprets and infers meaning. Understanding the nature of the structure of linguistic behavior in its segmental and non-segmental aspects can provide us with valuable information about the condition of language as a system and about its restructuring and adaptation processes.

5. References

Bleser, R., Poeck, K. (1985). Analysis of Prosody in the Spontaneous Speech of Patients with CV-Recurring Utterances. In *Cortex*, 21, pp. 405--416.

Blanken, G., Marini, V. (1997). Where Do Lexical Speech Automatisms Come From? In *Journal of Neurolinguistics* 10 (1), pp. 19--33.

Blanken, G. (1991). The Functional Basis of Speech Automatisms (Recurring Utterances). In *Aphasiology*, (5) 2, pp. 103--127.

Code, C. (1989). Speech Automatisms and Recurring Utterances. In Code, C. (Ed.), *The Characteristics of Aphasia.* London: Ed. Taylor and Francis.

Code, C. (1994). Speech Automatism Production in Aphasia. *Journal of Neurolinguística*, (8) 2, p. 135--148.

Lebrun, Y. (1993). Repetitive Phenomena In Aphasia. In Blanken *et al.* (Eds.), *Linguistics Disorders and Pathologies: An International Handbook*, Berlim; Walterde Gruyte.

Gandour, J. Petty, S.H. and Dardarananda, R. (1988). Perception and Production of Tone in Aphasia. Im *Brain and Language*, 35, p. 201--240.

Halliday, M.A.K. (1970). *A Course in Spoken English: Intonation*. London: Oxford University Press.

Rizzo, J.F.P. (1981). O Papel da Entonação do Português Brasileiro na Descrição de Atos de Fala. Master thesis. Campinas: Department of Linguistics, Institute of Language Studies, Universidade Estadual de Campinas.

Blanken, G., Wallesh, W. and Pagano, C. (1990). Dissociation of Language Functions in Aphasics with Speech Automatisms (Recurring Utterances). In *Cortex*, 26, pp.41-63.

Análise de processos multimodais na interação multipartilhada entre afásicos e não afásicos

Heloísa MACEDO, Ana Lucia TUBERO, Sandra CAZELATO

COGITES, IEL, UNICAMP

R. Sérgio Buarque de Holanda, 571, 13083-859 - Campinas - SP - Brasil

sandracazelato@afasia.com.br; heloisamacedo@afasia.com.br; anatubero@afasia.com.br

Abstract

This study investigates the relationship between language and cognition and discusses the importance of multimodal processes in the construction of meaning in multiparty interaction between aphasic and non-aphasic participants in the Centro de Convivência de Afásicos (CCA – IEL, UNICAMP). Aphasics exhibit impairments in language, as much in terms of expression as understanding, as a result of a brain injury. Nevertheless, these impairments do not mean that the aphasic isn't able to interact linguistically in the construction of meaning. From a multimodal perspective, speaking and writing are not the only ways that bring relevance to interaction. Gestures, gaze, voice, prosody, facial expression, mime, head and hand movements, posture, distribution of persons within a space of interaction and the context of interlocution characterize themselves as other modes that are brought into action and co-occur the other aspects of language reference in the construction of meaning, dislocating language as the most relevant mode in the *continuum* proposed by Norris (2006). Analyzing data from the interaction of aphasics and non-aphasics from a socio-cognitive perspective with a textual-interactive base, we are seeking to build a sufficiently accurate *corpus* to give heightened visibility to the co-occurrence between verbal and nonverbal processes in the construction of meaning.

Keywords: multimodality; aphasia; multiparty interaction.

1. Introdução

Este trabalho se inscreve na agenda atual de questões teórico-metodológicas relativas ao campo de estudos neurolinguísticos – que investiga as relações entre linguagem, cérebro e cognição em contextos normais e patológicos – e, especificamente, discute a relevância de processos multimodais na construção da significação na interação multipartilhada entre afásicos e não afásicos. Nossos dados – verbais e não verbais – compõem um *corpus* bastante especial uma vez que trabalhamos com linguagem afásica em práticas de grupo, o que impõe um rigor metodológico na coleta, constituição, transcrição e análise dos dados.

2. Objetivo e justificativa teórica

Tendo como objetivo discutir as questões metodológicas acionadas na composição desse *corpus* específico, focamos nosso trabalho nos processos multimodais interatuantes na comunicação de afásicos e não afásicos.

A afasia se traduz em alterações da linguagem oral e escrita, tanto em relação à expressão quanto à compreensão, o que não significa que o afásico não possa interagir linguisticamente na construção da significação. No caso das afasias, o sujeito costumeiramente enfrenta no campo mesmo da linguagem dificuldades metalinguísticas (reparos, reformulações, prosódia, repetições, hesitações, *promptings* orais do interlocutor, *etc*.) e, além disso, lança mão de semioses não verbais (como gestos, direcionamento do olhar, postura corporal, *etc*.) que atuam de maneira solidária à linguagem na configuração ou na interpretação da referência.

De acordo com Norris (2006), a multimodalidade, em uma perspectiva discursiva e interacional, implica a noção de mediação semiótica (de inspiração bakhtiniana e vygotskiana), de densidade modal "*that makes up a specific higher-level action*" (Norris, 2006: 402) e de continuidade entre figura e fundo nas atividades de atenção e conhecimento (*op.cit.*: 401) – que, reunidas, poderiam ser identificadas como o que tem sido chamado de contexto em perspectivas textuais-interativas (*cf*. Koch, 2002). Ainda que não se oponha ao papel relevante reivindicado para a linguagem na constituição das interações pelos estudos da conversação e do discurso, Norris chama a atenção para o caráter semioticamente plural da comunicação.

Portanto, a adoção de uma abordagem multimodal da linguagem não implica apenas admitir que os processos linguísticos estão ligados a recursos semióticos, mas sim e sobretudo que estes seriam desprovidos de sentido se fossem tomados de maneira descontextualizada e alheia às rotinas ou práticas simbólica e socialmente significativas.

Consideramos, então, que a linguagem verbal não é necessariamente o único modo que carrega a relevância na interação (Norris, 2006). Fala e escrita são modos de linguagem verbal, mas também os gestos (dêiticos, icônicos, metafóricos), o olhar, a voz (risadas, ruídos, entonação), a prosódia, a expressão e a mímica faciais, os movimentos da cabeça e das mãos, a postura, as posições das pessoas em relação umas às outras, a distribuição das pessoas no espaço da interação (Mondada em 2008, por exemplo, apontou a importância da disposição dos corpos no espaço para a criação de um território de interlocução) e o contexto da interlocução se caracterizam como outros modos que são mobilizados e coocorrem com os demais aspectos referenciais da linguagem na construção do sentido (Norris, 2006; Mondada e Markaki, 2006; **Holler e Beattie, 2006**). A abordagem multimodal permite dar visibilidade a estes outros modos também relevantes para a significação, seja em contextos patológicos ou normais, em interações específicas.

Ao observarmos, nos encontros do Centro de Convivência de Afásicos (CCA – IEL/UNICAMP) e,

portanto, no contexto de interação entre afásicos e não afásicos, a ocorrência de diferentes semioses configurando diferentes processos multimodais é possível afirmar que tanto os afásicos quanto os não afásicos lançam mão de vários processos multimodais, de maneira conjugada ou não à própria fala ou à de seu interlocutor, na busca de uma melhor construção do elemento referencial.

Em nosso *corpus* – caracterizado pela ocorrência e coocorrência de processos multimodais – selecionamos alguns episódios cuja análise permite observar que os processos multimodais vão desde os gestos mais estandartizados, formulaicos, como os gestos dêiticos e apontamentos que se conjugam – ou não – com a fala (aqui, lá) ou os movimentos de cabeça indicando negação, até gestos elaborados (icônicos, pantomímicos e metafóricos) com tal completude de sentido que tornam desnecessárias as palavras (mesmo que essas sejam ditas por um outro – o interlocutor). Além disso, a análise dos dados permite considerar os aspectos entoacionais, as posições ocupadas pelos interlocutores no espaço da enunciação, o direcionamento do olhar, dentre outros elementos multimodais, como reconhecidamente parte da cena enunciativa. A decisão metodológica por uma abordagem multimodal do *corpus* levou-nos a considerar tantos modos quantos necessários para mostrar a coreografia das interações entre afásicos e não afásicos.

Os diferentes processos multimodais que participam na construção de objetos de discurso mostram-se altamente frequentes e presentes, sendo mesmo fundamentais na compreensão da significação pretendida, na manutenção do tópico discursivo, na introdução de novo tópico, na tomada de turno, nos processos de referenciação e de inferenciação, mas não por isso devem ser tomados como compensatórios, estratégicos ou simplesmente complementares das dificuldades linguísticas dos afásicos. Neste contexto, questionamos a noção que descreve os processos multimodais como elementos não linguísticos – extralinguísticos ou paralinguísticos – e apostamos numa relação de *continuum* (Marcuschi, 2003; Koch, 1998, 2002) entre as partes que constituem o discurso, em que qualquer dos elementos pode ocupar, a depender das condições de interlocução, uma determinada relevância na construção da significação veiculada no contexto comunicacional.

3.   Metodologia

Para ilustrar nossa discussão e, sobretudo, para dar visibilidade aos diferentes processos multimodais que participam na construção de objetos do discurso, selecionamos dois episódios extraídos de encontros realizados no CCA, que foram recortados e nomeados de acordo com o tópico discursivo neles desenvolvido (introdução, manutenção e desenvolvimento tópico).

Os dados que compõem o *corpus* pertencem ao *AphasiAcervus*. Para sua constituição, i. selecionamos 5 encontros videogravados no CCA; ii. identificamos os processos multimodais coocorrentes, atribuindo nomes aos quadros enunciativos construídos; iii. selecionamos

excertos cujo tratamento multimodal permitiu incorporar modos comunicativos relevantes na análise das interações em foco e iv. refinamos a transcrição para discussão e análise multimodal.

4.   Apresentação dos dados

Para exemplificar nossa reflexão, analisamos e discutimos um mesmo gesto realizado por SP, JC e EM em duas cenas enunciativas distintas em que participam os sujeitos afásicos SP e MS, e os sujeitos não afásicos HM, EM e JC. O gesto – *esfregar repetidas vezes o dedo polegar contra o dedo indicador, com a palma da mão posicionada para cima e os demais dedos fechados contra a palma* – apresenta um sentido convencional, teoricamente cristalizado nas práticas conversacionais cotidianas brasileiras.

4.1   Dado 1: *AphasiAcervus* (07/04/2005) – *Hospital Particular*

Neste episódio, SP explica ao grupo (e mais especificamente à EM e à HM, coordenadoras das atividades do grupo) que provavelmente fará uma cirurgia para a retirada de um cálculo renal e por isso não sabe se poderá ou não participar de uma atividade de fisioterapia. SP quer desenvolver um pouco mais este tópico, informando aos demais onde realizará a cirurgia ou os exames que irão decidir pela necessidade ou não de intervenção.

SP usa a conjugação dêitica do gesto de apontamento com o dedo indicador e a produção de "*lá*" (em algum lugar outro que não "aqui", Unicamp ou Campinas) para se referir ao local da provável cirurgia, produzindo depois *São Paulo lá lá* também conjugado ao gesto de apontamento com o dedo indicador.

```
EM   o senhor [não sabe °se vai operar
     ou não/°]
SP           [então entã:o] lá: é:
     +justamente (0,7) e:: e-e-+
sp   +movimento de afirmação com a cabeça
     e com o dedo indicador levemente
     para cima e para baixo+
SP   +então lá o:::\ são paulo lá lá/+
sp    +com a mão fechada e o dedo indicador
     aberto da mão esquerda e depois faz
     movimento para a direita+
EM   ahn/
```

Tabela 1: Excerto de *AphasiAcervus* (07/04/2005) – *Hospital Particular*

Em seguida, SP tenta construir um novo referente mobilizando vários gestos que, apesar de conjugados à fala, não são suficientes para a construção da significação pretendida. HM entende, a partir dos gestos mobilizados por SP, que ele está se referindo a exames. Mas a produção verbal de SP permite a HM mobilizar o referente Hospital Sírio Libanês. HM demonstra ter um conhecimento prévio a respeito de São Paulo e de hospitais de São Paulo, pois acede ao referente implícito

na fala de SP: *Sírio Libanês*, o hospital. SP, por sua vez, sabe que HM é paulistana e, apesar de atual moradora de Campinas, viveu sempre em São Paulo. É por isso que ele direciona seu olhar e volta seu tronco para HM, marcando com isso seu interlocutor. Em seguida, sua postura é mais uma vez relevante para determinar a troca de interlocutor, sugerindo que a sequência se daria com EM.

```
HM os exames//
SP +°nã-nã na te-tem no: no:\° (1,5)
   no: ai: tem: o::\ são paulo:+
sp +volta-se para HM e com a mão esquerda
   aberta verticalmente faz
   movimento de cima para baixo em
   menor extensão e depois fecha a mão
   e utiliza o dedo indicador em direção
   à direita+
SP +(1,5) °é::\lá° +sírisíri-li
   sírea:::\ [lá lá:]+
sp +movimenta o dedo indicador da mão
   esquerda sobre a mesa repetidas
   vezes na sequencia da fala e
   volta-se para HM+
HM          [sírio libanês//]
SP +[°lá: lá::°]+
sp +gesto com o dedo indicador esquerdo
   para a direita,direciona-se para EM+
HM [o hospital/ ah tem convênio/]
   (0,6)
SP +/e'za/+
sp +volta-se para HM com movimento de
   afirmação com a cabeça+
SP não não num-é::\ +isso aí não\+
sp +esfrega repetidas vezes o dedo
   polegar contra o dedo indicador, com
   a palma da mão posicionada para cima
   e os demais dedos fechados contra a
   palma e depois abre a mão e a
   movimenta de baixo e para cima+
SP +d-d- lá lá: porque lá:+
sp +gesto com a palma da mão esquerda
   aberta verticalmente em direção à
   direita, direcionando-se para EM+
EM  tá legal (0,6)t-t-t num sei Xr
```

Tabela 2: Excerto de *AphasiAcervus* (07/04/2005) – *Hospital Particular*

Podemos afirmar que SP, ao realizar o gesto que convencionalmente significa "dinheiro" – *esfregar repetidas vezes o dedo polegar contra o dedo indicador, com a palma da mão posicionada para cima e os demais dedos fechados contra a palma* – promove uma recategorização do mesmo, levando à construção do referente "hospital ou consulta particular", e faz com que a materialidade do gesto ganhe uma nova significação referencial, construída na interação, por meio de processos inferenciais explicitados e mobilizados pela ocorrência conjunta entre o gesto, a verbalização, as trocas de olhares e o conhecimento partilhado entre os sujeitos em interação na cena conversacional.

4.2   Dado 2: *AphasiAcervus* (07/04/2005) – *Paraíso Fiscal*

Neste segundo episódio, o grupo conversa sobre a morte do Príncipe Rainier, tópico introduzido por SP a partir de uma notícia no jornal. Aqui, o mesmo gesto que convencionalmente significa **dinheiro** – *esfregar repetidas vezes o dedo polegar contra o dedo indicador, com a palma da mão posicionada para cima e os demais dedos fechados contra a palma* – é usado por três sujeitos diferentes, JC, SP e EM acionando, porém, distintas significações.

JC, sentada em uma das extremidades da mesa, conjuga semioses verbal e não verbal ao fazer um comentário sarcástico sobre o *status* econômico do Principado de Mônaco. Refere-se a Mônaco como – *insignificante* – fazendo uso da prosódia para marcar a ironia e produz "*econômico*" conjugado ao gesto referente a *dinheiro*, acionando aqui o sentido de riqueza.

(Continua na próxima página)

```
JC  EU num sei nã::0\ môna[co é
    insignificante do ponto de vista
    +econômico:\]+
jc +esfrega repetidas vezes o dedo
    polegar contra o dedo indicador, com
    a palma da mão posicionada para cima
    e os demais dedos fechados contra a
    palma+
SP                    +[s:e se num
    me engano: lá]+
sp +apontando com o dedo indicador para o
    jornal sobre a mesa à frente de EM+
SP +lá no outro:::\+
sp  +movimento com a mão esquerda fechada
    e com o indicador para frente+
SP +semana lá te::m o::: [corrida]+
sp +apontando com o dedo polegar para o
    jornal sobre a mesa à frente de EM+
JC                    [então] é
    isso que conta\
(...)
EM é::\ famosa né/
MS -é\ ahn/
HM nas ruas//
ms ((voz imitando o barulho do motor de
    um carro e gestos da mão esquerda
    aberta verticalmente fazendo
    movimentos como curvas))
EM +de mônaco//+
em +faz gestos de curvas com a mão
    direita aberta como MS+
SP  é XX ((o mesmo gesto da mão aberta
    descrevendo curvas, como o gesto de
    MS))
MS  isto\ ((imita, novamente, o barulho
    do carro de corrida))
EM AH\o AYRTON SEnna: né-/ (.) ganhou um
    prêmio lá num foi//
MS MUito:\ ahn\ ((levanta o polegar da
    mão esquerda em positivo e faz
    movimento de afirmação com a cabeça))
EM +ELE MORAVA também PA:rte [da: a vida
    dele\]+
em +volta-se para SP e faz movimento com
    o dedo indicador direito+
```

```
SP                         [é:\ tudo
    tudo\]
EM parte do tempo dele do a:no (.) ele
    [morava lá:\]
SP [ah é:\+muito:]muito do::do: do:+
sp  +faz três vezes o gesto com o polegar
    voltado para trás+ ((esfrega
    repetidas vezes o dedo polegar contra
    o dedo indicador, com a palma da mão
    posicionada para cima e os demais
    dedos fechados contra a palma))
SP +tinha lá é lá:::::\
    (.)°de:::::\°corrida né/+
sp  +gesto com o polegar voltado para trás
    e depois com os dedos polegar e
    indicador abertos em direção ao
    jornal+
EM °ahan\°
SP +jo:ga lá na na:::+
sp +aponta para trás com o polegar+
EM ah::\ JOGA nos cassinos\ é isso//
SP +na na la ne lenã:o\+
sp +movimento com o polegar da mão
    esquerda para trás repetidas vezes+
SP +ja- joga no:: ahn: no::: banco\°lá
    na:\°+
sp +gesto com a mão esquerda fechada com
    movimento para baixo como se
    "depositasse" algo+
EM  AH: tá:\ é [como se a-]
SP            +[lá ele num] tem [NAda:\]+
sp +movimento com a mão esquerda aberta
    para baixo da direita para a
    esquerda+
EM                         [como um
    PARA]ÍSO FISCA:L
em [esfrega três vezes o dedo polegar contra
    o dedo indicador, com a palma da mão
    posicionada para frente e os demais
    dedos fechados contra a palma]
MS +[I:::::SSO:::\]+
ms +apontando o dedo indicador esquerdo em
    direção à EM+
EM  +muiTA gente de +dinheiro+
    (0,6)tinha dinheiro em BANco lá\+
em +esfrega três vezes o dedo polegar contra
    o dedo indicador, com a palma da mão
    posicionada à sua frente e os demais
    dedos fechados contra a palma, em
    seguida aponta com a mão direita  para
    o jornal na sua mão esquerda+
ms ((risos))
EM  XXX muito dinheiro- como na suíça
```

Tabela 3: Excerto de *AphasiAcervus* (07/04/2005) – *Paraíso Fiscal*

A progressão tópica acima se dá a partir da semiose não verbal acionada por MS. É a partir de seus gestos que HM produz "*aquela curvinha*", acionando o conhecimento de mundo relativo a uma certa curva do Circuito de Mônaco que foi determinante para a vitória do campeão brasileiro Ayrton Senna sobre seu rival Alain Prost. EM reconhece o sentido veiculado na fala de HM, pois evoca, então, o nome de Ayrton Senna. É neste contexto da interação que SP vai fazer uso da mesma semiose não verbal referente a **dinheiro**: *esfregar*

repetidas vezes o dedo polegar contra o dedo indicador, com a palma da mão posicionada para cima e os demais dedos fechados contra a palma. Mas a significação pretendida por SP com o uso deste gesto – que EM retoma em dois momentos distintos da interação – só será compreendida ao longo da progressão tópica: Mônaco é um paraíso fiscal.

5.   Análise dos dados

Analisamos os dados de acordo com a abordagem multimodal proposta por Norris (2006), buscando compreender o significado e a relevância das ocorrências multimodais nas cenas enunciativas em que foram produzidas. Observamos que:

1. as ocorrências dos processos multimodais vão desde os gestos mais estandartizados, formulaicos, como os gestos dêiticos e apontamentos que se conjugam com a fala (aqui, lá) ou os movimentos de cabeça indicando negação, a gestos elaborados (icônicos, pantomímicos);
2. os aspectos entoacionais, as posições ocupadas pelos interlocutores no espaço da enunciação, os direcionamentos de olhares dentre outros elementos multimodais são reconhecidamente parte da cena enunciativa;
3. o mesmo gesto – *esfregar repetidas vezes o dedo polegar contra o dedo indicador, com a palma da mão posicionada para cima e os demais dedos fechados contra a palma* – mobiliza sentidos diversos que se deslocam pela interlocução de maneiras diferentes, construindo objetos de discurso (Mondada, 2001) distintos nas atividades de referenciação e inferenciação ou introduzindo novo tópico discursivo. O gesto de SP no dado 1, desloca a linguagem como modo mais relevante e se reveste de alta densidade modal, tornando-se foco de atenção e figura – e não mais fundo – no *continuum* proposto por Norris. Já o mesmo gesto feito por JC e EM, no dado 2, tem baixa densidade modal, sendo apenas fundo com função de enfatizar a fala que acompanha.

6.   Comentários e conclusão

O levantamento e a análise dos processos multimodais coocorrentes na referenciação, aqui apresentados, nos permitem refletir sobre a relação das semioses verbais e não verbais na construção da significação. Se se reivindica para a linguagem um papel relevante na constituição das interações e se as semioses não verbais são tidas como elementos não linguísticos, nossa análise deixa entrever a relação solidária entre as semioses verbais e não verbais na referenciação.

Semioses verbais e não verbais, como a fala, a escrita, o gesto, o olhar, a prosódia, a expressão e a mímica facial, os movimentos de cabeça e das mãos, as posições das pessoas em relação umas às outras, o contexto da interlocução *etc*., são produzidas e interpretadas no processo de referenciação,

desenvolvendo-se e transformando-se a partir dos contextos e através de operações linguístico-cognitivas realizadas pelos sujeitos na interação.

Observamos que os processos multimodais são mobilizados e coocorrem com outros aspectos referenciais na construção do sentido, sendo fundamentais na compreensão da significação pretendida. Os gestos dos sujeitos afásicos e não afásicos – SP, MS, EM, HM, JC – deslocam a linguagem como modo mais relevante e se revestem de alta densidade modal, tornando-se foco de atenção e figura – e não mais fundo – no *continuum* proposto por Norris. Portanto, uma abordagem teórico-metodológica que não considere a multimodalidade – tanto na constituição quanto na análise de um *corpus* – possivelmente encobrirá ou distorcerá as múltiplas ações nas quais os sujeitos em interação estão simultaneamente envolvidos (Norris, *op.cit.*).

Enfim, podemos afirmar que uma perspectiva sociocognitiva de base textual-interativa que considere os processos multimodais permite construir um *corpus* suficientemente acurado para dar visibilidade à coocorrência entre os processos de significação verbais e não verbais na construção do sentido, como observado neste estudo nos episódios de interação entre afásicos e não afásicos.

7.   Referências

Holler, J., Beattie, G. (2006). Gesture use in social interaction: how speaker's gestures can reflect listener's thinking. In L. Mondada, V. Markaki (Eds), *Interacting Bodies*. Online proceedings of the 2d ISGS Conference. Available at: <http://gesture-lyon2005.ens-lsh.fr/article.php3?id_article=259>.

Koch, I.G.V. (1998). *A Inter-Ação pela Linguagem*. 4ª ed. São Paulo: Contexto.

Koch, I.G.V. (2002). *Desvendando os segredos do texto*. São Paulo: Cortez.

Marcuschi, L.A. (2003). Do código para a cognição: o processo referencial como atividade cognitiva. In *Veredas* 13: pp. 43--62.

Mondada, L., Markaki, V. (Eds). (2006). *Interacting Bodies*. Online proceedings of the 2d ISGS Conference. Available at: <http://gesture-lyon2005.ens-lsh.fr/article.php3?id_article=259>.

Mondada, L. (2001). Gestion du topic et organization de la conversation. In *Cadernos de Estudos Linguísticos*, Campinas, (41): Jul. / Dez. 2001.

Mondada, L. (1998/2). Production du savoir et interactions multimodales. Une étude de la modélisation spatiale comme activité pratique située et incarnée. In: *Revue d'anthropologie des connaissances* 2008/2 (Vol. 2, n° 2).

Norris, S. (2006). Multiparty interaction: a multimodal perspective on relevance. In *Discourse Studies.* Vol 8(3): pp. 401--421.

8.   Apêndice

Notação utilizada na transcrição (baseada no Sistema de Notação do *AphasiAcervus*):

i.   Iniciais em maiúsculas (SP) – identificam os participantes, correspondendo às iniciais dos nomes e indicam os turnos de fala

ii.   Iniciais em minúsculas (sp) – descrições de aspectos não verbais sincronizados aos turnos de fala

OCORRÊNCIAS	NOTAÇÃO
overlap	[início do overlap] fim do overlap
micro pausas < 0,3s	(.)
pausas	(0,4), (1,0), (2,3)
alongamento silábico	:
truncamento de palavras	-
entonação ascendente	/
entonação descendente	\
ênfase particular	<u>sublinhado</u>
volume forte de voz	MAIÚSCULA
volume baixo murmúrio de voz	° °
comentários do transcritor e fenômenos e atividades não transcritos, como risos, leitura, mudança de lugar, saída da sala, conversas de fundo não transcritas *etc*. são indicados em *itálico* e entre parênteses	((*comentários*))
segmentos inaudíveis ou incompreendidos são indicados com a letra X, correspondendo, sempre que possível, ao número de sílabas produzido	X XX
+ + delimitam o tempo de duração dos aspectos não verbais sincronizados aos turnos de fala	+ +

Tabela 3: Notação utilizada na transcrição

Correlatos acústicos e perceptivos de qualidade vocal e dinâmica vocal: dados a partir da fala de criança com deficiência auditiva e de criança ouvinte

Aline PESSOA, Zuleica CAMARGO

Laboratório da Análise Acústica e Cognição (LIAAC) do Programa de Linguística Aplicada e Estudos da Linguagem - PUC-SP

aline.nevespessoa@gmail.com, zcamargo@pucsp.br

Resumo

Esta pesquisa origina-se da clínica fonoaudiológica com crianças com DA e usuárias de implante coclear (IC). Utilizou a análise fonética (correlatos acústicos e perceptivo-auditivos) para descrição dos ajustes de qualidade vocal e elementos de dinâmica vocal, com enfoque dos elementos prosódicos da fala. Estudou-se a fala espontânea de irmãos gemelares (um com DA e outro ouvinte). Ambos apresentaram variação suave de f0 e valores aproximados de derivada de frequência fundamental –f0 - e de espectro de longo termo-ELT. Apresentaram valores de semi-amplitude entre quartis de f0 com média de 121 Hz e ambos revelaram segregação dos valores médios de derivada de f0 numa única classe na análise aglomerativa hierárquica. Ajustes com redução de área de cavidades ressoadoras, identificadas em criança usuária com DA, destacaram-se pelas tendências `a diminuição de extensão do trato vocal (corpo de língua, mandíbula e lábios), ajustes de voz crepitante, *pitch* habitual elevado, falsete, hipofunção laríngea, ponta de língua avançada, que correlacionaram-se com as medidas de desvio padrão e mediana de f0, além de assimetria de intensidade. Os resultados de validação cruzada a partir da análise discriminante revelaram a análise perceptivo auditiva possibilitou a segregação das amostras da criança com DA (66,67%) e da ouvinte (91,67%).

Palavras-chave: qualidade vocal; dinâmica vocal; implante coclear; percepção da fala; acústica da fala.

1. Introdução

A análise fonética (acústica e perceptiva) tem se configurado como uma ferramenta clínica auxiliar no entendimento das características de fala de crianças com deficiência auditiva (DA). A descrição dos ajustes de qualidade vocal e dos aspectos de dinâmica vocal pode levantar inferências sobre o processo de aquisição de linguagem oral nesta população e, especialmente, sobre a intervenção terapêutica.

Esta pesquisa origina-se de questões clínicas do atendimento fonoaudiológico de crianças com DA, usuárias de implante coclear (IC), que visa aquisição de linguagem oral-verbal (Yoshinaga-Itano, 2003; Xu *et al*., 2009; Tobey *et al*., 2003; Novaes & Mendes, 2011). As investigações têm relacionado as esferas da percepção e da produção de fala, diante das interações que se estabelecem entre elementos segmentares e prosódicos (Albano *et al*., 1997; Benninguer, 2011), a partir de *corpus* estruturado em coletas seriadas em situação de terapia (Pessoa *et al*., 2010a; Pessoa *et al*., 2011; Pessoa *et al.*, 2012).

Neste contexto, instrumentos de análises perceptivo auditiva e acústica têm sido utilizados. Tais análises têm permitido correlações com detalhamentos em instâncias de longo termo da fala.

Do ponto de vista perceptivo-auditivo, o roteiro Vocal Profile Analyses Scheme- for Brazilian Portuguese –VPAS-PB (Camargo & Madureira, 2008, 2009, 2010) adaptado para o português brasileiro, permite a descrição perceptiva dos elementos prosódicos a partir de dois módulos: qualidade vocal e dinâmica vocal. Considera-se, nesse instrumento, a qualidade vocal como resultado da ação conjunta da laringe e do trato vocal supralaríngeo, emergindo da combinação dos ajustes de longo termo na fala (Laver, 1980; Mackenzie-Beck & Laver, 2007; Abberton, 2000). Ou seja, busca descrever as tendências de longo termo que caracterizam um falante em particular, produtos das atividades respiratória, laríngea/fonatória,

supralaríngea/articulatória e de tensão muscular (Hammaberg & Gauffin, 1995; Camargo & Madureira, 2010). O módulo de dinâmica vocal oferece a possibilidade de julgamento dos parâmetros de *pitch, loudness*, uso de pausas, taxa de elocução e suporte respiratório.

Do ponto de vista acústico, aspectos de qualidade e de dinâmica vocal têm sido explorados por meio da combinação de um grupo de medidas acústicas (Barbosa, 2006, 2007, 2009) referentes à freqüência fundamental (f0), primeira derivada de f0, intensidade, declínio espectral e espectro de longo termo (Camargo & Madureira, 2010; Madureira & Camargo, 2010; Rusilo *et al*., 2011; Pessoa *et al*., 2010a, Pessoa *et al*., 2010b; Pereira *et al*., 2010; Pessoa *et al*., 2012a; Pessoa *et al*., 2012b; Camargo *et al*., 2012).

Tais correlações, pautadas em modelos dinâmicos e procedimentos metodológicos de Fonética Experimental, remetem ao conhecimento da produção da fala em contextos de falantes com e sem alteração na aquisição de linguagem.

Além disso, podem prover `a aplicabilidade dessas ferramentas como instrumento de acompanhamento da evolução de linguagem oral do sujeito no processo terapêutico, bem como para aprofundamento do conhecimento de marcos de desenvolvimento de fala (aquisição dos sons da língua e da estruturação dos elementos prosódicos) também em crianças ouvintes.

2. Objetivo

Caracterizar a qualidade vocal e dinâmica vocal de criança com DA usuária de IC em comparação a uma criança ouvinte, a partir de correlatos acústicos e perceptivo-auditivos.

3. Material e Método

A gravação do *corpus* de fala em contexto terapêutico (em curso) acontece em sala de atendimentos fonoaudiológicos. As coletas ocorreram de forma a registrar em um contexto

lúdico, as vocalizações e as produções de fala típicas do espaço terapêutico, de maneira que a coleta foi planejada para promover o mínimo de interferências na situação em questão. No caso da criança ouvinte, o mesmo espaço foi utilizado com os mesmos materiais, porém de forma lúdica, sem o respaldo de um plano de terapia.

O Quadro 1 apresenta dados da caracterização audiológica dos sujeitos participantes da pesquisa.

Sujeito	Dados audiológicos
Criança ouvinte	Limiares auditivos melhores do que 15dB nas frequências de 0,25; 0,5; 1; 2; 3,;4; 6 e 8 KHz, em cabine audiomérica.
Criança com DA usuária de IC unilateral e Aparelho de amplificação sonora individual (AASI) na orelha contralateral.	DA congênita, diagnóstico no primeiro ano de vida com adaptação de AASIs bilateralmente e cirurgia para inserção de IC aos 2 anos de idade. Respostas mínimas auditivas em cabine audiométrica com IC: limiares melhores do que 15dB nas frequências de 0,25 ; 0,5; 1; 2; 3; 4; 6 e 8 KHz.

Quadro1: Caracterização audiológica dos sujeitos

Para o presente estudo foram selecionadas amostras de fala de duas crianças (uma com DA - usuária de IC e outra ouvinte) do sexo masculino, irmãos gemelares, de 6 anos de idade. O instrumental utilizado refere-se a microfone unidirecional ML 70-D Lapela (*Le son*) e a gravador digital *MD Sony modelo MZ-R70*. Os processos de edição, tratamento e análise das amostras foram realizados no Laboratório Integrado de Análise Acústica e Cognição (LIAAC) da PUC-SP. O material é digitalizado na freqüência de amostragem 22050 Hz e 16 *bits*, extensão *wav*, a partir do *software Sound Forge Edit* (versão 7.0) e analisado por meios acústico e perceptivo.

A análise perceptivo-auditiva foi realizada por meio do roteiro VPAS-PB (Camargo & Madureira, 2008), por dois juízes experientes, a partir dos itens de qualidade vocal e de dinâmica vocal.

A análise acústica foi realizada a partir da aplicação do *script ExpressionEvaluator* (Barbosa, 2009) ao *software Praat,*. O *script* gera dados de mediana, semi-amplitude entre quartis, quantil 99,5% e assimetria de freqüência fundamental (f0); média, desvio padrão e assimetria de primeira derivada de f0; assimetria de intensidade; média, desvio padrão e assimetria de declínio espectral; desvio padrão de ELT (espectro de longo termo).

Tais dados multivariados foram correlacionados estatisticamente (Lattin *et al*., 2011), enquanto tendências de agrupamentos na análise aglomerativa hierárquica de *cluster,* bem como de correlações aos dados da esfera perceptiva (roteiro VPAS-PB) por meio da análise de correlação canônica e de análise discriminante (Rusilo *et al*., 2011), buscando-se comparar a distribuição das informações dos dois falantes (com e sem DA e uso de IC).

A pesquisa em questão foi aprovada pelo Comitê de Ética em Pesquisa da Instituição onde é realizada (nº 135/2009).

4. Resultados e Discussão:

Para esta etapa de apresentação de dados, a análise pautou-se na correlação entre achados acústicos e perceptivo-auditivos, aplicada em pareamento entre dados advindos de criança com DA, usuária de IC, e de criança ouvinte.

Os resultados da análise perceptivo-auditiva são apresentados na Figura 1.

Figura 1: Análise perceptivo-auditiva – Roteiro VPAS-PB: fala da criança ouvinte (X) e da fala da criança com DA usuária de IC( )

Na análise aglomerativa hierárquica de *cluster* para dados perceptivo-auditivos (Figura 2 e 3, para criança ouvinte e com IC, respectivamente) verificou-se, para criança ouvinte foram agrupados em quatro classes: classe 2 (ponta de língua avançada), classe 3 (corpo de língua abaixado, continuidade interrompida), classe 4 (denasal) e demais mobilizações agruparam-se na classe 1. No caso dos dados de criança usuária de IC, os julgamentos foram agrupados em seis classes: classe 2 (extensão diminuída de lábios, extensão diminuída de mandíbula), classe 3 (ponta

de língua avançada), classe 4 (corpo de língua abaixado, nasalização, ajuste denasal, hipofunção laríngea, falsete, voz crepitante, continuidade interrompida, suporte respiratório inadequado), classe 5 (extensão diminuída de corpo de língua), classe 6 (*pitch* habitual elevado) e demais ajustes agruparam-se na classe 1.
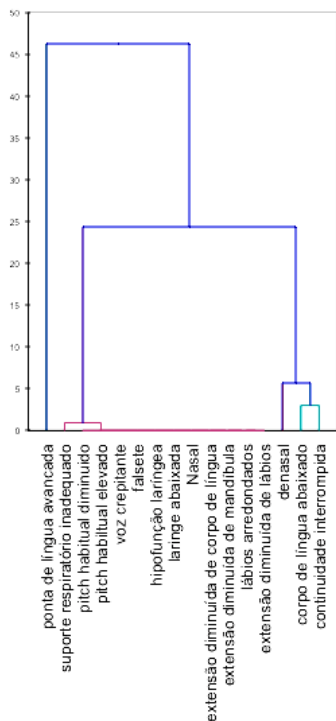
Figura 2: Criança ouvinte – dendrograma dos dados perceptivo-auditivos de fala
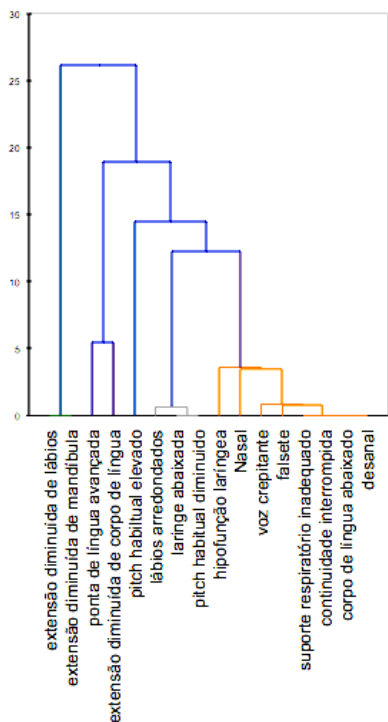
Figura 3: Criança usuária de IC – dendrograma dos dados perceptivo-auditivo de fala

Os trabalhos anteriores revelaram o agrupamento de ajustes de maior grau de hiperfunção laríngea e de aumento do *pitch* habitual, combinados `a diminuição de amplitude de movimento de articuladores, especialmente de lábios, mandíbula e de língua (Pessoa *et al*., 2010a; Pessoa *et al*., 2011a; Pessoa *et al*., 2012b; Ubrig *et al*., 2011), com exceção dos ajustes de tensão laríngea, hipofunção em lugar de hiperfunção, e no plano supraglótico, ponta de língua avançada, em lugar de ponta de língua recuada. Neste estudo, tais combinações foram notórias em ambos os casos. Tais dados encontram respaldo na literatura de estudos de inteligibilidade que têm como enfoque o plano segmental, especialmente para mobilizações de língua em vogais e consoantes (Ubrig *et al*., 2011; Coelho, 2011).

Neste estudo a distribuição dos julgamentos perceptivos revelou que a criança usuária de IC diferenciou, em relação ao ouvinte, em número maior de classes de julgamentos, especialmente no que se refere aos ajustes de diminuição de movimento de articuladores (lábios, mandíbula e corpo de língua). No plano da analise perceptivo-auditiva, foi possível identificar maior especificidade na descrição diferenciada dos falantes (com IC e ouvinte), especialmente a partir da de descrição dos graus de manifestação do mesmo ajuste.

Os resultados da análise acústica são apresentados na Tabela 2, quanto aos valores gerados por meio do *script ExpressionEvaluator*.

Criança ouvinte

Variável	Mínimo	Máximo	Média	Desvio padrão	Valores Absolutos
mediana de f0	0,450	0,700	0,567	0,074	299
semi amplitude entre quartis de f0	0,500	1,100	0,828	0,237	121,6566667
quantil 99,5% de f0	0,490	1,420	1,204	0,293	X
assimetria de f0	-0,200	0,160	0,078	0,097	X
média de derivada de f0	-3,100	0,430	-0,688	0,989	-0,159005
desvio padrão de derivada de f0	0,050	0,140	0,108	0,028	0,0248325
assimetria de derivada de f0 / 10	-0,460	0,530	0,000	0,297	-0,004166667
assimetria de intensidade	0,190	1,010	0,475	0,242	4,75
média de declínio espectral	0,210	0,340	0,283	0,034	2,833333333
desvio padrão de inclinação espectral	0,230	0,360	0,312	0,035	X
assimetria de declínio espectral	1,200	1,340	1,282	0,044	X
desvio padrão de LTAS	1,090	2,090	1,428	0,288	14,275

Criança com DA usuária de IC

Variável	Mínimo	Máximo	Média	Desvio padrão	Valores absolutos
mediana de f0	0,130	0,840	0,382	0,219	276,8
semi amplitude entre quartis de f0	0,430	1,290	0,868	0,276	121,735
quantil 99,5% de f0	0,660	1,510	1,218	0,255	X
assimetria de f0	-0,020	0,370	0,163	0,109	X
média de derivada de f0	-5,310	4,580	-0,533	3,193	-0,1230075
desvio padrão de derivada de f0	0,070	0,170	0,117	0,037	0,02695
assimetria de derivada de f0 / 10	-0,430	0,520	0,003	0,335	0,025
assimetria de intensidade	-0,090	0,730	0,326	0,262	3,258333333
média de declínio espectral	0,200	0,360	0,290	0,047	2,9
desvio padrão de inclinação espectral	0,230	0,390	0,315	0,046	X
assimetria de declínio espectral	1,200	1,410	1,272	0,072	X
desvio padrão de LTAS	1,100	2,410	1,659	0,359	16,59166667

Tabela 1: Valores de medidas acústicas de f0 (mediana, semi-amplitude entre quartis, quantil 99,5% e assimetria), primeira derivada de f0 (media, desvio padrão e assimetria), declínio espectral (media, desvio padrão e assimetria) e espectro de longo termo (desvio padrão) da criança ouvinte (acima) e da criança usuária de IC (abaixo)

Os valores de mediana de f0 apresentaram-se próximos aos valores de dados de f0 de crianças ouvintes brasileiras, ouvintes, saudáveis, de 6 e 7 anos do sexo masculino (258Hz, com desvio padrão de 25Hz), conforme apresentado por Andrade (2009). Os valores de f0 médio da criança ouvinte (299Hz) encontram-se aumentados em relação aos da criança com DA (276,8Hz).

Os valores obtidos em média de primeira derivada de f0, que representam a taxa de variação do parâmetro,

CORRELATOS ACÚSTICOS E PERCEPTIVOS DE QUALIDADE VOCAL E DINÂMICA VOCAL: DADOS A PARTIR DA FALA DE CRIANÇA
COM DEFICIÊNCIA AUDITIVA E DE CRIANÇA OUVINTE

303

sinalizam para variações suaves de f0 no fluxo da fala, e não abruptas. Assim, ambos os sujeitos apresentaram variação suave de f0. Os valores de derivada de f0 e de ELT foram parecidos nas amostras das duas crianças. Diferente desses dados, na literatura são referidas comumente variações extremas e abruptas de f0, tanto para falantes usuários de AASI como de IC (Cukier *et al*., 2005; Baudonck *et al*., 2011).

Os valores de semi-amplitude entre quartis de f0 (com média de 121 Hz em ambos os casos) revela aspectos de variabilidade de julgamentos de extensão de *pitch* comumente descritos na fala de DAs com ou sem IC. Variações extremas ou restritas são descritas para esta população (Stuchi *et al*., 2007, Ubrig *et al*., 2011).

Dados de ajustes de qualidade vocal e mobilizações de dinâmica vocal foram descritos pela análise aglomerativa hierárquica de cluster aplicada aos dados de medidas acústicas da criança ouvinte (Figura 4) e da criança com DA usuária de IC (Figura 5) revelaram diferenciação na distribuição de medidas das duas crianças.   As   medidas acústicas das amostras da criança ouvinte segregaram-se em: classe 3 (desvio padrão de ELT, quantil 99,5% de f0, média de declínio espectral) e classe 2 (média de derivada de f0) e demais medidas agruparam-se na classe 1. Já as medidas acústicas dos dados da criança com IC revelaram a formação de 4 classes: classe 1 (mediana de f0, semi-amplituide entre quartis de f0, assimetria de f0, desvio padrão de primeira derivada de f0, assimetria de primeira derivada de f0 e assimetria de intensidade, média e desvio padrão de declínio espectral), classe 2 (quantil 99,5% de f0, assimetria e declínio espectral e desvio padrão de ELT), classe 3 (média de primeira derivada de f0).

Ambos os falantes revelaram segregação dos valores médios de derivada de f0 numa única classe. Tais dados reforçam a importância do enfoque na variabilidade de f0 no fluxo da fala.
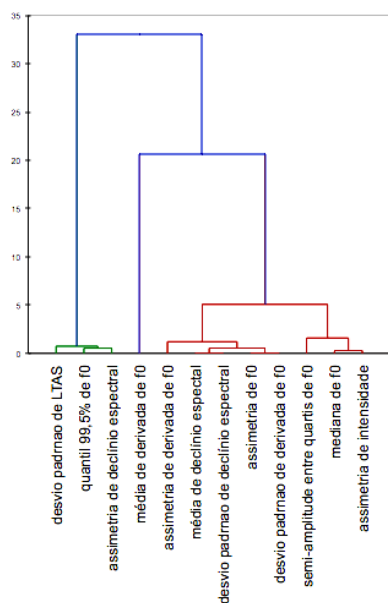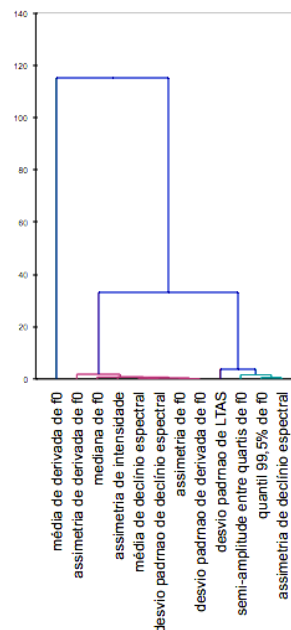
Figura 5: Criança usuária de IC: dendrograma dos dados acústicos de fala

A análise de correlação canônica dos dados acústicos e perceptivo-auditivos da criança ouvinte (Figura 6) e da criança usuário de IC (Figura 7) revelaram que nas amostras de fala da criança com IC, os ajustes de lábios arredondados, laringe abaixada e pitch habitual diminuído correlacionaram-se com as medidas de assimetria de f0, media de derivada de f0, semi-amplitude entre quartis de f0 e quantil 99,5% de f0, assimetria, mediana e desvio padrão de declínio espectral e desvio padrão de ELT. Neste grupo, destacaram-se tendências diminuição de extensão do trato vocal. Os ajustes de voz crepitante, *pitch* habitual elevado, falsete, hipofunção laríngea, ponta de língua avançada, diminuição de extensão de corpo de língua, mandíbula e lábios correlacionaram-se com as medidas de desvio padrão e mediana de f0, além de assimetria de intensidade. Neste grupo as mobilizações concentram-se em ajustes com redução de área de cavidades ressoadoras.

Nas amostras de fala da criança ouvinte, as medidas de declínio espectral (assimetria e desvio padrão), de intensidade (assimetria) e de mediana de f0 agruparam-se com os ajustes de corpo de língua abaixado, continuidade interrompida e suporte respiratório inadequado. As medidas de média de declínio espectral, desvio padrão de ELT, média de derivada de f0, desvio padrão de f0, assimetria de f0, quantil 99,5% de f0 e semi-amplitude entre quartis e de f0 agruparam-se com os ajustes denasal e ponta de língua avançada, apesar de apresentarem ajustes de qualidade vocal similares em sua natureza, a combinação deles no fluxo da fala, bem como o seu grau de manifestação puderam diferenciar os falantes em termos de combinação com medidas acústicas.

Figura 4: Criança ouvinte: dendrograma dos dados acústicos de fala
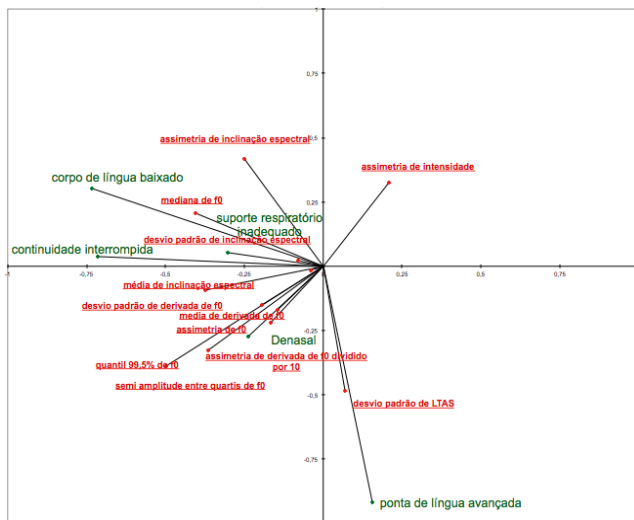
Figura 6: Análise canônica: correlatos perceptivo-auditivos e acústicos (sublinhados) da fala da criança ouvinte
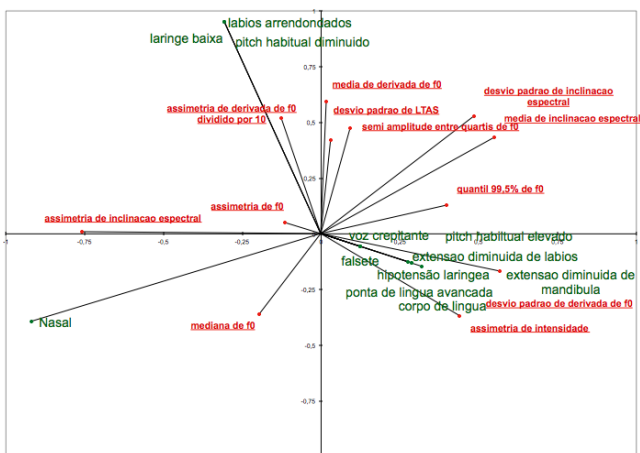
Figura 7: Análise de correlação canônica: correlatos perceptivo-auditivos e acústicos (sublinhados) da fala da criança usuária de IC

Finalmente, na análise discriminante, os resultados de validação cruzada revelaram que os julgamentos por meio do roteiro VPAS-PB possibilitaram a segregação das amostras da criança usuária IC (66,67%) e da ouvinte (91,67%). O ajuste de ponta de língua avançada foi o que apresentou significância (p=0,001) em relação aos outros ajustes utilizados pelas duas criança. Assim, na análise discriminante, as medidas acústicas não segregaram as emissões dos dois falantes.

Particularidades das combinações de ajustes de qualidade vocal e aspectos da dinâmica vocal, além das medidas acústicas, foram identificadas para cada falante estudado. Salienta-se que os achados de julgamentos perceptivos permitiram segregação das amostras de ambos os falantes, com maior potencial para detecção da criança ouvinte.

5.   Conclusão

Ressalta-se a descrição dos ajustes de qualidade vocal, aspectos de dinâmica vocal e medidas acústicas em correlação, cuja composição do *corpus* de fala se dá em situação terapêutica. Tais dados poderão colaborar de forma a estimular o enfoque dos elementos prosódicos no estudo da fala de crianças usuárias de IC desde idades precoces.

6.   Agradecimentos

7.   Referências

Abberton, E. (2000). Voice Quality of deaf speakers. In R.D. Kent, M.J. Ball (Eds.), Voice Quality Measurement. San Diego: Singular, 22: pp. 449--59.

Albano, E., Barbosa, P., Gama-Rossi, A., Madureira. S. and Silva, A. (1997). A interface fonética-fonologia e a interação prosódia-segmentos. In *Estudos Linguísticos XXVII - Anais do XLV Seminário do Grupo de Estudos Linguísticos do Estado de São Paulo-GEL'97*. Campinas. pp. 135--43.

Andrade, F.V. (2009). Análise de parâmetros espectrais da voz em crianças saudáveis de 4 a 8 anos. 72p. Dissertação de Mestrado em Fonoaudiologia. Faculdade de Fonoaudiologia, Universidade Veiga de Almeida, Rio de Janeiro.

Barbosa, P.A. (2006). *Incursões em torno do ritmo da fala*. Campinas: Pontes/FAPESP.

Barbosa, P.A. (2007). From Syntax to acoustic duration: a dynamical model of speech rhythm production. Oxford: Speech Communication, Sept. 49(9): pp. 725--42.

Barbosa, P.A. (2009). Detecting changes in speech expressiveness in participants of a radio program In *Proc. of Interspeech*. v.1, pp. 2155--2158.

Baudonck, E.D., Dhooge, I. and Lierde, K.V. (2011). Objective vocal quality in children using cochlear implants: a multiparameter approach. In *J Voice*, vol. 25, n 6, 2011, pp. 683--691.

Benninguer, M.S. (2011). Quality of the Voice Literature: What is There and What is Missing. In *Journal of Voice* 2011 Nov; 25(6): pp. 647--52.

Camargo, Z., Madureira, S. (2008). Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS) In *Proc. of the 4th Conf on Speech Prosody* v.1, pp. 57--60.

Camargo, Z., Madureira, S. (2009). Dimensões perceptivas das alterações de qualidade vocal e suas correlações aos planos da acústica e da fisiologia. In *DELTA - Documentação de Estudos em Linguística Teórica e Aplicada*. PUCSP, 2009. 25(2): pp. 285--317.

Camargo, Z., Madureira, S. (2010). The acoustic analysis of speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS): long-term f0 and intensity measures. In *Proc.of the Workshop on Experimental Linguistics* v. 1. pp. 33--6.

Coelho, A.C.C. (2011). Efeito de diferentes estratégias de codificação dos processadores de fala na voz de crianças usuárias de implante coclear / Ana Cristina de Castro Coelho. – Bauru. 151 p.

Cukier, S., Camargo, Z. (2005). Abordagem da qualidade vocal em um falante com deficiência auditiva: aspectos acústicos relevantes do sinal de fala. Revista CEFAC, Jan-Mar. 7(1): pp. 93--101.

Hammaberg, B., Gauffin, J. (1995). Perceptual and acoustics characteristics of quality differences in pathological voices as related to physiological aspects. In O. Fujimura, M. Hirano (Eds.), *Vocal foldphysiology*. San Diego: Singular, pp. 283--303.

Lattin, J., Carrol, Douglas J.D. and Green, P.E. (2011). Análise de dados multivariados. São Paulo: Cegrage Learning.

Laver, J. (1980). The phonetic description of voice quality. Cambridge: Cambridge University Press.

Laver, J., Wirz, S.L., Mackenzie-Beck, J. and Hiller, S.M. (1981). A perceptual protocol for the analysis of vocal profiles. Edinburgh University Department of Linguistics Work in Progress, pp. 139--155.

Laver, J., Mackenzie-Beck, J. (2007). Vocal Profile Analysis scheme-VPAS. Edinburgh: QMUC, Speech Science Research Centre.

Laver, J., Mackenzie-Beck, J. (2007). Vocal Profile Analysis Scheme -VPAS. Queen Margareth University College-QMUC, Speech Science Research Centre, Edinburgh.

Mackenzie-Beck, J. (2005). Perceptual analysis of voice quality: the place of vocal profile analysis. In W.J. Hardcastle, J. Mackenzie-Beck (Eds.), *A figure of speech: a festschrift for John Laver*. Mahwah: Lawrence Erlbrum Associates, pp. 285--322.

Novaes, B.C.A.C., Mendes, B.C. (2011). Habilitação Auditiva: Intervenção em Bebês e Crianças. In S.C. Neto, J.F.M Júnior, R.H.G. Martins and S.S. da Costa (Eds.), *Tratado de Otorrinolaringologia* - Aborlccf. 2 ed. São Paulo: Edit Roca, pp. 371--80.

Pereira, L.C.K., Pessoa, A N; Camargo, Z A; Fontes, S M; Novaes, B. An Analysis Of Voice Quality And Voice Dynamics In The Speech Production Of A Cochlear Implant User 13th Meeting of the International Clinical Linguistics and Phonetics Association - ICPLA, Oslo-Norway, June 23-26, 2010, pp. 286.

Pessoa, A., Novaes, B.C.A.C. and Camargo, Z. (2011). A fala em usuários de implante coclear: evidências de análise perceptivo-auditiva e acústica. In *19º. Congresso Brasileiro e 8º Internacional de Fonoaudiologia*, São Paulo. *Revista da Sociedade Brasileira de Fonoaudiologia* - Suplemento Especial. São Paulo: Sociedade Brasileira de Fonoaudiologia, v. 16. pp. 1247--1247.

Pessoa, A.N., Novaes, B.C.A.C., Mendes, B. de C. and Camargo, Z.A. (2010a). Voice Quality And Dynamics Of Child-Cochlear Implant User During Therapy Session. In *XXXth International Congress of Audiology (ICA 2010)*, 25o Encontro Internacional de Audiologia (EIA), São Paulo. pp. 83--83.

Pessoa, A.N., Pereira, L.K., Novaes, B., Camargo, Z.A. and Mendes, B. (2010b). Using Acoustic Analysis To Follow-Up The Language Development Of A Brazilian Deaf Child With Cochlear Implant. Pôster.11th International Conference on Cochlear Implants and Other Implantable Auditory Technologies, Stockholm- Sweden, 30jun-03jul, pp. 371.

Pessoa, A.N., Pereira, L.K., Novaes, B. and Camargo, Z. (2012). Dados de dinâmica e qualidade vocal: correlatos acústicos e perceptivo-auditivos da fala em criança usuária de implante coclear. Journal of Speech Sciences, 1(2),pp. 17--33.ISSN: 2236-9740.

Pessoa, A.N., Pereira, L.K., Novaes, B.A.C., Camargo, Z.A. (2012a). Correlação acústica e perceptivo-auditiva: a fala de criança usuária de implante coclear. In *Encontro Internacional de Audiologia*, USP-Bauru, 14-17 abril, pp. 185, ISSN: 1983-179X. Trabalho premiado: Prêmio Excelência em Audiologia 2012.

Pessoa, A.N., Novaes, B.C.A.C., Madureira, S and Camargo, Z.A. (2012b). Perceptual and acoustic correlates of a speech in a bilateral cochlear implant user. In Q. Ma, H. Ding and D. Hirst (Eds.), *Abstract Book Speech Prosody*, 6th International Conference, Tongji University Press, Shanghai, China, May 22-25, vol 1, pp. 51--54, ISBN 978-7-5608-4869-.

Rusilo, L.C., Madureira, S., Camargo, Z. (2011). Evaluating speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS). In *Proceedings of the 4rd ISCA Workshop ExLing*, 25-27 May, Paris, France, pp. 51.

Rusilo, L.C., Madureira, S., Camargo, Z. (2011). Evaluating Speech samples for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS). In *Proceedings of the 4rd ISCA Workshop ExLing 2011*; May 25-27; ISCA: Paris, pp. 51.

Stuchi, R.F., Nascimento, L.T., Bevilacqua, M.C. and Brito Neto, R.V. (2007). Linguagem oral de crianças com cinco anos de uso do implante coclear. Pró-Fono. Abr-Jun;19(2): pp. 167--76.

Tobey, E.A., Geers, A.E., Brenner, C.B., Altuna, D. And Gabbert, G. (2003). Factors associated with development of speech production skills in children implanted by age five. Ear & Hearing. Feb; 24(1). pp. 36--45.

Ubrig, M.T., Goffi-Gomez, M.V.S., Weber, R., Menezes, M.H.M., Nemr, N.K., Tsuji, D.H. and Tsuji, R.K. (2011). Voice Analysis of Postlingually Deaf Adults Pre- and Postcochlear Implantation. Journal of Voice, Nov; 25(6): pp. 692--99.

Wirz, S. The voice of the Deaf. (1986). In M. Fawcus (Ed.), *Voice Disorders and their Management*. Croom Helm.

Xu, L., Zhou, N., Chen, X., Li, Schultz, Zhao et al. (2009). Vocal singing by prelingually-deafened children with cochlear implant. Hearing Research, Jun. 255: pp. 129--34

Yoshinaga-Itano, C. (2003). From Screening to Early Identification and Intervention: Discovering Predictors to Successful Outcomes for Children With Significant Hearing Loss. J Deaf Stud Deaf Educ; 8(1): pp. 11--30.

SPEECH AND PRAGMATICS

The voice and emotion: a tribute to rhetoric

Eliana Amarante de M. MENDES

Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627 - Belo Horizonte/MG - Brasil
mendes@ufmg.br

Abstract

Various authors today are interested in the quality of the voice and, mainly, in the relation between emotion and voice. It is known that the human voice is an extremely flexible medium and one of the most important forms of transmitting and exchanging information between people and that the voice's messages tend to be more or less "colored" by emotional meanings which constitute an important source of voice variation. This topic has been widely researched, both theoretically and empirically, from diverse foci, but rarely in these studies is there found any mention of the great contribution of the classics to this theme. In relation to emotion, one cannot forget that Plato, Aristotle and the Stoics were the precursors of this study. But, mainly the rhetorical studies on *pathos* and persuasion and on the importance of the voice in the transmission and reception of emotions were chiefly important. In this study, I intend to return to the contributions of classical rhetoric on this theme, through a survey of the primary sources of this ancient art, with the aim of showing the great importance and opportunity of the classical studies today.

Keywords: voice; emotion; rhetoric.

1. Paper

We live in a period of great developments in linguistic studies on orality, phonetics, phonology and prosody, which, allied with the parallel developments in the studies of cognition, pragmatics, the corpora, the contributions of phono-audiology and the new technologies, are demonstrated to be more and more exhaustive, complex, and sophisticated.

In this context, important as well are the studies on the quality of the human voice. It is known that the human voice is an extremely flexible medium and one of the most important forms of transmitting and exchanging information between people. It is also known that the voice's messages tend to be more or less "colored" by emotional meanings, positive or negative, subtle or strong, which constitute an important source of voice variation. Thus, the voice acts like a powerful messenger not only of the linguistic content of speech, but also of the physiological and psychological state of the speaker.

The notion that changes in the expression of the voice can be caused by emotions is normally attributed to Charles Darwin. According to Darwin, as he demonstrated in his work *The Expression of the Emotions in Man and Animals* (1998: 235), emotional expression externalizes an individual's reaction and action propensity and passes this information for the social environment. Emotion is found in many species, particularly in mammalian and in species which have a complex social life based in interactions among their members. Body posture, facial features and vocalization are involved in emotion communication.

Concerning facial expression, Eckman (1973) garthered evidence on the universality across cultures. Likewise Izard (1971) and Ekman; Friesen; Ellsworth (1972) found in their studies rich information content of emotion in facial expressions.

According to Scherer (1995), research on animal communication developed by many scientists "demonstrated that in many species affective states, generally linked to changes in physiological arousal, are externalized in vocalizations and serve specific communication functions, often involving acoustic patterns that are similar across species. In close parallel to animal affect vocalizations, we still find rudiments of nonlinguistic human affect vocalizations, often referred to as 'interjections', such as 'ouch', 'ai', 'oh', 'yuck', etc".

Kleinpaul, in 1888, had already claimed that these reflexive "nature and feeling sounds" sound very much the same when uttered by speakers in different cultures. He distinguishes between interjections or exclamations expressing an emotional state and calls or shouts intentionally uttered for communicative reasons. (Kleinpaul, 1972 [1. ed. 1888] apud Scherer, 1995)

More recently, great interest for "emotion" arouse, as well as for its history, in reason of the development of areas as philosophy, sociology, communication studies, cultural studies, psychoanalysis, linguistics, and phonetics, among others.

But the current theories of the emotions do not share a consensus. For example, there are theories that divide the emotions into primary (basic) and secondary. Others include factors such as valence and activity; still others distinguish emotion from affect. They understand that affect is bio-physiological, is a more primitive response to a stimulus, and that emotion is of a cognitive nature.

Whether primary or secondary, affect or emotion, these manifestations, as mentioned, emerge in different forms: by facial expression, by gestures and by the quality of the voice.

Various authors today are interested in the quality of the voice and, mainly, in the relation between emotion and voice. This topic has been widely researched, both theoretically and empirically, from diverse *foci*. For example, there are studies that are dedicated to the study of the relation of voice, emotion and culture; of voice, emotion, and personality; of voice, emotion, and smile, of

voice, emotion, and gender. There are studies related to the production and reception of emotion in the voice and to the phonetic description of emotion in the voice. There are studies that analyze the interaction between acoustic data and linguistic data for the recognition of emotion in the voice; experiments for the automatic recognition of emotion in the voice, studies for emotion in speech variation, the simulation of emotion of the voice in speech synthesizers.

Despite all this wealth, rarely in these studies is there found any mention of the great contribution of the classics to this theme. Its primary reference to vocal sound as the vehicle of human utterance dates to at least the fourteenth century BC.

In relation to emotion, one cannot forget that Plato and mainly Aristotle were the precursors of this study, with the latter, due to his contribution, being considered the father of human psychology. Nor can the important contribution of the Stoics to the emotions be forgotten.

According to Plato, in his *The Republic* (Book X Part 1) the soul consists of three parts, three basic energies − reason, emotion, and appetite. Reason is the most valuable. Emotion and mainly appetite are considered "lower passions". For Plato, the soul that is governed by reason controls the emotions and appetites, that is, the lower passions must submit to reason.

In Plato's time the Sophists were philosophers who invented moral subterfuges to get people out of obligations or to excuse what was considered immoral behavior. Plato's theory of the soul, in contrast, defends that people must live morally.

The aristotelian rhetorical studies on *pathos* and persuasion were chiefly important. Aristotle defines rhetoric as "[…] the faculty of discovering in any particular case all of the available means of persuasion (Aristotle, *Rhetoric*, I, 2).

For him, there are three means of persuasion: appeals to *logos*, to *ethos* and to *pathos*.

Concerning *pathos* Aristotle says:

> "*The Emotions are all those feelings that so change men as to affect their judgements, and that are also attended by pain or pleasure.* Such are anger, pity, fear and the like, with their opposites. We must arrange what we have to say about each of them under three heads. Take, for instance, the emotion of anger: here we must discover (1) what the state of mind of angry people is, (2) who the people are with whom they usually get angry, and (3) on what grounds they get angry with them. It is not enough to know one or even two of these points; unless we know all three, we shall be unable to arouse anger in any one. The same is true of the other emotions" (Aristotle, *Rhetoric*, II, 1, emphasis added).

> "[…] *persuasion may come through the hearers, when the speech stirs their emotions.* Our judgements when we are pleased and friendly are not the same as when we are pained and

hostile. It is towards producing these effects, as we maintain, that present-day writers on rhetoric direct the whole of their efforts" (Aristotle, *Rhetoric* ,I,2, emphasis added).

Beyond these two great philosophers, also the Stoics, contemporaries of Aristotle, had interested for the emotion. But, differently of Aristotle, they thought that the emotions must be prevented, and according to this point of view, that the language would have to be neutral.

The Stoics were the first philosophers that defined passion. Considering the different facets of the term, they defined passion as:

1. An excessive impulse;
2. An impulse disobedient to reason;
3. A false judgment or opinion;
4. A fluttering of the soul.

The first two definitions saw passion as a kind of impulse. The first of these focuses on force. The second, as Chrisippus said, "passion is like a person running downhill and unable to stop at will." The third and fourth definitions emphasize the logical side of the term. According to these definitions, passions are contrary to reason because they are unruly, based on equivocation or erroneous opinions (Schmitter, 2010).

These earlier sources deeply influenced the early modern studies of the passions. Particularly Aristotle was very important influencing many theories of emotion in this period. But Stoicism and the neo-Stoicism (16th century) also influenced the early modern theories of emotion.

In this period, the philosophers used diverse terms for discussing the emotions. Perhaps because of the influence of Descartes (*Passions of the Soul*, 1649) the most used term was "passion". But others terms were also common: 'affect', 'sentiment', perturbation' and 'emotion' (Schmitter, 2010).

The practice of creating long lists of emotions and the many forms of classification are also indebted to these early sources – "*all without anything like citation of sources.*" (Schmitter, 2010, emphasis added).

But, concerning the relation voice and emotion − the importance of the voice in the transmission and reception of emotions − , the rhetorical studies were undoubtedly the most important.

For Aristotle,

> "*It is, essentially, a matter of the right management of the voice to express the various emotions* -- of speaking loudly, softly, or between the two; of high, low, or intermediate pitch; of the various rhythms that suit various subjects. These are the three things -- volume of sound, modulation of pitch, and rhythm -- that a speaker bears in mind. It is those who do bear them in mind who usually win prizes in the dramatic contests; and just as in drama the actors now count for more than the poets, so it is in the contests of public life, owing to the defects of

our political institutions" (Aristotle, *Rhetoric*, III, 1.4, emphasis added).

And Aristotle advises, in order to persuade the audience:

"[…] if your words are harsh, you should not extend this harshness to your voice and your countenance and have everything else in keeping. If you do, the artificial character of each detail becomes apparent; whereas if you adopt one device and not another, you are using art all the same and yet nobody notices it. (To be sure, if mild sentiments are expressed in harsh tones and harsh sentiments in mild tones, you become comparatively unconvincing.) Compound words, fairly plentiful epithets, and strange words best suit an emotional speech. We forgive an angry man for talking about a wrong as 'heaven-high' or 'colossal'; and we excuse such language when the speaker has his hearers already in his hands and has stirred them deeply either by praise or blame or anger or affection, as Isocrates, for instance, does at the end of his Panegyric, with his 'name and fame' and 'in that they brooked'. Men do speak in this strain when they are deeply stirred, and so, once the audience is in a like state of feeling, approval of course follows. This is why such language is fitting in poetry, which is an inspired thing" (ARISTOTLE, *Rhetoric*, III, 7).

However, although the undeniable importance of Aristotle, it was mainly Cicero, in his work *De Oratore*, who applied the aristotelian ideas, showing how the orator can use the resources to move an auditorium, including the role of the orator's voice.

"Now nothing in oratory, Catulus, is more important than to win for the orator the favour of his hearer, and to have the latter so affected as to be swayed by something resembling a mental impulse or emotion, rather than by judgement or deliberation. *For men decide far more problems by hate, or love, or lust, or rage, or sorrow, or joy, or hope, or fear, or illusion, or some other inward emotion, than by reality, or authority, or any legal standard, or judicial precedent, or statute*" (Cicero, *De Oratore*, II, 178, emphasis added).

And Cicero continues:

"Now, since the emotions which eloquence has to excite in the minds of the tribunal, or whatever other audience we may be addressing, are most commonly *love, hate, wrath, jealousy, compassion, hope, joy, fear or vexation,* we observe that love is won if you are thought to be upholding the interests of your audience, or to be working for good men, or at any rate for such as that audience deems good and useful. For this last impression more readily wins love, and the protection of the righteous; and the holding-out

of a hope of advantage to come is more effective than the recital of past benefit". (Cicero, *De Oratore*, II, 206, emphasis added)

He advises us:

"For it is not easy to succeed in making an arbitrator angry with the right party, if you yourself seem to treat the affair with indifference; or in making him hate the right party, unless he first sees you on fire with hatred yourself; nor will he be prompted to compassion, unless you have shown him the tokens of your own grief by word, sentiment, tone of voice, look and even by loud lamentation. For just as there is no substance so ready to take fire, as to be capable of generating flame without the application of a spark, so also there is no mind so ready to absorb an orator's influence, as to be inflammable when the assailing speaker is not himself aglow with passion" (Cicero, *De Oratore*, II, 190).

Also Quintilian, in *Institutio Oratore*, following Cicero, but dedicated mainly to the teaching of rhetoric, mentions more than 130 times the term "voice": its importance for the orator and its importance for the persuasion through the *pathos*.

"Now I ask you whether it is not absolutely necessary for the orator to be acquainted with all these methods of expression which are concerned firstly with gesture, secondly with the arrangement of words *and thirdly with the inflexions of the voice, of which a great variety are required in pleading. But eloquence does vary both tone and rhythm, expressing sublime thoughts with elevation, pleasing thoughts with sweetness, and ordinary with gentle utterance, and in every expression of its art is in sympathy with the emotions of which it is the mouthpiece*" (Quintilian, *Institutio Oratore*, I,24, emphasis added).

And Quintilian goes on:

"It is by the raising, lowering or inflexion of the voice that the orator stirs the emotions of his hearers, *and the measure, if I may repeat the term, of voice or phrase differs according as we wish to rouse the indignation or the pity of the judge.* For, as we know, different emotions are roused even by the various musical instruments, which are incapable of reproducing speech" (Quintilian, *Institutio Oratore*, I,25, emphasis added).

As can be seen, rhetorical works such as these by Aristotle, Cicero and Quintilian provided (and still provides) a great deal of material for taxonomizing and manipulating the emotions.

But rhetoric was also soundly rejected by some of the most famous philosophers, starting with Descartes. Under the influence of the positivism of Descartes, in

discussions of the mind people have believed that logic can function well only in the absence of emotion, that emotion interferes with reasoning ability. Many philosophers and scientists, even today, are dubious about the role of emotion in the mind. (Pfeifer; Scheier, 1999).

Although the early rhetoricians have claimed that powerful emotional oratory, using voice effects beyond verbal appeal is able to induce emotion, and such effects seem evident, modern scientists require empirical evidence that, indeed, listeners are able to correctly recognize the speaker's emotional state from vocal cues alone, independent of information from situational context or other expressive cues, such as facial expressions, gestures, or posture. So far, these scientists placed the emphasis on the recognition of a speaker's emotion from the voice. They assume that there is a clear criterion for the nature of the emotion present (or, as in most research studies, of an actor's encoding intention). (Scherer, 1995)

According to Copeland (2012), we must recognize that the history of rhetoric opens another window onto the historicized understanding of the emotions − a window into the past :

> "[…] and current interest in historicizing emotional responses underscores the continuing relevance of rhetorical thought, whether in its pre-modern formations or the broader cultural constructions of rhetoric in our own era. The opportunities are wide open for thinking concretely and historically about rhetoric's role in mobilizing and giving formal expression to the passions" (Copeland, 2012).

As could be seen, it is indisputed the importance of the contribution of the classical rhetoric for the studies on emotion and on voice and emotion.

But it was not my intention to disparage the present state of art – there are very important studies, mainly the ones that subsidize the development of the voice sinthezisers and recognizers and the ones concerning emotional intelligence. My intention was to contribute to the recognition of the importance and current relevance of the work of the ancients, showing that there are many today, to a certain extend, "reinventing the wheel."

In the words of Kelly, (1969):

> "There has been a vague feeling that modern experts have spent their time in discovering what other have forgotten; but as most of the documents are in Latin, [and Greek and not all documents are translated into our modern languages] moderns find it difficult to go to original sources. In any case, *much that is being clamed as revolutionary in this century is merely a rethinking and renaming of earlier ideas and procedures*" (emphasis added).

2. References

Abelin, A. (2008). Anger or fear? Cross-Cultural multimodal interpretation of emotional expressions. In: K. Izdebski (Ed) *Emotions in the human voice*. Vol. I. Plural Publishing. San Diego. pp. 65--73

Aristotle. (2000). *Art of rhetoric*. Trad. de J.H. Freese. Cambridge, London: Harvard University Press.

Cicero. (1942). *De oratore* . 2 vols. Trans. H. Rackham. The Loeb Classical Library. Cambridge: Harvard University Press.

Cicero. (1949). *De inventione*. Trans. H.M. Hubbell. Loeb Classical Library.

Copeland, Rita. (2012). *Ancient and Medieval Rhetorical Thought about the Emotions.* Available at: <http://www.bbk.ac.uk>. Access: 20/02/2012.

Darwin, Charles, (1998) [1896]. *The Expression of the Emotions in Man and Animals*, introduction, Notes and Commentaries by Paul Ekman. London: Harper Collins.

Ekman P., Friesen, W.V. and Ellsworth, P.C. (1972). *Emotion in the human face: guidelines for research and an integration of findings*. New York: Pergamon Press.

Ekman P. (1973). Darwin and cross-cultural studies of facial expression. In P. Ekman, (Ed.), *Darwin and facial expression.* New York: Academic Press, pp.1--83.

Guerrero, L.K., Andersen P.A. and Trost, M.R. (1998). Communication an emotion: basic concepts and approaches. In P.A. Andersen, L.K. Guerrero (Eds.), *Handbook of communication and emotion. Research, theory, applications and contexts.* USA: Academic Press, pp. 3--27.

Izard C.E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.

Kleinpaul R. (1972). [1.ed. 1888]. *Sprache ohne Worte. ldee einer allgemeinen Wissenschaft der Sprache.*In P. Ekman (Ed.), *Emotion in the human face*. Cambridge and New York: Cambridge University Press.

Kelly, L.G. (1969). *25 Centuries of Language Teaching: an Inquiry into the Science, Art and Development of Language Teaching Methodology, 500 B.C. – 1969.* Howley, MA: Newbury House. .

Murray, I.R., Arnott, J.L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of The Acoustical Society of America,* pp. 1097--1108.

Pfeifer, R., Scheier, C. (1999). *Understanding Intelligence.* Cambridge, Massachusetts: MIT Press.

Salomon, R.C. (1998). Philosophy of emotions. In E.Craig (Ed.), *Routledge Encyclopedia of Philosophy.* Vol. III.London and New York: Routledge, pp. 285--290.

Scherer, K.R. (1995). Expression of Emotion in Voiceand Music. *Journal of the voice*. Vol. 9, No. 3,.Lippincott-Raven Publishers. Philadelphia, pp. 235--248.

Schmitter, A.M. (2010). 17th and 18th Century Theories of Emotions. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2010 Edition). Available at: <http://plato.stanford.edu/archives/win2010/entries/emotions-17th18th/>.

Italians speaking English: the contribution of verbal and non-verbal behavior

Maria Grazia BUSÀ, Luca ROGNONI
Università di Padova, Italy
mariagrazia.busa@unipd.it, luca.rognoni@studenti.unipd.it

Abstract

This paper reports the preliminary findings of an investigation of the transfer and interpretation of non-verbal features from the L1 to the L2, focusing in particular on Italian speakers of English. The following hypotheses were tested: 1) Italian speakers of English transfer non-verbal features (i.e., gestures) from their L1 into the L2; 2) the transferred non-verbal features are not understood correctly by non-native Italian speakers. The paper also presents a protocol for eliciting the production and evaluation of emblems in L2 communication. 10 (Northern) Italian speakers of English were filmed during two speech tasks that were expected to elicit their use of emblems, that is: 1) the retelling of a fable; and 2) the enactment of a short dialogue. From these audio-video recordings, short video clips were extracted to create the stimuli for a two-part visual perception study aimed at getting evaluations of the speakers' gestures. In the first part, Italian native speakers (INS) and English native speakers (ENS) watched muted productions of INSs and were asked to tell what language was spoken in the clips. In the second part, the same subjects were asked to choose the correct meaning of selected gestures presented in the clips. The results suggest that INS recognize and correctly understand the meaning of the gestures produced by Italians when speaking English. ENS, however, do not interpret the meaning of Italian emblems correctly. This may lead to misunderstandings in L2 communication.

Keywords: Emblems; transfer of non verbal-features; Italian L1; English L2.

1. Introduction

In communication, a great deal of meaning is exchanged through non-verbal language. This includes prosodic aspects of the speech signal (pitch, voice quality, tone of voice, volume, etc.), as well as body language (eye gaze, facial expressions, hand gestures and body movements) (Mehrabian, 1972).

While there may be a universal basis that cuts across cultural and linguistic differences, non-verbal behavior is, to a large extent, culture specific. Thus, individuals learn it as part of the process of learning to communicate in a socio-linguistic community (Ekman, 1972; Feldman & Rime, 1991; Gudykunst & Mody, 2001; Harper *et al*., 1978; Kendon, 1981). It is therefore not surprising that speakers should transfer the non-verbal behavior acquired during their first language acquisition to the second language when they learn it and use it. In fact, recent research has proposed that non-verbal behavior should be studied as part of the interlanguage of an L2 learner (e.g., Gullberg, 2006; Pika *et al*., 2006).

As with any aspect of linguistic behavior, non-verbal behavior that is not congruent with the one of the target language may have an effect on the outcome of cross-linguistic communication. This is because cultures differ in the semantic meaning attributed to body postures, interpersonal space, and all other components of non-verbal behaviors, which comprise an important part of the communication process (Burgoon & Bacue, 2003; Matsumoto, 2006; Wang & Li, 2007). Also, the use of heavy gesturing during speech may be common and/or accepted in some linguistic communities but be considered distracting, cause annoyance to the listener, or project an image of the speaker of which the latter may be unaware of (Axtell, 1991; Efron, 1972; Ekman & Friesen, 1969; Graham & Argyle, 1975; Okada, & Brosnahan, 1990).

However, much is still to be learned about how non-native speakers' non-verbal behavior contributes to the meaning and interpretation of crosslinguistic communication, and to what extent it may affect it. To shed light on this important issue, more research is needed to investigate the interplay of linguistic and non-linguistic features in interlinguistic communication. Also, protocols should be devised to study the interpretation of non-verbal language experimentally.

The aim of this paper is to provide a preliminary investigation of how Italian non-verbal behavior in English L2 is interpreted by English native speakers. The following hypotheses were tested: 1) Italian speakers of English transfer non-verbal features (i.e., gestures) from their L1 into the L2; 2) the transferred non-verbal features are not understood correctly by non-native Italian speakers. This study also presents an experimental protocol that can be used for eliciting the production and evaluation of emblems in L2 communication.

2. About Italian Gestures

Italian has been defined as a high frequency gesture language (Pika *et al*., 2006). This means that gestures play a crucial role in conveying meaning and pragmatic force. Italians especially use emblems, that is, gestures that have an arbitrary connection with a meaning (i.e., substitute for words or expressions) (Poggi & Magno Caldognetto, 1997; Kendon, 2004). Emblems are culture- and language-specific, and so are unlikely to be interpreted correctly by people that are not familiar with them. The richness of the repertoire of Italian emblems is evidenced by the wide variety of "Italian gesture dictionaries" (available both online and on paper) aimed at helping the traveler to Italy to understand the spoken language.

Italians also use Italian emblems when speaking an

L2, assuming that the meaning of their gestures will be understood by their interlocutors. It is possible, however, that the use of Italians' emblems in the L2 may not be understood by non-native Italians. In addition, it may contribute to reinforcing the stereotypes of Italians being people who gesture a lot when they're speaking.

The aim of this study is to apply an experimental protocol to test whether Italians can be recognized as such for their gesturing and whether Italian gestures in English are in fact understood by English-native speakers.

3.   The experiment

A study was conducted to test the following hypotheses: 1) Italians transfer culture- and language-specific emblems from their L1 to the L2; and 2) the use of culture- and language-specific emblems is not understood correctly by speakers of different cultures/languages.

3.1   Selected emblems

Based on the first author's observation of her students' gesturing patterns when speaking English L2 in class, two frequently used emblems were targeted for the experiment. These were:

- The "Once Upon A Time" gesture (OUAT) (Fig. 1);
- The "What Are You Doing?" gesture (WAYD) (Fig. 2)

Figure 1: A speaker using the emblem meaning "Once upon a time" (OUAT)

Figure 2: A speaker using the emblem meaning "What are you doing?" (WAYD)

Both of these emblems can be considered part of the Italian language, and have been described in the literature on Italian gestures (Diadori, 1991; Poggi & Magno Caldognetto, 1997; Caon, 2010).

3.2   Emblem elicitation

To study emblems experimentally, the first problem to face is how to elicit a reasonably large number of any single type of emblem so that this can be part of a structured corpus and can be used in production and/or evaluation studies. A widely accepted elicitation protocol that has been used in gesture studies is based on the narration of the events seen in a short cartoon (McNeill, 1992). However, while this protocol is suitable to elicit iconic and co-speech gestures, it is not very effective to elicit emblems. In addition, this method is best suited to be used with native speakers or highly proficient L2 speakers, while L2 speakers with low levels of proficiency may not have the linguistic skills necessary to tell the details of a story they have watched.

Thus, to elicit the target emblems, two tasks were used. In the first task, the speakers were asked to learn, and re-tell aloud, a version of the Aesop's fable "The Fox and the Crow", adapted by the authors. This task was used to elicit the OUAT emblem, triggered by the narration of the events in the past. In the second task, the speakers had to learn and enact a short dialogue picturing an everyday situation ("A meeting at the pub") written by the authors. This task was used to elicit the WAYD emblem, triggered by the question-and answer exchanges in the dialogue. In both tasks, the speakers were instructed to speak and act as naturally and expressively as possible.

Both sets of productions were recorded using a digital video camera and were then transferred onto a computer.

The subjects were 10 graduate female students from the University of Padua. They were all Italian native speakers, born and living in the Veneto region, in North-Eastern Italy. Their average age was 23.

3.2   Evaluation of overall gestures and emblems

Two experiments were created to test: 1) whether an Italian speaker's overall gesturing may look 'foreign' to non-native Italian speakers; and 2) whether the Italians' emblems are understood correctly by non-native Italian speakers.

3.2.1.   Stimuli

The video recordings obtained in the elicitation task were used to create clips (with Final Cut Pro) for the two evaluation tasks described below.

The first clip, used in the first evaluation task, consisted of one muted 19-second video showing two speakers interacting with gestures in a dialogue.

The second clip, used in the second evaluation task, consisted of two repetitions in a row of each of the following muted stimuli: 3 samples of the OUAT emblem, 3 samples of the WAYD emblem, and 3 gestures that were used as distractors in the stimulus presentation sequence. The gestures that were selected to work as distractors were iconic gestures recurrent in the data, as they had been produced by some of the subjects to describe the landing of the crow on the cheese in the fable "The Fox and the Crow". The resulting set consisted, thus, of a total

of 9 stimuli produced 2 times (9x2) by 9 different speakers. The total duration of the clip was approximately 5 minutes.

3.2.2. Procedure and subjects

The two clips were cropped together and presented as part 1 and part 2 of a short video, embedded in the web-based survey and test presentation tool *eSurveysPro* (http://www.esurveyspro.com/). The evaluations were obtained via web in Italy and abroad.

The clips were evaluated by a group of 30 English native speakers (INS, average age: 36) and a group of two 30 Italian native speakers (INS) for control (average age: 27). In both groups, the subjects were either university students or professionals.

3.2.3. Evaluation Task 1

In this task, designed to test whether an Italian speaker's overall gesturing may look 'foreign' to non-native Italian speakers, the subjects were presented with the muted 19-second video clip showing two speakers interacting with gestures in a dialogue.

After watching the clip, the subjects were asked to guess the language spoken by the people in the video by choosing between 5 options: "Italian", "Spanish", "German", "English", "I don't know".

3.2.4. Evaluation Task 2

In this task, designed to test whether the Italians' emblems are understood correctly by non-native Italian speakers, the subjects were presented with the clip showing the 2 target emblems and the distractor. After each stimulus, the subjects were asked to select the meaning of the speaker's gesture from 5 options: "A long time ago", "I'm hungry", "It's hot in here", "What's the problem?", "No meaning".

4. Results

4.1 Transfer of emblems

The procedure we used to elicit emblems proved successful. The target emblem OUAT was obtained in 4 out of 10 instances, while the WAYD emblem was produced in 4 out of 5 dialogues. Because emblems are used in connection with a particular meaning, to trigger emblems it is necessary to create elicitation tasks where the situation will make specific reference to the targeted meaning. Thus, in our case, the fable's beginning 'Once upon a time' created the condition for the production of the emblem meaning 'a long time ago'. On the other hand, in the mini-dialogue, the subjects were instructed to ask each other questions related to why they were in the pub at that particular time and day; the amount of questioning involved in the dialogue triggered the production of the emblem meaning 'why/what'. In both cases, our previous attempts at eliciting emblems using the widely accepted protocol for the elicitation of iconic gestures (McNeill, 1992) had not been successful.

The results of the elicitation tasks show that, as expected, the Italian subjects did use Italian emblems when speaking English. Also as expected, the subjects did not seem to be aware that they were using Italian gestures in English whose meaning might not be understood by non-Italian speakers.

4.2 Transfer of emblems

4.2.1. Evaluation Task 1

The results of the first evaluation task show a clear difference in the responses given by the INS, on the one side, and the ENS, on the other side. While 50% of the INS thought that the muted speakers in the video clip were speaking Italian (although the dialogue was, in fact, in English), the ENS gave their answers randomly. The percentages of answers given for each category by the INS and the ENS are shown in Figures 3 and 4.
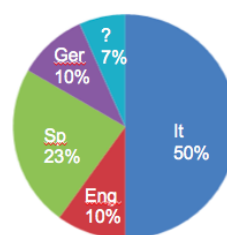
Figure 3: Italian Native Speakers' responses, by percentage, in the Evaluation Task 1
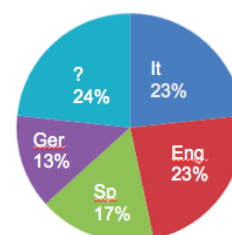
Figure 4: English Native Speakers' responses, by percentage, in the Evaluation Task 1

The results of the Evaluation Task 1 give support to the hypothesis that a speaker may identify correctly other speakers of his/her native language, based on their use of gestures; conversely, speakers using gesturing following rules that are not those of the native language are identified as foreigners.

4.2.2. Evaluation Task 2

The results of the second evaluation task also show a clear difference in the evaluations made by the INS, on the one side, and the ENS, on the other side. These results are shown in Figure 5. The INS identified the correct meaning of the OUAT and the WAYD emblems in 91% cases, and identified the distractor correctly as carrying no particular meaning in 80% cases. The ENS gave much lower percentages of correct responses for both the emblems (the OUAT was identified correctly in 31% cases, the WAYD in 68% cases), and the distractor (53% correct responses). The difference in the general accuracy scores

for the performances of the INS' and the ENS' evaluation tasks was significant to a paired t-test (mean ENS: 60.55556, mean INS: 87.44444, t = 4.8634, df = 8, p-value = 0.001250).
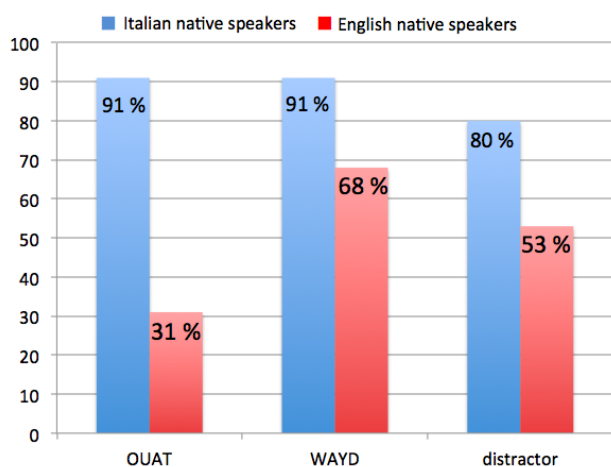
Figure 5: Percentages of correct identification of the emblems by the two speakers' groups

The results of this task, showing that the ENS perform far below the INS, provide support for the hypothesis that English speakers do not understand the Italian gestures that are transferred in the L2.

However, it was expected that the difference, between the INS and the ENS, in the percentages of correct emblem identification would be greater for both the WAYD emblem and the distractor. We suspect this result is due to some glitch in the methodological procedure used for this evaluation task. In the first place, for the WAYD emblem, the clip showed two speakers interacting and discussing with each other, whereas for both the OUAT emblem and the distractor the clip showed only one person gesturing. This may have led the subjects to choose the correct response -'What's the problem?'- for the emblem WAYD even when they did not in fact know its meaning. As for the distractor, different results might have been obtained if the choice 'I don't know' had been a selection option instead of 'No meaning'.

In spite of these glitches, we do believe that the procedure we devised for eliciting the interpretation of the meaning of the emblems can be used successfully in the analysis of L2 gestures. Future research will correct for the methodological problems encountered in the present study.

5. Conclusions

In a global world, the importance of non-verbal language in intercultural and interlinguistic communication should not be underestimated. However, there is a great deal that we still do not know about the meaning L2 speakers convey, inadvertently and unintentionally, through the gestures they transfer from the L1 into the L2. More studies are needed to understand the meaning of L2 gestures in L2 communication.

This study shows that Italian speakers transfer non-verbal features from their L1 into the L2, and that the transferred non-verbal features are perceived as foreign, and are not well understood by the target language speakers. This may have consequences in interlinguistic communication by affecting the successful outcome of interactions between speakers of different mother tongues. Thus, non-verbal behavior should be taught and learned in L2 courses as part of the learners' attainment of a complete linguistic competence.

This study also suggests that the use and interpretation of emblems can and should be studied experimentally. A protocol for the elicitation and evaluation of emblems is proposed here, which, with some corrections, appears suitable to be used in experimental research on gestures.

6. References

Axtell, R.E. (1991). *Gestures – The Dos and Taboos of body language around the world*. New York: John Wiley & Sons.

Burgoon, J.K., Bacue, A.E. (2003). Nonverbal communication skills. In J.O. Greene & B.R. Burleson (Eds.), *Handbook of communication and social interaction skills*. Mahwah, NJ: Erlbaum, pp. 179--219.

Caon, F. (2010). *Dizionario dei gesti degli Italiani. Una prospettiva interculturale*. Perugia: Guerra edizioni.

Diadori, P. (1991). *Senza Parole. 100 gesti degli italiani*. Roma: Bonacci editore.

Efron, D. (1972). *Gesture, race, and culture*. Hague: Mouton.

Ekman, P. (1972). Universal and cultural differences in facial expression of emotion. In J.R. Cole (Ed.), *Nebraska symposium on motivation*, 1971. Lincoln: Nebraska University Press, pp. 207--283.

Ekman, P., Friesen, W.V. (1969). The repertoire of non-verbal behavior: Categories, origins, usage, and coding. *Semeiotica*, *1*, pp. 49--98.

Feldman, R.S., Rime, B. (eds.) 1991. *Fundamentals of non-verbal behavior*. Cambridge: Cambridge University Press.

Graham, J.A., Argyle, M. (1975). A cross cultural study of the communication of extra-verbal meaning by gestures. In *International Journal of Psychology*, 10(1), pp. 57--67.

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). In *International Review of Applied Linguistics*, 44(2), pp. 103--124.

Gudykunst, W.B., Mody, B. (Eds.). (2001). *Handbook of international and intercultural communication*. Newbury Park, CA: Sage.

Harper, R.G., Wiens, A.N. and Matarazzo, J.D. (1978). *Nonverbal communication: The state of the art*. New York: John Wiley & Sons.

Kendon, A. (Ed.). (1981). *Nonverbal Communication, Interaction, and Gesture: Selections from Semiotica (Approaches to Semiotics)*. The Hague: Mouton and Co.

Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.

McNeill, D. (1992). *Hand and mind.* Chicago: Chicago University Press.

Matsumoto, D. (2006). Culture and nonverbal behavior. In V. Manusov, M. Patterson (Eds.). *Handbook of nonverbal communication*. Thousand Oaks, CA: Sage, pp. 219--235.

Mehrabian, A. (1972). *Non-Verbal Communication*. Chicago, Illinois: Aldine-Atherton.

Okada, T., Brosnahan, L. (1990). *Japanese and English gesture: Contrastive nonverbal communication*. Tokyo: Taishukan Press.

Pika, S., Nicoladis, E. and Marentette, P.F. (2006). A cross-cultural study on the use of gestures: Evidence for cross-linguistic transfer? *Bilingualism: Language and Cognition, 9*, pp. 319--327.

Poggi, I., Magno Caldognetto, E. (1997). *Mani che parlano. Gesti e psicologia della comunicazione*. Padova: Unipress.

Wang, D.-H., Li, H. (2007). Nonverbal language in cross-cultural communication. *Sino-US English Teaching*, 4(10), pp. 66--70.

Paired fiction writing: the dialogal text as a structure that triggers "verbal erasure"

Eduardo CALIL, Cristina FELIPETO

Federal University of Alagoas

Rua Presciliano Sarmento, s/n, residencial Oceanis, quara B, lote 3, 57044-130, Maceió, Alagoas, Brazil

eduardocalil@hotmail.com, crisfelipeto@hotmail.com

Abstract

Considering contestation from a dialogic and socio-historical point of view, this paper describes some types of comments made by a student about lexical items proposed by her partner during paired fiction writing processes. The nature of this investigation is quantitative, qualitative and longitudinal. For two years we followed the teacher's proposals of text production in the classroom. We adopted ethnolinguistic methodological procedures. Once a month, we filmed two students (6 to 7 years old) who were good friends and had recently become literate and our corpus was composed of 16 proposals of text production. We identified the occurrence of comments with structures of autonymic modalization enunciations in which the pupils return to a term expressed earlier and comment on it, justifying why it could or could not be written in the current text. Our results indicate that the meta-enunciative characteristic of the comments focuses on specific elements of the narrative, such as story titles, character names and terms related to the characterization of these characters. In addition, we found that the contestation between the students, expressed by the comment that follows the word spoken by the other, highlights the meaning that a term has for each of them.

Keywords: school; writing; narrative; dialogism; autonymic modalization; memory; text generation.

1. Introduction

Investigations into collaborative writing in the school context (Daiute & Dalton, 1993; Vass, 2002; Vass *et al.*, 2008; Dale, 1996; Calil, 2008, 2009; Felipeto, 2008) highlight the importance of the social context and the preservation of its ecological conditions for the analysis of its core components (planning, formulation and revision), as well as its creative processes, in real situations of use. Among the different types of didactic situations, those that choose paired collaborative writing argue that peer interaction differs in many aspects from teacher-student interaction, mainly because the pairs do not intentionally and deliberately assume the position of "teacher," the one who will teach and assess her students.

Another significant difference lies the fact that collaborative writing promotes "contestation," i.e., the emergence of a confrontation of points of view, when students reflect on what was said, questioning their partner. This may elicit a variety of comments involving explanations, arguments, and justifications about the text that is being written. As Daiute noted, "The partner would then participate in constructing an opening sentence, for example, or raise questions about it – whether such a sequence should be there at all or whether it should be phrased in some other way" (Daiute & Dalton, 1993: 320).

"Contestation"[1] presupposes "dispute," "dissension," or "controversy," and points to the negotiation of meanings between students. Although this confrontation, in this specific interactional situation, may indicate what each student is thinking, some types of comments refer to the meaning of what was said. Therefore, considering the importance of contestation in

the collaborative writing process, but delimited by the dialogic and socio-historical field (Bakhtin, 1986), our interest lies in the genetic processes of fictional writing by beginning writers. The comments made in these co-enunciative conditions are of paramount importance in understanding these processes.

Our studies[2] (Calil, 2003; Calil & Felipeto, 2006; Felipeto, 2008, among others) on Textual Genetics (Grésillon, 1994) and Enunciation Linguistics (Authier-Revuz, 1995, 2004), discuss writing in real time, in the context of the classroom, based on these didactic practices of collaborative writing. By focusing on the process of text creation, we value the written erasures, and above all the oral erasures[3] left throughout the manuscript in progress. Through the filmed record (videotape) of the ecological situation in which two newly literate students make up fictional narratives together, we highlight the importance of spontaneous speech in the dialogic text (Bres, 2005) that is established. In this paper we discuss specifically how a dyad, followed for two years, comments about the meaning of some terms that emerge as they make up these fictional stories. We will begin by

---

[1] In her work, Felipeto (2008: 17) calls this moment "altercation," but defends its importance in the production of "language misunderstanding" (Milner, 1978).

[2] These studies are linked to the School Writing Laboratory (L'ÂME) located at the Federal University of Alagoas (Brazil), whose objective is the documentation, archiving and preservation of school manuscripts and writing processes originating from different school contexts.

[3] As described in Calil (2012), the verbal erasure is characterized by linguistic operations of "substitution," "addition" or "displacement" of the elements that may be part of the manuscript that is being produced. These erasures may involve the speech of the speaker herself or that of the interlocutor, accompanied or not by different kinds of comments. The peculiarity of this type of rephrasing stems from the fact that the properties of the written text genre interfere in the enunciative act of students that *say* something to be *written*. (Calil, 2003: 31-32).

presenting the frequency of these occurrences, and then analyze some forms of comments that these terms are given.

Based on this longitudinal *corpus*, our first hypothesis related to the qualitative study of the data was that, during the process of collaborative writing of fictional narratives, students produce verbal erasures linked to the meaning of a term. We call this type of erasure "Semantic Verbal Erasure", or simply "SVE." With respect to the quantitative nature of our data, our second hypothesis assumed that these comments would appear with greater frequency as the students appropriated the linguistic and formal properties of the genre in question. These hypotheses led us to describe and analyze this type of verbal erasure, indicating its occurrences in each writing process, the objects of discourse to which they referred, and the linguistic and enunciative structures presented by the students involved.

2. Dialogic text, spontaneous speech and autonymic modalization

From the enunciative standpoint, "dialogic text" (Bres, 2005; Bres & Nowakowska, 2006) – taken as a unit of analysis in these paired writing processes – is directly related with spontaneous speech. The interchange *in praesentia* of spoken exchanges, the successiveness of the statements, their breaks, digressions, pauses, hesitations, syntactic threads and thematic resumptions of the highlighted objects of discourse, marked primarily by the voice of each speaker in the here and now of his utterance, in a real, everyday and immediate situation, not planned or premeditated, are constituent elements of dialogic text. Add to these the immediate context and the conditions of production given socio-historically, the idiosyncratic expressive elements of each of the interlocutors (body movement, gestures, glances, facial expressions...) sitting face-to-face and engaged in shared and collaborative writing.

The "dialogical" condition, in which each speaker responds directly or indirectly to the utterance of the speaker, is proposed by Bres (2005) from a rereading of Bakhtinian dialogism. Related to interlocutive dialogism, the dialogic text created in the flow of speech of the interlocutors would include, among its multiple dialogical characteristics, the speaker's comments about what was said prior to his own or the other's utterance.

From the dialogic text recorded by camcorder, we will highlight the co-enunciative threads marked by the emergence of a term, its resumption, denial and comments, which are structured as follows:

a) Speaker A: [X].
    Speaker B: [X] (NO) + Z

b) Speaker A: [X].
    Speaker B: [X]?
    Speaker A: [X] (NO) + Z

In these structures, the formula "[X] (NO) + Z"

formalizes the statement that may be made about an uttered word. "X" represents a word or expression related to the object of discourse (OD) highlighted by one of the speakers.   The denial, which may or may not be linguistically marked, is usually followed by a comment. "Z" is the comment or gloss referring to the term uttered previously and therefore to the OD in question.

As we showed in our analysis of writing processes of fictional stories in Calil (2008), the OD refers to the elements of various orders (linguistic, narrative, textual, orthographic, communicational...), while the resumption of these elements by the interlocutor and his commentary may express a reflexive position about it. The resumption and semantic comment about what was said by the interlocutor indicates the recognition of a difference between the "sense of what was said" and the "sense of what was heard" and indicate, through the questioning and suspension of the use of X, the discovery, by the enunciator, "of 'something' that does not go unnoticed and to which his comment responds" (Authier-Revuz, 1995: 29). In other words, the SVE may elicit a type of comment whose structure resembles the enunciative non-coincidences   identified   and   described   by Authier-Revuz as forms of autonymic modalization in which the interlocutor recognizes the enunciative heterogeneity and seeks to mitigate it, in a deliberate effort of negotiation starting from the contestation of what was stated[4].

Thus, autonymic modalization, which is one of the forms of manifestation of the constitutive heterogeneity of speaking, has to do with the way in which the subject represents   and   demarcates   the   phenomena   of non-coincidence, which may appear in four different forms:

i.   Non-coincidence of words with themselves, in which the subject, in a number of ways, eliminates or admits other meanings of a word or of other words that, through the play of polysemy,   homonymy,   etc.,   affect   his utterance;

ii. Non-coincidence of the discourse with itself, in which the words of other(s) discourse(s) "present themselves," "invade" the discourse of the subject;

iii. Interlocutory non-coincidence in which the subject, in his relation with the utterance of the other, highlights in his own enunciation non-shared meanings, a distancing between an utterance that "is mine" and one that "is not mine" or, if convenient, that can be accepted, shared;

iv. Non-coincidence between words and things,

---

[4] Figueira (2003), in his study about the reflexive property of language in the speech of children, identifies some initial forms of autonymy at around age 4.

when it involves indicating that the words employed do not correspond exactly to the reality they should represent, culminating in the impossibility of an object being totally "completed" by the play of the designation.

As will be seen in the presentation and analysis of our data, verbal erasure may also occur through autonymic modalization, through repetition (resumption of another's words or one's own, involving the use of the term), with an additional comment about this use (reflective comment in which the mention of the use of "X" intervenes). Thus, we believe that SVE resembles the phenomenon of autonymic modalization in that its enunciation comprises two main components of modality: use and mention.

3.   Fictional stories and paired writing: didactic guidelines and methodological procedures

The choice of the paired writing process in the classroom context as an object of study requires an approximation between the investigative objectives and the didactic context [5] of which the school and the participating classroom are part. In this case study, a private school[6] in the city of São Paulo was selected, located in a middle class neighborhood whose residents have high purchasing power and access to cultural and consumer goods. The parents were architects, lawyers, university professors, businessmen, and liberal professionals (dentists, medical doctors, psychologists...) linked to the artistic (musicians, plastic artists, actors...) or political milieu.

A group of students were learning to read and write and were observed for two years. Among these students, we selected two girls (Isabel and Nara) who met the three criteria for their choice: they were friends inside and outside school; they were extroverted and articulate; and they were newly literate. In April 1991, when we started collecting data, Isabel was 6 years and 5 months old and Nara was 5 years and 10 months old. In November 1992, when we recorded the last proposal, Isabel was 8 years and 1 month old and Nara was 7 years and 5 months old. Sixteen text production proposals were filmed, six during the first year and ten during the second year[7], with the video recordings taking place on average every 30 days.

In all the proposals, the 1st and 2nd grade teachers

both followed a similar procedure: they usually talked about the stories that had already been written, pointed out some learning contents [8], and lastly presented the text production proposal. The genre chosen for the production of text was fictional narrative, which the teacher referred to as a "made up story." The majority of themes were free, without any indication of title, character or plot. The didactic procedures sought to encourage planning of the story, asking the students to agree about what they would write. After that, they would ask the teacher for pens and paper to write down the text.

The video recordings were later transcribed using the ELAN program, a tool that facilitates the synchronization of captured images and sound, and allows for the definition of tracks with linguistic types related to the chosen object of study. Considering dialogic text and the co-enunciative nature of verbal erasure, we sought to identify the semantic comments made by the dyad during the recorded writing processes.

4.   SVE, between quantity and quality

The two students participated actively in all the writing processes that resulted in their respective manuscripts. They discussed, invented and agreed upon character names, titles, plots, outcomes... narrative elements typical of traditional fictional narratives, such as the presence of "fairies," "stepmothers," "magic," "happy endings", mixed with other elements related to contemporary fictional narratives (comics, TV commercials, and modern children's literature). The articulation of these elements revealed some surprising and creative aspects, as shown in Calil (2009[9]).

SVE is one of those phenomena that reveal the text creation process, in that it highlights the competition between terms, occupying the same position in the syntagmatic chain to be written or indicating problems of unity of meaning when they refer to previous elements. A good example of this is Isabel's contestation of the term "*Zumbacalabumba!*" suggested by Nara to represent the noise a fairy makes. Immediately after Nara's statement, Isabel says: "*It's like this, listen! Let's make a more beautiful one, OK?! Zabumbacalabumba... for a fairy?*"

This SVE, accompanied by Isabel's comment, indicates that the value of "*Zumbacalabumba*" is not fitting, not suitable for the type of character, a fairy. It marks the difference between Nara's words and those of Isabel, causing the latter not only to reflect upon the relationship between the character and what characterizes it, but also and especially to look for a word that can

---

[5] We understand the "didactic context" as all that which characterizes a school, from its infrastructure to the school community involved, and including its socioeconomic and cultural conditions. Specifically, this context involves equally the didactic practice established between the teacher and her students.

[6] It should be noted that this school adopts "constructivist pedagogy" based on the ideas of Piaget and Vygotsky, and particularly so in regard to the teaching of reading and writing, according to the studies of Emília Ferreiro and Ana Teberosky in the 1980s (Ferreiro & Teberosky, 1985).

[7] The smaller number of recordings in the first year was because our data collecting started only in April 1991 and the fact that we missed three recordings due to technical sound problems.

[8] Mainly in the second year of data collection, these contents had to do with spelling, punctuation, separation of words, etc.

[9] We refer to the stories "The gluttonous queen," (original title in Portuguese: "A rainha comilona"); "The three chocolate milks and madam flavor" (original title in Portuguese: "Os três todinhos e a dona sabor"; where madam flavor stands for mother) and "The muddled F family" (original in Portuguese: "A família f atrapalhada", where F stands for the names of the father, mother and son in the story, which are "Fumo", "Fina" and "Fim", respectively), whose analysis revealed the wealth of these aspects.

ensure the unity of this relationship and the naming of the character's action.

It is this type of SVE and these forms of comments that we attempt to identify during the videotaped and transcribed proceedings. The graphs below indicate the number of SVEs per writing processes in each year of the data collection.
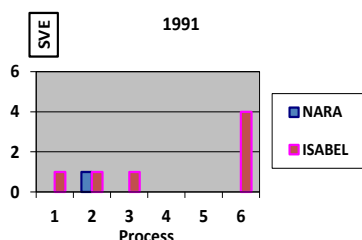
Graph 1: Comments per writing process in 1991.
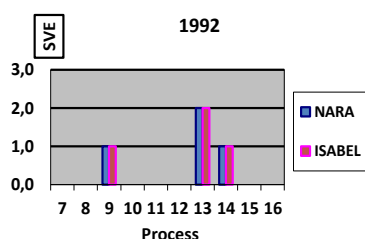
Graph 2: Comments per writing process in 1992.

An analysis of these graphs allows for a few significant considerations. First, we note that this type of verbal erasure with gloss is neither frequent nor systematic. Its occurrence is low, i.e., ranging from one and four events per writing process. In addition, SVE did not occur in most cases, i.e. the 4th, 5th, 7th, 8th, 10th, 11th, 12th, 15th and 16th processes were devoid of SVEs.

Three points should be noted in the processes in which the presence of SVEs was identified:

1. There was no increase in SVEs related to the learning time, i.e., there does not seem to be a direct relationship between the increase in the mastery of rules of grammar and text in written production, such as punctuation marks, paragraphing, use of uppercase and lowercase letters, assimilation of the spelling system, differentiation between direct and indirect discourse (the narrator and characters' lines), teaching objects valued by the school (and the teacher), emphasized over the two years, and this type of verbal erasure. In fact, from one year to the next, we find that the occurrence of SVE decreased from seven occurrences in six writing processes (1991) to five events in ten

processes (1992).

2. Unlike this trend, three SVEs were recorded in the first three writing processes (Nara and Isabel approximately 6 years old). In a single process recorded at the end of that year, there were four SVEs, all uttered by Isabel

3. Upon determining which of the two students produced more SVEs, we did not find a consistent predominance of one over the other. During the first year, Isabel made six SVEs compared to one by Nara, but in the following year Nara made three of the five SVEs.

5. Conclusion

The dialogue between the dyad favors contention, debate, confrontation, and also potentiates reflexivity about the word put into play, producing meta-enunciations, and thus indicating some important metalinguistic operations to understand the process of text creation by beginning students.

In general, the interaction between this dyad proved very useful in the production of verbal erasures. Specifically with respect to those that focus on the meaning of a term, semantic verbal erasures, we did not find a large number of erasures. However, the number of SVEs produced by the students appears to be related to the complexity that a reflexive comment involves as well as to the school period when formal issues such as grammar and textual rules, punctuation, paragraphing, etc., become relevant.

Peer interaction during these writing processes not only favors reflection about narrative elements, but also allows for the rediscovery of significant moments in the genetic processes of text creation by beginning students.

6. Acknowledgements

7. References

Authier-Revuz, J. (1995). *Ces mots qui ne vont pas de soi. Boucles réflexives et non coïncidences du dire*. Tome 1. Paris: Larousse.

Authier-Revuz, J. (2004). *Entre a transparencia e a opacidade: um estudo enunciativo do sentido*. Porto Alegre: Editora da Pontifícia Universidade Católica do Rio Grande do Sul.

Bakhtin, M.M. (1986). *Speech Genres and Other Late Essays*. Austin, TX: University of Texas Press.

Bres, J. (2005). Savoir de quoi on parle: dialogue, dialogal, dialogique; dialogisme, polyphonie. In J. Bres, P.P. Haillet, S. Mellet, H. Nolke and L. Rosier (Eds.), *Dialogisme, polyphonie: approches linguistiques*. Paris, De Boeck.duculot, pp. 47--61.

Bres, J., Nowakowska, A. (2006). Dialogisme: du principe à la matérialité discursive. In *Recherches*

linguistiques, 28, pp. 21--48.

Calil, E. (2003). Processus de création et ratures: analyses d'un processus d'écriture dans un texte rédige par deux écoliers. In *Langages & Société*, 103, pp. 31--55.

Calil, E. (2008). *Escutar o invisível: escritura & poesia na sala de aula*. São Paulo: Editora da Universidade do Estado de São Paulo.

Calil, E. (2009). *Autoria: a criança e a escrita de histórias inventadas*. Londrina: Editora da Univresidade Estadual de Londrina.

Calil, E. (2012). La rature orale en processos d'écriture en acte: lieu de tension et production du sens. In *Oralia*, 6, pp. 215--230.

Calil, E., Felipeto, C. (2006). Quand la rature (se) trompe: une analyse de l'activité métalinguistique. In *Langage & Société*, 117. Paris, pp. 63--86.

Daiute, C., Dalton, B. (1993). Collaboration between children learning to write: Can novices be masters? In *Cognition and Instruction*, 10, pp. 281--333.

Dale, H. (1996). The influence of co-authoring on the writing process. In *Journal of teaching writing*, 15(1), pp. 65--79.

Felipeto, C. (2008). *Rasura e equívoco no processo de escritura em sala de aula*. Londrina: Universidade Estadual de Londrina.

Ferreiro, E. & Teberosky, A. (1985). *Psicogênese da língua escrita*. Porto Alegre: Artes Médicas.

Figueira, R. A. (2003). La Propriété Réflexive du Langage dans le Parler de l´Enfant. Quelques Manisfestations du Fait Autonymique dans l´Acquisition du Langage. In J. Authier-Revuz, M. Doury and S. Reboul-Touré. (Eds.), *Parler des Mots. Le Fait Autonymique en Discours*. Paris, pp. 193--204.

Grésillon, A. (1994). *Eléments de Critique Génétique: lire les manuscrits modernes*. Paris: Presses Universitaires de France.

Milner, J.-C. (1978). *L'Amour de la langue*. Paris: Seuil.

Vass, E. (2002). Friendship and collaborative creative writing in the primary classroom. In *Journal of Computer Assisted Learning* 18, pp. 102--110.

Vass, E., Littleton, K., Miell, D. & Jones, A. (2008). The discourse of collaborative creative writing: Peer collaboration as a context for mutual inspiration. In *Thinking Skills and Creativity*, 3(3), pp. 192--202.

Il progetto LIRA: un *repository* multimediale per lo sviluppo delle competenze pragmatiche in parlanti non nativi d'italiano

Elena NUZZO, Greta ZANONI

Università di Verona; Università di Bologna
elena.nuzzo@univr.it, gzanoni@sslmit.unibo.it

Abstract

This paper discusses some of the issues concerning the preparation of a set of e-learning modules on how to use the Italian language appropriately from a pragmatic point of view. These modules are part of a wider project called LIRA – Lingua/cultura Italiana in Rete per l'Apprendimento (Italian language/culture for learning on the Net) involving four Universities (Bologna, Modena and Reggio Emilia, Perugia, and Verona). This project mainly aims at creating a multimedia repository of materials that can help the recovery, preservation and development of linguistic, pragmatic and cultural competences by second and third generation of Italians living abroad. After analysing the characteristics of the target users, this paper addresses one crucial issue associated with the teaching of pragmatics, namely, how to combine the intrinsic variability of this area with the need to resort to a standard reference system and to provide learners with clear corrective feedback. Then it briefly presents the materials and the activities included in the modules in order to show how LIRA deals with this and other issues related to the teaching of pragmatics.

Keywords: Multimedia repository; pragmatics; L2 Italian.

1. Il progetto[1]

Il progetto LIRA (*Lingua/cultura Italiana in Rete per l'Apprendimento*), cui partecipano le università di Perugia Stranieri, Bologna, Modena e Reggio Emilia, e Verona, ha l'obiettivo di favorire il recupero, il mantenimento e lo sviluppo di competenze pragmatiche e culturali da parte di italiani di seconda e terza generazione residenti all'estero tramite la creazione di un *repository* multimediale, ossia un ambiente intelligente di contenuti digitali. Una volta ultimato, questo strumento, fondato sulla condivisione di risorse multimediali, sulla costante interazione fra i membri della comunità virtuale e sulla loro partecipazione alla creazione dei contenuti, permetterà agli utenti di accedere a materiali adatti al loro profilo e altamente rappresentativi della lingua e della cultura italiane, e di auto-valutare i progressi nell'apprendimento. In questo contributo intendiamo presentare alcuni nodi teorici – e le relative ricadute didattiche – legati all'insegnamento di aspetti pragmatici dell'italiano L2 così come sono emersi nell'ambito del progetto e in particolar modo nel corso del lavoro svolto dalle unità di Bologna e Verona; non saranno invece trattati gli aspetti legati alla cultura italiana e al testing, di cui si occupano le altre due unità del progetto.

LIRA è un *repository* di materiali multimediali misto, ma prevalentemente orientato verso l'uso orale della lingua: anche se non mancano esempi di lingua scritta, i testi raccolti, accuratamente selezionati con lo scopo di mostrare alcune specificità linguistico-pragmatiche dell'italiano, sono costituiti soprattutto da brani di parlato. Si tratta di un *repository* con elevato grado di generalità: non è stata scelta una tipologia specifica di testi, perché si cerca di offrire all'utente una gamma il più possibile variegata di usi linguistici e di contesti. Anche se la grande varietà di testi presenti nel *repository* permette di considerare i materiali raccolti come rappresentativi di molti tratti e proprietà dell'italiano, rispettando quindi una delle caratteristiche dei corpora di linguistica ovvero la rappresentatività, LIRA non può essere considerato un corpus in quanto non ne soddisfa un altro requisito fondamentale: l'estensione. Inoltre, vale la pena di ricordare che i testi raccolti nel *repository* LIRA non sono codificati omogeneamente per essere interrogati in modo avanzato all'interno della piattaforma, a differenza di quanto accade per i corpora.

2. I destinatari

Si è detto che i destinatari principali di LIRA sono gli italiani di seconda e terza generazione residenti all'estero. Questi utenti sono per molti versi più assimilabili ad apprendenti intermedi o avanzati dell'italiano L2 che ai parlanti nativi, e, via via che il momento dell'insediamento nel nuovo paese si allontana nel tempo, per molti di loro la lingua degli antenati risulta essere sempre più una lingua da imparare ex novo piuttosto che da consolidare o da arricchire dopo l'apprendimento in casa (per una riflessione sul rapporto tra lingua seconda e lingua etnica cfr. Montrul, 2008). Questa tendenza è stata riscontrata anche per l'italiano: i risultati dei numerosi studi dedicati all'argomento (ricordiamo, tra i più recenti, i lavori condotti da Scaglione, 2000 e De Fina, 2003 negli Stati Uniti; da Krefeld, 2004 in Germania; da Ciliberti, 2007 e Bettoni, 2008 in Australia) evidenziano uno stato di perdita funzionale e di erosione formale dell'italiano sempre più avanzato tra le nuove generazioni nate all'estero. Ciò rende dunque proficuo anche il paradigma di indagine scientifica dell'acquisizione di una seconda lingua, oltre a quelli dell'interferenza strutturale e del *code switching*, tradizionalmente adottati nell'analisi dei fenomeni linguistici legati all'emigrazione. Se questi paradigmi infatti insistono negativamente su quello che si va perdendo, quello dell'acquisizione valorizza positivamente quanto può ancora venire recuperato da parte dei parlanti delle nuove generazioni.

Pur nell'ampia e variegata casistica che si può ricondurre alla definizione di lingua etnica, è possibile

---

[1] Sono da attribuirsi a Elena Nuzzo i §§ 2, 3 e 6, e a Greta Zanoni i §§ 1, 4 e 5.

individuare alcune caratteristiche linguistiche ricorrenti nei discendenti di immigrati. Si osserva per esempio che molti di loro non sviluppano completamente la gamma dei registri padroneggiata dai parlanti nativi e che, anche quando sono fluenti nell'eloquio, non dominano alcuni di quegli aspetti della lingua che vengono generalmente appresi tardi, tra cui elementi di semantica e di pragmatica (Clyne, 1994).

3. Insegnare la pragmatica

Quando si desidera insegnare una struttura grammaticale è generalmente possibile fare riferimento a una o più regole che definiscono in maniera univoca le relazioni tra le forme linguistiche e le loro funzioni. Si possono incontrare delle difficoltà nel rendere tali relazioni comprensibili agli apprendenti, ma per l'insegnante il punto di riferimento nella lingua d'arrivo è chiaro. Quando invece ci si occupa di insegnamento della pragmatica, il riferimento alla "norma" è una questione molto più delicata e complessa. Per fare un esempio, ogni insegnante sarebbe in grado di dire come funziona in italiano l'accordo di genere e numero tra gli elementi nominali, ma forse non di spiegare come si fa una protesta o un complimento, perché i modi sono tanti quanti sono i contesti in cui ci si può trovare a compiere questi due atti linguistici: nonostante i vincoli legati alla salvaguardia della "faccia", i parlanti possono scegliere in quale misura attenuare o intensificare un atto anche in base al peso che personalmente attribuiscono alle variabili contestuali. Si possono naturalmente individuare alcuni schemi ricorrenti nelle situazioni più comuni, nonché alcuni strumenti linguistici che hanno una funzione pragmatica prevalente – per esempio il condizionale o le espressioni dubitative sono spesso usati in italiano per attenuare l'intensità di un atto linguistico –, ma non è possibile compilare un manuale di pragmatica così come si può creare un manuale di grammatica o un vocabolario. Il punto di riferimento più affidabile è quindi costituito da documenti autentici che mostrino l'uso effettivo della lingua nel contesto di reali interazioni.

4. Le fonti dei materiali didattici per LIRA

Partendo da questi presupposti, per la creazione del materiale didattico sugli atti linguistici destinato a LIRA si è scelto di utilizzare prevalentemente dati provenienti da corpora di parlato spontaneo e (semi)spontaneo (come ad esempio riprese video e registrazioni di role-play guidati) o da frammenti di trasmissioni radiofoniche e televisive (soprattutto fiction). L'ampio ricorso a materiale video consente di focalizzare l'attenzione non solo sulle strutture più propriamente linguistiche, ma anche sulle componenti paraverbali e ambientali della comunicazione. I video sono accompagnati nella maggior parte dei casi da trascrizioni, che sono pensate per aiutare gli utenti a comprendere le scelte linguistiche dei parlanti piuttosto che i tratti formali del parlato come gli aspetti fonetici e prosodici. Tali scelte sono in linea con le finalità didattiche – e non di ricerca – del sito. Come già sottolineato, LIRA è un *repository* misto: ai numerosi campioni di parlato si alternano esempi di lingua scritta,

come brevi estratti di articoli di giornale o di romanzi, ma anche messaggi tratti da forum, chat e blog. Questi ultimi sono stati volutamente inseriti nel *repository* perché, pur essendo testi in forma scritta, presentano spesso, come è noto, tratti e caratteristiche del parlato spontaneo. Su questo materiale autentico vengono proposte varie attività, il cui formato s'ispira sia ai test più frequentemente usati negli studi sull'apprendimento e sull'insegnamento della pragmatica – come il *Discourse Completion Task* (DCT), le scale di appropriatezza e le simulazioni di ruolo più o meno guidate (cfr. per es. Ishihara & Cohen, 2010) –, sia agli esercizi comunemente impiegati nell'insegnamento delle lingue seconde, come i questionari a scelta multipla, gli abbinamenti, i *cloze*, il riordino di parole o frasi, il completamento di schemi o tabelle con elementi tratti dal testo ecc.

Il progetto prevede anche lo sviluppo di funzioni, attualmente ancora in fase di elaborazione, che consentano di far caricare direttamente agli utenti (apprendenti, ma anche insegnanti di italiano per stranieri) ulteriori contenuti, in modo da favorire la partecipazione attiva degli utenti alla vita della piattaforma e il continuo incremento del materiale disponibile.

5. Struttura e contenuti del *repository*

I materiali LIRA per lo sviluppo delle competenze linguistico-pragmatiche sono raggruppati in 7 macro aree tematiche. Nell'individuare le tematiche da sviluppare si è cercato di comprendere le funzioni e gli usi linguistici maggiormente presenti nelle situazioni comunicative ma allo stesso tempo problematici dal punto della gestione delle variabili contestuali. Le aree tematiche affrontate comprendono l'uso delle forme di cortesia e le forme pronominali Tu e Lei, le espressioni cristallizzate in routine comunicative legate a particolari situazioni o eventi (saluti, auguri, condoglianze ecc.), le routine comunicative che seguono formule meno standardizzate (come ad esempio i complimenti, le scuse, le modalità per iniziare una conversazione con sconosciuti o per offrire il proprio aiuto), la funzione comunicativa legata alle richieste (come richiedere qualcosa, come accettare o rifiutare), le modalità per mettersi o non mettersi d'accordo (inclusa la fase di negoziazione tra gli interlocutori che spesso può risultare complessa), tutte le funzioni relative alla conflittualità tra i parlanti (dalla critica all'accusa, dalla protesta alla minaccia, dal litigio all'insulto) e infine un'area dedicata in generale alle modalità per esprimere le proprie opinioni, per mostrare e osservare alcuni tratti della conversazione (introdurre e chiudere un argomento di conversazione) introducendo elementi come lo scherzo e l'ironia. Ogni macro area è strutturata in modo da presentare inizialmente il contenuto generale oggetto dei percorsi e poi articolare il problema con specificità in grado di far comprendere gli usi linguistici attuali, compresi quelli più atipici. Se si considera ad esempio l'area dedicata alle forme di cortesia, troviamo sia attività e approfondimenti dedicati all'uso standard delle forme pronominali Tu e Lei sia usi meno frequenti dei pronomi con valenza ironica o

offensiva; nell'area dedicata alla conflittualità, accanto ai materiali esplicativi delle offese e degli insulti si trovano anche testi che mostrano l'uso delle stesse strutture lessico-grammaticali in senso scherzoso, amichevole e ironico. Ogni macro area è articolata in più percorsi che contengono un contenuto-stimolo culturalmente e linguisticamente significativo (ad esempio un breve filmato, un brano tratto da fonte scritta o un'immagine) e un numero variabile di attività, che hanno l'obiettivo di rendere consapevole l'utente della varietà e della variazione degli usi linguistici proposti nei diversi percorsi. Alcune delle attività si focalizzano specificamente sul contenuto pragmalinguistico del percorso didattico, mentre altre hanno una funzione di supporto alla comprensione, sia globale sia di singole strutture lessico-grammaticali. La struttura del *repository* così articolata permette all'utente sia una navigazione lineare, e quindi più controllata, secondo la sequenza suggerita dagli autori, sia una navigazione libera, con passaggio immediato da un percorso all'altro ed eventualmente anche da una macro area all'altra. Per consentire questa modalità di navigazione meno lineare, la piattaforma offre la visualizzazione simultanea e gerarchizzata dei contenuti principali e di quelli correlati, permettendo all'utente, attraverso il ricorso a un sistema ragionato di *tagging*, di muoversi agevolmente tra i contenuti tra loro collegati.

6.   L'interazione con gli utenti

Poiché il *repository* di LIRA è uno strumento pensato prevalentemente per l'autoapprendimento, il riscontro fornito dal computer dopo lo svolgimento dell'attività rappresenta per l'utente un aiuto essenziale per capire e imparare. Sebbene infatti l'ambiente preveda anche degli spazi dedicati a brevi spiegazioni ed esemplificazioni dei diversi fenomeni, è soprattutto dalla correzione delle attività che l'apprendente può cogliere il legame tra forme e funzioni nei diversi contesti. Poiché, per la natura stessa della pragmatica di cui si è discusso prima, non è possibile fornire all'apprendente un'unica soluzione corretta, occorre piuttosto offrire alcuni modelli di riferimento sulla base di ciò che vari parlanti nativi, magari provenienti da regioni diverse, hanno effettivamente detto nelle situazioni presentate all'interno delle attività, invitando l'apprendente a riflettere sui mezzi linguistici che consentono di attribuire agli enunciati diverse sfumature pragmatiche. Da questo punto di vista un validissimo contributo è offerto dalle potenzialità della Rete e in particolare dal tipo di ambiente in cui si muovono gli utenti di LIRA, che si propone come un *social network* più che come un semplice magazzino di contenuti e attività. I percorsi didattici sono integrati in spazi di condivisione (*forum*) nei quali gli apprendenti sono stimolati a discutere, porre quesiti e offrire opinioni sui documenti e sulle attività proposti. L'utente ha la possibilità di confrontare la sua risposta non solo con le soluzioni proposte dagli autori, ma anche con le risposte fornite dagli altri membri della comunità virtuale e da parlanti nativi, operando tra queste una selezione in base

alle caratteristiche socio-biografiche fornite al momento della registrazione. Effettuando il primo accesso a LIRA, infatti, gli utenti sono invitati a completare un breve questionario che consente al sistema di associare a ogni utente un profilo contenente dati anagrafici, interessi, conoscenze e abitudini relative all'uso della lingua italiana. Il controllo delle proprie conoscenze pragmalinguistiche è dunque rappresentato da un confronto con opinioni diverse piuttosto che da una tradizionale correzione. L'apprendente non è quindi soltanto un utente di materiali didattici on line, ma anche un membro della comunità virtuale che condivide l'interesse per l'uso concreto dell'italiano nei diversi contesti. Discussioni e riflessioni collettive offrono la possibilità di acquisire quella consapevolezza sui fenomeni pragmatici della lingua che possiamo considerare l'obiettivo fondamentale dell'apprendimento della pragmatica di una lingua seconda (Bettoni, 2006).

7.   References

Bettoni, C. (2008). Migrazioni e competenze linguistiche. In *Lingua, cultura e cittadinanza in contesti migratori. Europa e area mediterranea. Atti dell'8° Congresso dell'Associazione Italiana di Linguistica Applicata, Malta, 21-22 febbraio 2008*. Perugia: Guerra, pp. 17--31.

Bettoni, C. (2006). *Usare un'altra lingua. Guida alla pragmatica interculturale*. Roma - Bari: Laterza.

Ciliberti, A. (Ed.). (2007). *La costruzione interazionale di identità. Repertori linguistici e pratiche discorsive degli italiani in Australia*. Milano: Franco Angeli.

Clyne, M. (1994). *Inter-cultural communication at work. Cultural values in discourse*. Cambridge: Cambridge University Press.

Ishihara, N., Cohen A.D. (2010). *Teaching and learning pragmatics. Where language and culture meet*. Harlow etc.: Pearson.

Krefeld, T. (2004). *Einführung in die Migrations-linguistik von der* Germania italiana *in die* Romania multipla. Tübingen: Narr.

Montrul, S. (2008). Second language acquisition welcomes the heritage language learner: opportunities of a new filed. In *Second Language Research*, 24, pp. 487--506.

Scaglione, S. (2000). *Attrition. Mutamenti sociolinguistici nel lucchese di San Francisco*, Milano: Franco Angeli.

BP *obrigado eu* and *obrigado você* in counter service utterances

Bruna Karla PEREIRA

Universidade Federal de Lavras

Department of Human Sciences, Campus Universitário, Lavras, Minas Gerais, 37200-000

brunaufmg@yahoo.com.br

Abstract

This paper aims at investigating reanalysis and analogy in two common responses to thanks in Brazilian Portuguese: *obrigado eu* ('thank I') and *obrigado você* ('thank you'). A spoken corpus of commercial encounters was recorded and transcribed for this. My main interest is concerned with the utterances used to close these encounters by attendants and costumers. In order to understand pragmatic issues on thanking as a discursive device for closing commercial encounters, I will take a look at the assumptions made by Aston (1995). Differently, in order to understand their formal configuration and the changes undergone by them in a synchronic perspective, the discussion will be based on theoretical assumptions made by Hopper & Traugott (1993) and Harris & Campbell (1995). *Obrigado* is used in two contexts either when thanking or when replying to thanks. I hypothesize that *obrigado*, as an interjection, has been reanalized from the past participle of *obrigar* ('to obligate'). In response to "thanks", *obrigado* shows verbal valences usually attributed to *agradecer* ('to thank'). This feature probably rises by analogy with *agradeço eu* ('thank-1SG I') and *agradeço você* ('thank-1SG you').

Keywords: Brazilian Portuguese *obrigado eu/você*; reanalysis; analogy; counter service utterances.

1. Introduction

In BP, there are many different ways to reply to 'Thanks!': *De nada!* (of nothing), *Por nada!* (for nothing), *Às ordens!* (to-the orders), *À disposição!* (to-the disposition), *Disponha!*, *Estamos à disposição!* (be-PRES-1PL to-the disposition), *Estamos aí pra isso!* (be-PRES-1PL to-the disposition), *Eu é que agradeço!* (I is that thank-PRES-1SG), and so forth.

In English, we find several options as well: "You are welcome!", "No problem!", "Not at all!", "My pleasure!", "No worries!", etc. Nonetheless, in contrast to English in which "Thank you!" is used with a pronoun, BP *Obrigado!* is closer to French *Merci!* which is independent from pronominal categories. Recently though, in BP, variants other than those listed above have drawn attention especially due to their pronominal make up, as follows in the examples with *Obrigado você!* (thank you) and *Obrigado eu!* (thank I) where "A" stands for attendant and "C" for customer.

(1) A: É agora tá tranquilo.[1]
    'Yeah! It is easy now ...'
  C: É.
    'Yeah.'
  A: … pra tirar saldo, extrato ... Tá bão?
    '... to have your balance, bank account statement ... All right?'
  C: Brigadu.
    'Thanks.'
  A: Brigadu eu, tchau!
    'You are welcome, bye!'
(2) A: Deix'eu te dá um recibinho, aqui. Só isso mesmo?

'Let me just give your receipt. Is there anything else I can do for you?'
C: Só. Brigadu.
  'No. Thanks!'
A: Brigado ocê[2].
  'You are welcome!'

According to Aston (1995: 59), thanking

> "may function more as formal marker of discourse structure than as an indication of genuine gratitude [...] Rubin (1983) assigns it a ritual 'role' in closing service encounters".

As such, *Obrigado você!* and *Obrigado eu!* which mean respectively "It is **to you** that I have to say 'thanks'" and "It is I who have to say 'thanks'" play "an important role in conversation management" (Aston, 1995: 59).

However, some doubts come up when these structures are studied:

- Why is that *obrigado* ('thanks') sometimes shows an intransitive argument (*obrigado a você* – 'thanks to you'), but has been used without preposition, as in (2A)?
- Which is the syntactic status of 'I' in *obrigado eu* ('thank I')? Is it a subject?
- Is there anything beyond a relationship of synonym between *obrigado* ('thanks') and *agradecer* ('to thank')?
- What are the sintactic and semantic differences between *obrigado* and *agradecer*?
- How can *obrigado* ('thanks') be past participle of *obrigar* ('to obligate'), its cognate, and assume arguments of *agradecer* at the same time?

---

[1] The dialogues in (1) and (2) were taken from Pereira (2012).

[2] *Cê* and *ocê* are spoken variants of *você* ('you') while *brigadu* is a variant of *obrigado* ('thanks').

- How can a corpus-based analysis be helpful in answering these questions?

This paper will discuss these queries and investigate explanations for them.

2. Theoretical review

According to Harris & Campbell (1995: 61),

> "Reanalysis directly changes underlying structure, which we understand to include information regarding at least (i) constituency, (ii) hierarquical structure, (iii) category labels, (iv) grammatical relations, and (v) cohesion [...] Semantic change is involved also in many of these reanalyses" (Harris & Campbell, 1995: 61).

It will be shown that the changes undergone by *obrigado* are related to: (i) category labels, such as past participle and interjection, (ii) grammatical relations, such as valence and argument position, and (iii) semantic change indicating thankfulness or simply a discursive device for ending a commercial encounter.

Analogy is

> "a process whereby irregularities in grammar […] were regularized. The mechanism was seen as one of 'proportion' or equation. Thus, given the singular-plural alternation cat-cats, one can conceive of analogizing child-children as child-childs" (Hopper & Traugott, 1994: 56).

According to Hopper & Traugott (1994:57), "Kiparsky (1968) [...] views analogy as generalization or optimization of a rule from a relatively limited domain to a far broader one". My hypothesis is that, having the meaning of thankfulness, just like *agradecer*, *obrigado* has borrowed the argument structure from *agradecer*, surfacing with either an accusative pronoun or a post-verbal subject.

A traditional example of reanalysis and analogy is the Romance perfect which has developed from an adjectival form (3). In (3), accusative agreement is overt (*vos ... fatigatos*). In (4), however,

> "there is indeterminacy whether there is or is not agreement, since neuter singular (nihil […]) is the 'default' gender/number marker in Latin" (Hopper and Traugott, 1993: 57).

It turns out that lack of agreement between object and participle is extended to other contexts, as in (5). "These unambiguously non-agreeing forms presumably arose by analogy (=rule generalization) from neuter singular contexts to other contexts" (Hopper & Traugott, 1993: 57).

(3) *Metuo enim ne ibi **vos** habebam fatiga**tos**.*
    *fear-1SG for lest there you:ACC:**PL** HAVE-1sg*
    *tired-ACC:**PL***

'For I fear that I have tired you' (Hopper & Traugott, 1993: 57).

(4) *Promiss**um** habeo **nihil** [...].*
    *Promised-**NEUT/SG(?)** have-1SG nothing-**NEUT/SG***

'I have promised to do nothing' (Hopper & Traugott, 1993: 53).

(5) *Haec ominia probat**um** habemus*
    *Those-ACC-PL all-ACC-**PL** tried-**PART** have-1PL*

'We have tried all those things' (Hopper & Traugott, 1993: 57).

Concerning *obrigado* ('thanks') and its translation into English, it is appropriate to point out that, while 'thanks' and 'to thank' are cognate words, *obrigado* and *agradeço* are not. Despite this, it seems that BP speakers have been attributing grammatical patterns of *agradecer* to *obrigado* by analogy after it has undergone reanalysis as an interjection.

3. Methodology

This work was carried out by collecting data in commercial conversations, transcribing their final excerpts and providing them with a formal description of the phenomenon.

In a commercial establishment of a small city, in Minas Gerais state, three attendants conceded authorization to have their utterances recorded. Having got a corpus with 2 hours of counter service utterances, I have found more tokens with *brigado cê* than with *brigado eu*, which was restricted to the responses of only one of the three attendants recorded. The customers generally prefer *brigado cê* when they reply to attendants' thanks. That is why there was no occurrence of *obrigado eu* among the customers.

With this methodological approach, I am interested in data effectively produced by speakers. In Kennedy's (1998: 271) words,

> "In contrast to Chomskyan approaches to language, corpus-based descriptions are based on non-elicited linguistic performance as the source of evidence for theories of language, and so far have largely focused on particular languages rather than universals of language. However, although the goals and focus of study have typically differed, the two approaches can be seen as complementary rather than conflicting".

Therefore, a spoken corpus will be used to study the structures above mentioned, though intuition data will not be excluded.

4. A possible analysis

Some dictionaries attribute to *obrigado* a meaning like obliged or grateful, as in (6), in the sense that a person is obligated to someone else. This is the meaning derived from its cognate verb *obrigar* ('to oblige').

(6) "Fico-lhe muito obrigado pelo que me fez"
     (Ferreira, 1999).
     Stay-1SGnominative-3SGdative very oblidged
     by what me did
     'I am much obliged for what you did for me.'

However, in contemporary Brazilian Portuguese, speakers do not understand *obrigado* as the past participle of *obrigar* ('to oblige') at all. That is why other dictionaries point out the neutralization in gender and number agreement, with the forms *obrigada(s)* (thank-FEM-PL) and *obrigados* (thank-MASC-PL) out of use in the vernacular. Following Luft (2007: 357),

> "the insistence in calling attention to this rule of agreement [in gender and number] proves that the invariability is common, usual: _(Muito) obrigado, meu querido (_Much thank-0, my-MASC-SING darling-MASC-SING); _Vamos bem! (Muito) obrigado (GO-PRES-1PL well! Much thank-0). In this case, we have an interjective and invariable expression[3] (my translation)".

In addition to the lack of agreement, another evidence for the fact that *obrigado* ('thanks') is not understood as the past participle of *obrigar* ('to oblige') is its meaning. *Obrigado* is much closer in meaning to thankfullness, like *agradecer* ('to thank'), rather than to obligation. Though it is true, the participial configuration of *obrigado* in BP gives us a clue for understanding its intransitive argument in (8), because the past participle of *obrigar* has intransive valence 'obliged to'. Nonetheless, its participial configuration does not explain the postposition of *eu* ('I'), which is allowed in (1), repeated below as (7), but not in (9).

(7) C: Brigadu[4].
        'Thanks'.
     A: Brigadu eu, tchau!
        Thanks I, bye!
        'You are welcome, bye!'
(8) (Estou) obrigado a você.
     BE-PRES-1SG thank-past to you
     'I am obliged to you'.
(9) *Estou obrigado eu.
     BE-PRES-1SG thank-past I

A very plausible explanation for the configuration of structures like (10) and (11) is assuming that "say" and "say to" were left out.

(10) Obrigado (digo) eu.
     Thanks (say-PRES-1SG) I
     'It is I who say 'thanks'.'
(11) Obrigado (digo a) você.
     Thanks (say-PRES-1SG to) you.
     'It is to YOU that I have to say 'thanks'.'

However, what we intend to investigate in this paper is whether *obrigado* undergoes any kind of reanalysis and analogy in the responses *obrigado eu* and *obrigado você*. We have already seen that historically *obrigado* ('thanks') derives from *obrigar* ('to oblige'), but nowadays it is used as an interjection, having its agreement neutralized. In addition, *obrigado* ('thanks') has independent status, being able to surface alone in a sentence, like other interjections, such as: *olá* ('Hi!'), *oi* ('Hi!'), *saúde* ('Blessings!'), etc. Therefore, this is one of the linguistic changes undergone by *obrigado*: that is, the past participle of *obrigar* was reanalyzed into an interjection.

A second change taken place is the use of *obrigado* in responses to 'thanks', as seen in the examples (1) and (2) given in the introduction and discussed so far. According to Hopper & Traugott (19994: 61),

> "Reanalysis and analogy (generalization) have different effects. Reanalysis essentially involves linear, syntagmatic, often local, reorganization and rule change. It is not directly observable. On the other hand, analogy makes the unobservable changes of reanalysis observable".

An unobservable change is the fact that, having gratefulness more than obligation meaning, *obrigado* ('thanks') becomes interchangeable with *agradeço* ('thank-1SG'), as given below in the comparison between (12) and (13).

(12)   A: Só isso mesmo?
           'Is there anything else I can do for you?'
        C: Só. Brigadu.
           'No. Thanks!'
        A: Brigado ocê.
           'You are welcome!'
(13)   A: Só isso mesmo?
           'Is there anything else I can help do for you?'
        C: Só. Eu agradeço.
           'No. Thanks!'
        A: (Eu que) agradeço você.
           'You are welcome'.

Through analogy, irregularities in grammar are regularized (Hopper & Traugott, 1994: 56). Therefore, because *obrigado* becomes interchangeable with *agradecer*, *obrigado* may probably be used structurally like *agradecer* having either complement (followed by preposition), as in (14b), or a post-verbal subject, as in (15b). As a consequence, *eu* in (15a) looks like a grammatical subject because its position rejects the dative

---

[3] "a própria insistência em alertar para essa regra de concordância prova que a invariabilidade é frequente, usual: _(Muito) obrigado, meu querido; _ Vamos bem! (Muito) obrigado. Trata-se neste caso de expressão interjetiva, invariável" (Luft, 2007: 357).

[4] The dialogues in (7-9, 12-13) were taken from Pereira (2012).

mim and the accusative *me*, as seen in (16).

(14) a. Obrigado (a) você!
Thanks (to) you
'You are welcome!'
b. Agradeço (a) você!
Thank-PRES-1SG (to) you
'You are welcome!'
(15) a. Obrigado eu!
Thanks I
'You are welcome!'
b. Agradeço eu!
Thank-PRES-1SG I
'You are welcome!'
(16) *Obrigado mim/me.
Thanks to-me/me

Therefore, by analogy, *obrigado*, in responses to 'thanks', seems to follow the rules of *agradecer* argument structure. As a result, *obrigado* just like *agradecer* may have different pronouns as arguments, such as in *agradeço (vo)cê* (thank-1SG you), *agradeço vocês* (thank-1SG you-PL) and *agradeço o senhor* (thank-1SG the sir - 'You are welcome, sir'). The examples (17), (18) and (19) below show *obrigado* with all these pronouns and without the preposition *a* ('to').

(17) A: Deix'eu te dá um recibinho, aqui. Só isso mesmo?[5]
'Let me just give your receipt. Is there anything else I can help you with?'
C: Só. Brigadu.
'No. Thanks!'
A: Brigado ocê.
Thank you.
'You are welcome!'
(18) A: Sessenta e três. Mais alguma coisa?
'Sixty-three [Reals]. Something else?'
C: Só. Beleza.
'No. It is fine!'
[...]
A: [...] Então, falô. Brigadão.
'So, it is ok. Thanks.'
C: Então, beleza. **Brigadu oceis aí.**
So, nice. Thank YOU-PL there
'So. It is fine. Thank you all'.
A: Até mais.
'Bye!'
(19) A: Mais alguma coisa, seu L.?
'Something else, Mister L.?'
C: Só isso.
'No. It is fine'.
A: Muito obrigado.
'Thanks'.
C: **Muito obrigado o senhor**, então.
Much thank the sir, so.
'Thank YOU, sir'.

So far, I have been investigating two mechanisms of change probably operated on *obrigado*. The first one is its reanalysis from the past participle of *obrigar* to an interjection. The second one is the analogy with the verb *agradecer* which makes *obrigado* surface with post-verbal arguments either nominative or accusative.

According to Harris & Campbell (1995: 72),

"the conditions necessary for reanalysis to take place are that a subset of the tokens of a particular constructional type must be open to the possibility of multiple structural analyses, where one potential analysis is the old one [...] and the other potential analysis is the new one".

Considering the first mechanism above mentioned, *obrigado* is open to a reading where it is a varible participle of *obrigar* meaning obligation, as in (20), and to another reading where it is an invariable interjection meaning thankfulness, as in (21).

(20) "Ficamos-lhe muito obrigadas pelo que nos fez."
Stay-**1PL**-**3SG**dative very obliged-**FEM-PL** by what us did
'We are much obliged to what you did for us'.
(21) "_Vamos bem! (Muito) obrigado."
GO-PRES-**1PL** well! Much thank-**0**
'We are fine, thanks.'

Considering the second mechanism above mentioned, *obrigado*, as a response to thanks, shows argumental structure of *agradecer*. For convenience, I show in the next page a table with a summary of these processes of change.

It is interesting to mention that the structures studied in this paper are also productive in European Portuguese. Having done a very brief research on the *Reference Corpus of Contemporary Portuguese*[6], I found nine sentences with *obrigado eu*, as seen in the following examples:

(22) O Orador: Muito obrigado, Sr. Presidente. Assim sendo, terminei.
'The speaker: Thank you so much, Mr. President. Being so, I have just finished it'.
O Sr.Presidente: Muito **obrigado eu**, Sr. Deputado.
'The President: You are welcome, Mr. Deputy.'

(23) Vozes : - Muito bem! Muito obrigado!
'Voices: - Congratulations! Thank you!'
O Orador: - Muito **obrigado eu**, e seria assim, volto a agradecer a V. Ex.ª, a todos os Srs.

---

[5] The dialogues in (17-19) were taken from Pereira (2012).

[6] "The CRPC contains texts from the second half of the 19th century up until 2006, but most of the texts have been produced after 1970" (information taken from the *Reference Corpus* website).

'The speaker: - You are welcome. That is all. Once again, I thank Your Excellency and Gentlemen'.

(24) O deputado: Muito obrigado, Sr. Presidente.

'The deputy: Thank you very much, Mister President.'

O Sr.Presidente: -Muito **obrigado eu**, Sr. Deputado.

'The president: - You are welcome, Mister Deputy.'

	Reanalysis	Analogy
past participle of *obrigar*	> interjection (Thanks!)	> interjection (You are welcome!)
(6) and (20)	(1C, 2C, 17C, …)	(1A, 2A, 17A, …)
dependent form (auxiliary plus main verb)	independent form	independent form
variable (agreement)	invariable (neutralization of agreement)	invariable (neutralization of agreement)
indirect argument	without arguments	direct arguments
obligation meaning	thankfullness meaning	used in responses to thanks in order to close service encounters

Table 1: Summary of the changes undergone by *obrigado*

5.   Conclusions and further developments

With spoken data collected in counter service utterances, I have investigated the hypothesis according to which *obrigado* has undergone reanalysis while *obrigado eu* and *obrigado você* has undergone analogy. The first mechanism changed the past participle into an interjection. The second one changed syntactic properties of *obrigado* which shows accusative arguments and also postposition of *eu* like a post-verbal subject. This hypothesis is still very preliminary, but it seems to apply not only to BP but also to EP which have similar data.

It is also worth pointing out that there are other structures in BP where the regency of certain verbs seems to be, in a certain way, transferred to another verb. For instance, when a speaker says something like (25), where the verb *comentar* is used unexpectedly with a direct object, he is transferring the valence of *contar* (*me contou* –me told) or *dizer* (*me disse* – me said) para *comentar* (*me comentou* – me commented). This happens through analogy, because *dizer* ('to say') and *contar* ('to tell'), both speech verbs, have a pronominal direct object. The same seems to happen in (26) where the valence of verbs bearing company meaning, such as *casar com* ('marry

with') and *ficar com* ('stay with'), seems to be transferred to *namorar* ('date').

(25) Ele **me comentou** que você estava namorando.
    He me comment that you were dating
    'He told me that you are hanging out with someone'.

(26) Quando eu **namorava com** o João, não podia vestir saia curta.
    When I dated with João, not could wear short skirt
    'When I dated João, I was prevented from wearing short skirts'.

Kurilowicz (1945 apud Hopper & Traugott, 1994: 57) considers analogy or generalization as a "tendency to replace a more constrained with a more general form". Therefore, examples (1) and (2), as well as (25) and (26), should be viewed as a trend of BP to have either verbal or nominal valences regularized.

6.   References

Aston, G. (1995). Say 'thank you': some pragmatic constraints in conversational closings. *Applied linguistics*, 16 (1), pp. 57--86.

Ferreira, A. (1999). *Novo dicionário Aurélio - século XXI*. Rio de Janeiro: Nova Fronteira.

Harris, A., Campbell, L. (1995). *Historical syntax in cross-linguistic perspective*. Cambridge: University Press.

Hopper, P., Traugott, E. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Kennedy, G. (1998). *An introduction to Corpus Linguistics*. New York: Longman.

Lehman, C. (1995). *Thoughts on grammaticalization*. Munchen; Newcastle: Lincom Europ.

Luft, C. (2007). *Dicionário prático de regência nominal*. São Paulo: Ática.

Pereira, B. (February 2012). *Counter service recordings*. Perdões.

Reference Corpus of Contemporary Portuguese. Available in:<http://alfclul.clul.ul.pt/CQPweb/portugal/context.php?batch=2&qname=duud9pf2hrP&uT=y>. Acessed in: 3rd July 2012.

You're so not talking to me like that: analysing conflict talk in a corpus of sitcom discourse

Barbara Malveira ORFANÒ

Federal University of Sao Joao Del-Rei

Campus Dom Bosco, Praca Dom Helvecio, 74- Fabricas- CEP:36301- Sao Joao del-Rei

bmalveira@yahoo.com.br

Abstract

Koester (2006) explains that it is difficult to analyse arguments due to the fact that usually participants do not feel comfortable in allowing their arguments to be recorded and that may be the reason for the sparse amount of research on the subject. However, arguments have been addressed by many scholars in a variety of contexts within different approaches including: sociolinguistics, pragmatics, discourse analysis and conversation analysis. In the present study dialogues containing an argument will be analysed from two different perspectives: (i) Muntigl and Turnbull's (1998) model for the study of arguments and (ii) politeness (hedges). By combining the two approaches, we can determine how speakers in the sitcom orient themselves in the dialogues containing arguments in the narrative of the show. We concluded that in *Friends* speakers use more *contradiction* and *counterclaim utterances* which results in a high frequency of arguments that contain a low cost of face to participants. Even when act combinations are used, the least face aggravating type of arguments are preferred by speakers. The results together with a close examination of the examples present in the data contribute to the ongoing discussion on the representation of real language in media discourse.

Keywords: argumentation; politeness and corpus; media discourse.

1. Introduction

According to Grimshaw (1990), arguing is a common practice among humans, and any adequate account of the nature of spoken interaction needs to be able to describe how arguments are produced and managed. When analysing the dialogues from the sitcom *Friends*, it is observed that the main structure of the sitcom implies that arguments are in a certain way part of the show. The the classical structure of a sitcom involves: *familiar situation-disruption- and refamiliarisation with the current situation* . This suggests that arguments are likely to be part of the *disruption* phase of the show. Examples from the *Friends* corpus will be analysed focusing on the types of arguments found in the sitcom and also on the ways in which a resolution is negotiated by speakers in the data. It is likely that negative politeness will be of importance in this study, reinforcing the claim that the sitcom discourse is influenced by its global audience.

1.1 Definition of Argument

Argumentation theory has its roots in classical Graeco-Roman writings on rhetoric, legalistic reasoning and persuasion. The term argumentation derives from this formulaic and rationalistic approach. Within conversation analysis and related perspectives, a different notion of argument has developed. While studies of argumentation and rhetoric see arguments as a function of reason, an activity of the intellect, conversation analysis views arguments as events unfolding in a real time flow of turn-taking, in which adversary positions evolve in the light of utterances as they are emitted into the interactional space (Hutchby 2001: 124). Although Hutchby's (2001) view of arguments is of importance, it is important to emphasise here that dialogues in the sitcom are already written and decided by scriptwriters, thus, argument dialogues in *Friends* are carefully chosen by the show's writers who ultimately decide the outcome of each argument considering the main purpose of each episode of the show.

2. Literature Review

Arguments have been addressed by many scholars in a variety of contexts within different approaches including: sociolinguistics, pragmatics, discourse analysis and conversation analysis. Koester (2006) explains that it is difficult to analyse arguments due to the fact that usually participants do not feel comfortable in allowing their arguments to be recorded and that may be the reason for the sparse amount of research on the subject. Conversation Analysis has provided a good framework for the study of arguments and we shall rely on the most prevalent studies to support the analysis in section 9.3. Pomerantz' (1984: 64) work on agreement and disagreement in assessment sequences gave interesting insights to the study of arguments. She distinguishes a preferred-action turn shape from a dispreferred-action turn shape and concluded that disagreements were a dispreferred activity and their occurrences were often minimized through delays in the production of a disagreement and prefaces that mitigated the disagreement (see also Levinson 1983 and Sacks 1987). In Kotthoff's (1993) study, he observes that initially disagreements with dispreferred turn shapes occur, but as arguments develop, disagreements are expressed in a more unmodulated way, thus becoming the preferred response. However, Goodwin (1990), analysing children's disputes in a multiparty setting, observes that participants organise their talk highlighting opposition. Rather than being preceded by delays or hedges, turns containing oppositions are produced immediately. In addition, such turns frequently contains a preface which announces right at the beginning that an opposition is being produced (see Goodwin, 1990: 145). Coulter (1990) examines the structure of arguments and states that

arguments have a minimal adjacency pair structure consisting of an assertion and a counter-claim. In another study, Muntigl and Turnbull (1998) propose a minimal three-part structure consisting of a claim, a disagreement and a counter-claim. Up to this point, we have surveyed the most prevalent studies on argumentation and it is fair to say that CA has brought interesting insights to the study of conflict dialogues. In this article dialogues containing an argument will be analysed from two different perspectives: (i) Muntigl and Turnbull's (1998) model for the study of arguments and, of particular importance in this chapter, (ii) politeness as emphasised in chapter eight (hedges and boosters). By combining Muntigl and Turnbull's (1998) framework for the analysis of arguments in casual conversation with Brown and Levinson's (1987) study on politeness, we can determine how speakers in the sitcom orient themselves in the dialogues containing arguments in the narrative of the show. Before we move to the analysis, we briefly comment on the two.

2.1   Muntigl and Tumbull's model (M-T model)

Muntigl and Turnbull's (1998) research on arguments focuses on naturally occurring conversational data from two sources. The first is from ten hours of taped discussion of university students in naturally occurring conversation. The second consists of the recording of twenty-one families in which parents discuss a moral issue with their sons or daughters. In total their data comprises 155 dialogues and 4 types of disagreement utterances were identified: *Counter claims*, *contradictions*, *challenges* and *irrelevancy*.

  i. **Counter claims**: They are usually preceded by pauses, prefaces, and mitigating devices. Muntigl and Turnbull (1998) consider them the least face threatening of all types of disagreement acts. When using *counterclaims*, speakers can propose an alternative claim that does not directly contradict or challenge another's claim allowing further negotiation of the previous claim.
  ii. **Contradictions**: They are considered less aggravating than *irrelevancy claims* and *challenges* due to the fact that they do not directly attack the competency and rationality of the other speaker. Contradictions often occur with a negative particle such as *no* or *not*, signalling that the contradiction of the previous turn is true.
  iii. **Challenges:** They are often introduced by reluctant markers that display disagreement with the prior turn and they often have the syntactic form of an interrogative, co-occurring with wh-questions such as *when*, *what, who, why, where* and *how*. Challenges usually question an addressee's prior claim. They expect that the addressee will provide evidence for his/her claim, while suggesting that he/she cannot do so.
  iv. **Irrelevancy claims:** They are, according to

Muntigl and Turnbull (1998), the most face threatening type of conflict act. Irrelevancy claims express extreme opposition that limits any further discussion. Muntigl and Turnbull (*ibid.*) explain that in uttering an *irrelevancy claim* the speaker asserts that the previous claim is not relevant to the discussion, by disagreeing in overlap or without pauses to the preceding.

Muntigl and Turnbull (*ibid*: 230) claim that the type of disagreement acts used by speakers can be determinant to participants' face. They put forward the idea that disagreements are inherently face-threatening as many times they can convey disapproval of another person. Thus, face concerns can be expected to influence the conversational structure of arguing exchanges. Brown and Levinson (1987) developed a theory of politeness that acknowledges *positive politeness* and *negative politeness* .

Throughout the analysis sections we will pay attention to the role that both *positive* and *negative politeness* play in determining the kinds of disagreements and resolutions found in the sitcom.

3.   Data and Methodology

The *Friends* corpus consists of transcripts of fourteen shows from the seventh season (2000-2001) and amounts to approximately 40,000 words. The episodes were transcribed by many online fan clubs after being aired. The transcripts from (http:members.lycos.nl/frtrk/) comprise the data present in this study. Generally, the transcripts were correct containing detailed information of the scenes and actor's performance in parentheses. After downloading the episodes and saving them in a text file, the dialogues with the actual videos of the shows were checked and the mistakes were corrected (see Orfano, 2010). The *Friends* corpus was searched manually for dialogues that contained a dispute. These dialogues were isolated for analysis and classified under Muntigl and Turnbull's (1998) framework for the analysis of arguments in casual conversation. From the 27 dialogues containing an argument, 22 contain only one type of argument utterance and 5 dialogues contain more than one type of argument utterance and were classified as act-combination argument utterances following Muntigl and Turnbull's (*ibid*.) framework.

4.   Analysis

In this part of the analysis, we focus on the types of disagreements found in *Friends* according to the type of utterances used by speakers. Figure 1, in the next page, shows the distribution of disagreement utterances in the sitcom in comparison to Muntigl and Turnbull's (1998) model.

As can be seen in figure 1, speakers in *Friends* use more *contradiction* claims when arguing than speakers in the M-T model. There is also a difference between the number of *counter claims* used by the sitcom and the data

used by Muntigl and Turnbull's (1998) framework. In the sitcom, speakers use fewer counter claims when compared to Muntigl and Turnbull's (1998) data. This might be an indication that when arguing in order to sound more assertive speakers in *Friends* prefer to contradict their opponent's turn, while speakers in the casual conversation data prefer to use counter-claims. This needs to be further investigated when analysing the dialogues in the subsequent sections taking into consideration the issue of politeness. After examining the types of argument utterances present in the sitcom, we have classified the arguments in *Friends* according to the face cost imposed for participants during the argument as *lower face cost*, *moderate face cost* and *high face cost*.

(i) Lower face cost: Dialogues containing *counter-claims* and *contradiction utterances*

(ii) Moderate face cost: Dialogues containing *challenge* utterances

(iii) High face cost: Dialogues containing *irrelevancy claims*
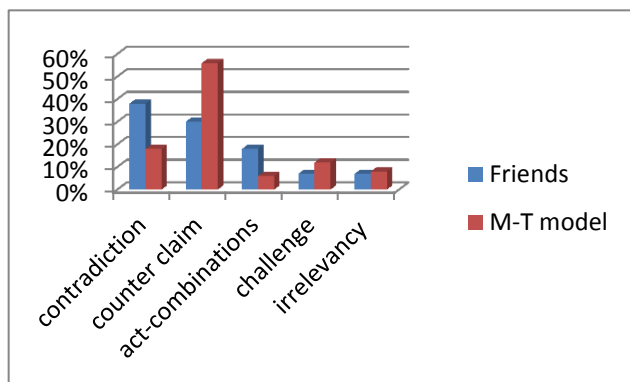
Figure 1: Distribution of disagreement utterances in *Friends* and in the M-T model

Significantly important to the analysis of arguments in the present chapter is the issue of face. Figure 2 shows the distribution of the types of disagreement in the sitcom and in the M-T model.
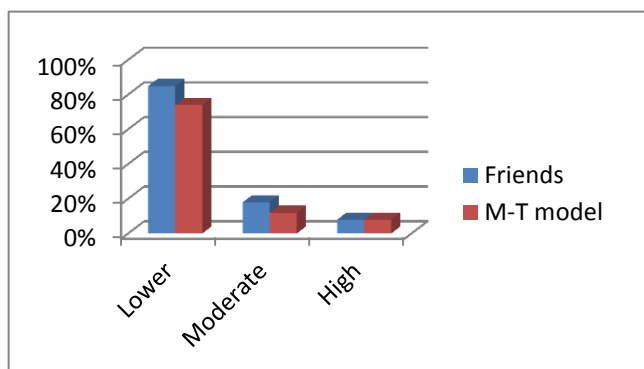
Figure 2: Types of disagreement in *Friends* and in the M-T model considering the issue of face

As can be seen in figure 2, 85% of the arguments in *Friends* belong to the lower category which means that participants in the sitcom when arguing prefer to be less assertive and are aware of face issues. In the M-T model, speakers are also concerned with face issues, 74% of the types of disagreement utterances in the M-T model belong to the lower face aggravating category. This preliminary finding supports the discussion carried out in chapter eight that hedges and negative face are critical politeness markers in casual conversation. If they were removed, the dialogues in the sitcom would look very unreal and therefore, the audience would not authenticate the show.

4.1 Lower cost of face

In this category we analyze the dialogues containing *counter-claims* and *contradictions*. The two types of argument utterances combined account for 85% of the arguments in the sitcom. Predominantly the show is comprised of arguments that present a low cost of face to participants which reinforces the claim that negative politeness is important in the sitcom. Thus, the use of contradiction utterances in conflict dialogues is an indication that speakers in the sitcom try to avoid strong face threatening acts while involved in verbal conflicts and when they do use threatening acts, they are often preceded by mitigation devices like hedges.

4.2 Moderate cost of face

In this section looks at the dialogues containing *challenge utterances*. As Muntigl and Turnbull (1998: 244) observe, 'they are highly face aggravating since, by implicating that the other cannot back up his/her claim, they attack the competency of the other'. Maybe for that reason they are not frequent in the sitcom. Speakers in *Friends* are very concerned about their interlocutors face and avoiding face threatening acts against participants during a conversation is common among characters in the show.

4.3 High cost of face

Muntigl and Turnbull (1998) state that the most face threatening type of disagreement occurs when speakers use *irrelevancy claims*. As we can see from figure 1 speakers in *Friends* do not use much of *irrelevancy claims*, instead, they prefer to use much lesser face threatening acts by using *contradiction* and *counter- claims utterances*. There are only 2 examples of irrelevancy claims in the sitcom (8%).

4.4 Act combination acts

In extended conflicts where there are more than two people arguing we find examples of what Muntigl and Turnbull (1998) call act combination conflicts. Their study shows that the most frequent act combination is contradiction followed by a counter claim (CT+ CC).

Although Muntigl and Turnbull (*ibid.*) have not analysed any other type of act combination in their data, after searching for argument dialogues in *Friends,* we

found five different act combinations in the sitcom. Table 1, in the next page, suggests a different organisation of act combination types of argument utterances as found in the sitcom data.

As can be seen in table 1, the sitcom follows a different organizational framework in relation to the acts found in Muntigl and Turnbull (1998). This might be due to the fact that the sitcom needs to comply with its audience who need to understand and ratify the dialogues of the show.

| 1-counter- claim + challenge |
| 2-irrelevancy claim + counter claim |
| 3-contradiction + challenge |
| 4-contradiction + challenge + irrelevancy claim |
| 5-challenge + contradiction |

Table 1: Act combinations in *Friends*

5.   Conclusion

The analysis above suggests that in *Friends* the most common type of argument utterance used by speakers usually imply a low cost of face for participants. This is represented by the prevalent use of *counter-claim* and *contradiction* utterances in the dialogues containing an argument in the sitcom. The reason for this might be the fact that both of them carry the least face aggravating acts. This indicates that negative politeness plays an important role in the arguments present in the sitcom. Muntigl and Turnbull's (1998) model shows that speakers also prefer to use the least face aggravating types of argument utterance in their dialogues in order to lessen the impact of their utterances on their interlocutors. This suggests that the sitcom follows a similar structure to the one used in casual conversation. However, there are differences regarding the type of utterances used in each study. In the M-T model speakers prefer to use counter-claims while in the sitcom speakers show opposition using contradiction utterances. This might be due to the fact that contradiction utterances portray an argument in a better way for the TV medium making the argument clear to the audience.

6.   References

Coulter, J. (1990). Elementary properties of argument sequences. In, G. Psathas (Ed.), *Ethnomethodology and Conversation Analysis.* RID: University Press of America, pp. 181--203.

 Goodwin, M.H. (1990). *He-said-she-said: Talk as Social Organization among Black   Children*. Bloomington: Indiana University Press.

Hutchby, I. (2001). Oh, Irony and Sequential Ambiguity in Arguments. In *Discourse and Society,* 12, pp. 123--141.

 Koester, A. (2009) .*Investigating Workplace Discourse*. London: Routledge.

Kotthoff, H. (1993). Disagreement and concession in disputes: On the context sensitivity of   preference structures. In *Language in* Society. 22, pp. 193--216.

Levinson, S. (1993). *Pragmatics*. Cambridge: Cambridge University Press.

Muntigl, P., Turnbull, W. (1998). Conversational structure and faceworking in arguing. In *Journal of Pragmatics*. 29, pp. 225--256.

Orfano, B. (2010). The Representation of Spoken Language: a corpus based study of sitcom discourse. PhD. Thesis in Applied Linguistics. Mary Immaculate-College University of Limerick.

Sacks, H., Schegloff, E.A. and Jefferson, G. (1974). A simplest systematic for the organization of turn-taking for conversation. In *Language*, 50(4), pp. 696--735.

How a story is told in Italian and in Italian Sign Language. Deictical, anaphoric and gestural strategies in Italian and LIS

Maria ROCCAFORTE[1], **Giulia PETITTA**[2], **Alessio DI RENZO**[2], **Paola Maria GIUNCHI**[1]

[1]Università "Sapienza" di Roma, Dipartimento di Scienze Documentarie, Linguistico-filologiche e Geografiche;
[2]Istituto di Scienze e Tecnologie della Cognizione (ISTC), CNR, Roma

maria.roccaforte@uniroma1.it, giulia.petitta@tiscali.it, alessio.direnzo@istc.cnr.it, paola.giunchi@uniroma1.it

Abstract

The present study derives from a FIRB research project which was designed to implement a e-learning environment for Italian deaf learners both teens and adults. Our task was to investigate which aspects of LIS and Italian are comparable and which are idiosyncratic. The aim was to assess to what extent salient and distinctive features of LIS can interfere and hinder the process of learning Italian by deaf learners. We focused on narrative texts in Italian and LIS and more specifically on deictic and anaphoric features which allow for the textual cohesion of these texts. We asked six subjects to watch a story and then tell it to other people. This story was simple and short and it required the narrators to resort to a variety of communicative strategies. The study showed how deixis/anaphora overall appeared and how it was linked to a gesture in LIS and in Italian verbal narrations.

Keywords: Italian Sign Language; deixis and anaphora; gesture; LIS-Italian comparison; speech.

1. Introduction

In last ten years, studies on deixis[1] and anaphora have been conducted both on signed and spoken languages looking at person reference, co-verbal gestures, discourse organization and cohesion devices.

The study of discourse organization both in spoken and sign language provide crucial findings about semiotic issues related to human language. It is important to note that speech and signed discourse share properties and organization features related to the face-to-face modality. Sign languages are indeed not written languages representing a means to understand more about "oral" communication and speech.

In comparing spoken and signed performances, we have to face some methodological and theoretical issues. First of all, in spoken face to face narratives we find two ways of expression, saying (by words) and saying-while-showing (by gestures, among others).

In sign languages, deictic-anaphoric reference can be carried out by means of complex manual and nonmanual units. These are marked by specific eye-gaze patterns, and exhibit highly iconic features. These units are often used in simultaneous signed units, representing a challenge in comparing spoken and signed languages (Volterra *et al*., 2005; Pizzuto 2007).

In signed languages two major types of units have been identified: "conventional", or "frozen" signs (which are comparable to lexems in spoken languages) and productive signs, described by researchers with a variety of compositional and highly iconic labels. The latter type of structures display a mode of saying which "show" how an action, a process or a state manifest themselves. This showing mode, with a depictional intent and demonstrative expression, is intralinguistic: signs can say and show at the same time (and signers use gestures too). For example, a speaker could say "pear" pointing up to express the position of a pear on a tree. A signer could instead sign modulating space and position of the reference, providing some spatial information while articulating the sign meaning "pear".

Signs perform two distinct functions. They can convey a specific meaning or can provide information about size, shape, spatial relations, and/or process. When signs express meaning they are called frozen signs, they provide the dictionary definition without expressing size, shape and aspect. When signs provide information about size, shape, spatial relations, and/or process, they are called Highly Iconic Structures (HIS). HIS are only partially comparable to gestures in spoken languages and are unavoidable cohesion devices. They are indeed frequently used in signed discourse and, as Pizzuto (2007) pointed out, deixis, anaphora and person reference strategies include different distribution of these signs: HIS are frequently used both with an anaphoric role and to express person reference, while LU are commonly used to introduce an object for the first time in the discourse.

In verbal languages, deictic-anaphoric reference can be carried out through verbal units, a combination of word + gesture and gestures only. Like in signed languages it could be are marked by specific eye-gaze patterns, and through highly iconic gestures.

2. Aims

The aim of this paper is to provide elicited data to compare structures in relation to the cohesion devices used in face-to-face narratives, both spoken and signed.

Our aim was to study deictic and anaphoric strategies concerning language, body movements,

---

[1] When a story is told, it occurs in a specific location, at a specific time, is produced by a specific person and is (usually) addressed to some specific other persons. Deictic terms such as personal pronouns (I, you, s/he, ...) and demonstratives (this/that) refer to a particular entity which is only given by the context. According to Levinson, deixis shows how the relationship between language and context is reflected in the structure of languages themselves. It concerns two things: the ways in which languages encode features of the context of utterance, and the way in which the interpretation of utterances depends on the analysis of that context of utterance (Levinson, 1995).

gestures, and gaze adopted in the act of telling in a cross-language and cross-modality perspective to highlight both functional and structural similarities in deixis and anaphora in signed discourse and speech.

In fact while in oral speech we can use gestures and words to express different sense units, Sign languages are so structured as to allow to simultaneously express actions subjects and objects. Using HIS is for example in signed languages it is possible to simultaneously coarticulate signs with hands and non manual elements which are frequently used as cohesion devices .

In order to investigate the discourse organization, we asked six subjects (three deaf and three hearing italian speakers with an age range from 35 to 50) to watch twice the Chafe's Pear Story[2] and then tell it to other people in Sign language (deaf) and in Italian (hearing). The storytellings were videotaped (Chafe, 1980). We created a common Excel file table with relative percentages of the deictical and anaphorical occurrences of references in the two linguistic systems and annotated the various modalities in which the information was expressed.

We chose to transcribe speech using the Jeffersonian transcription system (Jefferson,1984) which allows to take into account breaks, shooting, hesitations and false starts and to take note of extralinguistic behavior.

To transcribe and annotate LIS stories, we chose Sign Writing (hereafter SW), a specific writing system designed for signed languages. This is a sort of "iconic alphabet" (Sutton, 1995) not only allows for an adequate representation and observation of signs features but also a form-meaning multilinear notation which covey specific sign language properties (Antinoro Pizzuto, Chiari & Rossini, 2010). SW glyphs can indeed encode both manual and non-manual components (facial expression, eye gaze, mouthing and mouth gestures 3 , shoulder

orientation, etc.), providing accuracy of description, multilinear organization of signed units, representation of discourse organization and face-to face modality features.

In spoken narratives we have analyzed gestures breaking them down into two distinct categories, deictic gestures and representational gestures. Deictic gestures are those that refer to something in the narrative - pointing, showing an object, or reaching for something. Representational gestures have meaning independent of the objects. (Iverson *et al.*, 2008).

We compared UL deictical/anaphorical occurrence expressing the vocal deictic reference with the verbal + gestural or gestural explanation only (Table1).
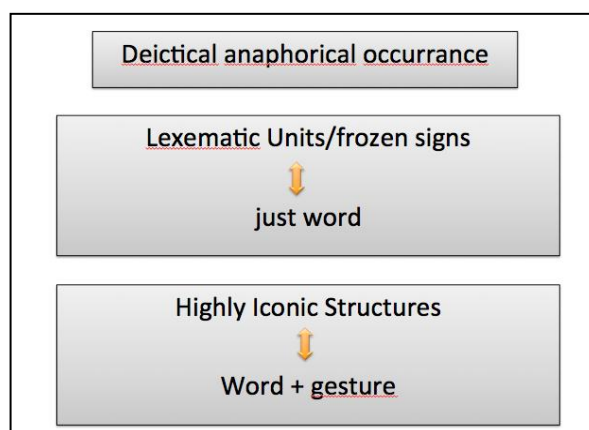
Table 1: ITA-LIS comparison

Three signers produced a face-to-face signed rendition of the Pear Story (recounted to another experienced signer). This text was subsequently transcribed with the help of the SW system. Analyses were performed on the SW-encoded transcript, checking the original video recorded narrative as needed. The analysis focused on the different strategies adopted by signers in telling a story they had seen. We observed the linguistic devices used by signers to introduce for the first time in discourse people and objects they were talking about, their position and their spatial-temporal characteristics (deictic reference) and to refer, later in their signed narratives, to the same people and objects (anaphoric reference) specifying their actions, states, locations (reference maintenance). While HIS are frequently adopted to express anaphoric reference and reference maintenance, they can also be used to convey deictic reference. Instead, frozen signs can only express deictic or anaphoric reference and are more frequently used for deictic reference.

---

[2] A farmer with a red bandana around his neck, carefully collects pears on a tree. A boy passing by, steals a bike and a basket of pears. While cycling on the country road he falls off the bike. Walking on the country road three boys see what has just happened to him and immediately decide to help him gather the pears from the ground. The boy he gives them one each and goes away. The three boys pass beside the tree where the farmer, incredulous, is counting the baskets of pears and gives them puzzled looks while they are eating the pears.

[2] Sign language research provides evidence on a bifurcation in mouth movements (both independently articulated and coarticulated with manual components of signs). Mouthing is a word, or a part of it, borrowed from a spoken language, while mouth gestures are specific movements with no relation with any word. Mouth gestures can be articultated using lips, mouth, cheek, and are not related to co-verbal gestures (Boyes Braem & Sutton Spence, 2001).

[3] Sign language research provides evidence on a bifurcation in mouth movements (both independently articulated and coarticulated with manual components of signs). Mouthing is a word, or a part of it, borrowed from a spoken language, while mouth gestures are specific movements with no relation with any word. Mouth gestures can be articultated using lips, mouth, cheek, and are not related to co-verbal gestures (Boyes Braem &

Sutton Spence, 2001).
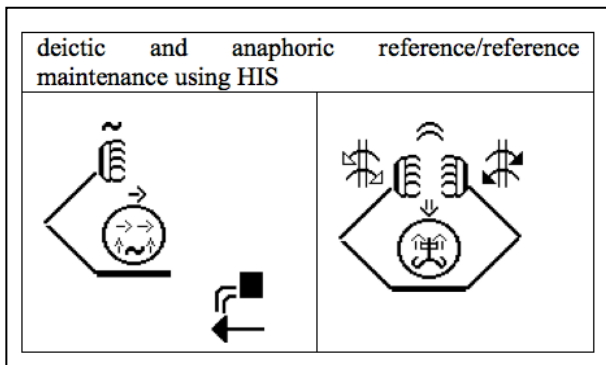
Figure 1: On the left there is a spatial deictic reference and the first appearance of HIS in the telling of this story. The meaning is, "Someone comes on the right while someone else is picking up the pears." On the right there is an anaphoric reference expressed by HIS, meaning "The man, previously introduced, is picking up the pears
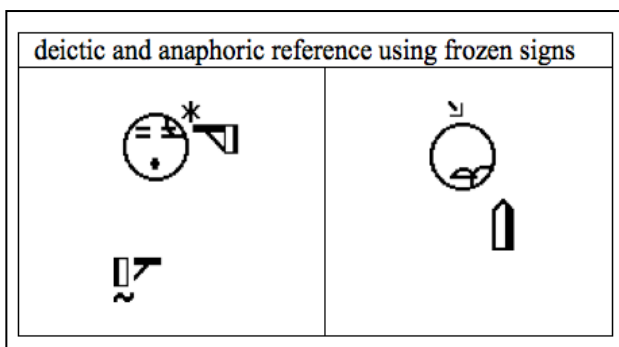
Figure 2: On the left is the frozen sign for 'man' from the first introduction. On the right is another frozen sign of an anaphoric reference

3. Results

The collected data show the prominence of HIS as referring expressions in signed discourse. Although HIS seem to function primarily as text cohesion device ('specialized' for anaphoric reference and reference maintenance, both animate and inanimate) they are also used for deictic introduction of referents in discourse.

Although in spatial deixis we find both frozen signs and HIS, it is important to note that sign language use often requires a spatial information addict. It is impossible to articulate a sign without moving in space, and there are constraints related to direction, verse and space. These constraints make signers articulate their discourse with a lot of spatial marked points, so, the phenomenon of deixis regards on average 7% of spoken Italian and 21% for the LIS one.

Furthermore, there are some crucial issues regarding the units of analysis and the differences between spoken and signed discourse. It is likely that the multilinear organization of signed discourse exhibits two or more sense units per sign, each including deictic or anaphoric reference. On the other hand, spoken speech exhibits only one sense unit per word, except for coverbal gesture coarticulated units. It is important to note that

visual-gestural linguistic multilinearity affects the units count, and further research is required in an across modalities perspective to understand discourse organization.

In the three spoken tellings we have 195 gesture manifestations. As the table (Table 3) shows more than 40% of these gesture occurrences are linked with deixis. In narrations approximately 8 minutes long we have in average 42 extralinguistic manifestations.

Furthermore each hearing teller produced 240 deictic/anaphoric references against the 230 occurrences in LIS. While the result appears similar in both languages, but in LIS we found a huge lack of homogeneity in comparison with ITA narrators.
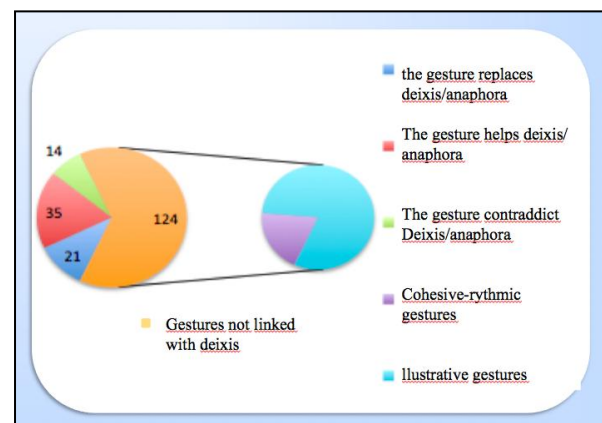
Table 2: ITA-speakers gestures: 124 occurrences not linked with deixis 70 linked with deixis

4. Conclusions

The Pear Story by Chafe allowed us to compare the differences in communicative strategies used in LIS and Italian. We found similar results in the use of deictic and anaphoric devices adopted in the oral narration by our LIS and Italian subjects. However when the LIS subjects told the Pear Story, they used a more accurate and functional set of communicative devices to refer to space and people. The high number of occurrences of these linguistic features in LIS seemed to fill the information gap which is usually counterbalanced by the use of gestures in Italian. This phenomenon applied to 20% of the cases. We observed that in some crucial instances LIS speakers adopted HIS strategies while Italian speakers relied on gestures. It is as if words and frozen signs would not be good enough to fully render the message. The percentage of anaphorical personal references (both animate and inanimate) was very high in comparison with deixis because of the constant reference to the person in the speech. Maintaining this reference is a hallmark of some of the marked structures, such as transfers of person. The majority of deictic-anaphoric references consists of HIS, in line with the results of Antinoro Pizzuto *et al.* (2008). Many units are simultaneous with the co-articulated expression of several referents. For instance, this is the case when the narrators needed to refer to one of the boys who help the character of our story to pick up the pears he

has stolen from the ground. The only feature that distinguishes the three boys is that only one of them is playing paddleball, a game played with a paddle attached to a little ball by a string. In this specific instance LIS speaker relied on HIS and specifically on the transfer of person (TP) where the signer embodied the boy playing with the paddle to refer to him anaphorically. Speakers who had to tell the story in Italian had to mime the ball play while saying, "When of the three boys …." The challenge was due both to the difficult task of referring to a specific person, out of three, and to the fact that none knew the name of the game. In addition, when the speakers had to identify the farmer, they made use of the bandana which he wore around his neck and had taken off to clean the pears; they introduced the bandana in their narrative commenting that it was around the farmer's neck and used to clean the pears. All three of the LIS narrators presented the bandana by means of HIS while only one of the three Italian speakers was able to achieve this communicative goal (Figure 3)

Figure 3: Italian speaker shows just trough the gesture the farmer's bandana position

These were only two examples where LIS was shown to be a more accurate system to convey special deixis, person and situational references, when compared with Italian where the speakers tended to rely on extralinguistic means such as gestures.

5. References

Antinoro Pizzuto, E., Rossini, P., Sallandre, M.-A. and Wilikinson, E. (2008). Deixis, anaphora and Highly Iconic Structures: Cross-linguistic evidence on American (ASL), French (LSF) and Italian (LIS) Signed Languages, 9th. In R.M. de Quadros (Ed.), *Theoretical Issues in Sign Language Research Conference, TISLR 9*. Florianopolis, Brazil, December 2006.

Bellugi, U., Klima, E.S. (1982). From Gesture to Sign: Deixis in a Visual-gestural Language. In R.J. Jarvella, W. Klein (Eds.), *Speech, Place and Action: Studies in Deixis and Related Topics*. Chichester: John Wiley &

Sons Ltd, pp. 297--313.

Antinoro Pizzuto, E., Chiari, I. and Rossini, P. (2010). Representing Signed Languages: Theoretical, Methodological and Practical Issues. In M. Pettorino, A. Giannini, I. Chiari and F. Dovetto (Eds.), *Spoken Communication*. Newcastle, U.K: Cambridge Scholars Publishing, pp. 205--240.

Chafe, W. (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Jefferson, G. (1984). Transcript notation. In J. M. Atkinson, & J. Heritage (Eds.) *Structures of Social Action: Studies in Conversation Analysis* (9-16). Cambridge: Cambridge University Press, Petròpolis/RJ, Brazil, Editora Arara Azul, pp. 475--495.

Iverson J. M., Capirci, O., Volterra, V. and Goldin-Meadow, S. (2008). Learning to talk in a gesture-rich world: Early communication in Italian vs. American children. In *First Language,* 28 (2), pp. 164--181.

Levinson, S.C. (1995). *Pragmatics*. Cambridge: Cambridge University Press.

Pizzuto, E. (2007), Deixis, anaphora and person reference in signed languages. In E. Pizzuto, P. Pietrandrea and R. Simone (Eds.), *Verbal and Signed Languages: comparing structures, constructs and methodologies*. Berlin: Mouton De Gruyter, pp. 275--308.

Sutton-Spence R. (1995). *The role of the manual alphabet and fingerspelling*. In *British Sign Language. Doctoral dissertation*. University of Bristol, Bristol.

Volterra, V., Caselli, M.C., Capirci, O., Pizzuto, E. (2005). Gesture and the emergence and development of language. In M. Tomasello, D.I. Slobin (Eds.), *Beyond Nature-Nurture - Essays in Honor of Elizabeth Bates*. Mahwah, N.J.: Lawrence Erlbaum, pp. 3--40.

Resonance, subjectivity and intersubjectivity in Brazilian Portuguese everyday talk

Maria Elizabeth Fonseca SARAIVA

Faculdade de Letras da Universidade Federal de Minas Gerais

Av. Antônio Carlos, 6.627 - 31270-901 - Belo Horizonte – MG

bethsaraiva@uol.com.br

Abstract

One of the assumptions of functionalist approaches is that form tends to respond to communicative or cognitive functions. Thus, this paper aims at finding motivations that would justify the emergence of resonant utterances in spontaneous conversations in Brazilian Portuguese. By resonance I mean, following Du Bois (2001), a speaker's retake of linguistic devices that have just been used by the interlocutor. Such phenomenon causes the establishment of lexical-structural and cognitive mapping relations between both utterances. In search of the motivations for this phenomenon, first I focus on the manifestation of the speaker's subjectivity by means of the resonant utterances. The next step consists of demonstrating that, beyond subjectivity, resonance iconically reveals the moments of greatest interpersonal involvement of the interlocutors. This intersubjective alignment, in turn, subsumes various degrees of tuning in (or not) between the co-participants' perspectives in the spontaneous dialogue.

Keywords: Resonance; subjectivity; intersubjectivity.

1. Introduction

One of the principles shared by all functionalist approaches is that form is mostly motivated by communicative and cognitive functions. Assuming this to be the case, in this paper I take up again the study of lexical-structural resonances in spontaneous conversations in Brazilian Portuguese, trying to answer this question: what motivates speakers to produce resonances?

Before that, it should be understood what I mean by *resonance*, a term introduced by Du Bois (2001). In face-to-face dialogue interactions, it can be noted that, at times, the speaker reuses, in his/her utterance, linguistic devices (patterns, structures, lexical items, etc.) that have just been used by the interlocutor, thus creating formal and conceptual mapping relations between both utterances, as suggests the data in boldface in example (1),[1] whose translation follows in (1'):

(1) (Pedro e sua noiva Bia estão vendo fotos de paisagem)

1 – Pedro: qual que ocê quer ver primeiro?
2 –   vão ver das paisagens...
3 – Bia: **nó que lin::do né**?
4 – Pedro: **nossa ficou lin::do...**
5 – Bia: **nossa essas andorinhas aí tão maravilhosas...**

Translation:[2]

(1') (Pedro and his fiancé, Bia, are seeing photos of landscapes)

1 – Pedro: which (one) do you wanna see first?
2 –   let's see (the ones) of the landscapes...
3 – Bia: **wa how beau::tiful, isn't it?**
4 – Pedro: **wow (it) turned out beau::tiful…**
5 – Bia: **wow these swallows there are wonderful…**

In the example above, Bia manifests her appreciation of a photo, especially through the following linguistic devices: interjection / admiration marker – "nó" ("wa"), a reduced form of "nossa" ("wow"); adjective of evaluation-affection with vowel stretching – "lin::do" ("beau::tiful"); and a tag-question "né?" ("isn't it?"), which indicates a search for approval in discourse. Pedro, in 4, retakes Bia's utterance (see the use of the same interjection in full and the repetition of the adjective with vowel stretching), to demonstrate his agreement with his interlocutor's evaluation. Upon such stimulus, in the utterance of line 5, she notes another detail in the photo – "essas andorinhas aí" ("these swallows there"), completing her evaluation with the same linguistic devices used before by herself and Pedro. This time, however, the chosen adjective is "maravilhosas" ("wonderful"), which has a more expressive power than "lindo" ("beautiful").

A noteworthy fact is that the quantification of lexical-structural resonances in spontaneous dialogues in Brazilian Portuguese shows a frequency of 24,5% (Matta, 2010). Therefore, we can attest the prominence of such utterances in discourse, following Givón (1995: 64): "(...) salient experience is clearly the less frequent *figure*, standing out on the more frequent *ground*." Thus, the question raised in the first paragraph is justified, for which

---

[1] The data in this paper were obtained from transcriptions of four spontaneous conversations in Brazilian Portuguese, which are part of the database of the *Grupo de Estudos Funcionalistas da Linguagem (CNPq – Conselho Nacional de Pesquisa).* The transcriptions were made according to the norms of the NURC-SP project (Castilho & Pretti (Eds.). 1986), being divided into semantic-intonational units. In the data presented, the following conventions should be noted – omission of a passage: (...); any pause: ...; voice superposition: [; question: ?; the transcriber's descriptive comments: ((laughs)); vowel streching: ::.

[2] In this paper, an approximate translation of each example into English will follow its introduction. In the translation, the elements in parentheses do not appear in the original.

an answer will be searched in the next section, based on the socio-cognitive notions of subjectivity and intersubjectivity.

2. Resonance, Subjectivity and Intersubjectivity

A first tentative answer to the question of the motivation that leads speakers to resort to the linguistic device of lexical-structural resonances has already been suggested in Saraiva (2008), following Thompson and Hopper (2001): in spontaneous conversations among friends and acquaintances, it is not our main goal to speak objectively about events and actions. Rather, we are interested in expressing our values, points of view, feelings and emotions, in evaluating people, attitudes and situations, weighing our perspectives against those of our dialogue partners. In short, in that study the emphasis was placed on the manifestation of subjectivity by means of resonant utterances. We tried to list a number of the linguistic marks that manifest subjectivity in those utterances, such as: use of evaluative-subjective adjectives; interjections showing surprise, admiration, reproach, etc.; modal verbs, adverbs and epistemic fragments; affective invocation; the use of verbs that describe internal situations of the participants in an interaction (evaluative, affective, cognitive, etc.), etc. However, in that article, nothing was mentioned in relation to the various devices that speakers of Brazilian Portuguese use to create a light environment of humor and play. As I see it, though, these are situations where subjective intentionality manifests itself very clearly, since they distance from the ordinary, the predictable. Note the example below:

(2) (Pedro, sua noiva e sua sogra Dalva estão vendo fotos)

1 – Pedro: isso aí é um jatinho né?
2 –   que eu deixei um jatinho lá fora agora
3 –   pra sempre que a gente for passear lá...
4 – Dalva: ah então eu vou ter... cadeira cativa?
5 – Pedro: lógico...
6 –   aí quando tiver lá em cima o que eu faço?
          ((risos))
7 – Dalva: **abre a janela e me joga...**
          [
8 – Pedro: **abro a porta e jogo ela pra fora**...

Translation:

(2') (Pedro, his fiancé and his mother-in-law Dalva are seeing photos)
1 – Pedro: this is a jet, isn't it?
2 –   'cause I left a jet outside now
3 –   for whenever we go there…
4 – Dalva: ah so I'll have… a permanent seat?
5 – Pedro: of course...
6 –   so when (you)'re up there, what do I do?
          ((laughs))

7 – Dalva: **open the window and throw me (out)…**
          [
8 – Pedro: **(I) open the door and throw her out…**

In (2), the mood of play and laughter permeates the whole example, having been set since the beginning with Pedro's turn from line 1 to 3. For our purpose, however, I emphasize the fact that the climax of the playful mood happens at those moments in which resonance emerges (see 7 and 8). Pedro's rhetoric question (line 6) about what he intended to do with his mother-in-law once they were up high, in a jet, uttered with laughter, gave her the opportunity to anticipate a humorous answer in the utterance in line 7 – "abre a janela e me joga..." ("open the window and throw me (out)…"). Pedro, in turn, resonates Dalva's answer in voice superposition (see line 8), stretching the mood of intimacy and play. Thus, we can see that humor is a creative way of revealing subjective affection.

On the other hand, the data in (2) gives me the opportunity to demonstrate that, besides expressing subjectivity, resonance reveals, iconically, as I see it, the great intersubjective involvement of the interlocutors. In fact, in spontaneous dialogues, intersubjective and subjective relations permeate the whole interaction. However, the point I want to make is that their materialization is brought to full potential at those moments when the interlocutor retakes the other's "words". In the example above, Pedro and Dalva get aligned in the interaction itself by means of the humor they co-create. This is then a local activity of the participants of that interaction, which constitutes one of the aspects of intersubjectivity. But intersubjective relations also show another facet: that of the system of beliefs, values and socio-cultural expectancies shared by co-participants in a dialogue. In (2), this dimension can be noted by the emergence of a cultural stereotype (the one, according to which, mothers-in-law are undesirable), "against" which the interlocutors react when they use it to create humor. As we know, humor is a light and creative form to manifest disagreement with a position, belief, value, etc.

Finally, according to Du Bois (2007), we note that the intersubjective *alignment* materialized by the resonances subsume a number of pragmatic/discourse functions. Although the author mentions the fact without exploiting it further, the analysis of the data in Brazilian Portuguese revealed a gradient in the weighing of perspectives, which range from less predictable and expected functions, such as the creation of play, humor, irony, etc., as in (2), to more conventional and predictable ones, as in the case of the use of resonances to respond to a question, to ask for clarification, or to manifest that an interlocutor is following the other's train of thought (phatic function), etc. Note the following data:

(3) (Fred e Carla, dois amigos, estão conversando enquanto preparam um lanche)

((música do vizinho ao fundo))

1 – Fred: ((risos)) tá rolando um karaokê...
2 –   **cê tá sacando**?
3 – Carla: uhn... uhn...
4 –   **tô ouvindo...**

Translation:

(3') (Fred and Carla, two friends, are talking while preparing a snack)
((the neighbor's music in the background))

1 – Fred: ((laughs)) a karaoke is taking place…
2 –   **dig that?**
3 – Carla: uhn...uhn...
4 –   **(I) can hear (it)...**

The resonance exemplified in (3) can be classified as one of the responsive kind (Matta, 2010), so it is one of those functions of greater predictability. However, we can add that, in this example, there is more than a mere information request (through a "yes/no question"), which is attended to by the interlocutor. When Fred asks Carla if she "tá sacando" (digs) the neighbor's karaoke, he demonstrates his care towards her at the same time. Carla feels moved by such an interest, and thus responds affirmatively. Notice that the consent markers "uhn... uhn..." already function as an affirmative answer. But Carla prefers to "qualify" them, emphasizing them with the resonant utterance "tô ouvindo" ("(I) can hear (it)"), in which the structure of the predicate "tá sacando" ("dig"), by Fred, is maintained (auxiliary + perception verb in the gerund). By means of a resonance, she aligns with her interlocutor's interest interactively.

The phatic function mentioned above can be illustrated by example (4):

(4) (Bia está explicando a sua sogra, Vera, a razão de não poder assistir à apresentação de um ballet)

1 – Bia: que **é amanhã à noite**...
2 – Vera: **é... de noite...**
3 – Bia: não tem jeito...

Translation:

(4') (Bia explains to her mother-in-law, Vera, the reason why she cannot watch a ballet presentation)

1 – Bia: which **is tomorrow night…**
2 – Vera: **right…night…**
3 – Bia: there is no way…

In the context of this dialogue, Vera's retake of Bia's utterance means to signal that she is attentive to her daughter-in-law's argumentation, that she follows it.

The intersubjective alignment between the participants of an interaction, materialized in the resonances, still includes varying degrees of tuning in or not between their perspectives. This fact is illustrated by example (1), shown earlier, and the data in (5) below, respectively:

(5) (Bia e Vera discutem qual seria o melhor horário para ir a uma feira de moda)

1 – Bia: **oito horas também é vazio...**
2 – Vera: **oito horas é cheio...**
Translation:

(5') (Bia and Vera argue about what would be the best time to go to a fashion fair)

1 – Bia: **eight o' clock is empty too…**
2 – Vera: **eight o' clock is full…**

Example (5) illustrates the use of the linguistic device of resonance to express divergence in opinion. The context of the utterances is that of two interlocutors arranging a time to visit a fair when fewer people would be present, so that it would be more convenient. In line (1), Bia suggests 8 a.m. as a good time: "oito horas é vazio..." ("eight o'clock is empty…"). Vera, however, disagrees, by retaking Bia's own "words" and replacing the adjective "vazio" ("empty") by its antonym "cheio" ("full"): "oito horas é cheio" ("eight o' clock is full").

As for the data in (1), they illustrate the convergence of the interlocutors' evaluation by means of the device in focus in this paper: lexical-structural mappings, as already mentioned.

In short, the data analyzed in this section confirm the gradient of intersubjective alignment materialized by resonances. In one end of this "scale", there are the more predictable and expected functions, such as the function of offering an answer to a question. Next in this "scale', there are the varying degrees of convergence or divergence between the interlocutors' perspectives. Finally, in the other end of the less conventional and least expected functions, there are the cases of creation of irony, humor, play, etc.

3. Conclusion

Assuming the functionalist principle that very often form is iconically motivated by communicative (or cognitive) functions, in this paper I defended the idea that the linguistic device of resonance (i.e. insertion of the interlocutor's utterance in one's own utterance, partially or totally) reveals, in a transparent fashion, the moments of greatest intersubjective involvement of the co-participants in an interaction.

4. References

Castilho, A.T.de; Pretti, D. (1986). *A Linguagem Falada Culta na Cidade de São Paulo: Materiais para seu Estudo*. São Paulo: T. A. Queiroz.

Du Bois, J.W. (2001). *Towards a Dialogic Syntax*. Ms., Department of Linguistics, University of California, Santa Barbara.

Du Bois, J.W. (2007). The Stance Triangle. In E. Englebretson (Ed.), *Stancetaking in Discourse*. Amsterdam: John Benjamins Publishing Company, pp. 139--182.

Givón, T. (!995). *Functionalism and Grammar*. Amsterdam: John Benjamins Publishing Company.

Matta, B. (2010). As funções discursivas das ressonâncias. In M.E.F. Saraiva., J.H.C. Marinho (Eds.), *Estudos da Língua em Uso: da Gramática ao Discurso*, Belo Horizonte: Editora da UFMG, pp. 191--230.

Saraiva, M. E. F. (2008). Marcas de subjetividade em enunciados ressoantes em português. *Alfa*, 52 (1), pp. 157--166.

Thompson, S. A., Hopper, P. J. (2001). Transitivity, clause structure, and argument structure: evidence from conversation. In J. Bybee, P.J. Hopper (Eds.), *Frequency and the Emergency of Linguistic Structure*. Amsterdam: John Benjamins Publishing Company, pp. 27--60.

Charisma perception in political speech: a case study

Rosario SIGNORELLO[1,2], **Francesca D'EERRICO**[2], **Isabella POGGI**[2], **Didier DEMOLIN**[1], **Paolo MAIRANO**[1]

[1]GIPSA-Lab, UMR 5216 CNRS, Grenoble INP, UJF, U. Stendhal (Grenoble, France); [2]Dipartimento di Scienze dell'Educazione, Universita` degli Studi Roma Tre (Rome, Italy)

rosario.signorello@gipsa-lab.fr, didier.demolin@gipsa-lab.fr, fderrico@uniroma3.it, poggi@uniroma3.it, paolomairano@gmail.com

Abstract

The charisma of the leader is conveyed through multiple aspects: his ideas and vision and his perceivable verbal and non verbal behaviors. Among these perceivable behaviors there are the acoustic characteristics of speech. We present here a study on the perception of charisma in political speech. We collected speech statements with different illocutionary value taken from two speeches given by Umberto Bossi, the leader of an Italian party, before and after a stroke which caused him a voice disorder. Stimuli from the two condition differed significantly in the acoustic-prosodic features. In the first part of the study 40 French listeners rated normal speech stimuli (20 pre- and 20 post-stroke) and in the second part 22 French (11 pre- and 11 post-stroke) and 31 Italians (15 pre- and 16 post-stroke) rated the de-lexicalized version of the same stimuli. Results for the first part of the study show that pitch contour in Bossi's pre-stroke speech positively influence the perception of his speech as charismatic, as opposed to those some years after the stroke. Results for the de-lexicalized speech confirm for French listeners our hypothesis of the influence of the pitch contour in Bossi's charisma perception but they are controversial for Italian participants that seem to perceive Bossi as more charismatic in the post-stroke condition.

Keywords: charisma; political speech; intonation; illocution; voice disorder; speech synthesis.

1. Introduction

Charisma was firstly described by Weber as an "extraordinary quality" of a person who is believed to be endowed with superhuman properties thanks to which s/he gets ac- knowledged as a leader Cavalli, 1995: 5). Though no specific objective description of the "extraordinary quality" was given in Weber's studies, some works started to study the perceivable behaviors of charismatic leaders: some, e.g., (Boss, 1976), focus on what we called the "charisma of the mind" (Signorello *et al.*, 2012), that dwells in the strength of a leader's ideas, others, e.g., (Atkinson, 1984) try to find visually or acoustically perceivable aspects of a leader's behaviors that we called "charisma of the body" (Signorello *et al.*, 2012). We suggest that both aspects of charisma, either jointly or independently, are responsible for its conveying and perception.

In the present study we focus on one aspect of the charisma of the body: the speech. We assume here that some of the perceivable acoustic-prosodic characteristics of a leaders speech are specifically responsible for conveying charisma. Our general goal is to characterize acoustically and distinguish perceptually a charismatic speech from a non- charismatic one.

Within previous work investigating the relationship be- tween the acoustic-prosodic characteristics of a political leader's speech and the perception of his/her charisma, Rosenberg and Hirschberg, 2009) studied the correlation between acoustic, prosodic, and lexico-syntactic characteristics of political speech and the perception of charisma; Touati (1993) investigated the prosodic features of rhetoric utterances in French political speech in pre and post- elections discourses. Other works examined the relationship between prosodic features and the perception of a speaker as a "good communicator" (Strangert & Gustafson, 2008) or analyzed the pitch contour of French political leaders' speech and its idiosyncratic and contextual variations (Martin, 2009).

2. A hypothesis about charisma

According to Poggi (2005), in persuasive discourse the speaker tries to convince the audience to do some action by exploiting the three strategies posited by Aristotle (2011): *Logos* (the rational argument), *Pathos* (the appeal to the audiences emotions), and *Ethos* (the character of the speaker). According to the theory of Poggi (2005) and Poggi *et al*. (2011), the dimension of *Ethos* also includes, for the political leader, three sub-dimensions: *Benevolence* (the tendency to act in the interest of the audience), *Competence* (the capacity for rational foreseeing and planning), and *Dominance* (the power to prevail in a competition).

The notion of charisma we proposed in (Signorello *et al*., 2012) is based on this theoretical framework. We de- fined charisma as a set of characteristics of a leader that include his "having a vision" (a goal towards which he wants to lead his followers), a "high level of dominance" (look strong, persistent and fighting) and "emotional intelligence" (the ability to feel and transmit emotions, and to be and look empathic). The combination of these features makes a leader charismatic, and is displayed by his/her non-communicative and communicative behavior..

3. What makes a speech charismatic?

To investigate the perception of charisma in political speech we analyzed the acoustic and prosodic characteristics in the speech of Umberto Bossi, an Italian politician who in 2004, during his political career, had a stroke that resulted in severe speech impairment. We collected two samples taken from two speeches performed, respectively, in 1994 (the pre-stroke condition, PRE) and in 2011 (the post-stroke condition, POST). Our hypothesis was that the important differences in

acoustic-prosodic characteristics of Bossi's speech, in samples of political speeches preceding and following the stroke, give rise to a different perception of charisma. If this hypothesis is validated perceptually we might conclude that information about charismatic qualities are borne by the acoustic-prosodic characteristics that differ in the two samples.

In order to describe the charisma phenomenon through common language adjectives we conducted a qualitative study collecting adjectives describing what charisma is and what it is not (a brief summary is presented in section 3.1.. For the extensive study see (Signorello *et al*., 2012). We then analysed Bossi's acoustic-prosodic features in the PRE and POST and conducted a language-independent perceptual study on French participants (section 3.2.4.). We then de-lexicalized our stimuli by synthesis only preserving the pitch contour, the duration and the intensity and conducted a perceptual study on French and Italian listeners. In isolating the pitch contour we could verify if this is the aspect that influences the perception of charisma in Bossi's speech (section 3.3.).

3.1 Describing charisma

In a previous work (Signorello *et al*., 2012) we constructed a questionnaire aimed to assess the perception of charisma in the samples of Bossi's speech required to previously make up a list of adjectives that express charismatic and non-charismatic qualities. To find out such adjectives in an empirically grounded way, we administered a questionnaire through Internet to 58 French participants (42 female, 16 male, mean age 30), asking to freely generate adjectives connected to the idea of what charisma is and what it is not. We obtained a list of French adjectives, 106 describing charisma positively and 105 describing what charisma is not. In order to make a manageable questionnaire, we further selected 67 adjectives (Table 1) retaining only those occurring more than once, 42 positively and 20 negatively related with charisma. We then classified those adjectives in a multidimensional scale of charisma under five dimensions describing this phenomenon. An extended report of this multidimensional scale of charisma and on how adjectives describing charisma are classified in it can be founded in (Signorello *et al*., 2012).

3.2 Normal Speech

3.2.1. Stimuli

Previous works about the perception of a speaker as a good (Strangert & Gustafson, 2008) or charismatic speaker (Rosenberg & Hirschberg, 2009) rely on the acoustic analysis and the perceptual evaluation of stimuli classified per speaker, topic and genre of speech. Our approach is different. We chose 3 stimuli per condition (PRE and POST) according to their illocutionary value: an assertion, an in- citation and a rhetorical wh- question. As we know the speaker shapes prosody differently in relation to different speech acts (Firenzuoli, 2001). Our

hypothesis is that all three types of speech acts are perceived as more charismatic in the PRE condition thanks to prosodic features. Further we argue that incitation might be perceived as more charismatic than rhetorical question which in turn might be perceived as more charismatic than assertion. Below we describe the acoustic-prosodic features of our stimuli.

DIMENSION	PRE	POST
Pathos	passionate, empathetic, enthusiastic, reassuring	cold, indifferent
Ethos Benevolence	extraverted, positive, spontaneous, trustworthy, honest, fair, friendly, easygoing, makes the others feel important	untrustworthy, dishonest, egocentric, individualistic, introverted
Ethos Competence	visionary, organized, smart, sagacious, creative, competent, wise, enterprising, determined, resolute, who propose, seductive, exuberant, sincere, clear, communicative	inefficient, inadequate, uncertain, faithless, unclear, menacing
Ethos Dominance	dynamic, calm, active, courageous, confident, vigorous, strong, leader, authoritarian, captivating, who persuade, who convince	apathetic, timorous, weak, conformist, unimportant, who scare
Emotional Induction Effects	charming, attractive, pleasant, sexy, bewitching, eloquent, influential	boring

Table 1: The 67 positive and negative adjectives related with charisma collected among the naïve French participants (in English for clarity purposes). Reprinted from Signorello *et al*., 2012)

3.2.2. Overall F0 measures

The PRE speech presents higher F0 means than the POST speech: PRE (F0 mean 178.89 Hz; min 101.84 Hz; max 241.10 Hz), POST (F0 mean 120.20 Hz; min 91.78 Hz; max 155.99 Hz). All means from the PRE differ significantly from the POST ($p<0.0001$). Our findings confirm and extend (Murry, 1978)'s findings on significant differences in F0 measures between normal and disordered voice. We argue that F0 values might be positively correlated to charisma perception.

3.2.3. Pitch contour description

The assertion in the PRE condition (Figure 1a below) presents a syntactic focus on "questo" [this], emphasized by a high fall and separated by a pause from the rest of the sentence. The right-side part of the tonal unit presents a falling contour with a small peak on the last tonic syllable. Instead, in the POST condition (Figure 2a below) the

sentence presents a moderate falling and flat pitch contour with a peak on the third lexical word. The incitation in the PRE condition (Figure 1b) includes two parts, each with a

(a)

(b)

(c)
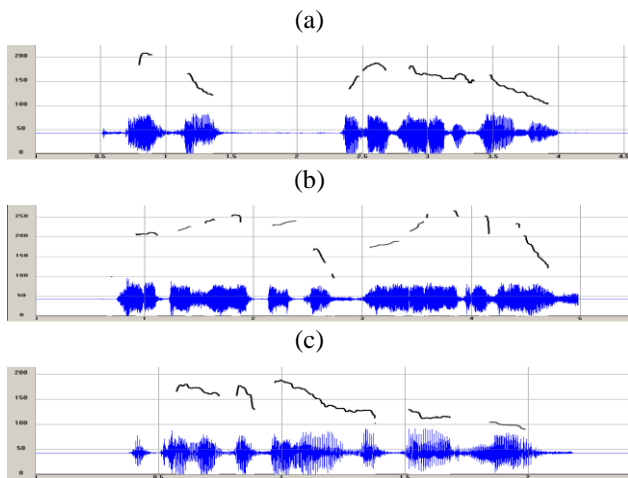
Figure 1: Intonation contour, transcription, translation, du- ration and F0 measures of PRE stimuli per speech act. (a): Assertion. "*Questo amici ereditiamo*" [This, my friends, is what we inherit]. 3,51s. F0 mean 52.62 Hz; SD 12.40 Hz; min 95.25 Hz; max 210.94 Hz; range 13 ST. (b): Incitation. "*Si ritorna all'attacco, fuori dalle trincee*" [Let's take up again the offensive, get out of the trenches]. 4.27s. F0 mean 225.51 Hz; SD 38.58 Hz; min 107.74 Hz; max 270.36 Hz; range 16 ST. (c): Rhetorical wh- question. "*E come facevamo a farlo?*" [How could we have done it?]. 1.81s. F0 mean 138.28 Hz; SD 27.98 Hz; min 96.07 Hz; max 189.39 Hz; range 11.72 ST. Spectrogram and pitch contour graphics obtained with WinPitch software (Martin, 2011)

pitch contour starting with high frequency and falling sharply in the last tonic syllable. In the POST condition instead the incitation (Figure 2b) presents two rising-falling contours in the first part and falls gradually in the right part of the tonal unit. The rhetorical wh- question in the PRE condition (Figure 1c) presents two contiguous pitch contour movements: the rising part corresponds to the wh- element, the falling part corresponds to the verb. A gradual falling movement comes on the right side of the tonal unit. In the POST statement (Figure 2c) a falling contour corresponds to the wh- element and a rising contour to the verbal element, with a gradual falling movement on the right side of the tonal unit.

3.2.4. Perception experiment

Forty French participants with no knowledge of Italian rated the stimuli presented in the section above via a HTML/PHP browser-based interface. Twenty of them listened to the PRE condition and twenty to the POST condition stimuli. The test took place in an anechoic chamber and participants wore a Sennheiser HD 25-13 headphone. After listening to each stimulus a participant had to answer to some check questions to verify that the

perception of the acoustic signal was optimal and that the semantic content was not understood. Then they had to express their judgment about the stimuli through our 67-adjective inventory on a 7-point Likert scale

(a)

(b)

(c)
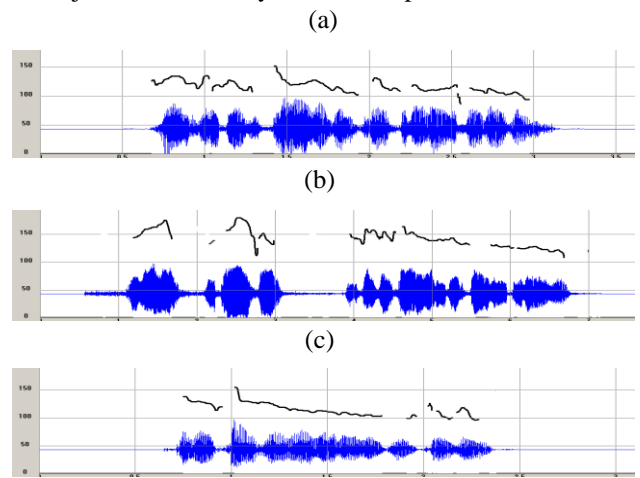
Figure 2: Intonation contour, transcription, translation, du- ration and F0 measures of POST stimuli per speech act. (a): Assertion. "*Noi siamo schiavi del centralismo romano*" [We are slaves of the Roman centralism]. 2,46 s, F0 mean 116.77 Hz, SD 10.74 Hz, min 86.64 Hz, max 146.45 Hz, range 9 ST. (b): Incitation. "*La Lega e` pronto per conquistare la liberta` della padania*" [The Lega is ready to conquer the freedom of padania]. 6.61s, F0 mean 142.02 Hz, SD 38.58 Hz, min 86.2 Hz, max 182.08 Hz, range 12 ST. (c): Rhetorical wh- question "*E come fanno a lavorare questa gente?*" [How can these people work?]. 1.89 s, F0 mean 117.93 Hz, SD 15.54 Hz, min 90.56 Hz, max 192.99 Hz, range 13 ST. Spectrogram and pitch contour graphics obtained with WinPitch software (Martin, 2011)

(0 = "totally disagree", 7 = "totally agree"), with some adjectives from the list substituted by their reverses (e.g., *warm* instead of *cold*) to avoid answer habituation. The average duration of the test was of 20 minutes.

3.2.5. Results

From our check questions it resulted that perception was good and there was no semantic comprehension. Hence, the differences between PRE and POST, that are mostly significant (t-test, $p<0.05$), must be due only to acoustic and not to semantic features. Out of the 67 adjectives used to measure the perception of charisma, about 33 adjectives obtained significantly different values (t-test, $p<0.05$) between PRE and POST speech, and most of them were rated higher for the PRE condition (Table 2 below). This is consistent with our hypothesis about the PRE speech as more charismatic than the POST thanks to its acoustic features. The PRE speech is positively correlated with most adjectives describing charismatic qualities (Table 1 below). In the dimension of *Pathos* the speaker is perceived as *passionate*, *eloquent* and *enthusiastic* in the PRE and as *indifferent* in the POST. As

to *Ethos Benevolence* results are quite inconsistent: the adjectives attributed to the PRE speech include *egocentric*, *dishonest* and *individualistic*, which in our previous qualitative study (Table 1) are non-charismatic qualities.

DIMENSION	PRE	POST
Pathos	passionate (5.02), enthusiastic (3.25)	indifferent (2.83)
Ethos Benevolence	egocentric (4.51), dishonest (3.95), makes the others feel important (3.68), individualistic (4.29)	trustworthy (3.51), introverted (2.41)
Ethos Competence	competent (4.83), smart (4.52), organized (4.75), determined (5.51), exuberant (4.57), faithless (3.57), clear (4.65), communicative (4.25), seductive (3.17)	wise (3.90), unclear (3.37)
Ethos Dominance	dynamic (5.13), authoritarian (5.73), confident (5.89), leader (5.87), captivating (3.57), convincing (4.40), captivating (4.78)	calm (4.29)
Emotional Induction Effects	attractive (3.10), eloquent (4.68), charming (4.78)	boring (3.63)

Table 2: Adjectives describing the perception of charisma in the Bossi's speech by condition with rating values (t-test, p<.001)

	PRE			POST		
ADJECTIVES	A	I	Q	A	I	Q
dynamic	5	**5.45**	4.42	2.09	3.09	2.14
authoritarian	6.19	**6.42**	4.57	3.61	4.04	3.61
calm	2.66	1.76	3.42	**4.61**	4.14	4.09
extraverted	1.66	1.23	2	2.14	**3.14**	2.9
timorous	3	2.38	**3.95**	3.76	3.38	3.42
wise	2.85	2.28	3.38	3.95	**4.23**	3.52
individualistic	**4.81**	4.61	3.42	3.28	3.33	3.71
active	4.9	**5.66**	4.28	2.28	2.81	3.52
introverted	1.52	1.14	2.33	**2.81**	2.19	2.23
menacing	4.57	**5.33**	3.33	3	2.9	2.85
energic	5.14	**6.09**	4.52	2	2.9	3.38

Table 3: Adjectives describing the perception of charisma in the Bossi's speech by speech act (A=assertion, I=incitation, Q=rethorical wh- question) and condition with rating values and one-way ANOVA's values (p<.001). Higher rates in bold

As for the dimensions of *Ethos Competence* and *Ethos Dominance* our hypothesis is almost completely validated: the speaker is perceived as *competent*, *smart*, *clear*, *seductive*, etc. in the PRE and as *unclear* in the POST; as *dynamic*, *authoritarian*, *confident*, *leader* in the

PRE and as *boring* in the POST speech. These results validate our hypothesis on the attribution of charismatic qualities to the PRE as opposed to the POST speech.

Taking into account the different types of speech act both in the PRE and in the POST speech the different illocutionary act elicitates a different perception. The incitation is the one that influences the most the perception of charisma. In particular for the dimension of *Ethos Competence* the incitation elicitates adjectives as *competent* ($F_{(2, 123)}=3.114$; $p<0.048$), *resolute* ($F_{(2, 123)}=6.767$; $p<0.002$), *enterprising* ($F_{(2, 123)}=8.515$; $p<0.001$), *clear* ($F_{(2, 123)}=3.046$; $p<0.05$), *exuberant* ($F_{(2, 123)}=4.232$; $p<0.017$) and *communicative* ($F_{(2, 123)}=2.705$; $p<0.05$).   More than other speech acts the incitation has a significant effect on the perception of the speaker's emotional state (see adjectives as *passionate* ($F_{(2, 123)}=2.999$; $p<0.05$), *influential* ($F_{(2, 123)}=9.359$; $p<0.001$) and *enthusiastic* ($F_{(2, 123)}=4.765$; $p<0.010$)). The assertion on the other hand evokes more non-charismatic qualities like *indifferent* ($F_{(2, 123)}=3.459$; $p<0.035$) and *unclear* ($F_{(2, 123)}=3.662$; $p<0.029$). Finally the rhetorical question seems to not influence a specific dimension of charisma. However, if we consider effect of both condition and a particular speech act the results are quite different. Through a one-way ANOVA we crossed the results of the condition (PRE vs. POST) and the different types of speech act (assertion, incitation and rhetorical wh- question) to study the influence of the different illocutionary acts on the perception of Bossi's charisma (see Table 3). The incitation makes Bossi to be perceived as more *dynamic*, *authoritarian*, *active*, *menacing*, and *energic* in the PRE condition and as *extraverted* and *wise* in the POST condition. Through the Assertion he has been perceived as *individualistic* in the PRE speech and as *calm* and *introverted* in the POST speech. As for the rhetorical wh-question the only significantly results is *timorous* in the PRE speech.

3.3   Synthesized speech

3.3.1.   Stimuli

We decided to carry out a perceptive test on de-lexicalized stimuli in order to further validate our hypothesis that the pitch contour is a relevant element for the perception of charisma. In fact, our de-lexicalization procedure enables us to isolate the pitch contour of a sentence from the semantic content, segmental features and voice quality characteristics. In this way, listeners are therefore forced to give their judgments solely on the basis of intonation, all other linguistic information being eliminated. The de-lexicalized procedure we chose has been developed for the AMPER (*Atlas Multimedia Prosodique de l'Espace Roman*) project developed by Albert Rilliard on the basis of scripts originally elaborated by Antonio Romano (see Contini *et al.*, 2002 for details). It consists in synthesizing a periodic waveform with the original pitch, intensity and duration values of the actual sentence (this is done by taking three measures per vowel, respectively at the onset,

middle and offset-consonants are replaced with silence). This procedure has been used by several authors working on the AMPER project and has already proved its efficacy.

3.3.2. Perception experiment
Twenty-two French (11 PRE, 11 POST) and thirty-one Italian (15 PRE, 16 POST) listeners participated to a perception analysis with the same methodology described in section 3.2.4. Thus the only differences were the de-lexicalized stimuli.

3.3.3. Results
The first results for the de-lexicalized stimuli perception, compared to results for normal speech perceptions, confirm in one hand our hypothesis of the influence of the pitch con- tour in Bossi's charisma perception for French participants but they are, in the other hand, controversial for Italian participants. In fact French listeners describe Bossi as *charming*, *who propose*, *timorous*, *confident*, *pleasant*, *introverted* in the PRE speech and as *inadequate*, *spontaneous*, *active*, *leader* in the POST speech (t-test, p<0.05). For Italian participants we only performed the perceptual test of de-lexicalized stimuli in order to avoid semantic and ideology influence on the perception of Bossi's speech. Italian listeners perceived the speaker as *boring*, *indifferent* and *unimportant* in the PRE speech and as *attractive*, *visionary*, *sexy*, *cold*, *passionate*, *seductive* in the POST speech (t-test, p<0.05). From these preliminary results it seems that the pitch contour-only stimuli elicit a different type of of Bossi's charisma for Italians listeners. In fact the POST speech is described with adjectives positively related with charisma and the PRE speech with adjectives describing charisma negatively, a trend in results that goes against our theory of pre-stroke speech as more charismatic than the post-stroke.

4. Conclusion

In this study we aimed to demonstrate that the perception of charisma in political speech is partly determined by the acoustic characteristics of speech. To do so, we first analyzed samples from the speech of the Italian politician Umberto Bossi before and after a stroke; through a qualitative study we singled out 67 adjectives describing charismatic and non-charismatic qualities. finally we run a perception study asking participants to rate Bossi's samples in terms of those adjectives. As resulted from the acoustic analysis, the PRE speech, with its intonation features as focus words, tonal jumps, and higher values, dramatically differs from the POST. And since the results of the perception study validate our hypothesis that Bossi's speech after the stroke is perceived as less charismatic than before, we may reasonably conclude that the characteristics of intonation that differentiate Bossi's PRE and POST speeches are an important factor in the perception of charisma. And this hypothesis has been validated once more through a perceptual experiment in which we only tested the intonation contour influence on

the Bossi's charisma perception. We also de-lexicalized stimuli and preserved original pitch, intensity and duration values and we tested French and Italian participants. Results validate our hypothesis on the intonation contour relevance on charisma perception of the PRE speech for French participants but are controversial for Italians. In any case our results on synthesized speech are preliminary and they will be statistically analyzed more in depth. Naturally we are aware that the acoustic characteristics of speech also include voice quality, which we think is relevant too. In future work we will investigate the importance of voice quality in determining the perception of charisma, while trying to distinguish it from the contribution of intonation, also through synthesis of speech fragments.

5. Acknowledgements

6. Acknowledgements

Aristotle. (2011). *Rhetoric*. The Internet Classics Archive.

Atkinson, M. (1984). *Our Masters' Voices: Language and Body Language of Politics*. Routledge, London, UK.

Bligh, M.C., Robinson, J.L. (2010). Was Gandhi "charismatic"? Exploring the rhetorical leadership of Mahatma Gandhi. In *The Leadership Quarterly*, 21(5), pp. 844--855.

Boss, P. (1976). Essential attributes of charisma. In Southern *Speech Communication Journal*, 41(3), pp. 300--313.

Cavalli, L. (1995). *Carisma: la qualita` straordinaria del leader*. Laterza, Rome, Italy.

Contini, M., Lai, J.-P., Romano, A., Roullet, S., de Castro Moutinho, L., Coimbra, R., Bendiha, U. P. and Ruivo, S.S. (2002). Un projet d'atlas multime´dia prosodique de l'espace roman. In B. Bel, L. Marlien (Eds.), *Speech Prosody 2002*, Aix- en-Provence, France, pp. 227--230.

Firenzuoli, V. (2001). Le forme intonative di valore allocu- tivo dell'italiano parlato. Analisi sperimentale di un cor- pus di parlato spontaneo (LABLITA). PhD thesis, Uni- versita` degli Studi di Firenze.

Goleman, D.P. (1995). *Emotional Intelligence: Why It Can Matter More Than IQ for Character, Health and Lifelong Achievement*. Bantam Books, New York, USA.

Martin, P. (2009). Autres styles: quelques variantes id- iosyncratiques. In *Intonation du Français*, Armand Colin, pp. 228--232.

Martin, P. (2011). WinPitch W7 by Pitch Instruments Inc. Computer program. Available at: < http://www.winpitch.com/>.

Mohammadi, G., Mortillaro, M., and Vinciarelli, A. (2010). The voice of personality: Mapping nonverbal vocal behavior into trait attributions. In *International Workshop on Social Signal Processing*, pp. 17--20.

Murry, T. (1978). Speaking fundamental frequency characteristics associated with voice pathologies. In *Journal of Speech and Hearing Disorders*, 43, pp. 374--379.

Poggi, I. (2005). The goals of persuasion. In *Pragmatics & Cognition*, 13(2), pp. 297--336.

Poggi, I., D'Errico, F., and Vincze, L. (2011). Discrediting moves in political debate. In P. Ricci-Bitti (Ed.), *Proceedings of UMMS 2011*, Springer LNCS, pp. 84--99.

Rosenberg, A., Hirschberg, J. (2009). Charisma perception from text and speech. In *Speech Communication*, 51(7), pp. 640--655.

Signorello, R., D'Errico, F., Poggi, I., and Demolin, D. (2012). How charisma is perceived from speech. A multidimensional approach. In *ASE/IEEE International Conference on Social Computing 2012*, Amsterdam, The Netherlands.

Strangert, E., Gustafson, J. (2008). What makes a good speaker? subject ratings, acoustic measurements and per- ceptual evaluations. In *Interspeech 2008*, pp. 1688--1691.

Touati, P. (1993). Prosodic aspects of political rhetoric. In *ESCA Workshop on Prosody*, Lund, Sweden, pp. 168--171.

Collaboratively built utterances in the C-ORAL-ROM-Corpus: temporal organization, prosodic design and forms of participation

Britta THÖRLE
University of Siegen
Adolf-Reichwein-Str. 2, D-57076 Siegen
thoerle@romanistik.uni-siegen.de

Abstract

This paper focuses on the collaborative production of single utterances, that is, utterances that are begun by one speaker and, before being syntactically, semantically or pragmatically completed, are continued by one or more different speakers. When regarding a coproduction defined in this way as a product, one notices that in general such coproductions are syntactically coherent entities that satisfy the criteria of grammatical well-formedness and as such, when the change in speakers is disregarded, hardly differ from monologically given utterances. However, when regarding them as a process, it becomes clear that coproducing is an ordered conversational process where the interaction partners place their spoken activities in relation to and in coordination with each other. Speakers utilize structural resources, such as syntactic or prosodic projections, that allow the communication partners to anticipate the continuation of the utterance as well as the moment when they can make their own contribution to the production. In addition, speakers command a repertoire of means by which they locally coordinate their activities. Depending on how they negotiate this local organization, different forms of participation within collaboratively built utterances, such as "helping out", "pre-empting" or "speaking in chorus" with the current speaker, can be distinguished.

Keywords: conversation analysis; coproduction; dialogic syntax; list construction; projection; prosody; timing.

1. Introduction: the coproduction of talk

The term coproduction is best known from the film industry where it refers to a film project in whose production more than one producer is involved. Similarly, in the history of literature the phenomenon of co-authorship can be found in numerous cases, as seen for example, in collaborative fiction or in the writing games of the Dada movement. In the new media, collaborative writing is common practice as can be observed with Wikipedia. In all these instances, a common text product is created in coproduction that on the surface does not show any distinction from a text that would have been produced by a single author. The same phenomenon exists in spoken language. Here the joint production of (oral) texts can perhaps even be considered as the normal way. A text is created through alternating contributions of the participating speakers, whereby the roles of producers and recipients cannot be strictly separated from each other. It is, for example, common practice in oral storytelling that those whose original role assignment is that of listener also participate – quite independent of whether they know of or were involved in the event being talked about. This joint text production goes so far that a single oral utterance is created by several speakers together. This is the case in the example below, an excerpt from a conversation between two men talking about contaminated meat and the role of the media:

Example 1: 🔊

1	*SRB:	[…] questo problema che è sempre esistito /
2		e esiste / e su tanti altri settori / tuttora //
3		però / &he / su tante cose / cioè nessuno le
4		prende in considerazione // perché no [/]
5		non fa audience / non fa interesse della
6		gente / per cui cioè / magari si mangia un
7		qualcosa che [/] che può far male / che fa
8		schifo / però / cioè / nessuno se ne rende
9		conto / e nessuno lo prende in
10		considerazione / perché +
→11	*SMN:	non fa notizi<a> //
12	*SRB:	<non> fa notizia //

(C-Oral-Rom : ifamdl06 macellaio)

What we can observe here is an utterance of speaker SRB which is syntactically not complete: *e nessuno lo prende in considerazione / perché +* (l. 9-10) and a completion of this utterance by another speaker SMN: *non fa notizia* (l. 11) which is repeated and thus ratified by the current speaker SRB: *non fa notizia* (l. 12). In short, and put simply, we could say that a *coproduction* is the production of a single utterance by more than one speaker.

In previous studies this phenomenon has been called *locuteur collectif* ('collective speaker') (Loufrani 1984; Blanche-Benveniste *et al.*, 1990) – a term that focuses on the fact that collaboratively built utterances hardly differ from those produced by a single speaker. Indeed, when regarding a collaboratively produced utterance as a *product*, one notices that in general it represents a syntactically coherent entity that satisfies the criteria of grammatical well-formedness. However, once the *process* is also considered, it becomes clear that coproducing is an ordered conversational process in which the interaction partners contribute their spoken activities in relation to and in coordination with each other.

The question that is addressed in this article is how speakers succeed in coordinating their activities in the coproduction of an utterance. In the following, we will analyse some examples of a larger set of coproductions extracted from the Spanish and Italian subcorpora of C-ORAL-ROM (Cresti & Moneglia, 2005). In section 2 we argue that the shared knowledge about language structure is a resource of oral coproduction because it

allows speakers to project the possible continuation of the ongoing utterance. Section 3 analyses the way in which participants coordinate their activities when they coproduce an utterance. This section focuses on differences concerning the precision timing and prosodic design of the coproduced element according to which different forms of participation within the collaborative production of utterances can be distinguished.

2.   Language structure as a product and a resource of oral coproduction

Language structure can be regarded as a result as well as a resource of discourse production or, as Humboldt (1836, 1999: 63) puts it, "language belongs to me, because I bring it forth as I do; and since the ground of this lies at once in the speaking and having-spoken of every generation of men, so far as speech-communication may have prevailed unbroken among them, it is language itself which restrains me when I speak." This conceptualization of language is very similar to the perspective taken by Interactional Linguistics where language structure is regarded as being actively (re)produced and thus emerging in interaction and, at the same time, as a shared knowledge which serves as a resource for the construction of discourse: "Rather than conceptualizing language as an abstract and balanced system of pre-established discrete elements which are combined with one another into 'sentences' that are then realized in speech, interactional evidence suggests that language forms and structures must be thought of in a more situated, context-sensitive fashion as actively (re)produced and locally adapted to the exigencies of the interaction at hand. In this sense they can be conceived of as arising or emerging in use. [...] In this view, syntax, just like prosody and semantics, is resource that can be relied on as shared knowledge in the speech community and that can be 'distributed' across speakers in collaborative productions." (Couper-Kuhlen & Selting, 2001: 4f)

The coproduction of utterances provides obvious evidence of this double principle – language structure as a result and a resource of speech activity. Looking again at lines 9 to 12 in the excerpt above, the construction which results from the coproduction (e nessuno lo prende in considerazione perché non fa notizia) can on the one hand be seen as an interactive achievement. On the other hand, the emerging construction (or the construction so far) serves as a resource for the "second speaker" who processes synchronically the emerging construction and realizes what is said and done by the "first speaker" only with a minimal temporal delay (Auer, 2000). The synchronic processing of SRB's ongoing utterance as well as the shared knowledge about constructions in Italian establish "discourse expectations" (Langacker, 2001) or "projections" (Auer, 2005) allowing SMN to anticipate the possible continuation of the utterance and to coproduce it. So the perché at this moment of the utterance production can be interpreted as a subordinating conjunction which projects a subordinated clause. In accordance with the grammatical projection, the prosodic characteristics of the ongoing utterance mark the utterance as incomplete. Thus, syntactic and prosodic projections allow a possible "collaborator" to anticipate the potential continuation as well as to predict the moment at which a particular continuation has to be uttered.[1] However, this does not imply that he actually supplies this continuation, and it does not explain either how the current speaker will handle this contribution to his utterance. In the following, we therefore deal with the participant's methods of the local organization of coproduced utterances.

3.   Temporal organization, prosodic design and forms of participation

In this section we address the question of how speakers coordinate their contributions to one single utterance regarding in particular the temporal organization of these contributions as well as their prosodic design. As Jefferson (1973) shows, recipients of some ongoing talk have the technical capacity to produce their talk with precision in relation to that ongoing talk.[2] In the following, we argue that speakers display quite different forms of participation within the collaborative production of utterances according to the precise timing and design of their contributions. We will treat as examples of such forms of participation: *helping collaboration, pre-empting* and *choral-coproduction.*

3.1 "Helping" collaboration: saying something *instead* of the current speaker

The analysis refers again to excerpt 1. When we look at what happens before the "second speaker" starts, we observe a break of nearly one second. This break (in the C-ORAL-ROM transcription interpreted as "+", a prosodic break marking an interruption) can be interpreted as a hesitation of the current speaker and a signal for the interlocutor to participate in the construction of the utterance.

---

[1]   For syntax as a resource for the coproduction of utterances see Thörle (2011). In this example, projection is not the only resource. There is a number of constructions in the previous discourse which could possibly function as a model for the utterance under construction: The speaker SRB himself seems to take up a construction in lines 3-4 which he varies twice in line 8-9 and 9-10:

SRB: cioè nessuno le prende in considerazione // (l. 3-4)
SRB: cioè nessuno se ne rende conto (l. 8-9)
SRB: e nessuno lo prende in considerazione // (l. 9-10)

His interlocutor SMN does the same and takes up previous constructions of SRB:

SRB: perché   no / fa audience / (l. 4-5)
            non fa interesse della gente (l. 5-6)
SMN: non fa notizia (l. 11)

Du Bois (2010: 13) might have thought of examples like this when he wrote: "Again and again, we witness dialogic co-participants speaking as though they were drawing on paradigmatic alternatives within a semantic field, seemingly exploiting just the kind of structure described by the great structural linguists from Saussure on."
[2]   See also Müller & Klaeger (2010).

9   *SRB: [...] e nessuno lo prende in considerazione /
10   perché + (**break of 0.8 sec**)

The break is followed by the completion of the utterance by the interlocutor, ending with a conclusive prosodic break ("//").

11   *SMN: non fa notizia //

While the interlocutor is uttering *non fa notizia,* the "first speaker" does not continue. We have a very short overlap only at the very end of the interlocutor's contribution, when the original speaker starts repeating what SMN has said before.

11   *SMN: non fa notizi<a> // 🔊
12   *SRB: <non> fa notizia // 🔊

It is important to note that SRB repeats the completion provided by SMN in a prosodically very similar manner as regards rhythm and melody. His repetition can thus be interpreted as a ratification of the interlocutor's contribution to the utterance.

In this example the coproduced element is designed to "fill a gap" in the utterance of the current speaker and as being said *in his place*. This is perhaps the most typical case of collaborative utterance construction which Ferrara (1992:220f) calls "helpful utterance completions". A "second speaker" detects a difficulty of the speaker in accessing an item in the mental lexicon and offers a minimal contribution – often not more than one or two words in length – which the "first speaker" typically ratifies by repetition.

3.2 Pre-empting: saying something *before* the current speaker

The next example is taken from an informal conversation between two women who talk about the dental problems of the mother of one of them:

Example 2: 🔊

1   *INM: lo que le estaba dando problemas / es la
2   muela esa //
3   *PAT: la que no han quitado [/] la que le han
4   quitado el nervio //
5   *INM: la que no le habían quitado el nervio //
6   *PAT: la que no //
7   *INM: con lo cual / ahora tienen que volverle a [/] a
8   levantar / toda la dentadura / matarle el
9   nervio +
→10   *PAT: y ponérsela otra vez // ¡madre <mía!>//
11   *PAC: <es que> si a ti te
12   matan los nervios /
(C-Oral-Rom : efamcv06 las muelas)

What we are interested in here is the completion of an utterance, which is obviously not complete at the moment when the interlocutor provides her contribution. But, in contrast to the case analysed in the last section, the

current speaker does not seem to interrupt herself but to be pre-empted by her interlocutor. Although the transcription of the C-ORAL-ROM-Corpus interprets the transition again as a prosodic break marking an interruption ("+"), there is a fundamental difference between these two examples, which lies in the temporal and rhythmical design of the utterances.

In lines 7 to 9 INM is constructing what we could call with Jefferson (1990) a list construction. Lists are described by Sánchez-Ayala (2003: 325-332) as recurrent lexico-grammatical patterns in colloquial speech which are characterized, amongst other things, by prosodic features such as a robust parallelism between their prosodic and lexico-grammatical constituents, lengthening of the ultimate lexical stress of each intonation unit, the musical effect of "stylized intonation" as well as a coherent thematic structure in which the different parts of the list correspond to different stages in the rhetorical development of a point. Lists can therefore be considered as a holistic *gestalt* to which interlocutors orientate themselves in the construction of talk.

In example 2 INM has already produced two list elements: (*con lo cual ahora tienen que volverle a*) *levantar toda la dentadura* - 🔊 - and *matarle el nervio* - 🔊 - (l. 7-9). Both are infinitive phrases, uttered in a special rhythm which is produced by the stressing of syllables in *tOda* and *matAr* and characterized by a non-conclusive intonation structure. The two first list elements project – by virtue of their syntactic, semantic and prosodic characteristics – a third list element that the other participants hence are able to anticipate. [3] This third element *y ponérsela otra vez* - 🔊 - is provided by PAT, but – and this is important – *before* the original speaker is expected to realize it and without being "invited" by any hesitation marker. To understand this, we have to look at the temporal organization of the list construction:

After the first list element *levantar toda la dentadura,* there is a pause of 0.432 sec. The original speaker is constructing a rhythm structure for her list that would allow us to expect a break of more or less the same length after the second list element *matarle el nervio.*[4] Now, before the expectable "right" moment for the third element has come (that means after an interval of only 0.08 sec), PAT completes the list with *y ponérsela otra vez* (l. 10).[5] This means her contribution is not designed to be

---

[3]   As Jefferson (1990) shows, the three-partedness of lists appears to have "programmatic relevance" for its construction. Participants orient to this three-parted nature so that lists can become a conversational sequential resource. This means that a "list-in-progress is recognizable as a list prior to its completion" and that a second part of the list projects a third-as final part (Lerner 1991: 448).

[4]   The pausing between list elements cannot be interpreted as indicative of trouble. It is rather a "rest beat" in the rhythmical structure of the list (cf. also Lerner 1996: 242f).

[5]   Lerner (1996: 242) calls this kind of coproduction "anticipatory completion": "With anticipatory completion, onset occurs at a TCU-internal component completion, and therefore not at a place the turn itself could in most circumstances be

realized *with* the original speaker or in her place but *before* her. It seems to be a sort of friendly "competition" about who realizes the end of the story first as it may occur in the genre of "women's friendly talk" described by Coates (1997).[6]

Until now we have dealt with examples in which the contribution of the "second speaker" seemed to conform not only structurally to the ongoing utterance but also more or less semantically to what the current speaker would have uttered by himself. As the next excerpt shows the same procedure can be exploited to utter something obviously divergent from the current speaker's intention. This is an extract from a political debate in an Italian talk show:

Example 3: 🔊

```
1   *BER:  [...] il regime precedente / quello di Hoxha / era
2          / un regime / autoritario / chiuso / dispotico /
3   *BUT:  comu<nista> //
4   *BER:      <che ha> che si diceva comunista // e non
5          aveva / alcuna traccia / delle ragioni per cui
6          siamo comunisti // [...]
    (C-Oral-Rom : imedts 03 porta a porta)
```

What we can observe here seems at first to be very similar to the previous examples. BER (Fausto Bertinotti, at that time secretary of the communist party PRC) constructs a list which is characterized by a particular rhythmical and melodic pattern: *autoritario* - 🔊 - (falling intonation) – break of 0.44 sec – *chiuso* - 🔊 - (falling intonation) – break of 0.27 sec – *dispotico* - 🔊 - (rising intonation). The rising and non-conclusive intonation of *dispotico* makes us expect a continuation of the enumeration. Indeed, this continuation is produced by BER himself (*che ha ...*), but his interlocutor pre-empts him, proposing his continuation of the list (*comunista*) just an instant *before* the moment in which the next list element was expected to appear. In contrast to the previous examples, this time, the reaction of BER shows us that the contribution of the interlocutor obviously does not correspond to his own intentions: He does not complete his own next element of the list and interrupts himself to take up his interlocutor's contribution (*comunista*) that he subsequently reformulates (*che si diceva comunista ...*), relativizing by this means the validity of the resulting proposition. This shows that the procedure of anticipatory completion of utterances can also be used to distance oneself from what the other is saying (Mondada 1999:25f).

In the examples in this section, the contributions of the "second speakers" to ongoing list constructions are clearly designed to pre-empt the current speaker. They exploit the semantic, syntactic and prosodic features of

these lists to project the moment at which a probable last list element will be uttered in order to provide such an element *before* the current speaker does (or is expected to do so).

3.3 Choral coproduction – saying something *with* the current speaker

Finally, we present examples of coproductions where the element provided by the interlocutor is designed to be uttered simultaneously – in chorus – with the current speaker. According to Lerner (2002: 22) we call this phenomenon choral co-production which the author describes as "'voicing the same words in the same time' as another speaker – or at least demonstrating that one is aiming at that result".

Example 4: 🔊

```
1     *VIR:  entonces / como decías / el que / el VIH / el
2            virus del SIDA / sea capaz / de / atacar
3            específicamente a las células / fundamentales
4            del sistema de las defensas / del organismo
5            deja / al organismo / <indefenso> //
6     %alt: (19) toces
→7    *BLA:              [<] <indefenso> //
8     *VRI:  pero es que por otra parte / tiene un periodo
9            de incubación / muy largo // o sea desde que
10           una persona se infecta / hasta que desarrolla la
11           enfermedad / pasan ocho o diez años / como
12           término medio // con lo cual / cuando surge /
13           en mil novecientos ochenta y uno / la primera
14           descripción de / una enfermedad nueva / que /
15           luego / &eh / se llamó SIDA / etcétera y se / ha
16           investigado / enormemente / pues ya había
17           millones y millones / de personas <infectadas>
18           // no ? precisamente
→19   *BLA:  [<] <infectadas> //
20    *VRI:  por ese / período de incubación tan largo
             (C-Oral-Rom: emedts11 el virus del SIDA)
```

In this extract taken from a Spanish television interview the interviewer frequently coproduces the terminal items of the interviewee's utterances:

```
5     *VIR:  deja / al organismo / <indefenso> //
7     BLA:               [<] <indefenso> // 🔊)

17    *VRI:  pues ya había millones y millones / de personas
18           <infectadas> // no? precisamente por
19    *BLA:  [<] <infectadas> //
20    *VRI:  por ese / período de incubación tan largo 🔊)
```

If we look at the organizational features of this coproduction, we observe that there is no hesitation marker in the utterance of the interviewee, that the interviewee does not stop speaking so that the contribution of the interviewer produces an overlapping of speech, and that there is no ratification of the coproduced element. Focussing on the temporal organization, we note that the contribution of the interviewer seems to be designed to be realized not before

---

finished. That is, a next speaker begins speaking before the projected completion of a TCU and thus within the projected turn space of the still current speaker."

[6]   Competition here does not refer to competitive turn incomings as described by French & Local (1986) (cf. Szczepek 2000:26ff).

the current speaker but simultaneously: Even if the contributions of the interviewer do not start exactly at the same moment (infectadas a little bit earlier), the interviewer does not seem to try to pre-empt the current speaker. Rather, he speaks very calmly and adapts the projected conclusive intonation structure of the interviewee's utterance. What BLA is doing here when he coproduces VIR's utterances corresponds to a back-channel signal. He accompanies the discourse production of the current speaker showing that he is following and understanding the argumentation.[7]

4.   Conclusion

In this article, oral conversation has been analysed as a highly collaborative practice in which a single utterance can be produced by several speakers together. When doing so, "second speakers" exploit the syntactic, semantic and prosodic projections established by the utterance of their interlocutor to produce their own contribution with precision in relation to the ongoing talk. It has been argued that speakers use this general capacity for precise placement together with prosodic means to display quite different forms of participation, such as "helping", "competing" or "being in chorus" with the current speaker and, in so doing, achieve different pragmatic aims.

5.   References

Auer, P. (2000). On line-Syntax – Oder: was es bedeuten könnte, die Zeitlichkeit der mündlichen Sprache ernst zu nehmen. In *Sprache und Literatur*, 85, pp. 43--56.

Auer, P. (2005). Projection in interaction and projection in grammar. Text, 25 (1), pp. 7--36.

Blanche-Benveniste, C. *et al.* (1990). *Le français parlé. Etudes grammaticales*. Paris: CNRS.

Coates, J. (1997). The construction of a collaborative floor in women's friendly talk. In T. Givón (Ed.), *Conversation: cognitive, communicative and social perspectives*. Amsterdam / Philadelphia: Benjamins, pp. 55--89.

Couper-Kuhlen, E., Selting, M. (2001). Introducing Interactional Linguistics. In E. Couper-Kuhlen & Selting, M. (Eds.), *Studies in Interactional Linguistics*. Amsterdam: Benjamins, pp. 1--22.

Cresti, E., Moneglia, M. (2005). *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: Benjamins.

Du Bois, J.W. (2010). *Towards a Dialogic Syntax*. (Draft)

Ferrara, K. (1992). The Interactive Achievement of a Sentence. Joint Productions in Therapeutic Discourse.

Discourse Processes, 15, pp. 207--228.

French, P., Local, J. (1986). Prosodic Features and the Management of Interruptions. In C. Johns-Lewis (Ed.), *Intonation and Discourse*. Beckenham: Croom Helm, pp. 157--180.

Jefferson, G. (1973). A Case of Precision Timing in Ordinary Conversation. Overlapped Tag-Positioned Address Terms in Closing Sequences. Semiotica, 9 (1), pp. 47--96.

Jefferson, G. (1990). List-Construction as a Task and Resource. In G. Psathas (Ed.), *Interaction Competence*. Washington, D.C.: University Press, pp. 63--92.

Langacker, R.W. (2001). Discourse in Cognitive Grammar. Cognitive Linguistics, 12 (2), pp. 143--188.

Lerner, G.H. (1991). On the syntax of sentences-in-progress. In *Language and Society*, 20, pp. 441--485.

Lerner, G.H. (1996). On the "semi-permeable" character of grammatical units in conversation: conditional entry into the turn space of another speaker. In E. Ochs, E. Schlegloff and S.A. Thompson (Eds.), *Interaction and grammar*. Cambridge: University Press, pp. 238--276.

Lerner, G.H. (2002). Turn-Sharing. The Choral Co-Production of Talk-in-Interaction. In C.E. Ford, B.A. Fox and S.A. Thompson (Eds.), *The Language of Turn and Sequence*. Oxford: University Press, pp. 225--256.

Loufrani, C. (1984). Le locuteur collectif. Typologie de configurations discursives. In *Recherches sur le français parlé*, 6, pp. 169--193.

Mondada, L. (1999). L'organisation séquentielle des ressources linguistiques dans l'élaboration collective des descriptions. In *Langage et société*, 89, pp. 9--36.

Müller, F.E., Klaeger, S. (2010). Collaborations syntaxiques – Formes et fonctions de leur usage dans un groupe subculturel lyonnais. In *Pratiques*, 147/148, pp. 223--243.

Sánchez-Ayala, I. (2003). Constructions as Resources for Interactions: Lists in Spanish and English Conversation. In *Discourse Studies*, 5 (3), pp. 323--349.

Szczepek, B. (2000). Formal Aspects of Collaborative Productions in English Conversation. InList, 21.

Thörle, B. (2011). La sintaxis como recurso en la producción colectiva del discurso: a propósito de la construcción colaborativa de enunciados en español. In D. Jacob & A. Dufter (Eds.), *Syntaxe, structure informationnelle et organisation du discours dans les langues romanes*. Frankfurt am Main: Lang, pp. 153--171.

von Humboldt, W. (1999). *On the Diversity of Human Language Construction and its Influence on the Mental Development of the Human Species*. Ed. by Michael Losonsky. Cambridge: University Press.

---

[7]   In our example, the choral coproduction is clearly not turn competitive. Nevertheless, as Lerner (2002: 239ff) puts it, choral coproduction "can be used as a method for gaining sole speakership. [...] In this case, turn-sharing is a first step to sole turn occupancy." In our corpus this is sometimes the case in radio programs where hearers call in to and moderators co-produce terminal items in chorus to get the floor and to lead to the next caller.

A construção da cadeia referencial em sequências narrativas orais

Gustavo Ximenes CUNHA

Universidade Federal de Minas Gerais (UFMG/CNPq)
Av. Antônio Carlos, 6627 - Belo Horizonte/MG - Brasil
ximenescunha@yahoo.com.br

Abstract

The point of this paper is to study the informational continuity and progression in two narrative sequences extracted from an interview sociolinguistics. This study resulted in the mapping of referential chain of sequences, to understand the management of referents, as well as what are the linguistic clues (pronouns and nominal expressions) that signal this management. Following the method proposed by Modular Approach to Discourse Analysis, the study found that the progressions occur within each episode of the sequences. About the linguistic clues, each episode of the sequences featured many topical clues that facilitate understanding of referential chain. But the sequences did not show a predominance of full or empty expressions.

Keywords: referential chain, narrative sequence, modularity.

1. Introdução

O objetivo deste trabalho é investigar o processo de construção da cadeia referencial em sequências narrativas orais. Especificamente, o trabalho estuda o modo como ocorrem a continuidade e a progressão informacionais em duas sequências narrativas extraídas de uma entrevista sociolinguística, que integra o corpus do "Projeto Mineirês" (Ramos, 2007). Esse estudo implicou o mapeamento da cadeia referencial das sequências, na busca por compreender como a sua produtora, uma belo-horizontina de 54 anos com formação superior, faz a gestão dos referentes, introduzindo-os, preservando-os, modificando-os e reintroduzindo-os no discurso, bem como quais são as marcas linguísticas (pronomes e expressões nominais) que sinalizam essas diferentes ações.

O estudo foi feito com base na perspectiva teórica e metodológica do Modelo de Análise Modular do Discurso (Roulet, Filliettaz & Grobet, 2001). Seguindo o método proposto por esse modelo, a análise se desenvolveu em três etapas. Na primeira, os fragmentos selecionados foram caracterizados como sequências narrativas. Em seguida, na segunda etapa, analisou-se a forma como é feita a construção da cadeia referencial nas duas sequências. Por fim, os estudos realizados nas duas primeiras etapas foram combinados, na busca por compreender o modo como, nas sequências narrativas estudadas, ocorre a construção da cadeia referencial e a sua marcação linguística.

Neste artigo, realizamos inicialmente uma breve caracterização do gênero de discurso entrevista sociolinguística, ao qual pertencem as sequências estudadas. Em seguida, será apresentado o corpus de análise. Posteriormente, será feita uma apresentação do referencial teórico adotado, o Modelo de Análise Modular do Discurso. Por fim, o artigo expõe as três etapas da análise realizada.

2. O gênero de discurso entrevista sociolinguística

A entrevista sociolinguística é um gênero pertencente à esfera acadêmica, já que a sua função básica é permitir a um pesquisador da área de Linguística colher dados autênticos de língua oral com fins de pesquisa e análise.

A produção de um texto pertencente a esse gênero implica a participação de pelo menos dois interlocutores. De um lado, está o entrevistador, cuja função é propor os tópicos a serem abordados. Nessa interação, o entrevistador, diferentemente das entrevistas que ocorrem em outras esferas, como a jornalística, assume o papel social de pesquisador. Do outro lado, está o entrevistado, cuja função é desenvolver os tópicos propostos pelo entrevistador. Nesse gênero, a função social assumida pelo entrevistado é a de falante de uma dada língua natural. Nesse sentido e também diferentemente do que ocorre em entrevistas televisivas, por exemplo, importa mais a forma como o entrevistado utiliza a língua/linguagem para se expressar do que propriamente suas opiniões ou sua visão de mundo acerca dos fatos tratados (Tavares, 2004).

Na entrevista sociolinguística, há um certo grau de formalidade. Essa formalidade se deve a alguns fatores. O primeiro deles se refere à esfera acadêmica a que esse gênero pertence e em que se constituiu.

O segundo fator responsável pela formalidade da entrevista sociolinguística está ligado ao primeiro e diz respeito à imagem que o entrevistado pode construir acerca do entrevistador. Em nossa sociedade, o papel social que este exerce, o de pesquisador, é um papel considerado de prestígio (Mondada, 1995). O conhecimento que supostamente só o pesquisador e seus pares possuem e para o qual a fala do entrevistado será fonte de estudo pode ser um fator de inibição, que talvez leve o entrevistado a se comportar de maneira mais formal.

A formalidade da entrevista sociolinguística se deve ainda ao fato de que entrevistado e entrevistador não se conhecem. Em outros termos, entre eles, há pouca ou nenhuma intimidade, o que pode favorecer uma interação mais formal, menos espontânea.

Esses três fatores responsáveis pela formalidade da entrevista sociolinguística, bem como a função própria desse gênero terão impacto sobre a estruturação da entrevista. Assim, ao contrário do que ocorre, por exemplo, em conversações espontâneas entre amigos, os participantes de uma entrevista sociolinguística abordam fatos vivenciados apenas pelo entrevistado, raramente tematizam o contexto imediato em que se dá a interação e, caso os interlocutores mobilizem conhecimentos partilhados, estes serão informações introduzidas em momentos anteriores da própria entrevista ou informações compartilhadas de modo geral pelos membros da sociedade a que os interlocutores pertencem, exatamente porque estes não se conhecem ou se conhecem pouco.

Neste trabalho, conduziu nossas análises a hipótese de que essas características do gênero entrevista sociolinguística têm impacto sobre a forma como o produtor de sequências narrativas orais pertencentes a esse gênero realiza a ativação e a reativação de referentes e utiliza recursos linguísticos, como pronomes, expressões nominais e elipses, para sinalizar essas operações de ativação e reativação de referentes.

3. Corpus de análise

Tendo em vista os objetivos deste trabalho, a análise focalizou apenas um turno produzido por uma belo-horizontina de 54 anos com 3° grau completo. Na passagem selecionada para análise, a entrevista desenvolve o tópico "infância", já iniciado em turnos anteriores e sobre o qual a entrevistadora ainda pede esclarecimentos. A transcrição do par de turnos produzidos por entrevistadora e entrevistada segue abaixo[1].

Entrevistadora: Ah certo, i eram quantas mulheres assim, cê falou que eram dez irmãos.

(01) Eram seis mulheres i quatro homens (02) i era interessanti pelo siguinti, (03) porque igual os homens tinha brincadera deles, (04) mais, como eu já falei, (05) agenti brincava tamém com eles, (06) agora quando igual agenti ia brincá di buneca (07) agenti num pudia:: (08) agenti chamava, (09) quiria qui eles fossem pai, (10) qui batizassem i tudu, (11) mais eles não gostavam di bricá di buneca, (12) mais quandu as brincaderas davam errada (13) tamém eles criticavam, (14) eles riam muitu, (15) eu lembro muitu minha irmã mais velha ganhou uma buneca +, (16) ela era apaxonada com uma buneca grande (17) i a minha mãe num tinha condições di comprá buneca pra todo mundu, (18) intão compró, (19) i as amigas, nossas amigas todas tinham bunecas boas, bunecas famosas, im material bom i

tudu, (20) i a minha mãe num pudia dá seis bunecas, (21) intão compro uma buneca di papelão pra minha irmã + , (22) só qui a buneca era muitu bunita, (23) o rosto muitu bem pintado, (24) e::, pudia trocá as roupas dela (25) que ela tinha essa opção i tudu (26) purque os braçinhos moviam i tudu, (27) mais um dia (28) juntamos lá com otras amigas (29) pra [buscá] brincá di buneca (30) cada uma com uma buneca mais linda (31) fomos todo mundu brincá di buneca (32) i tudu qui uma fazia a otra fazia, (33) aí uma amiguinha nossa inventó di dá o banho, (34) nós tava brincanu num, (35) nós tínhamos ido num lá:: { }, (36) até existi ainda, (37) é uma área qui tem lá no hospital da baleia + (38) qui tinha água corrente, tinha as grutas qui as águas disciam, (39) i lá agenti pudia i, (40) a entrada era livre, (41) num pagava, (42) intão era um lugar qui a genti ia todo final di semana pa brincá por lá, (43) i lá num tinha pirigo, (44) num passava ônibus, (45) tinha segurança (46) pur causa do hospitali (47) i tinha uns riachozinhos ondi curria uma água, (48) i aí combinamos di brincá di dá banho nas bunecas +, (49) i aí foi todo mundu (50) i ta lá naqueli processo (51) cada uma arruma o banho da sua, (52) tira a ropa (53) e aquela confusão toda (54) i foi todo mundu pru riacho dá banho nas bunecas +, (55) quando a minha irmã pôs a dela na água, (56) a dela era di papelão, (57) ela não sabia, (58) a buneca começó a dismanchá +, (59) i ela começó a chorá (60) i aqueli disispero (61) i as otras meninas com dó (62) i os meninos riam riam riam (63) i lá foi a buneca si disfazendu toda. (64) Issu foi uma decepção muitu grandi pra ela, (65) ela choró muitu, (66) mais tamém quandu chego im casa qui nós comentamu, (67) contamu, (68) meu pai providenciô logo otra buneca, (69) aí já ele mesmo já num quis outra buneca di papelão + , (70) viu qui foi muita humilhação pra ela (71) i ai já deu uma buneca daqueli plástico, (72) era um material plástico, (73) mais um material duru i bom (74) tamém do rosto muitu bunitu, bem pintadu, (75) i issu foi mutivu assim di crítica dus meninos um tempo longo (76) purque toda vez qui falava das bunecas (77) a história da buneca di papelão surgia.

Para realizar a análise do turno produzido pela entrevistada, utilizaremos como referencial teórico e metodológico o Modelo de Análise Modular do Discurso, modelo que apresentamos no próximo item.

4. Referencial teórico e metodológico

Em sua versão atual (Roulet, Filliettaz & Grobet, 2001; Filliettaz & Roulet, 2002; Filliettaz, 2004; Marinho, Pires & Villela, 2007), o modelo modular constitui um instrumento de descrição e explicação da complexidade discursiva e compõe um quadro teórico e metodológico que visa a reunir, em uma mesma abordagem da complexidade da organização do discurso, as contribuições de pesquisadores que se centraram em aspectos isolados dessa organização.

---

[1] O trecho foi reproduzido da forma como está disponibilizado no site do projeto "Mineirês" (Ramos, 2007). Apenas a numeração não consta no texto original. Ela foi por nós inserida e indica que o trecho foi segmentado em atos. O ato é a unidade mínima de análise adotada pelo modelo modular.

Nesse modelo, identificam-se inicialmente os módulos que entram na composição dos discursos[2]. Na produção e na interpretação de toda forma discursiva, as informações de origem modular se interrelacionam em unidades complexas de análise, que são as formas de organização[3].

Neste trabalho, a análise do modo como se dá a construção da cadeia referencial em duas sequências narrativas orais será feita a partir da combinação do estudo de duas formas de organização elementares: a sequencial e a informacional.

Na forma de organização sequencial, estudam-se os tipos de discurso e as sequências discursivas. O objetivo aqui é, basicamente, o de segmentar as produções discursivas nas sequências (narrativas, descritivas e deliberativas) que as compõem. Sobre a sequência narrativa, o modelo, baseando-se nos trabalhos de Labov (1972, 1997) e Adam (1992), considera que a estrutura típica uma sequência narrativa se compõe dos episódios *sumário*, *estado inicial*, *complicação*, *reação (avaliação)*, *resolução* e *estado final* (Filliettaz, 1999, Cunha, 2010).

Na forma de organização informacional, estuda-se a construção da cadeia referencial, a fim de tratar a continuidade e a progressão informacionais do discurso. Mais particularmente, o objetivo é, valendo-se de contribuições de Danes (1974) e Chafe (1994), analisar a estrutura informacional de cada unidade mínima de referência (o ato), descrevendo como cada ato se ancora em uma informação previamente estocada na memória discursiva[4], o tópico. Nessa forma de organização, estuda-se ainda a inserção de cada ato na estrutura do discurso, com base na análise dos tipos de progressão informacional entre os atos. No modelo, os tipos de progressão considerados são *progressão linear*, *progressão com tópico constante* e *encadeamento à distância* (Grobet, 2000).

A seguir, serão apresentados os resultados da análise do corpus. Inicialmente, apresentamos a análise da forma de organização sequencial. Posteriormente, apresentamos os resultados da análise da forma de organização informacional. Por fim, serão combinados os resultados das análises das formas de organização sequencial e informacional.

5. Análise da forma de organização sequencial

A análise da forma de organização sequencial do turno produzido pela entrevistada revelou que esse turno constitui uma grande sequência narrativa. Do ponto de vista referencial, essa sequência narrativa atualiza uma estrutura praxeológica formada por todos os episódios componentes do tipo narrativo. Assim, essa sequência, que chamamos de sequência narrativa 1, apresenta sumário (01-14), estado inicial (15-54), complicação (55-63), avaliação (64-65), resolução (66-74) e estado final (75-77).

No sumário (01-14), a locutora resume o tópico que será abordado nos episódios seguintes: "brincadeiras infantis entre os irmãos" ou, mais especificamente, "brincadeiras com bonecas entre os irmãos". Em seguida, o estado inicial (15-54) traz muitas informações sobre as personagens envolvidas na história (a própria narradora, suas irmãs, sua mãe, suas amigas), bem como sobre o lugar onde se passou o fato principal da narrativa ("uma área lá no Hospital da Baleia"). Após o estado inicial, a locutora narra, na complicação (55-63), o acontecimento principal da narrativa, o acontecimento que justifica a sua própria ação de narrar: ao ser colocada na água, a boneca de sua irmã mais velha desmanchou, porque era feita de papelão. Depois, a locutora faz, na avaliação (64-65), um comentário, esclarecendo que esse acontecimento foi uma decepção muito grande para sua irmã. Feita avaliação, a locutora informa, na resolução (66-74), qual a consequência do acontecimento expresso na complicação: a irmã ganhou do pai outra boneca, mas agora uma boneca de plástico. No estado final (75-77), a locutora informa como tudo ficou após os acontecimentos centrais da narrativa, apresentando uma nova situação de equilíbrio.

A análise da forma de organização sequencial do turno revelou ainda que o estado inicial da sequência narrativa 1 constitui uma sequência narrativa encaixada, a qual chamamos de sequência narrativa 2. Do ponto de vista referencial, essa segunda sequência se constitui dos episódios estado inicial 1 (15-26), complicação 1 (27-33), estado inicial 2 (34-47), complicação 2 (48-53), resolução (54).

No estado inicial 1 (15-26), são informados os personagens que vão participar da história (a própria narradora, suas irmãs, sua mãe, suas amigas). Depois, vem a complicação 1 (27-33), episódio no qual é revelado o acontecimento central dessa narrativa encaixada: quando as meninas brincavam de boneca, uma das amigas inventou de dar banho nelas. Após essa complicação, um segundo estado inicial (34-47) descreve o local da brincadeira: "uma área lá no Hospital da Baleia". Depois desse estado inicial 2, que funciona como uma espécie de parênteses, a locutora dá sequência à complicação 1, informando, na complicação 2 (48-53), como se desenrolou o processo da brincadeira de dar banho nas bonecas. Finalmente, a resolução (54) informa o final desse processo, que foi a ida de todos para o riacho. Essa resolução é a etapa que antecede imediatamente a complicação da sequência 1, em que se encaixa toda essa sequência 2.

---

[2] Nessa abordagem, considera-se que cada dimensão do discurso se constitui de módulos. Assim, a dimensão linguística se constitui dos módulos lexical e sintático; a dimensão textual se constitui do modulo hierárquico; e a dimensão situacional se constitui dos módulos interacional e referencial.

[3] No modelo modular, as formas de organização são: fono-prosódica, semântica, relacional, informacional, enunciativa, sequencial, operacional, periódica, tópica, polifônica, composicional, estratégica.

[4] A memória discursiva é definida como "conjunto de saberes conscientemente partilhados pelos interlocutores" (Berrendonner, 1983, p. 230).

6. Análise da forma de organização informacional

Por motivo de espaço, não apresentaremos a análise completa da estrutura informacional de todo o turno produzido pela locutora. Abordaremos apenas os aspectos que nos pareceram mais relevantes.

Nesse turno, há uma grande concentração de traços tópicos, que são as marcas linguísticas, como expressões nominais e pronomes, que fazem referência ao tópico do ato em que ocorrem. Do ponto de vista quantitativo, verificou-se que 57 dos 77 atos do turno apresentam alguma marca linguística remetendo ao tópico. Essa marcação intensa pode ser ilustrada com a parte inicial do turno[5].

(01) Eram **seis mulheres i quatro homens** [irmãos]	Progressão linear
(02) (irmãos) i era interessanti pelo siguinti,	Tópico constante
(03) purque igual **os homens** [irmãos] tinha brincadera **deles**,	Tópico constante
(04) (agenti brincava tamém com eles) mais, como eu já falei,	Encadeamento à distância
(05) **agenti** [agenti – seis irmãs] brincava tamém com eles,	Tópico constante
(06) agora quando igual **agenti** ia brincá di buneca	Tópico constante
(07) **agenti** num pudia::	Tópico constante
(08) **agenti** chamava,	Tópico constante
(09) (agenti) quiria qui eles fossem pai,	Tópico constante
(10) (agenti quiria) qui batizassem i tudu,	Tópico constante
(11) mais **eles** [irmãos] não gostavam di bricá di buneca,	Progressão linear

Tabela 1: Excerto da entrevista com análise informacional

Em todo o turno, a locutora mobiliza uma grande quantidade de marcas, cuja função é permitir à interlocutora identificar o tópico do ato. Em outros termos, essas marcas têm como função guiar a interlocutora em seu processo interpretativo. Assim, no trecho acima, as várias ocorrências da expressão pronominal "agenti", de pronomes como "eles" e "deles" e de expressões nominais como "os homens" e "seis mulheres i quatro homens" permitem a compreensão de que os atos em que ocorrem se referem a objetos de discurso previamente estocados na memória discursiva.

A relevância desse resultado está em fornecer uma evidência que contesta a hipótese de que na linguagem

---

[5] Esse quadro apresenta o resultado da análise informacional de um texto. Na coluna esquerda, os atos são numerados e os traços que verbalizam o tópico são apresentados em negrito; o tópico assim verbalizado aparece entre colchetes, depois do traço. Quando o tópico é implícito, ou seja, não verbalizado por traço tópico, ele aparece entre parênteses, no início do ato. Na coluna direita, são apresentadas as progressões informacionais que ligam os atos aos seus tópicos.

oral o locutor não se preocuparia em explicitar os referentes mobilizados, por serem estes facilmente acessíveis ao interlocutor. Na verdade, a necessidade de explicitação de referentes parece decorrer mais das condições de produção do texto do que da modalidade (oral ou escrita) do texto.

Quanto ao tipo de progressão informacional, verificou-se que dos 77 atos 18 se ligam ao tópico por progressão linear, 20 por encadeamento à distância e 39 por tópico constante. Dessa forma, no turno analisado, há um predomínio de progressão por tópico constante, que ocorre quando uma série de atos se ancora em um mesmo tópico. Ou seja, nesse tipo de progressão, o locutor trata de um mesmo tópico em todos os atos, acrescentando informações a ele. Exemplo:

(35) **nós** tínhamos ido num lá:: { },	Tópico constante
(36) (lá) até existi ainda,	Progressão linear
(37) é **uma área** [lá] qui tem lá no hospital da baleia +	Tópico constante
(38) **qui** [uma área] tinha água corrente, tinha as grutas qui as águas disciam,	Tópico constante
(39) i **lá** agenti pudia i,	Tópico constante
(40) **a entrada** era livre,	Tópico constante
(41) (a entrada da área) num pagava,	Tópico constante
(42) intão era **um lugar** [uma área] qui a genti ia todo final di semana pa brincá por **lá**,	Tópico constante
(43) i **lá** num tinha pirigo,	Tópico constante
(44) (uma área) num passava ônibus,	Tópico constante
(45) (uma área) tinha segurança	Tópico constante
(46) (lá tinha segurança) pur causa do hospitali	Tópico constante
(47) (uma área) i tinha uns riachozinhos ondi curria uma água,	Tópico constante

Tabela 2: Excerto da entrevista com análise informacional

Nesse trecho, o local onde se passou a brincadeira ("uma área lá no Hospital da Baleia") é o tópico. A esse tópico a locutora acrescenta uma série de informações, que têm como fim caracterizar o local.

O predomínio de progressão por tópico constante se explica pelo fato de que, ao narrar fatos de sua vida, a locutora não propõe mudanças radicais de tópicos. Essa estratégia de construção textual é eficaz, porque, como a locutora aborda fatos não vivenciados pela interlocutora, a progressão por tópico constante trata de informações facilmente acessíveis a esta, o que permite a reconstrução adequada da cadeia referencial proposta.

7. Combinando as análises das formas de organização sequencial e informacional

Essa etapa da análise combina os resultados das análises sequencial e informacional, apresentadas nos itens anteriores, a fim de verificar como se dá o processo de construção da cadeia referencial no interior das duas

sequências identificadas. Essa etapa se guiou por algumas questões, cujas respostas somente poderiam ser alcançadas mediante a combinação das análises efetuadas anteriormente. Essas questões são:

- Quais e quantos tipos de progressão há dentro de cada episódio?
- Quantos traços tópicos há no interior de cada episódio?
- No interior de cada episódio, quantos traços tópicos são expressões referencialmente plenas e quantos são expressões referencialmente vazias?

A continuação deste item tem como fim oferecer respostas a essas questões.

7.1 Quais e quantos tipos de progressão há dentro de cada episódio?

No interior de cada episódio das sequências 1 e 2, verificou-se o predomínio da progressão por tópico constante. O predomínio de progressão por tópico constante no interior de cada episódio se explica pelo fato de que dentro de um episódio não costuma haver mudanças radicais de tópicos, e o locutor costuma tratar de informações facilmente acessíveis ao interlocutor.

A única exceção foi a complicação da sequência 1, que exibiu uma quantidade elevada de encadeamentos à distância, que é quando o tópico de um ato tem origem não no ato precedente, mas em um ato mais distante. Entretanto, em muitas ocorrências desse tipo de encadeamento na complicação, esses encadeamentos são bastante locais, isto é, as informações que funcionam como tópicos têm origem em atos localizados dentro do próprio episódio. Exemplo:

(55) quando **a minha irmã** [minha irmã mais velha] pôs a dela na água,	Encadeamento à distância
(56) **a dela** [boneca] era di papelão,	Progressão linear
(57) **ela** [minha irmã mais velha] não sabia,	Progressão linear
(58) **a buneca** começó a dismanchá +,	Encadeamento à distância
(59) i **ela** [minha irmã mais velha] começó a chorá	Encadeamento à distância

Tabela 3: Excerto da entrevista com análise informacional

Nesse trecho, que é parte da complicação da sequência 1, os atos (58) e (59) se ligam aos tópicos por encadeamento à distância, mas esses tópicos têm origem em atos muito próximos, (56) e (57) respectivamente.

7.2 Quantos traços tópicos há no interior de cada episódio?

Nas sequências narrativas estudadas, há uma grande concentração de traços tópicos em cada episódio. Na sequência narrativa 1, verificamos o seguinte resultado: sumário (11 traços em 14 atos), estado inicial (30 traços

em 40 atos), complicação (9 traços em 9 atos), avaliação (2 traços em 2 atos), resolução (3 traços em 9 atos), estado final (2 traços em 3 atos).

Na sequência 2, os resultados são: estado inicial 1 (10 traços em 12 atos), complicação 1 (5 traços em 7 atos), estado inicial 2 (9 traços em 14 atos), complicação 2 (5 traços em 6 atos), resolução (1 traço em 1 ato).

Como exposto na análise da organização informacional, esses resultados contrariam a crença de que na linguagem oral o locutor não se preocupa em explicitar os referentes mobilizados, por serem estes facilmente acessíveis pelo interlocutor.

Na entrevista, a locutora fala de uma situação não vivenciada interlocutora e não tematiza o contexto imediato em que se dá a interação. Por esse motivo, a locutora não pode contar com conhecimentos da ouvinte sobre a situação narrada. Essa propriedade interacional da entrevista sociolinguística é a responsável pelo uso intenso de marcas ou traços remetendo aos tópicos de cada ato.

7.3 No interior de cada episódio, quantos traços tópicos são expressões referencialmente plenas e quantos são expressões referencialmente vazias?

Na sequência narrativa 1, não se verificou o predomínio no uso de expressões referenciais plenas (expressões nominais) ou vazias (pronomes)[6]. Nela, 29 traços são expressões plenas e 28 são expressões vazias.

Na sequência narrativa 2, também verificou-se um equilíbrio no uso de expressões referenciais plenas e vazias. Nessa sequência, 17 traços são expressões plenas e 13 são expressões vazias.

Esse resultado contraria uma hipótese sobre a linguagem oral: a de que nessa modalidade usam-se mais expressões vazias do que plenas, tendo em vista a quantidade de conhecimentos compartilhada entre os interlocutores, permitindo ao locutor usar pronomes como traços tópicos por ser o referente facilmente recuperável pelo interlocutor.

Mais uma vez, o uso mais ou menos intenso de expressões referenciais plenas ou vazias tem a ver mais com as condições de produção do texto do que com a modalidade (oral ou escrita) do texto. Como foi dito, na entrevista, as interlocutoras se conhecem pouco, e a locutora conta uma história não vivida pela interlocutora, o que explica esse equilíbrio no uso de expressões referenciais plenas e vazias.

8. Considerações finais

Na etapa final da análise, a combinação dos resultados obtidos nas duas primeiras possibilitou extrair as observações a seguir sobre o processo de construção da cadeia referencial nas sequências narrativas estudadas.

---

[6] A distinção entre expressões plenas e vazias se refere à carga semântica do nome-núcleo dessas expressões. Enquanto nas expressões plenas esse nome apresenta um "conteúdo descritivo denso", nas expressões vazias esse nome apresenta um "conteúdo descritivo fraco" (GROBET, 1996, p. 84).

Sobre as progressões informacionais, os encadeamentos no interior de cada episódio das sequências são bastante locais, ainda quando há encadeamentos à distância. Nesses encadeamentos, os atos não se ancoram em tópicos ativados em atos mais distantes, localizados em outros episódios. A proximidade entre o ato e o tópico explica o predomínio dos encadeamentos com tópico constante nas duas sequências.

No interior de cada episódio das sequências narrativas, verificou-se a marcação intensa dos tópicos, facilitando a reconstrução da cadeia referencial por parte da interlocutora. Além disso, não se verificou um predomínio do uso de expressões referenciais plenas ou vazias.

Esses resultados são importantes, porque relativizam algumas "crenças" acerca da língua oral. Conforme apontado já há alguns anos por Marcuschi (2001), os gêneros de discurso se distribuem ao longo de um contínuo, que leva em conta os graus de formalidade de uso da língua e as condições de produção dos textos e não a modalidade escrita ou falada. Assim, tanto na modalidade oral quanto na escrita, existem gêneros mais ou menos formais, o que se reflete na linguagem empregada.

No gênero entrevista sociolinguística, as propriedades de suas condições de produção (a finalidade do gênero, a esfera acadêmica a que pertence, os papéis sociais dos interlocutores, *etc*) são as responsáveis, em grande medida, pela forma como a produtora da "história da buneca di papelão" constrói a cadeia referencial ao longo dos episódios das sequências narrativas.

9. Referências

Adam, J.M. (1992). *Les textes: types et prototypes*. Paris: Nathan.

Berrendoner, A. (1983). "Connecteurs pragmatiques" et anaphore. In *Cahiers de linguistique française* 5, pp. 215--246.

Chafe, W.L. (1994). *Discourse, Consciousness and Time*. Chicago: Chicago University Press.

Cunha, G.X. (2010). A atuação de sequências do tipo narrativo em um texto jornalístico impresso. *Revista do GEL*, v. 7, pp. 202--219.

Daneš, F. (1974). Functional sentence perspective and the organization of the text. In F. Daneš (Ed.), *Papers on functional sentence perspective*. Praga: Mouton, pp. 106--128.

Filliettaz, L. (1999). *Une approche modulaire de l'hétérogénéité compositionnelle du discours: Le cas des récits oraux*. Cahiers de linguistique française, Genebra, v. 21, pp. 261--327.

Filliettaz, L. (2004). Négociation, textualisation et action: le concept de négociation dans le modèle genevois de l'organisation du discours. In M. Grosjean, L. Mondada (Eds.), *La négociation au travail*. Lyon: Presses universitaires de Lyon, pp. 69--96.

Filliettaz, L., Roullet, E. (2002). The Geneva Model of discourse analysis: an interactionist and modular approach to discourse organization. In *Discourse Studies* 4(3), pp. 369--392.

Grobet, A. (1996). Phénomènes de continuité: anaphoriques et traces de points d'ancrage. In *Cahiers de linguistique française*, Genebra, v. 18, pp. 69--93.

Grobet, A. (2000). L'identification des topiques dans les dialogues. 513f. Thèse de Doctorat en linguistique. Faculté des lettres, Université de Genève, Genève, 2000.

Labov, W. (1972). The transformation of experience in narrative sintax. In W. Labov (Ed.), *Language in the inner city: studies in the black english vernacular*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1997). Some further steps in narrative analysis. In *Journal of narrative and life history*, 7, pp. 395-- 415.

Marcuschi, L.A. (2001). *Da fala para a escrita: atividades de retextualização*. São Paulo: Cortez.

Marinho, J.H.C., Pires, M.S.O. and Villela, A.M.N. (Eds.) (2007). *Análise do discurso: ensaios sobre a complexidade discursiva*. Belo Horizonte: CEFET-MG.

Mondada, L. (1995). La construction discursive des objets de savoir dans lécriture de la science. In *Réseaux*, v. 71.

Ramos, J.M. (2007). *Corpus do Dialeto Mineiro: textos orais; textos escritos dos séculos XVIII, XIX e XX*. Belo Horizonte: UFMG/FAPEMIG/CNPq/Núcleo de Pesquisa em Variação Lingüística.

Roulet, E., Filliettaz, L. and Grobet, A. (2001). *Un modèle et un instrument d'analyse de l'organisation du discours*. Berne: Lang.

Tavares, M.A. (2004). *Então* inferidor como marca de constituição de subjetividade e de instanciação de sentidos na entrevista sociolinguística. In *DELTA*, São Paulo, v. 20, pp. 77--95.

Face and head comments. Taking the floor without words

Francesca D'ERRICO, Laura VINCZE, Isabella POGGI

Department of Education, Roma Tre University

Via Milazzo 11 B - 00185 - Roma - Italy

fderrico@uniroma3.it, laura.vincze@gmail.com, poggi@uniroma3.it

Abstract

The paper aims to define the notion of comment as a communicative act not requested by the previous turn. In our corpus of political debates, comments convey a negative evaluation of the opponent and are generally addressed to the audience. Comments can be conveyed both through speech and body, through body posture, gestures or facial expression (body comments). In the present qualitative study we focused on facial and head comments, our aim being to single out the possible goals of the commenter when expressing negative evaluations of his opponent.

Keywords: comments; metacomments; head and face signals.

1. Introduction

Literature in Conversational Analysis has long studied the rules for turn taking. When people talk to each other, their utterances are felt as filling slots in a ping-pong game where a first throw is followed by another, and the latter responds to the former. Utterances lock with each other in a systematic sequencing, so much so that when an utterance does not fit the sequencing rules we clearly feel it odd, unrequested, out of the stream. Thus, a question is generally followed by an answer, a statement by an acknowledgement, and so on. And while peculiar cases of utterance sequences are allowed by particular roles in conversation, like for example the three turns sequence – question, answer, judgment (in teacher-student interaction), this sounds odd in other types of dialogue or discussion in which the role and status relationship between interactants does not imply one judging the other.

In political debates, debaters in principle are on the same level, and it is up to the Moderator to give them the floor and allow them to express judgments on the other's statements: i.e. to provide a comment concerning another's turn.

Yet, if a debater has something to add concerning another's communication, but s/he is not entitled or explicitly allowed to take the turn, s/he may comment on the present speaker's turn by exploiting another communicative modality: for instance, by making a gesture or a grimace, or by gazing at someone in a particular way. So, a comment that cannot be delivered through linguistic means can be expressed by body signals.

In a previous work we defined the notion of comment and analyzed cases of comments delivered during political debates by verbal or body modalities – gesture, gaze, face and posture. In this paper we focus on comments expressed only by a debater's head and face: head movements, facial expression, eye-gaze. After providing our definition of comment, we analyze a corpus of face and head comments in political debates and propose an account of their communicative and persuasive functions.

2. Comments

2.1 The notion of comment

We define a comment as (Poggi, D'Errico & Vincze, 2012)

a. a communicative act of an informative kind (i.e., an act aimed at providing information), with the information provided generally being

b. aimed at communicating an evaluation or at facilitating an interpretation on the object of previous turns

c. additional with respect to the previous turn,

d. pertinent but not requested by it, and somewhat unexpected.

Let us illustrate this definition.

a. A communicative act of *comment* has the goal of giving information, that is, providing an Addressee with some beliefs *bc* assumed by the Sender concerning some belief *bt* that is the topic of the communicative interaction at hand. The content of the comment – the set of beliefs *bc* provided – may be of two kinds, which make the communicative act be either an "interpretative" or an "evaluative" comment, respectively.

b. In an "interpretative" comment, the beliefs *bc* provided by the Sender are aimed at helping the Addressee to "interpret" belief *bt*: they are useful to understand *bt* better, by connecting it to other beliefs through inferential chains that set explanatory links (of space, time, class-example, cause, goal, condition) with each other. For instance, a literature critic who provides information about the author of a literary work, his biography, and the cultural milieu in which he operates (*bc*), provides an interpretative comment to the literary work, in that he helps the reader framing it within his time and culture.

In an "evaluative" comment, the beliefs *bc* provided by the Sender while dealing with the topic *bt* concern the Sender's opinion about *bt*: his subjective beliefs stemming from his peculiar point of view, that is

determined in its turn by his beliefs, goals and values. A particular type of opinion is an evaluation, that is, a belief concerning how much something may favor or prevent from the attainment of some goal. For example, the literature critic's interpretative comment may become an evaluative comment if he does not only provide factual information about the author's biography, but expresses his own sensations, opinions and judgments about the author's style or content.

c. Just like a communicative act of information becomes an answer if and only if the information it provides fulfils the request for information phrased by a previous question, in the same vein, a communicative act of information becomes a comment due to its relation with the communicative acts preceding it in the same interaction:

c.1. The beliefs provided by a comment are in some way "additional" information in that they are not by default presupposed as necessary in the context at hand: the information it provides is not foreseen nor requested by the typical structure of "adjacency pairs" (Sacks et al, 1974) such as question-answer, offer-acceptance or offer-refusal, greeting-greeting. A comment is a "third turn" unexpectedly added to an adjacency pair that is complete in itself: either it comes after the closure, or it is felt as an intrusion if it comes within the pair. So, we may consider a comment the "third turn" of the teacher in the typical triplet of teacher-pupil interaction (Fele & Paoletti, 2003):

(1) Teacher: When did Napoleon die?
    Pupil: In 1821.
    Teacher: Good!

Yet, we do not consider as a comment the expression of an evaluation that is explicitly requested: for example, an answer to an explicit request to judge some things, events or people, or the utterances that constitute a session of gossip.

d. The information provided by a comment, though not requested by the turn-taking structure, is however pertinent to the topic at hand. In any case, the new belief *bc* connects to belief *bt*, and the topic *bc* is shared by participants in the present communicative interaction: even if participants are not presently talking about it, it must be part of previously shared knowledge, and possibly evoked during the interaction, i.e., recalled in the participants' working memory. For example, see this comment by A:

(2) Two friends agreed to go for a picnic. The morning is sunny and A tells B: "Perfect for a picnic!".

For all these reasons, as we make a comment our interlocutor is highly aware that we are adding an unrequested information, and, depending on our current interaction, s/he may take it as a blatant violation of turn-taking rules, or as a rude intrusion, almost in the same vein as an overlapping or an interruption. Yet, if the comment is performed not by words but "simply" by

bodily signals, it may seem less intrusive and, at least formally, not be taken as an undesired contribution.

A such case is in TV broadcasted debates.

2.2 Body comments

As A has finished talking, B, the former Interlocutor takes the turn, so A now becomes the Interlocutor; if A has something to comment about B's talk, s/he is no more entitled to speak until B leaves the floor. Therefore A may comment not by words, but by body signals, since s/he knows that someone – whether the audience in studio or people viewing TV at home - can see his/her face, hands or body.

In TV mediated multimodal interaction, various participants that can hear and generally see each other are present at the same time: two or more debaters, with or without a moderator, interacting sometimes directly in studio, sometimes only through videorecording or phone call from home; and further, spectators at home and possibly in studio. While people in the studio have a reciprocal full-body acoustic and visual perception, those at home – both debaters and the audience – depend on what is caught by camera or microphone. In such a scenario, participants in a debate may often perform not only verbal but also bodily comments, relying on the fact that their gestures, poses or facial expression may be – and probably are – grabbed by malicious cameramen.

2.3 Meta-comments

That a gesture, grimace, gaze or posture can be a comment, and be acknowledged as such by the present speaker, moderator, or the audience, is witnessed by cases in which the present speaker, while seeing a body comment, in his turn meta-comments on it. Like in the following discussion between Francesco Boccia, a politician from the Democratic Party, and Marco Travaglio, a journalist of the newspaper "Il fatto quotidiano".

(3) Concerning the wiretapping of the Italian President during an important investigation in Palermo, the President has reminded that according to the Italian law, the contents of his phone calls must not be published, while the newspaper "Il fatto quotidiano" and its journalist Marco Travaglio have argued for a complete transparency. Francesco Boccia, whose party defends the President's position, is now provoking Travaglio, arguing that the investigation conducted by the judges from Palermo, Ingroia and De Matteo, is very similar to a further investigation over two other investigations for mafia: thus he is implying that Ingroia and De Matteo are superimposing themselves to other judges and other trials. Travaglio engages in a detailed answer, accompanied by iconic gestures, listing the misdeeds dealt with by the two trials of Caltanissetta and Firenze, and finally distinguishing the trial of Palermo from them:
    Travaglio: *L'inchiesta di Palermo si occupa di un àltro fatto, o meglio di una serie di fatti*

*(*Palermo investigation concerns another event, or better, a series of events)
Boccia: *di tutt'e due (* it concernes both events)
Travaglio: accaduti intorno (occurred around)
Boccia *makes a dental click ending with a small laughter, then he looks at the camera with an amused smile*
Travaglio: *Non s'indaga per strage. S'indaga su…* (They do not investigate about the massacre. They investigate on)….
(Then, seeing Boccia's smile): *Beh vedo che, vedo che… la mette di buon umore questo argomento. Complimenti, come se le avessi raccontato una barzelletta.* (Well, I see, I see that… this topic puts you in a good mood about. Congratulations, it's as if I had told you a joke)

In this fragment, Boccia smiles to display his skepticism about Travaglio's answer. And Travaglio is so aware of Boccia's smile, and of its being a comment on what he is saying, that he sarcastically congratulates Boccia for smiling. Thus he performs a "meta-comment", that is, a comment over Boccia's facial comment: an insinuation about Boccia being very cynical, given that he smiles about mafia massacres.

3.   Face and head comments. An observational study

We present an observational study on comments in political debates, focusing on a qualitative analysis of comments performed by head and face.

3.1   Method

In a corpus of 16 videorecordings from Italian political talk shows (interviews and political debates), we selected 46 fragments for a total of 150 minutes. 46 visible behaviours (face, hands and body movements or poses) were analyzed in the coding scheme of Table 1. In Table 1 (in Appendix), column 1 contains the time in the video and the name of the present Speaker, col. 2 contains the verbal message, col. 3 the sender of the non verbal comment and its addressee (interlocutor, audience, moderator), col. 4 the description of the commenter's body signals. In column 5 we specify the body modality used to convey the comment. In col. 6 we focus on the meaning of the comment; col. 7 illustrates whether the evaluation conveyed by the commenter concerns the person (the Speaker himself) or the content of the Speaker's turn; col. 8 contains the emotion possibly conveyed by the commenter through his body movement, col. 9 specifies the commenter's goal: discrediting, ridiculizing the opponent, or showing his own dominance.

A such case is in TV broadcasted debates.

3.2   Results

Among all the fragments analyzed, we collected 45 comments performed by face, head, or both head and face. .As to the object of the comments, we found out they are almost equally distributed between those concerning the person and those concerning the content of the other's

turn.. As to the communication modality,, the comments conveyed through both head and face are the most frequent (59% = n.27), followed by only face 28% (13) and 13% (6) only head. As to the comment goals, we found that 54% are clearly oriented to discredit the opponent, and within this total amount, 26% are performed by ridiculing the other (Poggi et al., 2012) . About 24% of analyzed facial and head comments are done in order to show that the sender has more power compared to the opponent (*dominance comments*).

Of course, there is quite of a subtle difference between a comment of discredit and one of dominance, in that discrediting the other is a way to lower his power and hence, indirectly, to enhance one's own. It is necessary to specify that in our work we distinguished the comments of dominance from the ones on discredit considering a first level of meaning in the face and head signals; of course in the second level of analysis the goal of discrediting someone is close to the expression of power but we can consider this level not necessarily useful at this stage of analysis. We also found out two more goals in our comments during political debates: *disagreement*, i.e., the expression of a negative evaluation of the other's opinion, and the simple *disconfirmation* of a fact stated by the other; 22% in total. (For a definition of opinion and fact see a recent work on agreement, Poggi et al., 2011).

As we can see from the example below, in these last two cases the most frequent head signal is the shake; but while in the case of a disconfirmation, it is short and performed in a simple horizontal direction, in the case of disagreement it is more emphasized.

The most common facial and head comments are those oriented to discredit someone or some content expressed during the debate. Discredit can be defined (D'Errico and Poggi, 2012) as the spoiling of the image of a person (B) in the eyes of other people (C), caused by a person (A) performing communicative acts that mention or point at actions or qualities of B that are considered negative by the third party C. . While in principle discredit may be cast either deliberately or not (A may mention some feature of B without knowing it is negative for C), in the comments we analyzed, those discrediting the other are presumably all deliberately aimed at doing so.

From this point of view we start to differentiate the spoiling of the opponent's image by means of direct attack to a person or to the debate's object.

An example of the first type is the taken from "Ballarò", an Italian political debate broadcasted in 2005 when the Right-Center party was in power but going through a critical moment. In fact, in the selected video, Berlusconi has just lost the regional elections and instead of explaining the reasons of this failure, is trying to defend himself by blaming the Left party, represented in studio by its leader D'Alema.

(4) Berlusconi says: *"La disoccupazione che abbiamo ereditato da voi che era al 21% oggi è al 16%".* (The unemployment we inherited from you, that was 21%, today is 16%) .
In correspondence to the sentence "we inherited

from you", D'Alema, recognizing this as a strategy for making the other guilty, typical of Berlusconi, on the verbal side says, with a *very low voice intensity*: *"Non ce la fa proprio"* (he really can't resist it). At the same time he makes a facial and head comment of discredit: he performs *very small head shakes* while *raising eyebrows up*, thus expressing his *disbelief* and *surprise* of how irresistible is for Berlusconi to refrain from accusing the others by making them guilty. Then he *lowers his head*, and makes a particular kind of *smile*: the "miserable smile" (Ekman, 1982) that conveys bitterness and resignation as it hides the sender's sense of impotence. All of these body actions – in accordance with D'Alema's typical body communication style – convey an ironic attitude. Both the surprise expressed by eyebrows up, and the sense of impotence and resignation conveyed by headshakes, head lowering and the miserable smile, are displayed ironically, thus communicating: "Oh poor thing, he really cannot refrain from doing so", and hence implying (antiphrastically) that Berlusconi is incorrect in accusing others.

Another way to discredit a person is by showing one's own impatience when s/he is speaking.

(5) During   "Tetris", a political talk show broadcasted by the Italian TV La7), while talking of the attitude of Muslims towards their women,   the   Right   politician Daniela Santanchè is praising her own feminist actions:*"Io che mi sono battuta per leggi di libertà, per liberare le donne mussulmane"* (I, who had fought for laws of freedom, to release Muslim women from repression).
The Leftist minister Fabio Mussi, while hearing such self-praise, suddenly *turns his head away* from Santanchè, *raises his eyebrows* and *shuts his eyelids*, while *raising lip corners with closed mouth*: he thus conveys a sense of smugness, (i.e., it is not important what you have done). ; afterwards he *liks his lips* and *nods faster* expressing his *impatience* to intervene and reply.

A very efficient way to cast discredit on the opponent's discourse and implicitly on the opponent himself is by communicating to the audience its dullness and incapacity to attract the listener's attention. In fact, one of the most disqualifying and discrediting emotions expressed by the interlocutor during the opponent's speech is boredom. Boredom emerges when the speaker provides information that are well-known to the listener. Moreover, if the listener is in overt disagreement with the Speaker's thesis, the fact of being obliged to hear it over and over again, provokes in the listener, altogether with impatience, inability to bear it. This is the case of the Italian   philosopher   Mario   Cacciari   (left-wing), interviewed from home, who has to keep silent and can't interfere in his interlocutor's turn, Roberto Cota, a member of the Lega Nord Party (right-wing). Nonetheless,

the audience can easily infer from Cacciari's non verbal behaviour, his emotions and states of mind.

(6) While looking at the Speaker, Cacciari is *leaning on the back of the chair*, hence communicating relaxation, his *eyelids are half-closed*, almost as if sleeping. The fact that he can't keep his eyes open communicates a total lack of interest in the speaker's saying. At the same time though, his *head and chin are high*, denoting superiority. Cacciari *snorts* loudly several times during Cota's turn, and *his head comes forward while snorting*, emphasizing his annoyance towards what he hears. While listening, he *shakes head* with *closed mouth* and *horizontally stretched lips with slightly pouched corners*, a facial behaviour which typically communicates "No way …".

Another example of communicating boredom in front of the speaker's sayings, though less emphasized, is the one of Brambilla, that we analyzed in the annotation scheme.
So far we have seen two cases of discrediting the *content* of the opponent's discourse by means of communicating   boredom.   But   through   the communication of boredom the listener can imply that he is bored by the *speaker himself*, besides by his sayings.
A very common way to discredit the opponent in political debates, also during a non requested turn is ridiculization.
Ridiculizing someone is in general performed by deliberately singling out a feature or an act of another person and pointing at it in front of other people as worth being laughed at (Poggi *et al.*, 2012).
In different cases a ridiculizing comment mostly done by the face is displaying surprise in a clear and marked way while listening to the opponent's words.

(7)   Matteo Renzi, the mayor of Florence, now candidate to primary elections in the Democratic Party, is talking of his electoral program, but he does so in such a complicated way that he is making himself incomprehensible and hence ridicule. Marco Travaglio, a journalist debating with him, promptly takes advantage of this and underlines Renzi's incomprehensible sentences: he makes a *large smile* and *opens his eyes wide* displaying surprise, then he *frowns* and *lowers his lip corners* while looking at the other participants present in the studio, as if stating: "Did you understand anything? I didn't!"

Travaglio in another debate makes grimaces to ridicule the opponent Daniele Capezzone.

(8) Capezzone, a former deputy of a Left-wing party who moved to a Right-wing party and became the spokesman of Berlusconi, is talking of this change proudly, saying it was in a sense a political suicide. Travaglio, to argue that this move was not at all against Capezzone's interest, as he tries to let the audience infer, but on the

contrary, it was a convenient opportunistic change, ridicules him through ironic grimaces: he suddenly *raises eyebrows and lip corners*, thus showing surprise and amusement, but then by *eyes gazing downward* he seems to imply "you cannot dupe me". The global meaning of these signals might be "For God's sake, don't exaggerate"; and *lack of eye contact* plus a sudden *restrained smile* indicate amusement but also a negative evaluation, that diminishes the nobleness of Capezzone's "sacrifice".

The goal of ridiculing is witnessed by more or less explicit laughter along the whole debate.

(9) Travaglio refers to Capezzone and Berlusconi as *"Tu e il tuo padrone"* (you and your master). Capezzone replies to the offence by threatening: *"Sei cascato male stasera, io non mi faccio insultare da questo signore"* (You are in the wrong place tonight, I will not stay here to be offended by this gentleman), and Travaglio, with *raised eyebrows*, makes a *loud laughter*. Laughing after a threat is a typical signal of ridiculization that conveys "I am not afraid of you", hence "I am stronger than you are". Further Travaglio *opens his mouth wide* as in surprise, thus making the serious thing the other is saying ridicule; finally his *tongue in cheek* conveys allusion to apparently serious but in fact comic.

Another possible goal of the commenter is to communicate his dominance over the speaker.

(10) Renata Polverini, a Right Party politician and current president of Lazio region addresses a direct reproach to Massimo D'Alema, the former national secretary of the Democratic Party of the Left. First D'Alema *interrupts eye contact* with the speaker and *looks down*, as if wanting to collect his thoughts before answering, then, still with lowered gaze, performs a *light smile* and the non verbal vocalization *"hm"* with *raising intonation* while simultaneously *shaking head*. The fact that he still does not stare at Polverini while going *"hm"* might be meant to signal that he is talking to oneself, expressing his irritation to himself and not trying to communicate it to the audience. Nonetheless, we know that signals of anger and irritation caused by the interlocutor's deeds or sayings are hardly ever meant to be kept secret. We may therefore interpret D'Alema's reaction as communicative and not a simple expression of his inner states. D'Alema is now ready to provide his answer in the form of a false act of praise: *"Tu sei straordinaria perché tu sei sempre all'opposizione anche quando stai al Governo"* (You are extraordinary because you are always on the opposition side even now when you make part of the Government). He then displays a false *laughter*, much *louder and more pronounced* than a normal sincere one, reminding us of the laughter purposely

introduced in the comedy sitcoms to signal that now it's time for the audience to laugh. In fact the audience does laugh and Polverini laughs as well, wanting to prove that she does not take it personally. D'Alema goes on: *"Quindi ti sei ritagliata un ruolo spettacoloare, vieni qui e fai l'opposizione pur essendo al Governo con i voti di Berlusconi. Ora non vorrei dire che è troppo comodo, diciamolo, ecco, troppo comodo"*. (You adopted a spectacular role, you come here and play the part of the opposition, although now you are a member of the Government with Berlusconi's votes on your side. I wouldn't like to say that it [your behavior] is too convenient, too convenient". While stating *troppo comodo* (too convenient), D'Alema performs a *headbutt* towards Polverini, a non verbal signal metaphorically disqualifying his opponent for not playing according to the rules of the game.

4. Conclusion

When people argue in debates, they sometimes do not give up opposing or maintaining the opposition even when they are not entitled to take the turn. They still do so by a particular kind of turn: comment, which allows them to be in a sense over and above the competition and to win over their opponent. Such a detour from the rules of conversation is more subtle but, if possible, even more effective when comments are conveyed by body signals like head and facial actions.

Although comments, both facial or verbal, may convey either positive or negative evaluations about people, opinions, behaviors or about the state of the world, in this particular context of political debates, comments generally convey negative evaluations of the opponent. In this paper we focused on the commenter's goals when communicating negative evaluations of the opponent and from a total of 46 face and head comments in our corpus, we observed that, excluding a few cases of (not requested) disconfirmation and disagreement, most of the body comments are made to discredit, ridicule and to express dominance. We noticed a certain tendency in our corpus to combine the communication of certain goals by displaying certain emotions, particularly: when discrediting, the commenter often shows boredom and impatience, when ridiculing: enjoyment and surprise, while when showing dominance: irritation and ironic enjoyment.

From our qualitative analysis of face and head comments it comes out that they are more frequently used to discredit the present speaker, also by making fun of him/her, and to display one's dominance also regarding to what s/he is saying. As Aristotle, Schopenhauer and more recently van Eemeren (2010) pointed out, in political debates there is a continuous tension between the dialectic and the rhetoric goal: one aims at finding the truth and the other at winning the contest, by showing that a participant is more intelligent, more competent than the other. From our work, the use of head and face comments seems more oriented to the latter than to the former.

Further analysis, conducted on a larger corpus, will

better explain the co-occurency of commenters' goals, associated emotions and power relations perceived between sender and addressee.

5.   Acknowledgements

6.   References

Aristotle (1961). *Retorica*. Bari, Laterza. Traduzione di Armando Plebe

D'Errico, F., Poggi, I. and Vincze, L., (2012). Discrediting signals. A model of social evaluation to study discrediting moves in political debates. Special issue in "Social signal processing". *Journal on Multimodal User Interfaces*. DOI: 10.1007/s12193-012-0098-4

Eemeren, F. H. van. (2010). *Strategic maneuvering in argumentative discourse. Extending the pragma-dialectical theory of argumentation.* Amsterdam/Philadelphia: John Benjamins Publishing Company

Ekman, P., Friesen, W.V. (1982). Felt, false and miserable smiles. *Journal of Nonverbal Behavior* 6, pp. 238--252.

Fele, G, Paoletti, I (2003). L'interazione in classe. Il Mulino, Bologna.

Poggi, I, D'Errico, F., Vincze, L. (2012) Comments in words, face and body. In *Journal of multimodal users interface.* DOI: 10.1007/s12193-012-0102-z

Poggi, I., D'Errico, F. and Vincze L. (2012). Ridiculization in public debates: making fun of the other as a discrediting move. In N. Calzolari et al. (Eds). *Proceedings of the 8th conference on International Language Resources and Evaluation* (LREC'12). Istanbul 21-27 May. European Language Resources Association (ELRA), pp. 44-50.

Sacks, H., Schegloff, E.A., Jefferson, G. (1974). A simplest semantics for the organization of turn-taking in conversation. In *Language*, 50 (4), pp. 696--735.

Schopenhauer, A. (1991). *L'Arte di ottenere ragione*. Adelphi Edizioni, Milano.

7.   Appendix

1. Speaker Timing	2. Verbal Message	3. Sender → Addressee (interlocutor, audience, moderator)	4. Body behaviour	5. Comm. channel	6. Meaning	7. Negative evaluation of the Object/ Person	8. Conveyed Emotion	9. Goal: (discredit, ridiculization, dominance)
S: Brambilla	*beh, non è un reato andare a una festa,* (well, it's not a misdeed to go to a party)	Brambilla → Audience	Rubs her forehead looking at S. obliquely with half-closed eyelids	Head, face	It is not a misdeed to go to a party I'm bored, what my interlocutor Seracchiani is saying is neither new nor interesting to me	Negative evaluation of Interlocutor's utterance (Content)	Boredom	Discredit

Table 1: Corpus coding scheme

SPEECH AND LINGUISTIC ANALYSIS

Properties of fronted direct object in Italian

Sandra AUGENDRE

UMR 5263 CLLE-ERSS(ERSSàB), Département de Sciences du Langage,
Université Michel de Montaigne Bordeaux 3, 33607 Pessac Cedex
augendre.sandra@wanadoo.fr

Abstract

The work presented in this paper focuses on a comparison of various occurrences of the same syntactic sequence in Italian: Object-Verb (OV). In this kind of utterance, the object occupies a "non canonical" position (preverbal position) and assumes the syntactic function of an object (no clitic is present). Classified among the so called "marked" (non canonical) structures in Italian grammars (cf. *Grande Grammatica Italiana di Consultatione*, 1988), OV order receives various names and descriptions from linguists. Based on a corpus of spontaneous productions, my study aims at reevaluating the properties attributed to OV order in Italian, for instance, the equivalence established between OV order, cleft sentence and narrow focus, the range of context possibilities for this structure or its pragmatic and prosodic characteristics.

Keywords: Italian; fronted direct object; syntax; pragmatics; prosody.

1.   Introduction

The work presented in this article is based on a corpus constituted to study different object constructions in Italian, and focuses on a comparison of various occurrences of the same syntactic sequence in Italian: Object-Verb (OV). In this kind of utterance, the object occupies a "non canonical" position (preverbal position) and fully assumes the syntactic function of an object (no object clitic is present):

Example 1:

IL DOLCE   ha mangiato.
THE CAKE   he ate
'(It is) THE CAKE (that) he ate / He ate THE CAKE.'

Unlike a dislocated object, here the preverbal SN is strongly connected to predication: it assumes the function of an object and there is no coreferent expression in the utterance.

In this paper, we first will give an overview of most of the previous studies that have been carried on OV order in Italian. Then, we will describe the data we have worked on and our methodology. Finally, we will present our analysis and results.

2.   OV order's description

Classified among the so called "marked" (non canonical) structures in Italian grammars (cf. *Grande Grammatica Italiana di Consultatione*, 1988), OV order has not attracted much attention (cf. Berretta, 1998 and Brunetti, 2009 for two works based on corpora) and receives various names and descriptions from linguists.

In relation to the object initial position and the comunicative status of the argument, the structure is often called *rhematic* (Stammerjohann, 1986) or *contrastive* (*GGIC*, 1988; Graffi, 1994; Ferrari, 2003) *topicalization*, *left rhematisation* (Berreta, 1998), *focus-background structure* (Brunetti, 2009), or more simply *NP preposing* (Abeillé, Godard & Sabio, 2008).

Retained as relatively infrequent in Italian by these linguists, OV order is described as limited to spoken dimension (Berretta, 1998; Brunetti, 2009), associated with a specific prosodic structure (peak of intensity on the object and fall of F0 after this argument, cf. Tamburini, 1998) and at a comunicative level, the object is described as assuming a contrastive focus function (Sornicola, 1981).

This work aims at evaluating these correlations on occurrences present in spontaneous data.

3.   Data and methodology

3.1   Corpora

Our corpora has been constituted in Sardinia and initially aimed at describing subject and direct object constituents in Italian utterances.

It is composed of spoken and written productions and divided in four parts: chat, e-mail, informal speech (spontaneous conversations) and formal speech (university lessons).

The entire corpora gathers 3000 utterances that contain a subject (realised by an independent element or in verb's ending), eventually associated with a direct object (640 cases).

3.2   Data collected

In our corpus, we listed only 11 cases of fronted direct object, the result that confirms the very low rate of use previously attributed by the linguists to OV utterances in Italian. The general properties of our OV occurrences are the following:

-   Only 3 of the 11 OV utterances come from the written corpus and 8 appear in spoken dimension. This repartition shows that this order is particulary related to prosody, that facilitates OV utterances' interpretation even if it is also available in writting.
-   All written OV utterances appear in chat, not in e-mails and all spoken OV utterances (except

one) appear in informal speech: these data indicates a close link between OV order, conversation and informality.

- Concerning the type of OV utterances, we have two (oral) interrogative structures and then exclamative ones.
- In 10 of the 11 OV utterances, the subject is not realised (utterance limited to O+V) and for the remaining case, the subject is postverbal (O+V+S).
- Finally, concerning the fronted objects, they are all directly followed by the verb (or are separated from it by clitics) and are short phrases (two words or less, except one case). Types of objects, divided in two classes, are the following:

  A. NP (6 cases): *l'ora*/the hour, *la finalità di parole*/the finality of words, *una torre*/a tower, *alcune parole*/some words, *un po'*/a little and a proper name.
  B. Proforms (5 cases): *qualcosina*/a little something, *questo*/this (three cases), *qualcosina*/something.

3.3 Strucrure of the analysis

Our analysis of OV utterances relies on three dimensions: syntactic (one specific syntactic structure: O+V), pragmatic (relation between OV and information structure) and prosodic (properties of OV utterances).

The analysis of OV utterances present in our corpus aims at showing if OV order in Italian has a specific domain of use or a given pragmatic value, more precisely, in which dimension(s) (spoken/written Italian) OV is reprensented, which communicative need(s) this structure responds to and which kind of prosodic structure it is associated with.

4. Analysis of OV utterances

The number of OV utterances available in the corpus confirms the weak degree of use of this order and the distribution of the occurrences proves that there is a close link between OV, conversayion (8/11 occurrences appear in spoken dimension) and informality (10/11 occurrences are present in spontaneous data).

By analyzing OV utterances, we aim at defining the domain of use of this structure, its information structure (focus domain, type of focus...) and also at distinguishing different prosodic structures according to each OV utterance properties (object's part of speech, type of referent, information structure, contextual data...).

4.1 Anaphoric vs non anaphoric fronted objects

Our analysis began with the classification of OV utterances according to the status of fronted objects' referents, in order to verify the distinction established by Benincà (1988) and resumed by Berretta (1998) between *left rhematisation* and *anaphoric anteposition*. The fronted object can be anaphoric or not:

- In the first case, it is a coreferent expression related to an element present in linguistic or extralinguistic contexts (simple anaphora) or a global resolution of a part of the previous speech (recapitulative anaphora). This type of OV utterance is analyzed by both linguists as cases of *anaphoric anteposition* because object's referent is contextually given and because OV order is here motivated by the will to leave postverbal/focal position available for another element, which is often the subject. Among the 11 OV utterances present in the corpora, 5 objects are anaphoric expressions, like in the following example:

Example (2):

A: *C'è anche <u>questo</u> che non ho capito*
There is also <u>this</u> that I don't understand
B: **Questo** *non hai capito ?*
You don't understant **this** [**this** (acc.) you don't understand] ?

- In the second case, the object is the element marked as the most prominent of an all focus utterance (emphasized object) or the element that constitutes the informative contribution of the utterance, that can be contrastive (contrastive focalisation) or not (completive focalisation). In this category, we find 6 of our 11 fronted objects, like in the following example, that represents a case of emphasized object in an all focus utterance:

Example (3):

*Hanno fatto anche il lavoro di trascrizione // naturalmente non su tutto perché // **un po'** facevano anche in classe // guidati dagli insegnanti*
They also did the transcription work // naturally not on all because // they did **a little** in class [**a little** (acc.) they did in class] // helped by the teachers

4.2 Substitution test by a cleft or by a presentational sentence

For all OV utterances, we also put in relation object referent status and information structure of the utterance. We thus tried to replace OV sequences by a cleft sentence (è X che / it is X that/who) and by a presentational sentence (c'è X che / there is X that/who), in order to verify the presupposed status (substitution by a cleft sentence acceptable) or non presupposed status (substitution by a presentational sentence acceptable) of the object and of what follows it in the utterance.

The results of this test are presented in the tables below.

Anaphoric objects	Cleft/Presentational Test
Questo non hai capito This you don't understand	// cleft sentence
Questo non riesco a capire This I don't manage to understand	// presentational sentence
L'ora non so The time I don't know	// presentational sentence
Questo vorrebbe dire This maybe it should mean	// cleft sentence
Qualcosa mi ricordo Something I remember	// presentational sentence

Table 1: Anaphoric fronted objects' substitution test

Non anaphoric objects	Cleft/Presentational Test
Qualcosa evito di chiedere Something I avoid asking for	// presentational sentence
Alcune parole non riusciva a leggere Some words she did not manage to read	// presentational sentence
Un po' facevano in classe A little they did in class	// presentational sentence
Una torre avevo fatto io A tower I had made	// cleft sentence
La finalità di parole vorrà dire The finality of words it should mean	// cleft sentence
Usandra mi hai detto Usandra you told me	// cleft sentence

Table 2: Non anaphoric fronted objects' substitution test

The substitution test allows us to show, on one hand, that contextual level and utterance level are relatively independent, and on the other hand, that the equivalence often established between OV order and the cleft sentence is only relative:

- Among anaphoric and non anaphoric objects, half (respectively 3 on 5 and 3 on 6) corresponds to a presentational sentence (wide focus) and half (respectively 2 on 5 and 3 on 6) to a cleft sentence (narrow focus). It is thus not possible to establish a clear relation between the status of fronted objects' referents to one of the two types of focalisation (wide and narrow).
- Among the 11 OV utterances of the corpus, more than half (6 cases) are equivalent to a presentational sentence (the subordinate clause is not presupposed) and only 5 to a cleft sentence (the subordinate clause is presupposed), data that reveals that in OV utterances, what follows the object is not inevitably presupposed, but

especially that this configuration (fronted object narrow focus) is even less frequent than the other one (wide focus).

4.3 More detailed analysis

After the presentation of all properties of our OV utterances, we will now concentrate on four representative examples and their analysis: a non focus anaphoric object (4), a fronted object in an all focus sentence (5), a fronted object focus (6) and a contrastive fronted object (7).

4.3.1. Anaphoric fronted object (5 cases)

In this first configuration, the object's referent is introduced in the linguistic or extralinguistic context and is then refferred to by a proform in preverbal position.

Example (4):

A: C'è anche <u>questo</u> che non ho capito
There is also <u>this</u> that I don't understand
B: **Questo** non hai capito ?
This you don't understand
'You don't understand **this**?'
(Is it) [**This** (acc.) (that) you don't understand]

In the example above, B's utterance is the identical repetition of what A says (questo + negation + capire / this + negation + to understand) but as a question. The informative content of OV utterance does not come from the elements' newness but only from the modality of the utterance (request of confirmation).
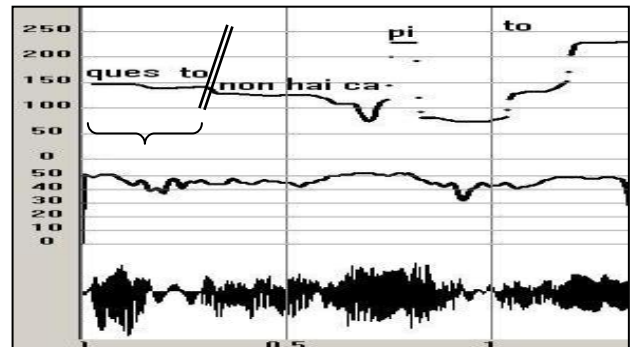
Figure 1: prosodic structure of the utterance "questo non hai capito?"

In Figure 1, we can observe that no considerable prominence is attributed to the preverbal proform (147 Hz, 51 dB and a duration of 267 ms for QUES(to)) and only the past participle, situated at the end of the question, is realised as prominent here (229 Hz and 52 dB on (ca)PI(to)).

4.3.2. All focus OV utterances (3 cases)

In this configuration, the fronted object is contextually new and represents the anchorpoint of a completely informative utterance.

Example (5):

A: Ho fatto qualcosa ?
  'Do I help in something?'
B: Sì grazie
  'Yes thanks'
C: **Alcune parole** non riusciva a leggere
  'She didn't succeed to read **some words**'
  (There are) [**some words** (acc.) (that) she didn't
  manage to read]

OV utterance aims here at closing a conversation by calling back the event which caused it: B and C asked A to read a document and C resumes in conclusion the cause of this recourse (they needed A because B did not manage to read some words).

If the utterance informs that B did not manage to read some words, it presents the object (alcune parole) as a major indication, thanks to the initial position of the object and to F0's fall between it and its right context. In fact, at prosodic level, the preverbal SN is marked as the utterance most prominent element, unlike what we observed previously for anaphoric objects.
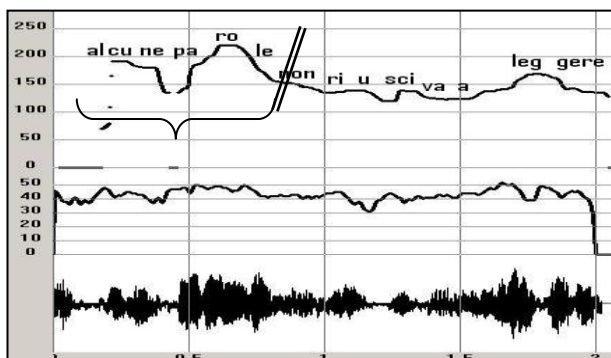
Figure 2: prosodic structure of the utterance
"alcune parole non riusciva a leggere"

In terms of F0, the curve's highest points correspond to the tonics of the adjective *alcune*/some and of the noun *parole*/words (192 Hz on (al)CU(ne) and 220 Hz on (pa)RO(le)). Furthermore, the melodic curve falls considerably from the tonic of the object phrase's noun (from 220 Hz on (pa)RO(le) to 151 Hz on non). At intensity level, we also observe a fall from the noun: we have three peaks on the three syllables of the noun (50 dB, 49 dB and 50 dB) and then lower values until verb's tonic (52 dB on LEG(gere)).

4.3.3.   Non constrastive fronted object (2 cases)
In the third configuration, the object constitutes the informational and prominent part of the utterance without being implicated in a paradigmatic opposition, whereas its right context is totally secondary at communicative level.

Example (6):

A: Ma è la "f" che non capisco.
  'But it is the "f" that I don't understand.'
B: **La finalità di parole** magari vorrà dire.
  'Maybe it should mean **the purpose of words**'
  (it is) [**the purpose of words** (acc.) (that) maybe
  it should means]

With respect to the linguistic context, the fronted object (la finalità di parole / the finality of words) is the informative contribution of the utterance (its focus), status confirmed by the possible substitution of this OV utterance by a cleft sentence (cf. 4.2).
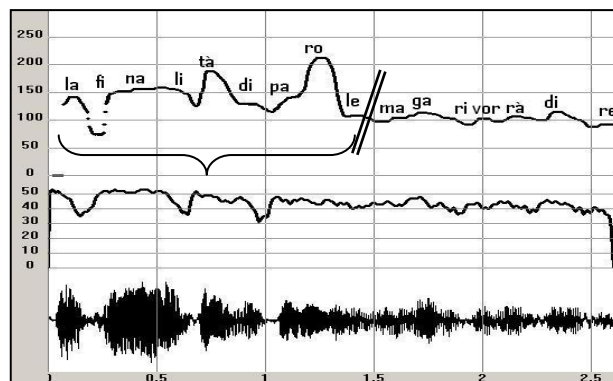
Figure 3: prosodic structure of the utterance
"la finalità di parole magari vorrà dire"

At prosodic level, we can note that the object is more prominent than its right linguistic context, whether at F0 level (that falls after the object), at intensity level (values superior to 50 Db on finaliTÀ) or at duration level (tonics of both preverbal nouns, finaliTÀ and paROle, occupy more space than the other syllables of the utterance).

4.3.4.   Contrastive fronted object (1 case, written)
In the last configuration, object's referent is introduced as both utterance's informational contribution and as a paradigm member. This case (fronted object narrow focus introduced in opposition to one or more other referents) corresponds to the one globally presented as prototypic by the linguists (cf. part 2). However, among our 11 OV utterances, only one of them is contrastive.

Example (7):

A: L'albero con la carta igienica, eri tu?
  'The tree with the toilet paper, it was you ?'
B: Albero??? Di carta igienica?????
  'Tree??? Of toilet paper?????'
B: No UNA TORRE avevo fatto io.
  'No it is a tower that I had made'
  (it is) [A TOWER (acc.) (that) I had made]

In this last example, the contrastive value of the fronted object is undeniable: to describe the same object,

A introduced the notion of *tree* and B replaced it by the concept of *tower*, kind of contrast called *replacing focus* by Dik (1997: 331-332): A says that B built a tree (assertion of *to make a tree (B)*) and B rejects part of A assertion by replacing object's referent by another one (negation of *to make a tree (B)* and assertion of *to make a tower (B)*). In this unique OV utterance, the only referent both contextually new and informative is the fronted object, as the fact that A built something is already presupposed in the previous discourse. What follows the object is presupposed and the utterance is equivalent to a cleft sentence (no, è una torre che... / no, it is a tower that…).

Finally, besides a focalisation of the fronted object, the utterance also contains a postverbal pronoun (OVSpr), whose presence is pragmatically motivated: the pronoun is not realized as an informational contribution but strengthens the contrastive value of the utterance by creating a second paradigm (io / I vs. someone else), connected to the first one (albero / tree vs. torre / tower), but that remains implicit. The effect obtained with the realization of the pronoun in final position is similar to the the one proposed by Blasco-Dulbecco (1995: 59) for the sequences *moi je* in French: " the tonic pronoun [...] seems to aim essentially the naming of an element distinguished among the others of its sort; as if it expressed a kind of contrast or of instigation. This is the case not only for the dislocation before the verb [...] but also for the dislocation after the verb ". Indeed, in our example, the subject is introduced as a contrastive topic as its presence can be interpreted in the following way: *to build a tower (me) involves to build a tree (not me / someone else)*.

5. Results and conclusions

To conclude, we will first sum up the properties of our corpus OV utterances and then the results of their analysis at pragmatic and prosodic levels.

Concerning the number and the distribution of OV utterances, our data confirms the weak degree of use of OV order (11 cases in the corpus) and the close link between OV order, conversation and informality. Indeed, the available occurrences are mostly present in speech dimension (2/3), rather conversational and informal.

Our fronted objects have the following formal properties: in terms of part of speech, we have 5 NP and 6 proforms and in terms of length, 10 of our fronted objects are short phrases ($\leq$ 2 words).

In terms of information, we distinguished first two types of object's referents: the anaphoric ones (5 cases) and the non anaphoric ones (6 cases). Among anaphoric fronted objects (a NP and 5 proforms), we isolated those that resume partially the previous speech and have only a single referent. Among non anaphoric fronted objects, we distinguished those present in all focus utterances (3 cases) and those that constitute the utterance informational contribution (3 cases).

Then, we tried to verify the link often established between OV order and focus-background information structure by using two substitution tests (OV / cleft sentence and OV / presentational sentence). These tests revealed that independently of the status of object's referent in the discourse (activated or not), the preverbal object of most of OV utterances does not constitute alone the utterance assertion (substitution by a cleft sentence impossible), in other words what follows the object does not tend to be presupposed.

Furthermore, only one of our fronted objects is clearly a contrastive focus, data that shows that OV order is neither limited to a narrow contrastive focalisation.

To conclude, OV order does not seem to be reserved to narrow focalisation at all (5 cases on 11) nor to contrastive focalisation (1 case on 11), and is more often connected to the will to mark the argument as the most prominent of a wider informational contribution (6 cases on 11).

Finally, at prosodic level, we first saw, with the three OV utterances present in written productions, that OV order, even if mostly used in spoken productions, does not inevitably need the prosodic marks to be interpreted.

In terms of realisation, we observed no net break between fronted objects and their right context but distinguished different prosodic structures according to OV utterance properties: object's part of speech and referential autonomy (proforms are less prominent than NPs), referent's type (anaphoric referents are perfectly integrated to the predication and are prosodically less prominent than non anaphoric ones), information structure (objects narrow focus are more prominent than objects that are part of a bigger focus unit)... At least, we have a small decline of F0 curve after the object and at most we have a net break between the object (focus) and its right linguistic context (background information). Fronted object's prominence is quite particularly marked at prosodic level when the object is the utterance focus: in these cases, prosodic structure clearly distinguishes the focus from the background, as all prominence marks are attributed to the first part of the utterance while the second part is pronounced as a sequence neither prominent nor informational (less audible, flat F0 curve and low values at F0, intensity and duration levels).

To conclude, our study allowed us to confirm the weak degree of productivity of OV order, but also to widen the use of OV order to written dimension or to observe some regularities concerning fronted objects' formal properties (part of speech, length...). At pragmatic level, our data and its analysis led us to reconsider the equivalence established between OV order, cleft sentence and narrow focus, which is only relative according to our data and at the same time, to widen the range of contextual possibilities for the structure by distinguishing different information and prosodic structures that can be associated to OV order in Italian.

6. References

Abeillé, A., Godard, D. and Sabio, F. (2008). Two types of NP Preposing in French. In S. Müller (Ed.), *Proceedings of the 15th HPSG Conference*, Stanford: CSLI on-line Publications, pp. 306--324. Available at:

<http://csli-publications.stanford.edu>.

Berretta, M. (1998). Valori pragmatici diversi dell'ordine OV (OVS/OSV) nell'italiano contemporaneo. In G. Ruffino (Ed.), *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza. Morfologia e sintassi delle lingue romanze* (vol.2). Tübingen: Niemeyer, pp. 81--90.

Brunetti, L. (2009). Discourse Functions of Fronted Foci in Italian and Spanish. In A. Dufter & D. Jacob (Eds.), *Focus and Background in Romance Languages*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 43--81.

Dik, S.C. (1997). *The theory of functional grammar. The structure of the clause*. Berlin: Mouton de Gruyter.

Ferrari, A. (2003). *Le ragioni del testo. Aspetti morfosintattici e interpuntivi dell'italiano contemporaneo*. Firenze: Accademia della Crusca.

Graffi, G. (1994). *Le strutture del linguaggio. Sintassi*. Bologna: Il Mulino.

Renzi, L., Salvi, G. and Cardinaletti, A. (1988). *Grande grammatica italiana di Consultazione*. La frase. I sintagmi nominale e preposizionale (vol.1). Bologna: Il Mulino.

Sornicola, R. (1981). *Sul parlato*. Bologna: Il Mulino.

Stammerjohann, H. (1986). *Tema e rema in italiano / Theme and Rheme in Italian.* Tübingen: Narr.

Tamburini, G. (1998). L'ordine dei costituenti e l'articolazione dell'informazione in italiano: un'analisi distribuzionale, *Studi di Grammatica Italiana* XVII, pp. 399-443.

Song lyrics and speech: similarities, differences and multi-dimension analysis of song lyrics from 1940 to 2009

Patrícia BÉRTOLI-DUTRA

UFMG
Av. Antonio Carlos, 6667
Belo Horizonte – MG cep: 31.270.901
patbertolid@gmail.com

Abstract

This paper shows the results of a research aiming at finding convergence of song lyrics speech and colloquial speech (general English) in order to highlight its relevance as a source for linguistic investigation. The second research goal was to find the dimensions of linguistic variation present in Anglo-American popular music lyrics. The study was theoretically based on Corpus Linguistics and the language views supported by it. Convergence was found by contrasting individual words and tri-grams (a sequence of three words) from a study corpus of over one million song lyrics to the British National Corpus and the American National Corpus. The most frequent 500 words occur in the three corpora and only three out of the 500 most frequent trigrams in the study corpus do not occur in the other corpora – such specific sequences of words reflect musical repetitions. After that, by following Douglas Biber's framework for a Multi-dimension analysis, we were able to find six linguistic dimensions and observe how those lyrics are close or different from each other according to their linguistic elements (parts of speech and semantics).

Keywords: Convergence; Corpus Linguistics; Multi-dimension Analysis; Song Lyrics.

1. Introduction

Seeing songs as a constant presence in people's everyday lives we have to consider the fact that the words people sing are also markedly relevant to the way people speak. In that sense we should consider song lyrics relevance as a source for linguistic investigation. Therefore, the first goal of the research presented here was to detect convergence points between Anglo-American song lyrics speech and colloquial speech. In other words, by considering song lyrics as a form of speech, linguistic characteristics present in song lyrics were contrasted to general English in order to highlight their similarities.

The second goal was to follow Douglas Biber's model for a Multi-dimension analysis (1988) aiming at finding dimensions of variations of Anglo-American popular song lyrics and how they could compare to the original dimensions found by Biber.

2. Research areas

Three different research fields comprise the theoretical framework of this study: 1) Studies about popular music and lyrics (Frith, 1993; Moore, 2003, Straw, 2003; Hall, 2006; Middleton, 2000; Starr & Waterman, 2007; Bértoli-Dutra, 2002); 2) Corpus Linguistics (Berber Sardinha, 2004a, 2004b; Halliday, 1991); and 3) Multidimensional Analysis (Berber Sardinha, 2004a, 2004b; Biber 1988; Kauffmann, 2005).

EFL teachers have long been using song lyrics mainly in order to either improve their learners listening skills or as a motivational asset for their classes. In fact, popular music is one of the few tools learners have to keep contact with English outside the classroom. Besides that, music also conveys social aspects as well as other aspects of the culture from where it was conceived. According to Frith (1993), music is connected to the identity of a people, "it isn't a way of expressing ideas; it is a way of living them." Thus, in a world that is getting more and more globalized exchanging music experiences is sharing identities (Hall, 2006), for music is the cultural means that best enables us to cross borders, to go where music can take us (Frith, 1993).

It is noticeable therefore that music, and most specifically its lyrics, should be used in the classroom in a more systematic way with all their linguistic information, their parts of speech and semantic aspects fully exploited. Hence, it shouldn't be considered only for its poetical or pronunciation aspects. In fact, we argue here that lyrics are not poetry with music but closer to actual conversation.

We have to highlight that for this study we considered popular music in a very comprehensive way, as the one highly disseminated by the media, sharing the view proposed by Starr and Waterman (2007): "we use the term 'popular music' broadly, to indicate music that is mass-reproduced and disseminated via the mass media (...) and that typically draws upon a variety of preexisting musical traditions (...) in which various styles, audiences, and institutions interact in complex ways."

Another important point taken into consideration for this study was the media categorization of music styles or genres. Even though we were looking at songs for their linguistics characteristics apart from their sound, it was expected that songs classified in a specific musical genre would also share the same linguistic characteristics. Among the most common musical genres present in popular music literature (Shuker, 1994; Brackett, 2000; Frith, Goodwin & Grossberg, 2003; Starr & Waterman, 2007) the following ones were present in our corpus: country (traditional country, country soul); pop (rhythm and blues); pop rock (pop rock; pop, alternative); rock (hard rock, rock, grunge, post-grunge, English rock, punk rock, heavy metal, blues rock, emo progressive); rock and roll; vocal pop (traditional pop music).

The theoretical touchstone of the whole research is Corpus Linguistics. It is an area that is based on

collecting and exploiting corpora, or a set of textual linguistic data carefully collected, in order to serve as source for the study of a language or linguistic variety (Berber Sardinha, 2004a: 3). The main concept underpinning Corpus Linguistics is viewing language as a probabilistic system (Halliday, 1991; Sinclair, 1991), that is, although there are a number of possible choices and lexical combinations they do not occur the same way or with the same frequency, not even randomly. In fact, each language follows certain patterns of lexical combinations, which represent each particular genre; thus the more words are considered for an analysis the bigger the chances of finding low frequency words and combinations. (Berber Sardinha, 2004a).

Finally, Multi-dimension analysis was used because we aimed at finding dimensions of variations of song lyrics according to Douglas Biber's model (1988), which presented a set of variation of General English. Biber's study assumes the probabilistic and functional characteristics of language (Halliday, 1991) and that linguistic variation occurs according to the context (Berber Sardinha, 2004a; Halliday; Hasan, 1989; Halliday & Webster; 2002; Sinclair, 1991). It also predicts that texts should be analyzed not only taking into account one but several linguistic features so as to determine their variation across linguistic functions. In other words, Biber states that "textual relations' among different kinds of texts" cannot "be defined unidimensionally" (1988: 20). The idea behind this methodology is to precisely quantify the frequency of each linguistic characteristic present in each text and compare every text to each other grouping them by the salience of characteristics.

In order to accomplish his goal, Biber used a corpus of 960 thousand words (mainly from the LOB-Corpus). The texts were tagged according to their parts of speech (POS). Each POS frequency was automatically calculated, normalized and submitted to statistical procedures of factorial analysis. Factorial procedure groups the most salient frequencies showing their medium, maximum, minimum and standard deviation scales. After that, the texts presenting the characteristics in each factor were checked for their relevance. It is important to highlight here that all the texts are present in all the dimensions, what makes them different in each dimension is the salience of the specific characteristics in each dimension.

Biber's analysis found six different dimensions of variation of the English Language: 1) Involved versus Informational Production; 2) Narrative versus Non-Narrative Discourse; 3) Situation-Dependent versus Elaborated Reference; 4) Overt Expression of Argumentation; 5) Non-abstract versus Abstract Style; and 6) On-Line Informational Elaboration Marking Stance.

Next section of this paper depicts the steps followed by each part of the study.

3.   Convergence study

The initial part of the study followed the principles of Corpus Linguistics (Berber Sardinha, 2004a; Bértoli Dutra, 2002; Hunston & Francis, 1999; Sinclair, 1991) first by describing the frequency of the words in the study corpus, then by describing the lexical-grammar patterns in the study corpus and finally by contrasting the patterns found in the study corpus with lexical-grammar patterns present in general English. A corpus of 1,078,882 words of song lyrics recorded originally in English by 30 different artists (American, British and Canadian) from different periods of time (from 1940's with Frank Sinatra to 2009's teen movies soundtracks, such as High School Musical and Hannah Montana).

After collecting the corpus, word lists were extracted and contrasted with word lists from the reference corpora BNC and the ANC[1] (single words and trigrams). Single words were analyzed aiming at verifying how the most frequent words in each corpus would match. After normalizing their frequency in the three corpora (so that they would be comparable), a sample of the 500[2] most frequent words in the study corpus was taken and manually contrasted to the other corpora.

Trigrams were analyzed considering they represent the best combination of words in use. According to Lafferty (Lafferty, Sleator & Temperly, 1992), "a usage of a word is determined by the manner in which the word is linked to the right and to the left in a sentence". The authors also point out that trigrams work so well for linguistic analysis "because they are firmly based on data" and because they "they reflect simultaneously syntax, semantics, and pragmatics of the domain question."

As a result of the contrastive analysis we found that the most frequent single words in the study corpus are also relevantly frequent in the general English corpora, as we can see at Table 1 below presenting the 15 most frequent words in the study corpus and their frequency in the reference corpora.

After analyzing single words we were able to conclude that song lyrics present high frequency of personal pronouns such as "I" and "YOU" which suggests interpersonal discourse. Besides that, we also noticed an overuse of the following words, when contrasted to the reference corpora: "baby"; "one"; "love"; "no"; "like"; "do"; "can"; "got"; "if"; "up"; "time"; "never" and "see".

A similar procedure was taken afterwards in order to analyze the trigrams. That is, from the 129.117 different trigrams extracted from the study corpus, 5.431.734 from BNC, 1.453.050 from ANC-spoken and

---

[1] It was used the BNC World Edition, with 100 million words available online at http://www.natcorp.ox.ac.uk/corpus/ and the online version of the ANC, available at http://www.americannationalcorpus.org/ with 22 million words. 2 Bearing in mind the amount of data we considered the most frequent 500 singles words and 500 trigrams as a representative sample.

4.236.030 from ANC written, the 500 most frequent were submitted to a manual contrastive analysis. Most of the trigrams were present in all three corpora (222), and only three out of the 500 most frequent trigrams in the study corpus do not occur in the other corpora, but they reflect something that we called "music language" (i.e. "c'mon c'mon c'mon"; "oooh oooh oooh"; "oo oo oo").

These results show that language present in song lyrics converges to everyday language, not only by the choice of individual words, but also when three words appear together. Such analysis also triggered the need for a more comprehensive analysis of lyrics speech. Thus, we chose Biber's model for a multi-dimension analysis.

WORD	FREQUENCY		
	Study Corpus	BNC	ANC
1. THE	4.02	6.02	5.44
2. YOU	3.33	0.58	0.80
3. I	3.33	0.73	0.85
4. TO	2.36	2.58	2.40
5. AND	2.28	2.61	2.68
6. A	2.14	2.17	2.21
7. ME	1.59	0.13	0.15
8. MY	1.35	0.14	0.24
9. IN	1.29	1.93	1.84
10. IT	1.21	0.91	1.15
11. OF	1.17	3.03	2.73
12. YOUR	0.99	0.13	0.11
13. ON	0.91	0.72	0.63
14. THAT	0.87	1.04	0.76
15.ALL	0.80	0.27	0.23

Table 1: Most frequent words in the study corpus compared to BNC and ANC

4.   Multi-dimension analysis

At this point of the study, the collected corpus (that never stops growing) consisted of approximately 1,200,000 words from 6,290 song lyrics originally written in English. The corpus was tagged for its parts-of-speech features and for its semantic groupings. These features and the most frequent lexical bundles (3-grams) in the corpus and in general English (Google N-Gram corpus) were considered as variables for the factor extraction at the SPSS program. Factor analysis reduces the huge number of variables, grouping them according to their co-occurrence. This procedure is done through the identification of the distribution patterns of variables. The 97 initial variables in our research were grouped into 13 grammar variables, 8 semantic variables, and 2 pattern variables (3-grams). Factor analysis resulted in three factors for each of the variable group.

The interpretation of the factors was conducted in order to find the main factors responsible for linguistic variation in song lyrics as so they would be interpreted as the dimensions they expressed. The dimensions were analyzed in search of how they were represented in relation to musical styles, to different artists and along the time.

The factor extraction resulted three factors that were accounted for their grammatical and semantic aspects. Grammatically they show the following oppositions: (1) infinitive, gerund and modals versus nouns; (2) personal pronouns and possessives versus qualifiers; (3) verbs in the past versus verbs in the present. Semantically the factors show the predominance of (1) movement/time/speech/people/object; (2) markers of emotion and social acts; (3) markers of music manifestation. From the interpretation of the factors emerged the following dimensions: (a) argumentative versus informative; (b) interactive versus descriptive; (c) past narratives versus immediate context; (d) personal acts; (e) emotion and society; and (f) musical manifestation.

The investigation of song lyrics on the dimensional scale showed how singers and bands, musical styles and the decade of the recordings are closer or more distant to each other in linguistic terms. The most representative style, artist and period of time for each of the dimensions, grammar and semantics, are as follows[3]: (a) Punk Pop, Simple Paln, 2000's; (b) Rock'n'roll, Madonna, 1940's; (c) Country, Johnny Cash, 1970's; (d) Surf Rock, Beach Boys, 1960's; (e) Heavy Metal, Metallica, 1940's; and (f) pop Vocal, Frank Sinatra, 1940's.

5.   Considerations

This study showed how close ordinary spoken and written English are to song lyrics speech. It also validated Biber's model for the research of contrast of linguistic features in functional terms. However, the Multi-dimension Analysis methodology cannot be considered as the only possible means for linguistic analysis of song lyrics or any other form of speech. We were able to observe how songs are close or distant, similar or different according to their linguistic elements and not only according to their rhythm and musical style generally imposed by the media.

6.   Acknowledgements

7.   References

Berber Sardinha, A.P. (2004a). *Linguística de Corpus*. Barueri: Manole.

Berber Sardinha, A.P. (2004b). Informatividade, interatividade e narratividade na reunião de negócios – Análise Multidimensional e palavras-chave. *DIRECT Papers* (52), São Paulo and Liverpool.

Bértoli-Dutra, P. (2002). Explorando a linguística de corpus e letras de música na produção de atividades pedagógicas. Master's Dissertation, Unpublished, LAEL, PUC-SP. Available at

---

3   For a comprehensive view of results, refer to http://www.sapientia.pucsp.br/tde_busca/arquivo.php?codArquivo=10985

<http://www4.pucsp.br/pos/lael/lael-inf/def_teses.html>

Bértoli-Dutra, P. (2010). Linguagem da Música Popular Anglo-Americana de 1940 a 2009. Doctorate's Theses, Unpublished, LAEL, PUC-SP. Available at: <http://www4.pucsp.br/pos/lael/lael-inf/def_teses.html>.

Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Brackett, D. (2000). *Interpreting Popular Music*. University of California Press.

Frith, S. Music and Identity. (1993). In S. Hall, D.G. Paul (Eds), *Questions of Cultural Identity*, London, UK: Sage publications. pp. 108--127

Frith, S., Goodwin, A. and Grossberg, L. (2003). *Sound and Vision: the music video reader.* London, UK: Routledge.

Hall, S. (2006). *A Identidade Cultural na Pós-Modernidade*. 11a ed. Rio de Janeiro: DP&A.

Halliday, M.A.K. (1991). Corpus studies and probabilistic grammar. In: K. Aijmer, B. Altenberg (Eds.), *English Corpus Linguistics: Studies in honour of Jan Svartvik*. London: Longman, pp. 30--43.

Halliday, M.A.K., Hasan, R. (1989). *Language, Context, and Text: aspects of language in a social-semiotic perspective*. 2nd edition. Deakin University Press/Oxford University Press.

Halliday, M.A.K., Webster, J. (2002). *On grammar*: *By Michael Alexander Kirkwood Halliday*. New York: Continuum.

Hunston, S., Francis, G. (2000). *Pattern Grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam/Phildelphia: John Benjamins.

Kauffman, C.H. (2005) Corpus do jornal: variação linguística, gêneros e dimensões da imprensa diária escrita. Master's dissertation, Unpublished. LAEL, PUC-SP. Available at:
<http://www4.pucsp.br/pos/lael/lael-inf/def_teses.html>.

Lafferty, J., Sleator, D. and Temperley, D. (1992). Grammatical trigrams: A probabilistic model of link grammar. In *Proceedings of the AAAI Fall Symposium on Probabilistic Approaches to Natural Language*, Cambridge, MA.

Shuker, R. (1994). *Understanding Popular Music*. London, New York: Routledge.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Starr, L., Waterman, C. (2007). *American Popular Music. From minstrelsy to MP3*. 2nd. ed. New York: OUP.

Straw, W. (2003). Pop music and postmodernism in the 1980s. In S. Frith, A. Goodwin and L. Grossberg. (Eds.), *Sound and Vision: the music video reader*. London, UK: Routledge. pp. 3--21.

The use of inflected infinitive in a spoken corpus

Fernanda CANEVER

Universidade de São Paulo (USP)

Av. Prof. Luciano Gualberto, 403 - Sala 16 - Cidade Universitária - 05508-010- São Paulo - SP

fernandacanever@gmail.com

Abstract

In light of the usage-based approach (Langacker, 1987, 2000; Bybee, 2006a, 2006b, 2010) and the theory of utterance selection proposed by Croft (2000), this study intends to contribute to the investigation of the continuous update of linguistic knowledge that occurs through language use. Building upon prior research done by Canever (2012), which quantified the usage of the inflected infinitive in a written corpus, the focus of this study is on the use of the inflected infinitive in Brazilian Portuguese in a spoken corpus, namely a sample of the corpus Nurc/SP. The results show the presence of inflected infinitive in some innovative constructions in the 1970s, suggesting that a quantitative study with the complete Nurc/SP corpus should be likewise revealing. It is also argued that more studies with large spoken corpora of Brazilian Portuguese are needed to confirm Canever's hypothesis that the infinitive inflection has received a positive social value, which, reinforced by the stigmatized lack of verbal agreement in Brazil and associated with the high frequency of occurrence of the infinitive inflection in other syntactic contexts, would be causing the inflection to spread to new infinitive constructions.

Keywords: Spoken Corpus; Usage-based Theories; Language Change; Inflected Infinitive; Automatic Data Extraction.

1. Introduction

Traditionally language use has not been the focus of linguistic investigation. Structuralism and generative grammar have given high priority to the *langue*, claiming that the linguistic system is self-contained and autonomous from other cognitive abilities and social factors (Croft, 2000). As a result, phenomena related to the *parole* such as variation have been considered peripheral.

Yet, Bybee (2006b) points that the interest for speech has increased in the last decades, and many theoretical approaches now claim that language structure should not be isolated from language use. Cognitive linguistics, which Langacker (1987, 2000) defines as usage-based, is one of them. According to this framework, language structure emerges from language use through general cognitive capabilities of the human brain, not because of an endowment exclusively related to language. But seen as symbolic, language represents a human biological adaptation for interactive goals (Tomasello, 2003). Thus, the role of experience in shaping both our linguistic knowledge and our concepts is highly emphasized in cognitive approaches to language studies.

Moreover, advances in computational and corpus linguistics have facilitated studies with real data. This means that those interested in capturing the more dynamic nature of language are now able to investigate linguistic phenomena by analyzing naturally-occurring data, and this is the realm this study belongs to. In light of the usage-based approach (Langacker, 1987, 2000; Bybee, 2006a, 2006b, 2010) and the theory of utterance selection proposed by Croft (2000), the aim of this study is to contribute to the investigation of how language use constantly shapes speaker's grammar by quantifying variation in speech. Building upon prior research done by Canever (2012), which quantified the usage of the inflected infinitive in a written corpus, this study focuses on the usage of inflected infinitive in a spoken corpus, namely Nurc/SP, as well as on the challenges involved in such a task.

2. Usage-based theories

Coined by Langacker (1987), the term usage-based model refers to a non-reductive approach that acknowledges the linguistic system as a collection of both rules and actual occurring expressions rich in semantic, phonological and symbolic details. The system comprises, therefore, not only "the schemas that emerge spring from the soil of actual usage" (Langacker, 2000: 3), but also instances of very specific occurrences of use in a storage of redundant information.

According to Langacker (1987), a language is a "structured inventory of conventional linguistic units" (p. 494). To understand how this inventory is structured, it is important to consider that in actual instances of language use, referred to by Langacker as *usage events*, the language user has to relate his linguistic system to these events. Either in order to produce an utterance with an intended meaning or to interpret someone else's utterance, the language user establishes a connection between the usage event and his inventory, trying to find a similar structure. In case a compatible structure is found, the schema instantiated in the utterance is taken to be conventional. When a good match is not possible, the schema instantiated is considered non-conventional.

According to Langacker, novel structures may gradually become conventional and be stored in our linguistic inventory depending on their frequency of occurrence. When a non-conventional structure gets into the system, it might be reinforced by frequent use or disappear due to non-use. What is crucial in this process is the cognitive ability of habit formation, which Langacker refers to as *entrenchment*: the more frequent an element is, the more entrenched it becomes. Repetition, thus, affects speakers' linguistic knowledge, and plays an important role in the characterization of a structure as being conventional.

The fact that the concrete use of language structures in the daily life of a speech community results in the emergence of new linguistic patterns may initially appear chaotic. However, it is undeniable that language is stable to a great extent. Such stability – or *convention*[1] – is what allows communication and all the other social-interactive goals involved in language use to be achieved.

Even though Langacker recognized the role of use in the shaping of linguistic structure, his work has not discussed why some utterances propagate while others disappear. Considering that when a novel structure emerges, its frequency of occurrence is low, Blythe & Croft (2012) state that all innovations are expected to disappear if only the frequency of occurrence is considered. For this reason, these authors claim that frequency alone cannot explain how novel structures may survive and even replace former conventional structures.

Croft (2000), who proposes a usage-based theory for language change that is directly connected to theories of language use such as the one developed by Clark (1996), claims that social factors need to be taken into account in the investigation of language change. In presenting his theory of utterance selection, which is based on Hull's generalized theory of selection (Hull, 1988), Croft (2000) proposes that language change is an evolutionary process, which is a model of change by replication. In this model, the replicator is a token of linguistic structure, which he calls a *lingueme*; the interactor is the speaker who replicates *linguemes* in interacting with other speakers; the population is a speech community, that is, a population of interactors; and the environment is the social context of the speech event, its goals as well as the other members of the population.

Based on the hypothesis that language change emerges from language use, the author claims that linguistic convention is central to the process of change. While interacting, when speakers are conforming to convention, they are doing what Croft called *normal replication*. However, even though speakers try to conform to convention, they often end up violating it by using non-conventional devices. Such non-conformity to convention is called *altered replication*, and is the first step to change – innovation. Once variation is generated through altered replication, different variants are made available for speakers to use, so they need to select among them, and this is called *differential replication*. To Croft, language change consists of these two steps: innovation and propagation/selection.

After innovations occur, they might be propagated or not. When propagation takes place, it means a new convention is established. As defended by Croft (2000), propagation is a social process, since it occurs according to the social values assigned to the variants, such as prestige, for example. However, in order to perpetuate,

the cognitive structures on which linguistic utterances depend need to be entrenched in the speaker's grammar.

The correlation between the degree of entrenchment and the social values assigned to linguistic variants in guiding language change posited by Croft seems to be the most appropriate way of approaching the issue, and therefore this idea underlies this investigation. Furthermore, since frequency of occurrence is crucial to determining the degree of entrenchment of linguistic constructions in speaker's grammars, frequency studies are presumed to play a vital role in the investigation of natural languages.

3.   The Portuguese inflected infinitive

According to Maurer (1968), the inflection of the infinitive has been documented since the first Portuguese documents, and has gradually spread to different constructions. Nowadays, the inflection is considered optional in numerous contexts, as in:

(1) Estudamos para   vencermos   na vida.
    study.1PL   to   succeed.INF.1PL   in life
    We study to succeed in life.

(2) Estudamos para   vencer   na vida.
    study.1PL   to   succeed.INF   in life
    We study to succeed in life.

Bechara (2009), for instance, states that the infinitive inflection is used when the speaker intends to emphasize the grammatical person, as shown in (1), and the uninflected form is used when the emphasis is on the action, as shown in (2).

Recently, though, examples[2] of the inflection of the infinitive in contexts where it is considered hypercorrection have been attested in spoken language, as in:

(3) Viemos para SP para podermos lançarmos …
    came.1PL to   SP to   can.INF.1PL launch.INF.1PL
    We came to SP to be able to launch ...

(4) Nós temos   que   nos   prepararmos…
    we have.1PL that   REFL.1PL   prepare.INF.1PL
    We need to prepare ourselves ...

Interested in infinitive constructions with optional inflection as well as in some more innovative contexts for the infinitive inflection, such as those illustrated by examples (3) and (4), Canever (2012) quantified the variation in a corpus of standard written language, more specifically a corpus of academic written Brazilian Portuguese that contained 11,000,000 words. The results

---

[1] Reformulating Lewis (1969 in Clark 1996: 71), Clark defines *convention* as a partly arbitrary regularity in behavior that is common ground in a given community, but even though it is stable, it is not static (Croft, 2000: 132).

[2] The examples (3) and (4) were collected by members of the LLIC/USP (http://www.linguistica.fflch.usp.br/llic), while the examples (5) to (9) were taken from Canever (2012). Because of space limitations, only excerpts of the examples are presented here.

reveal a high frequency of occurrence of the inflected infinitive, mainly in causal, final and temporal clauses, such as in:

(5) Tarefa que não podemos recusar, especialmente
    task   that not can.1PL   refuse   mainly

    para entendermos   a   falta de ...
    to   understand.INF.1PL the lack   of
    A task we cannot refuse, mainly in order to
    understand the lack of...

In constructions such as modal and aspect periphrases with an infinitive, Canever showed there is no preference for the inflection, as in:

(6) Podemos levantar   a   seguinte hipótese ...
    can.IPL   suggest.INF the following hypothesis
    We can suggest the following hypothesis...

(7) As mulheres começam a   ser   felizes …
    the women   start   to be.INF happy.PL
    Women start to be happy ...

However, a few occurrences of inflected infinitive were found in those constructions, such as in:

(8) Não poderiam serem   esquecidas …
    not could.3PL be.INF.3PL forgot.PL
    Couldn't be forgotten ...

(9) As virtudes começam a serem   tratadas …
    the virtues start.3PL to be.3PL.INF treated.PL
    The virtues start to be treated ...

Given the occurrence of such hypercorrect infinitive inflections in a written corpus of standard Portuguese, Canever claims that a positive social value might have been attributed to the inflected forms. Canever states that this positive value, reinforced by the stigma associated with the lack of verbal agreement in Brazil, and the high frequency of occurrence of infinitive inflection in other syntactic contexts could – together – be causing the inflection to spread to new infinitive constructions.

Although the results found by Canever suggest that in many constructions the inflected forms are highly entrenched in the grammars of the investigated speakers, further quantitative studies with spoken corpora are necessary to validate the hypothesis that the inflected infinitive is spreading in standard Brazilian Portuguese.

4.   Quantification in a spoken corpus

4.1   Methods

4.1.1.   Corpus
The spoken corpus used for this study was a sample of formal utterances – lectures, conferences, etc. – collected by the NURC project[3] in São Paulo, Brazil. The sample, with approximately 30,000 words, consists of utterances produced by six participants, and has been published in a book (Castilho & Preti, 1986).

4.1.2.   Data extraction
Because the original files were in .pdf format, they had to be converted to .txt format so the data extraction could be automatically done with the software R. In order to extract the occurrences of the infinitive inflection, a script containing the function *exact.matches* was used[4]. The script basically made R look for all the occurrences of words that ended either in *–rmos* or *–rem*, which are the infinitive plural inflections, and return the matches with some preceding and subsequent contexts. The output file was then handled in a spreadsheet program.

4.2   Results
Among the occurrences of infinitive inflection found, 20 were occurrences of the Third Person Plural (3PL) inflection *–rem*. Most of them occurred in contexts where a plural subject precedes the infinitive, such as in:

(10) (…) que levam as pessoas a demandarem …
        that lead.3PL the people to demand.INF.3PL
     (…) *that lead people to demand …*

As for the inflection of First Person Plural (1PL) – -rmos –, 8 occurrences were found, one of them being:

(11) Nós podemos utilizarmos   desta   reflexão …
     we can.1PL use.INF.1PL   of.this reflection
     We can use this reflection ...

4.3   Discussion
Given the small size of the sample, not many results were found. However, the quantification yielded some interesting results. The occurrence of an infinitive inflection after a modal verb such as in (11), for instance, suggests that the inflection of the infinitive in constructions such as modal periphrases, which Canever (2012) considered innovative and hypercorrect usage, already occurred in spoken language in the 1970s.

5.   Conclusion and future directions
This study quantified the usage of inflected infinitive in a sample of the spoken corpus (Nurc/SP) in order to contribute to the investigation of how usage is constantly

---

[3] NURC stands for *Norma Urbana Culta (*urban spoken standard language), and this project consisted of the investigation of spoken Portuguese in five state capitals in Brazil: São Paulo, Rio de Janeiro, Recife, Salvador and Porto Alegre in the 1970s.

[4] The script can be found in Canever (2012), and the function the function *exact.matches*, developed by professor Stefan Th. Gries (University of California Santa Barbara), is available at: <http://www.linguistics.ucsb.edu/faculty/stgries/exact_matches .r>.

shaping our linguistic knowledge. The results found are revealing and suggest that a quantitative study with the complete Nurc/SP corpus should be likewise relevant to the investigation of the spread of the inflected infinitive in Brazilian Portuguese.

In order to do to that, some methodological challenges will have to be dealt with, though. First of all, it is crucial that the corpus Nurc/SP be in a machine-readable format, ideally in a format that is compatible with software such as R. Once this is done, it will be important to decide what annotation should be kept, as well as what kind of cleaning will be necessary, mainly because some speech annotation might be a problem in data extraction.

To support Canever (2012)'s hypothesis that the inflected infinitive is spreading in Brazilian Portuguese not only because of its high frequency of occurrence in optional contexts, but also because the inflection has received a positive social value, the use of the inflected infinitive needs to be quantified in different spoken corpora. For this reason, after the study with the whole Nurc/Sp corpus is ready, it will be also important to contrast its results with data obtained from more contemporary spoken corpora of Portuguese.

Given the lack of large spoken electronic corpora of Contemporary Brazilian Portuguese, a solution might be to work with different corpora formed by different research groups in Brazil.

6.   Acknowledgements

7.   References

Barlow, M., Kemmer, S. (Eds.). (2000). *Usage-Based Models of Language*. Stanford: CSLI Publications.

Baxter, G., Blythe, R., Croft, W. and McKane, A.J. (2006). Utterance selection model of language change. In *Physical Review*, vol. E 73, pp. 046--118.

Baxter, G., Blythe, R., Croft, W. and McKane, A.J. (2009). Modeling language change: An evaluation of trudgill's theory of the emergence of new zealand english. *Language Variation and Change*, vol. 21(2), pp. 257-- 296.

Bechara, E. (2009). *Moderna gramática portuguesa*. Rio de Janeiro: Lucerna, 37 ed.

Blythe, R.A., Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, 88 (2), pp. 269--304.

Bybee, J. (2006a). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.

Bybee, J. (2006b). From usage to grammar: the mind's response to repetition. *Language*, 82 (4), pp. 711--733.

Bybee, J. (2010). Language, Usage and Cognition. Cambridge: Cambridge University Press.

Canever, F. (2012). *Evidências para um modelo de língua baseado no uso: o infinitivo flexionado em português brasileiro*. Dissertação de Mestrado. Universidade de São Paulo, Brasil. Available at: <http://www.teses.usp.br/teses/disponiveis/8/8139/tde-02082012-133430/>.

Castilho, A.T., Preti, D. (Eds). (1986). *A linguagem falada culta na cidade de São Paulo: materiais para seu estudo. v.I – Elocuções formais*. São Paulo: T.A. Queiroz, 1986.

Clark, H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Croft, W. (2000). *Explaining language change: an evolutionary approach.* Harlow, Essex: Longman.

Croft, W. (2008). Evolutionary linguistics. *Annual Review of Anthropology*, vol. 37, pp. 219--34.

Cunha, C., Cintra, L. (2008). *Nova gramática do português contemporâneo.* Rio de Janeiro: Nova Fronteira., 5a ed.

Hull, D. (1988) *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science.* Chicago, IL: Univ. Chicago Press.

Langacker, R. (1987). *Foundations of Cognitive Grammar, vol. 1, Theoretical Prerequisites.* Stanford: Stanford University Press.

Langacker, R. (2000). A dynamic usage-based model. In M. Barlow, S. Kemmer (Eds.), *Usage-Based Models of Language*. Stanford: CSLI Publications, pp. 1--63.

Maurer Jr, T.H. (1968). *O infinito flexionado português: estudo histórico-descritivo*. Imprenta. São Paulo: Cia. Ed. Nacional.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Harvard: Harvard University Press.

Team, R Develpment Core. (2011). R: A Language for Statistical Computing. R Foundation for Statistical Computing. Vienna, Áustria. ISBN 3-900051-07-0. Available at: <http://www.R-project.org/>.

A corpus-based analysis for superlative construction of body expression

Igor de Oliveira COSTA, Neusa Salim MIRANDA
Federal University at Juiz de Fora
Juiz de Fora, Minas Gerais, Brazil
igorsabo@yahoo.com.br, neusasalim@oi.com.br

Abstract

This work focuses on the corpus dimension of the Superlative Construction of Body Expression ("[...] solteirona e toda virgem, ignorava machezas, quase **morreu de vergonha** numa tarde de conversas"; "Padre Dito quase **estourou de rir** [...]"; "O Lúcio **rolou de rir** com a explicação, e como consequência acabou virando a vítima e a cobaia do seminário."), a major link in the network of constructions of Portuguese named by Miranda (2008a) as Superlative Constructions. The theoretical approach involves the Cognitive Linguistics and the Cognitive Construction Grammar. The corpus used is the Corpus do Português (http://www.corpusdoportugues.org/), composed of forty-five million words of fifty-seven thousand texts of the XIV-XX centuries. The results points, among other things, to the productivity of the construction under investigation, which instantiate, in the corpus investigated, 19 different types, and its conventionalization, outlined by the presence of 1.726 tokens, that corresponds to 43,9% of the usage of the searched verbs followed by the genitive preposition "de" in the *corpus* (3.929). The advantage in adopting a corpus based approach on the constructions' investigation is also highlighted, once it offers access to the comprehension of the construction's productivity and conventionalization in a language.

Keywords: cognitive linguistics; cognitive construction grammar; corpus-based approach; intensity; superlative constructions.

1. Introduction

The notion of degree is very rich to the grammar of languages. It is through scalar constructions that the language users denote the degree that speakers/writers can approach what they say/write what they saw, experienced or believe they have experienced, among other things.

There are many structures in the Portuguese language (as in other languages) that serve this purpose of intensifying a statement. But against what speakers/writers use, the Grammatical Tradition and even Linguistic Tradition, little or almost nothing, devoted to the study of this phenomenon. Some examples of degree modifier constructions present, for example, in normative grammars of Portuguese are: Comparative Constructions ("Ele é **tão** rápido **quanto** o Bolt"/"He is **as** fast **as** Bolt"; "Eu escrevo **melhor/pior do que** ele"/"I write **better**/**worse than** he"), Construction with Adverbs of Intensity ("Maria Fernanda Cândido é perfeita **demais**"/"Maria Fernanda Candido is **too** perfect"), pleonastic expressions ("Que jogada **linda, linda, linda**!"/"What a **pretty, pretty** move!").

In order to fill this gap, the present work, along with others, aims to expand the study of the manifestations of degree in Portuguese Language, as a way to contribute to a fuller description of the language. In this work, the object under investigation is the Superlative Construction of Body Expression (SCBE)[1]:

(1) <u>19:Fic:Br:Cony:Piano</u>   Enquanto o sábado não chegasse, ele podia se **fartar de ouvir** todos os discos que quisesse [...]

"While Saturday was not enough, s/he could **glut of listening** to all the discs he wanted [...]" (to glut of listening = to get enough of listening = to listen a lot)

(2) <u>19Or:Br:Intrv:ISP</u>   [...] o meu clown não consegue cruzar os braços. A platéia **morre de rir** do que é, na verdade, uma tragédia para o meu personagem.
"[...] my clown cannot cross his arms. The audience **die of laughing** about what is, indeed, a tragedy for my character." (to die of laughing = to die laughing = to laugh too much)

(3) <u>19:Fic:Br:Garcia:Silencio</u> [...] queria era apenas assustar, podemos telefonar para ele e dizer que eu **estou me borrando de medo**.
"[...] s/he just want to scare, we can call him and say I **am shiting of fear**." (to shit of fear = to scared shitless = to be very much afraid)

Because it is a very broad research (which, in addition to the formal description and semantic-pragmatic motivations, involves its conceptual motivation, its inheritance relations, its process of grammaticalization, among other issues[2]), this work cuts out the part of the SCBE study that is more directly related to the use of corpora.

This research is linked to the "Superlative Constructions of Brazilian Portuguese: a study about scale semantic" (Miranda, 2008 – CNPq), which, from its genesis to now, elucidated, with the study of the SCBE, seven nodes of this large network of constructions. Four other studies are still in progress.

The paper is organized as follows: the first section presents the theoretical perspective through which we develop our object; the following section discusses the research methodology chosen and the process of data

---

[1] All the English "versions" of the examples and SCBE types are just an attempting to clarify the phenomenon being studying, presenting the semantic nature of words that compose the construction.

[2] Costa (2010) covers most of these points.

collection; section 3, in turn, will bring the analyzes of SCBE, which involves the use of corpus; after that, we presented our conclusion, followed by the acknowledgments and the references.

2. Theoretical Bases

The theoretical framework of this study is composed of Cognitive Linguistics (Fauconnier, 1994; Fauconnier & Turner, 2002; Fillmore, 1982; Johnson, 1987; Lakoff, 1987; Lakoff & Johnson, 2002[1980], 1999; Miranda, 2002, 2008a, 2008b; Salomão, 1997, 2006; among others) and one of its models of grammar, the Cognitive Construction Grammar (Goldberg, 1995, 2006; Boas, in press).

The cognitive research program of language emerged at the end of the seventies last century, and strongly opposes to the Generative Grammar and Truth-conditions semantics. In general, Cognitive Linguistics considers (1) language as a non-autonomous cognitive faculty, governed by general cognitive apparatus; (2) advocates a central role for imaginative processes (metaphor, metonymy, blending) in human cognition and language; (3) sees grammar as conceptualization, as a way to profile a human scene; and (4) assumes that knowledge of language emerges from its use (Croft & Cruse, 2004: 1-4).

The Cognitive Construction Grammar (CCxG) (Goldberg, 1995, 2006; Boas, in press), defining constructs as pairs of form and function, gives these structures the status of basic units of language. Thus, the grammar and lexicon are defined as a network of constructions established by the use through culture. The description of such structures, therefore, is realized investigating not only their formal patterns, but also their dimensions of meaning and use.

A key point for the Goldberian model of grammar is the frequency of type and frequency of token variables, responsible respectively for the entrenchment of certain constructional pattern in the minds of speakers of a language and the conventionalization of a construction in a given language (that is, the capacity of a construction to be extended to new cases within the language). Once a corpus allows the verification of such data, the use of this tool in a study of an object like the one being investigated here is highly profitable and productive.

As a model of grammar fully immersed in the assumptions of Cognitive Linguistics, CCxG aims to provide psychologically plausible explanations for the language (Croft & Cruse, 2004: 272; Boas, in press: 12.) exploring the motivation and inheritance relations among constructions.

3. Methodology

Due to the importance of the use in the theoretical model adopted (CCxG is a use-based model of language, cf. Croft & Cruse, 2004: 291-327), we make use of a corpus-based approach (Aluísio & Almeida, 2006; Divjak & Gries, 2003; Sardinha, 2004; Stefanowitsch, 2006) in the investigation of the object.

The assembly of a database specifically for cases involving the SCBE is the first (and crucial) step in the study of a construction, because it is a way of letting the data speak, and not be hostage solely to our intuitions. Therefore, in order to be faithful to it, the search for cases of the construction was divided into two different phases: one in which we use different sources to get the most different types of the construction and another in which we make use of an annotated corpus for systematic study of the construction.

	Constructional types[3] (Y = rir)	CP	CE	Abril .com	Total
01	acabar(-se) de rir "to finish of laughing"	---	---	09	09
02	borrar(-se) de rir "to blot of laughing"	01	---	---	01
03	cagar(-se) de rir "to shit of laughing"	---	---	01	01
04	cair de rir "to fall of laughing"	---	---	01	01
05	cansar(-se) de rir "to be tired of laughing"	01	02	---	03
06	chorar de rir "to cry of laughing"	01	---	03	04
07	contorcer(-se) de rir "to contort of laughing"	---	01	01	02
08	dobrar(-se) de rir "to bend of laughing"	---	---	03	03
09	engasgar(-se) de rir "to choke of laughing"	---	01	---	01
10	esbaldar(-se) de rir "to splurge of laughing"	---	---	01	01
11	esborrachar(-se) de rir "to squash of laughing"	---	---	01	01
12	escangalhar(-se) de rir "to queer of laughing"	---	---	09	09
13	escrachar(-se) de rir "to shatter of laughing"	---	---	01	01
14	esganiçar(-se) de rir "to scream of laughing"	---	---	01	01
15	espremer(-se) de rir "to squeeze of laughing"	---	01	---	01
16	estourar(-se) de rir "to burst of laughing"	01	---	---	01
17	fartar(-se) de rir "to glut of laughing"	10	19	---	29
18	finar(-se) de rir "to die of laughing"	01	---	---	01
19	mijar(-se) de rir "to piss of laughing"	---	01	01	02
20	morrer de rir "to die of laughing"	14	20	185	219
21	não (se) aguentar de rir "to not hold of laughing"	---	---	01	01
22	passar mal de rir "to be sick of laughing"	---	---	02	02
23	rachar(-se) de rir "to crack of laughing"	---	---	08	08
24	rasgar(-se) de rir "to rip of laughing"	---	---	01	01
25	rebentar(-se) de rir "to burst of laughing"	01	---	---	01
26	rolar de rir "to roll of laughing"	---	08	52	60
27	torcer(-se) de rir "to twist of laughing"	---	---	01	01
	TOTAL	30	53	282	365

Table 1: SCBE Types

---

[3] The particle "se" presented between parentheses is a Portuguese reflexive pronoun demanded by one of the uses of some verbs in the construction.

First phase: having the results of Sampaio (2007) – which point "rir" ("laughing") as the most frequent Y element to the pattern 'X DE Y' ("chorar de rir"/"to cry of laughing", "fartar-se de rir"/"glut of laughing", "morrer de rir"/"to die of laughing", etc.) – as the start point, first we searched for the expression "de rir" in three different language database (the Corpus do Português, the Corpus Eye of the VISL project, and Abril.com) as a way to raise X elements of the constructional pattern being investigated. The initial hypothesis was that, starting from a most common form and therefore more conventional, it was possible to obtain wide and significant combinations of the variables which compose the construction. In fact, our hypothesis was confirmed. Table 1, below, shows the types collected in the searches.

	SCBE type	Results of the search	Tokens of SCBE	Productivity of the search
01	acabar(-se) de Y "to finish of Y"	252	08	3.2%
02	borrar(-se) de Y "to blot of Y"	08	04	50%
03	cagar(-se) de Y "to shit of Y"	03	02	66.7%
04	cair de Y "to fall of Y"	835	96	11.5%
05	cansar(-se) de Y "to be tired of Y"	437	372	85.1%
06	chorar(-se) de Y "to cry of Y"	196	112	57.1%
07	contorcer(-se) de Y "to contort of Y"	06	01	16.7%
08	dobrar(-se) de Y "to bend of Y"	75	01	1.3%
09	engasgar(-se) de Y "to choke of Y"	---	---	---
10	esbaldar(-se) de Y "to splurge of Y"	---	---	---
11	esborrachar(-se) de Y "to squash of Y"	---	---	---
12	escangalhar(-se) de Y "to queer of Y"	01	01	100%
13	escrachar(-se) de Y "to shatter of Y"	---	---	---
14	esganiçar(-se) de Y "to scream of Y"	---	---	---
15	espremer(-se) de Y "to squeeze of Y"	06	---	---
16	estourar(-se) de Y "to burst of Y"	27	17	63%
17	fartar(-se) de Y "to glut of Y"	401	381	95%
18	finar(-se) de Y "to die of Y"	18	05	27.8%
19	mijar(-se) de Y "to piss of Y"	02	01	50%
20	morrer de Y "to die of Y"	1.486	674	45.4%
21	não (se) aguentar de Y "to not hold of Y"	01	01	100%
22	passar mal de Y "to be sick of Y"	---	---	---
23	rachar(-se) de Y "to crack of Y"	18	01	5.6%
24	rasgar(-se) de Y "to rip of Y"	46	05	10.9%
25	rebentar(-se) de Y "to burst of Y"	52	34	65.4%
26	rolar de Y "to roll of Y"	29	---	---
27	torcer(-se) de Y "to twist of Y"	30	10	33.3%
	TOTAL	3,929	1,726	43.9%

Table 2: Data obtained in the second phase of the study

4.   Analysis

In the description and explanation of SCBE, some findings are more strongly linked to the adoption of corpus research. As explained to the introduction, these findings are topics of the next lines.

In view of the data obtained from the corpus, the SCBE appears as a very productive construction, instantiating 19 different types in the corpus investigated. The construction can also be considered conventionalized since 1,726 tokens of the construction were found in Corpus do Português. This corresponds to 43.9% of the use of the 19 verbs followed by the preposition "de" in the corpus (3,929).

There is, however, a variation in the conventionalization of each type: only "Morrer de Y", "Fartar(-se) de Y", "Cansar(-se) de Y", "Chorar de Y", "Cair de Y" had a number of tokens that could attest to their conventionalization, as shown in Table 3:

	SCBE Types	Tokens
01	morrer de Y "to die of Y"	674
02	fartar(-se) de Y "to glut of Y"	381
03	cansar(-se) de Y "to be tired of Y"	372
04	chorar de Y "to cry of Y"	112
05	cair de Y "to fall of Y"	96
06	rebentar(-se) de Y "to burst of Y"	34
07	estourar(-se) de Y "to burst of Y"	17
08	torcer(-se) de Y "to bend of Y"	10
09	acabar(-se) de Y "to finish of Y"	08
10	finar(-se) de Y "to die of Y"	05
11	rasgar(-se) de Y "to rip of Y"	05
12	borrar(-se) de Y "to twist of Y"	04
13	cagar(-se) de Y "to shit of Y"	02
14	mijar(-se) de Y "to piss of Y"	01
15	escangalhar(-se) de Y "to queer of Y"	01
16	contorcer(-se) de Y "to contort of Y"	01
17	dobrar(-se) de Y "to bend of Y"	01
18	não (se) aguentar de Y "to not hold of Y"	01
19	rachar(-se) de Y "to crack of Y"	01
	TOTAL	1,726

Table 3: Conventionalization of SCBE types in Corpus do Português

According to the occurrence of SCBE in the corpus, it was possible to more precisely understand the form of construction:

$$[X_V \text{ de } Y_{N/V}],$$

where X is filled with verbs that evoke the conceptual domains of physical impact ("acabar"/"to finish",

"cair"/"to fall", "rachar"/"to crack", "rolar"/ "to roll") or physiological impact ("cagar"/"to shit", "cansar"/"to be tired", "mijar"/ "to piss", "morrer"/"to die") and Y prototypically is an abstract name or a verb:

(4) <u>16:FMMelo:Letters</u>   Com as premissas de que haveria de seguir o Conde Ene ao Brasil, **me acabei de destruir**, empenhar e carregar de novas obrigações.
"With the assumptions that I should follow the Count Ene to Brazil, I **finished of destroying, engage and load** of new bonds." (to finished of destroying = destroy a lot; finished of engage = to engage in a superlative way; finished of load = load a lot)

(5) <u>18:Azevedo:Japão</u>   [...] dragonas de ouro e desses chapéus de pluma que fizeram **rebentar de medo** o Imperador da China nas profundezas empedradas de Pekin.
[...]gold epaulettes and these feather hats that made the Emperor of China **burst of fear** in the depths paved of Pekin. (to burst of fear = to have a lot of fear)

(6) <u>18:Álvares:Lira</u>   E quando eu **morra de esperar** por ela.../Deixai que eu durma ali […]
And when I **die of waiting** for her…/ Let me sleep here [...] (to die of waiting = to wait for a long, long time)

(7) <u>19N:Pt:Beira</u>   Maria do Carmo Borges, a presidente em exercício, **não se cansou de valorizar** esta festa, e tinha razões para isso.
Maria do Carmo Borges, the acting president, **wasn't tired of appreciate** this feast, and she had reasons for this. (to not be tired of appreciate = to appreciate a lot)

(8) <u>19Or:Br:Intrv:ISP</u>   Aí Cacá fez Ubu, estourou e eu **fiquei morrendo de inveja**.
Then Caca made "bang", he burst and I **was dying of envy**. (to die of envy = to have a lot of envy)

(9) <u>19:Fic:Br:Novaes:Mao</u>   Foi quando, quase **se mijando de medo**, o moleque o cutucou com a coronha do bacamarte [...]
That's when, almost **pissing of fear**, the boy nudged him with the butt of the blunderbuss [...] (to piss of fear = to have a lot of fear)

Corpus do Português, being a corpus consisted of more formal texts (cf. section 3) prevented the postulation of more broad generalizations about the habitat of the SCBE. Still, the data obtained allowed us to understand that SCBE is more pertinent to discursive contexts in which the speaker/writer has more freedom to express his subjectivity, since it is especially present in narrative sequences and dialogues (in fiction texts, 87.2% of its occurrence in the corpus used) and in excerpts of reports (other genres).

5.   Conclusion

It was our intention here to expose the corpus dimension involved in the research of SCBE. By doing so, we presented an effective form for investigating constructional patterns in a language and the advantages that a corpus-based approach can offer to researches investigating this kind of objects.

To form this framework, beyond a very brief presentation of the theories that underpin our way of looking at the object, we presented the method used in the research and also the findings directly related to the choice of use corpus in the work: the conventionalization and productivity of the SCBE in Portuguese, the description of the construction and the texts in which the construction appears.

The results show that, in fact, it is advantageous to use corpora in language research, not only for providing access to information inaccessible to introspection, but also to allow more precise descriptions, and actual, of a given object, since that arise naturally information data.
It is true that the use of corpus does not warrant a full analysis (in the study of the SCBE, for example, we found through the corpus research of common cases that we see in Portuguese, as "Pirar de rir", something like "freak out laughing"), but, as stated by Fillmore (1992: 35),

> "there can be any corpora, however large, that contain information about all of the areas […] that I want to explore; all that I have seen are inadequate. [But] every corpus that I've had a chance to examine, however small, has taught me facts that I couldn't imagine finding out about in any other way".

6.   Acknowledgments

7.   References

Aluísio, S.M., Almeida, G. M. (2006). O que é e como se constrói um *corpus*? Lições aprendidas na compilação de vários *corpora* para pesquisa lingüística. *Calidoscópio*, 4(3), pp.155--177.

Boas, H.C. (in press). Cognitive Construction Grammar. In G. Trousdale, T. Hoffmann (Eds.). *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

Costa, I.O. (2010). A Construção Superlativa de Expressão Corporal: uma abordagem construcionista. Dissertação de Mestrado em Linguística. Universidade Federal de Juiz de Fora, Juiz de Fora.

Croft, W., Cruse, A. (2004). *Cognitive Linguistics*. New York: Cambridge University Press.

Fauconnier, G. (1994). *Mental Spaces*. New York: Cambridge University Press.

Fauconnier, G., Turner, M. (2002). *The way we think*: conceptual blending and the mind's hidden complexities. New York: Basic Books.

Fillmore, C. (1982). Frame semantics. In Linguistic Society of Korea (Eds.). *Linguistics in the Morning Calm*: Selected Papers from SICOL-1981. Seoul, Hanshin, pp. 111--137.

Fillmore, C. (1992). "Corpus linguistics" vs. "computer-aided armchair linguistics". In *Proceedings from a 1991 Nobel Symposium on Corpus Linguistics*. Stockholm, Mouton de Gruyter, pp.35--66.

Goldberg, A. (1995). *Construction*: A construction grammar approach to argument structure. Chicago: The University of Chicago Press.

Goldberg, A. (2006). *Constructions at work*: The nature of generalization in language. Oxford: Oxford University Press.

Gries, S.T., Divjak, D. Behavioral profiles: A corpus-based approach to cognitive semantic analysis. In V. Evans, S. Pourcel (Eds.). *New directions in Cognitive Linguistics*. Amsterdam, Philadelphia: John Benjamins, pp.57--75.

Johnson, M. (1987). *The body in the mind*. Chicago: University of Chicago Press.

Lakoff, G. (1987). *Women, Fire and Dangerous Things*: What categories reveal about the mind. Chicago: The University of Chicago Press.

Lakoff, G., Johnson, M. (2002[1980]). *Metáforas da vida cotidiana*. Trad. Mara Sophia Zanotto (Ed.). Campinas: Mercado de Letras; São Paulo: Educ.

Lakoff, G., Johnson, M. (1999). *Philosophy in the Flesh*: The embodied mind and its challenge to western thought. New York: Basic Books.

Miranda, N.S. (2002). O caráter partilhado da construção da significação. *Veredas*, 5(2), pp. 57--81.

Miranda, N.S. (2008a). Construções Superlativas no Português do Brasil: um estudo sobre a semântica de escala. Projeto de pesquisa do Programa de Pós-Graduação em Letras – Mestrado em Linguística; GP "Gramática e Cognição", CNPq. Universidade Federal de Juiz de Fora, Juiz de Fora.

Miranda, N.S. (2008b). *Gramaticalização e gramática das construções*: algumas convergências. Um estudo de caso: as construções negativas superlativas de IPN. 2008. 110 f. Relatório de Pós-doutorado em Linguística. Universidade Presbiteriana Mackenzie, São Paulo.

Salomão, M.M.M. (1997). Gramática e interação: o enquadre programático da hipótese sócio-cognitiva sobre a linguagem. *Veredas*, 1(1), pp. 23--39.

Salomão, M.M.M. (2006). *Teorias da Linguagem*: A perspectiva sociocognitiva. Rio de Janeiro. Disponível em: <http://www.forumdelinguagem.com.br/textos/Texto%20Margarida%20Salom%C3 %A3o.pdf>. Acesso em: 05 out. 2008.

Sampaio, T.F. (2007). O uso metafórico do léxico da morte: uma abordagem sociocognitiva. Dissertação de Mestrado em Linguística. Universidade Federal de Juiz de Fora, Juiz de Fora.

Sardinha, T.B. (2004). *Lingüística de Corpus*. Barueri: Manole.

Stefanowitsch, A. (2006). Words and their metaphors: A corpus-based approach. In A. Stefanowitsch, S. Gries. (Eds.), *Corpus-based Approaches to Metaphor and Metonymy*. Berlin, New York: Mouton de Gruyter, pp. 61--105.

Past tense in Brazilian Portuguese: set of tense-aspect-modality features

Raquel Meister Ko. FREITAG

Universidade Federal de Sergipe, Centro de Educação e Ciências Humanas, Departamento de Letras Vernáculas
rkofreitag@uol.com.br

Abstract

In this paper, results from an investigation about the set of verbal features in Brazilian Portuguese are presented. Tense, aspect and modality features are described base on use of verbal forms in a sociolinguistic corpus of spoken Brazilian Portuguese. The verbal categories finding in the corpus are presented and the directions form > function and function > form. Results point that the IMP forms (simple and compound) are overlapping the most functions, specially the functions of modality domain, in *irrealis*.

Keywords: verbal categories; variation; Brazilian Portuguese.

1. Introduction

Normative grammars of Portuguese define the verbal paradigm as a tense: in the past scope there are the "pretérito perfeito" forms (simple and compound), "pretérito mais que perfeito" (simple and compound), "pretérito imperfeito" and future do pretérito), in indicative mode, and "pretérito imperfeito" in subjunctive mode. However descriptive and variacionist studies point that this forms pass for a) a semantic-discursive reset, with a single form expressing more than one function, losing the iconicity, and b) a morphosyntatic reset, with emergency and regularization of new forms and obsolescence of others. For example, there are evidences of obsolescence of simple "pretérito mais que perfeito" forms and the low frequency of compound "pretérito mais que perfeito" forms in context of anterior past; the simple "pretérito perfeito" forms assume this function (Coan, 1997). Other example is the emergency and regularization of form to expresses the imperfective progressive past, constituted by auxiliary verb "estar" + principal verb in gerund form, the compound "pretérito imperfeito" (Freitag, 2007). Still there are the switching between the "future do pretérito" and simple "pretérito imperfeito" forms (Costa, 1997), switching between "pretérito imperfeito" of indicative and subjunctive mode, and the specialization of compound "pretérito perfeito" form to expresses iterative perfect (Barbosa, 2008), and anymore. These switching contexts, emergency and regularization in verbal paradigm of Brazilian Portuguese are possibly due the reset processes of verbal paradigm, which origins are in the transition from Classical Latin to Vulgar Latin and to Romance languages. In this process language loses the aspectual distinction (''infectum'' and ''perfectum'' tenses), resulting in verbal paradigms in Romance languages that has an irregular paradigm as for the aspectual distinction. The emergency of compound forms, which codifies aspectual tense, is an evidence for this process.

In this paper, results from an investigation about the set of verbal features in Brazilian Portuguese are presented. Tense, aspect and modality features are described based on use's description of verbal forms in a sociolinguistic corpus of spoken Brazilian Portuguese (Banco de dados Falantes Cultos de Itabaiana/SE). The sociofuncionalist assumptions (Tavares, 2003) are adopted for the analysis: the emergency of forms (grammaticalization follows Bybee, Perkings and Pagliuca, 1994) and the use regularization (linguistic change follows Labov, 1972). This approach postulates that clines of linguistic change presuppose stages of more or left stability in system, in so far as there are overlapping functions for one form and/or overlapping forms for a single function. First, TAM domain is presented; follows forms and functions correlation is.

2. TAM Domain

To analysis, we assumed the postulate that verbal form accumulate the tense, aspect and modality (TAM) features, in a complex functional domain (Givón, 1995, 2001), in which the features interacting. The complexity of the functional domains is due the fact that the boundaries between each feature are not always clear or precise, locking the separation, in fact, of each feature. However to pick up nuances of emergency, switching and regularization processes must be analyzing the verbal features globally, observing the discursive features that locking or favor any verbal form in any contexts.

2.1 Tense

Tense notion refers at the ordaining events (experiences) in points and intervals at a sequence; this concept is based on Reichenbach (1947): verbal tenses are determined for the ordaining of event point in function of the reference point and speech point. Based on speech point is possible establish three basic temporal relations: past, tense and future. Fixate only one point allows diagraming only three temporal relations; but others two parameters – event point and reference point – amplifying the temporal possibilities. Event point is the point when the event occurs; and reference point is a parameter point, a temporal reference, to determinate the event point, which is established according to the speech point. The speech point becomes the reference point when there is not temporal reference contextually explicit.

2.2 Aspect

Aspect linguistic category refers at the different modes to perceive the internal tense of an event (Comrie, 1976). Aspectual notion involves the internal set tense in events

(initial, medial and final states/event presented as perfective/close or imperfective/open, and anymore possibilities). Perfective aspect is characterized for global perspective of event, which is expressed as closed, without internal reference, in a single united. Imperfective aspect focuses the internal constitution of events: its development (cursive, progressive imperfective aspect), or selecting stages of internal tense development (initial, medial and final), or expressing resultative states, and anymore. Imperfective aspect does not determine initial or final event points but focalizes its development, in contrast at perfective aspect, that emphasis the initial and the final points.

There is also other level of aspectuality: the inherent aspect of event. Bertinetto (2001) characterizes the event based on three aspectual proprieties: dynamicity durativity and homogeneity. Homogeneity refers at absence of inherent internal limit in any event: a [+ homogeneity] event is this that does not change its nature; yet [-homogeneity] event presents an inherent achievement point. Dynamicity is a propriety characterized according to observation of dynamic atoms, which corresponding at minimal granularity of event and hence these are not divisible indefinitely [+ dynamicity]; the statics atoms can be divisible indefinitely [- dynamicity]. Durativity is a concept strictly operational, since any event, for so soon as far, has certain duration; nevertheless is possible distinguished events whit duration [+ durativity] from instant events [- durativity].

2.3   Modality

Modality is usually defined as the grammaticalization of speaker attitudes as the propositional content. In the languages it possible recognizes a grammatical category (the modality) which is similar at tense, aspect, number, gender. Givón (1995) divides the modality in epistemic, which refers at truth, belief, probably, certainty and evidence, or deontic, which refer at preference, desire, intention, ability, obligation and manipulation.

Epistemic modalities from Aristotelian logic tradition, follows Givón, have communicative equivalents: at the necessary truth corresponds the communicative equivalent of presupposition; at factual truth corresponds the *realis* assertion; at possible truth corresponds the *irrealis* assertion; and at non truth correspond the negative assertion. The communicative redefinition of epistemic modalities takes the presupposition as a proposition assumes as truth for anterior concordance, cultural convention or obvious at all participants in context of interaction. *Realis* assertion takes a proposition strongly asserted as truth; *irrealis* assertion is a proposition strongly asserted as possible, probably or uncertain; negative assertion takes the presupposition strongly asserted as false, in contradiction with explicit or assumed belief by hearing.

3.   Prototypical tense features set in spoken Brazilian Portuguese

In a functionalist/cognitivist approach, the language structure reflects the experience structure, deriving from iconicity principle (cf. Bolinger, 1977; Givón, 1995). In a strong version of iconicity, model provides a one-to-one relation between form and function; however, in a moderate version the model provides the opacizition between codification and function, ant becomes possible the variation between forms and functions. In Brazilian Portuguese spoken the past tense domain presents non univocal relations between forms and functions: one single form codifies more than one function and one single function is codified by more than one form.

The verbal categories identified in *corpus* are presented, first in form > function approach and follow in function > form approach.

The mapping of *corpus* results the follow forms (in indicative mode):

- Simple "Pretérito Perfeito" (simple PP)
- Compound "Pretérito Perfeito (compound PP)
- Simple "Pretérito Imperfeito" (simple IMP)
- Compound "Pretérito Imperfeito" (compound IMP)
- Simple "Futuro do Pretérito" (simple FP)
- Compound "Futuro do Pretérito" (compound FP)
- Compound "Pretérito Mais que Perfeito" (compound +QP)

These forms codifying follows functions:

- Anterior past: a past event which reference is other past event;
- Iterative perfective past: a past event which occurs systematically to past into the present;
- Imperfective past: a past event which reference is other simultaneous past event;
- Perfective past: a past event which reference is the speech point;
- Habitual past: an irregular past event recurrent;
- Conditional past: an event due of other past event;
- Iminential past: an event which is presented at its pre-achievement.

Examples (1)-(12) illustrate the relation between forms and functions to expression of past tense in analysed corpus.

1) Inclusive conversei com alguns amigos meus que trabalham no escritório tal tudo e me ajudaram só a confirmar mesmo... que o curso era aquilo mesmo que eu já **ESTAVA ESPERANDO** *se ita mb lq 10*[1]

---

[1] The acronym in italics refers to source of data extrating from *Sociolingustic interview* sample from *Banco de dados Falantes Cultos de Itabaiana/SE*. Two first letters are the county (Sergipe) and the three follow letters are the city (Itabaiana); follow letters

'Also I talk with my friends which work in the office and they help me confirm the course was that even though I **WAS EXPECTING** (Compound IMP – Imperfective past)**'**

2) Olhe até ontem eu **ACHAVA** que seria um curso... né? que... dá as condições de emprego *se ita fp sq 02*
'Look until yesterday I **THOUGHT** (Simple IMP – Imperfective past) it would be a course... right? that... gives employment conditions'

3) Chegou um menino colega dele "me dê aí um geladinho" ele... "vá lá pegar por favor" ele foi pegar quando ele **ABRIU** a geladeira que **PEGOU** o geladinho *se ita mbh 08*
'Arrived a boy his colleague "Give me a chilled" he ... "Please come pick up" when he was caught he **OPENED** (Simple PP – anterior past) the fridge that **TOOK** (Simple PP – perfective past) the chilled'

4) Uma vez meu colega me **CONTOU** que a mãe dele **TINHA IDO** para a rua *se ita mbh 08*
'Upon time my friend **TOLD** (Simple PP – perfective past) me that yours mother **WENT** (compound PP – anterior past) out'

5) Se eu me formasse e visse que não que eu não dava pra ensinar que não era o meu ramo... eu não **FARIA**... eu não **EXERCIA** a profissão melhor dizendo *se ita fp sq 02*
'If I graduated and I see that I could not to teach because it was not my business ... I did not **DO** (Simple FP – conditional past) ... I did not **PURSUE** (Simple IMP – conditional past) the profession rather'

6) Se a prova trouxesse questões desse tipo questões relacionadas ao dia-a-dia das pessoas questões problemas todos os professores de escolas particulares **IAM** se **ADAPTAR** também né? *se ita mb sq 09*
'If the test brought issues matters to the day-to-day problems of people questions all private school teachers **WOULD ADAPT** (Compoud FP – conditional past) also right?

7) Ele achava que sendo universitário já era algo a mais que **IA ACRESCENTAR** no currículo dele *se ita mb lq 10*
'He thought that being university student was already something else that **WOULD ADD** (Compound FP – iminential past) to his resume'

8) Desde a oitava série do ensino fundamental eu já

tinha certeza de que a minha carreira seria na área da computação eu **ENXERGUEI** a área de tecnologia em geral como uma área bastante promissora e eu estava certo *se ita mp sI 01*
'Since eigth middle school I yet had 'certain that my career would be in computation area I **SAW** (Simple PP – iterative perfective past) the technological area as a promissory area and I was right'

9) Eu acho que eu vou conseguir colher os frutos que eu **TENHO PLANEJADO** *se ita mp sl 01*
'I think I will get to reap the fruits I **HAVE PLANNED** (Compound PP – iterative perfective past)**'**

10) Bom... eu pensei que o curso **SERIA** um curso voltado pra formação de professores né? *se ita mb sq 08*
'Well I guess the course **WOULD BE** (Simple FP – iminential past) a course to teacher formation right?'

11) É preciso saber escrever muito bem no idioma inglês e no seu próprio idioma inclusive pessoas de outros países a Google **COSTUMAVA** também contratar para fazer as traduções *se ita mp lq 10*
'You need to know how to write well in English and in your own language also people from other countries Google **USED HIRE** (Simple IMP – habitual past) to do the translations'

12) Como foi uma turma que **sempre ESTEVE ENVOLVIDA**... eu vejo que uma grande parte... né? está... realmente pensando e já criando os seus projetos... né? *se ita fp sq 02*
'As was a class that was **always WAS INVOLVED** (Compound IMP – habitual past)... I see that a large part ... right? is ... really thinking and already creating their projects... right?'

---

are the sex (F = feminine and M = masculine), age (P = 16 at 20 year old; B = 26 at 35 years old); school grading (S = college completed; B college in course) and the last numbers refer to informant identification.

Function	Temporal arrangement	Interval	Grammatical aspect	Inherent aspect	Modality	Forms
Anterior past	EP – RP – SP	-	-		*Realis*	Simple +QP Compound +QP Simple PP
Iterative perfective past	EP – SP, RP	Determinate	Perfective		*Realis*	Compound PP Simple PP
Perfective past	EP – SP, RP	-			*Realis*	Simple PP
Imperfective past	EP,RP – SP	Determinate	Imperfective		*Realis*	Simple IMP Compound IMP
Habitual past	EP,RP – SP	Indeterminate	Imperfective		*Realis/irrealis*	Simple IMP
Iminential past	EP,RP – SP		Imperfective inceptive/terminative	[- homogeneous]	*Irrealis*	Simple FP Compound FP Simple IMP Compound IMP
Conditional past	RP – SP – EP RP – EP – SP	-	-		*Irrealis*	Simple FP Compound FP Simple IMP Compound IMP

Table 1: Set of tense-aspect-modality

Each form and each function are analyzed separately ina a quantitative approach and after the general results was correlated, as in table 1. This summarization is based on the studies about these verbal categories in the corpus of "Variation in expression of past tense: concurrent functions and forms" project researchers' papers: Araujo & Freitag (2010, 2012), Cardoso & Santos (2011), Freitag & Araujo (2011), Freitag (2011), Freitag, Santos & Araujo (2011).

Results showed at table 1 point that the IMP forms (simple and compound) are polysemy, recovering a range of functions of imperfective aspect and *irrealis* modality. In perfective aspect, the actual verbal paradigm points the obsolescence of simple "pretérito mais que perfeito" form and the low productivity of compound "pretérito mais que perfeito" form; this form occurs in context of counter factuality. The realignment of verbal paradigm follow the specialization of forms based on distinction simple/compound: the IMP forms are distributed according the tendency simple IMP > habitual past and compound IMP > imperfective past.

The correlation between forms and TAM set contributes to elucidate the clines of grammaticalization of semantic-discursive functions which the verbal forms codify; these results contribute to the refinement of the theoretical model. The analyses also subsides the application in tagger *corpus* processes.

4. Conclusion

Empirical analysis of linguistic change phenomena in different grammatical levels provides reflections about the theoretical models of grammaticalization, and contributes to point the limits and limitations of theory, reinforcing interface approaches. If at first time the grammaticalization studies focus the design of clines change of constructions (forms), actually the functional domains (function) has been highlight also at object of investigation. In verbal categories domain this approach has been showed productive and evidencing the need of more studies to priming the model.

6. References

Araujo, A.S., Freitag, R.M.K. (2012). O funcionamento dos planos discursivos em textos narrativos e pinativos: um estudo da atuação do domínio aspectual. In *Signum. Estudos de Linguagem*, 15 (1), pp. 57--76. Available at: <http://www.uel.br/revistas/uel/index.php/signum/article/viewFile/11500/11174>.

Barbosa, J.B. (2008) Tenho feito/fiz a tese uma proposta de caracterização do Pretérito Perfeito no Português. Tese (Doutorado em Linguística e Língua Portuguesa). Universidade Estadual Paulista Júlio de Mesquita Filho.

Bertinetto, P. M. (2001). On a frequent misunderstanding in the temporal-aspectual domain: the 'perfective-telic confusion. In C. Cecchetto, G. Chierchia and M.T. Gausti (Eds.). *Semantic interfaces: reference, anaphora and aspect*. Stanford, CSLI, pp.177--210.

Bolinger, D. (1977). *Meanig and form*. London, Longman.

Bybee, J., Perkings, R. and Pagliuca, W. *The evolution of grammar: tense, aspect, and modality in the language of the world*. Chicago: The University of Chicago Press, 1994.

Cardoso, B.T., Santos, J.L.C. (2011). Variação na do tempo verbal passado na fala e escrita de Itabaiana/SE: formas de pretérito perfeito simples e pretérito perfeito composto na expressão do passado perfectivo iterativo. In *Littera Online*, 4 (2), pp. 22--42. Available at: <http://www.periodicoseletronicos.ufma.br/index.php/littera/article/viewFile/755/472>.

Coan, M. (1997). Anterioridade a um ponto de referência passado: pretérito (mais que) perfeito. Dissertação

(Mestrado em Linguística) – Programa de Pós-graduação em Linguística da Universidade Federal de Santa Catarina.

Comrie, B. (1976). *Aspect*. Cambridge, Cambridge University Press.

Costa, A.L. (1997) A variação entre formas de futuro do pretérito e de pretérito imperfeito no português informal no Rio de Janeiro. Dissertação (Mestrado em Linguística) – Programa de Pós-graduação em Letras/Linguística da Universidade Federal do Rio de Janeiro.

Freitag, R.M.K., Araujo, A.S. (2011). O passado condicional: formas e contextos de uso. In *Caligrama*, v. 16, pp. 199--228. Available at: <http://www.periodicos.letras.ufmg.br/index.php/caligrama/article/view/1624>.

Freitag, R.M.K. (2007). A expressão do passado imperfectivo no português: variação/gramaticalização e mudança. Tese (Doutorado em Linguística). Programa de Pós-graduação em Linguística da Universidade Federal de Santa Catarina.

Freitag, R.M.K. (2010). A expressão do passado iminencial em português: formas e contextos de uso. In *Anais do VII Congresso Internacional da Abralin*. Curitiba: Universidade Federal do Paraná, pp. 3654--3662. Available at: <http://www.abralin.org/abralin11_cdrom/artigos/Raquel_Freitag.PDF>.

Freitag, R.M.K. (2010). O domínio funcional tempo-aspecto-modalidade na expressão do passado imperfectivo no português falado no Brasil. In *Revista do GEL*, 7 (1), pp. 139--170. Available at: <http://www.gel.org.br/revistadogel/volumes/7/gel_7.2_t06.pdf>.

Freitag, R.M.K. (2011). Trajetórias de mudança do passado imperfectivo no português: entre o aspecto e a modalidade. In *Veredas*, 15 (1), pp. 148-163. Available at: <http://www.ufjf.br/revistaveredas/files/2011/05/ARTIGO-111.pdf>.

Freitag, R.M.K., Santos, A.M. and Araujo, A.S. (2011). O efeito gatilho e a continuidade tópica: atuação do domínio tempo - aspecto - modalidade. In *Signótica*, 23 (1), pp. 247--265. Available at: <http://www.revistas.ufg.br/index.php/sig/article/view/15221/10394>.

Givón, T. (1995). *Functionalism and grammar*. Amsterdam/Philadelphia, John Benjamins Publishing.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, University of Pennsylvania Press.

Reichembach, H. (1947). *Elements of symbolic logic*. New York, The MacMillan Company.

Tavares, M.A. (2003). A gramaticalização de E, AÍ, DAÍ, e ENTÃO: estratificação/variação e mudança no domínio funcional da sequenciação retroativo-propulsora de informações – um estudo sociofuncionalista. Tese (Doutorado em Linguística) Programa de Pós-Graduação em Linguística da Universidade Federal de Santa Catarina.

7. Appendix
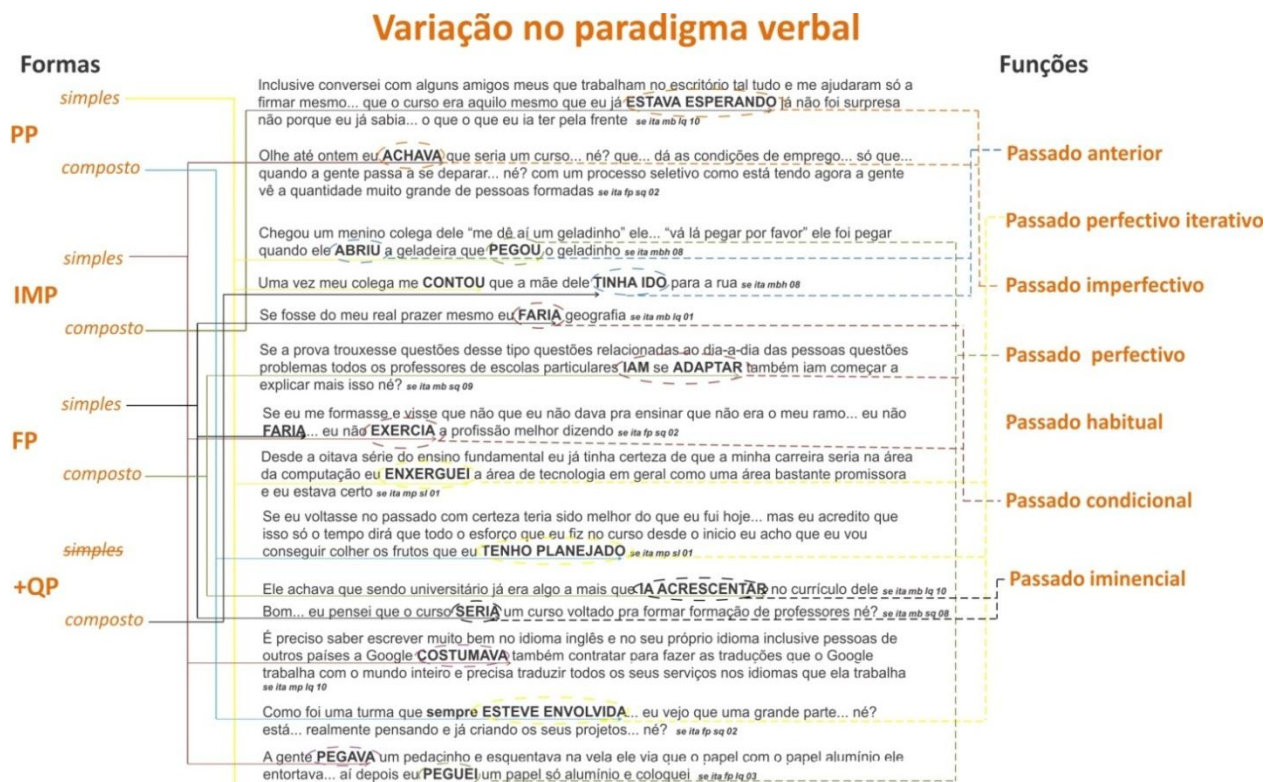
Figure 1: Form and function relations in past tense domain in spoken Portuguese

Lexical and grammatical features of spoken and written Japanese in contrast: exploring a lexical profiling approach to comparing spoken and written corpora

Itsuko FUJIMURA, Shoju CHIBA, Mieko OHSO

Nagoya University; Reitaku University; Nagoya University

Furo-cho, Chkusa-ku, Nagoya, Japan

fujimura@nagoya-u.jp, schiba@reitaku-u.ac.jp, ohsomk@ac.auone-net.jp

This paper statistically demonstrates the lexical and grammatical characteristics of conversational Japanese by comparing a 100 hour spontaneous spoken corpus: the NUCC (Nagoya University Conversation Corpus) with a written corpus: the Balanced Corpus of Contemporary Written Japanese (monitor version). 1) The conversation corpus contains more involved production than the compared written corpus. 2) The comparison between the spoken and written interactional corpora shows that the participants leave much more metalinguistic and illocutionary traces in their speech than their writing. This is explained by the difference of degree of elaboration of the emitted messages and the difference of degree of closeness between/among participants of exchanges. 3) Fragmented utterances are much more frequent in spoken conversation than written texts. In Japanese, because of its grammatical structure (=SOV type language; particles come after their head), fragmentation, omnipresent conversational phenomenon, easily causes a functional and grammatical change in the role of particles.

Keywords: conversation; internet exchanges; metalinguistic; norm; linguistic change; Japanese; fragmentation.

1. Introduction

In this paper, we describe the lexical and grammatical characteristics of Japanese face-to-face spoken conversation and show how they differ from written registers. The aim of this research is to elucidate the characteristics of spoken Japanese, so we can later compare them with the results piled in the literature of this domain (Blanche-Benveniste, 1990; Biber, 1995 among others). For this purpose, we compare a spoken corpus: the NUCC (Nagoya University Conversation Corpus) with a written corpus: the BCCWJ (Balanced Corpus of Contemporary Written Japanese, monitor version). The former is a corpus of 100 hours built by our research team. The latter is a 45 million morpheme-sized written corpus. Our method is mainly quantitative. We perform this research with a tool named Lexical Profiling System, devised by one of the co-authors of this paper.

2. Corpora and tool

2.1 NUCC

The NUCC was constructed between 2001 and 2003, and is available for research purposes from the site (https://dbms.ninjal.ac.jp/nuc/index.php?mode=viewnuc) free of charge. It is composed of transcriptions of 129 uncontrolled, natural conversations between or among friends, family members or colleagues. Each conversation has 2 to 4 participants and lasts 30 to 60 minutes. The participants are 198 native speakers of Japanese of various ages and from diverse academic backgrounds. Each conversation constitutes a file so that the corpus NUCC consists of 129 files.

Conversations were recorded and transcribed in standard Japanese orthography. The Japanese orthography currently used is quite phonemic, but suprasegmental features are not captured. Hence, accent, intonation, and prominence are not transcribed. Only the rising intonation that indicates questioning is marked with a question mark at the end of an utterance.

The corpus contains about 1.5 million morphemes ("short unit words" according to UniDic (cf. Ogiso *et al.*, 2012)), which shows that this is the largest corpus currently available of spontaneous spoken Japanese. As a caveat, there are more female participants (161) than male (37), and many of the participants are graduate students majoring in linguistic subjects. The lack of balance of the participants may be reflected in the data taken from this corpus.

2.2 BCCWJ (monitor version) [1]

The integral BCCWJ, published in 2012, includes about 170,000 samples of written texts, which are classified into carefully designed subcorpora (genres), namely books, newspapers, magazines, whitepaper texts, Internet texts, Diet minutes, among others. We see the BCCWJ as a good sample of written Japanese, because the corpus contains the samples from many genres, each of which is relatively large. It also utilizes unique sampling strategies so that the corpus represents the most recent status of contemporary written Japanese (Maekawa, 2007).

In this work, we used the monitor version of the BCCWJ earlier released in 2009, which is a part of the integral version. The monitor version consists of 4 subcorpora indicated in Table 1. We use the BCCWJ in two ways. One is the whole BCCWJ (monitor version) for the grammatical study in section 4, and the other, its subcorpora: Books (BK) and Internet Bulletin Boards (IBB) for the lexical studies in section 3. The BK is composed of 10423 samples taken from various genre of books published between 1971-2005. We used it because it is the largest part of the BCCWJ and for its standardized nature as written corpus. The IBB consists of "Questions and Answers" type written exchanges between anonymous writers and readers, published on Yahoo Japan's web site in 2005. The IBB is an interesting material to compare with the NUCC, because of their shared characteristics and for its novelty as a medium of communication. Both of them involve interaction

---

[1] Cf: http://www.ninjal.ac.jp/english/products/bccwj/. The BCCWJ refers to the BCCWJ (monitor version) from section 3 below.

between/among participants.   The relation between/among participants is different though; the participants in the latter have close relationships while those in the former are strangers. They made real-time interactions in the latter, while there is a time lag between questions and answers in the former.

Table 1 indicates the characteristics of the studied corpora.

Subcorpus of BCCWJ and NUCC	Number of morphemes (millions)	Characteristics
Books (BK)	36.0	No interaction Elaborated production
White Paper	5.8	
Internet Bulletin Boards (IBB)	6.7	Long-distance interaction Prepared production
Minutes of the National Diet	5.5	
NUCC	1.5	Close interaction Real-time production

Table 1: Subcorpora of the BCCWJ (monitor version) and the NUCC

2.3   Lexical profiling system

The Lexical Profiling System is designed to compare corpora of different size, genre, or even an individual part of a corpus with the whole. The data to be compared are morphologically analyzed by a GUI program Chamame (ver. 1.71) (composed by a part-of-speech and morphological analyzer: Mecab (ver. 0.98) and a dictionary: UniDic (ver. 1.3.12)), and the frequency of lemmas, word forms, bigrams are counted and stored in a database. The tool then computes the frequencies of these units using different statistical measures such as LLR (Log-Likelihood Ratio) among others.

3.   Lexical studies

3.1   60 Basic morphemes in the NUCC

First of all, we identified the 60 morphemes employed in all 129 conversations of the NUCC as in Table 2 in order to compare later the use of these morphemes in the NUCC and the IBB and the BK. We could say that these are basic morphemes of Japanese conversation. These consist of 6 adjectives, 4 adverbs, 1 conjunction, 4 interjections, 6 nouns, 18 particles, 1 prefix, 2 pronouns and 12 verbs[2]. Among the 18 particles, there are 4 utterance-final interactional particles, 13 sentence-internal casual or conjunctive particles and "*no*". "*No*", one of the most frequently used morphemes in Japanese, is

subcategorized into three according to the dictionary UniDic: genitive (*of* in English), quasi-nominal (*thing*, nominalizer) and interactional.   The first two are sentence-internal particles and the last one, utterance-final particle.

POS	No	Morpheme
ADJ	6	*nai (not to exist), yoi (good), you (to look like), sugoi (superb), sonna (that kind of), sono (that)*
ADV	4	*mou (already), dou (how), sou (so, in such a way), kou (this way)*
AUX	6	*da, desu* (DEC)*, reru* (PASS/POT/HON)*, ta* (PAST)*, nai* (NEG)*, teru* (PROG, PERF)
CONJ	1	*de (and)*
INTJ	4	*un (yeah, I see), ah , a! (wow),ano (well)*
NOUN	6	*koto (matter), hito (person), toki (time, when), hou (side), ato (behind, afterward), mono (thing)*
PRT	18	Utterance-final, interactional: *ne* (TAGQ*, you know)*, *yo (I tell you), ka* (Q)*, na (I tell you)* Sentence–internal: *wo* (ACC)*, ga* (SUB)*,wa* (TOP)*, ni* (DAT, LOC, TEMP, ADVL)*, to (and with), keredo (although), kara (from), mo (also), kurai (about) te, de (and* (V/ADJ Suffix*)) tte* (QUO)*, made (until), no*: GEN,QN (sentence-internal), INTA (utterance-final)
PREFIX	1	*o* (POLITE)
PRO	2	*nani (what), sore (that)*
VERB	12	*iru (to exist, to be), dekiru (to be able to), miru (to see, to look at), naru (to become), wakaru (to understand), omou (to think), aru (to exist), kuru (to come), suru (to do), yaru (to do), iku (to go), iu (to say)*
total	60	

Table 2: 60 Morphemes used in all 129 conversations of the NUCC[3]

The fact that there are no personal pronouns in the list should not be interpreted as lack of active interaction. In Japanese, one can speak even for 30 minutes long without mentioning "me" or "you". Especially the

---

[2] These are the output of the Analyzer Chamame. We only modified the result of the automatic analysis by grouping "Rentai-shi", "Keijo-shi" and "Keiyo-shi" in Adjective, since the major function of these three categories is noun modification.

[3] Glosses are approximate due to lack of space. The list of abbreviations is following. ADJ: Adjective, ADV: Adverb, ADVL: Adverbial, ACC: Accusative, AUX: Auxiliary, CONJ: Conjunction, DAT: Dative, DEC: Declarative, HON: Honorific, INTJ: Interjection, INTA: Interactional, NEG: Negation, GEN: Genitive, PASS: Passive, PAST: Past Tense, PERF: Perfect, POT: Potential,   PRO: Pronoun, PROG: Progressive, SUB: Subject, TAGQ: Tag-Question, Q: Question, TEMP: Temporal, QN: Quasi-Nominal, TOP: Topic, PRT: Particle, QUO: Quotation, V: Verb.

reference to the interlocutor with a personal pronoun meaning "you" is considered to be rude. The frequent uses of interactional particles like *ne*, *yo*, deictic verbs like *iku* (*to go*), *kuru* (*to come*) and honorific expressions fill the gap caused by the lack of personal pronouns.

3.2 NUCC compared with Books (BK)

The statistic measure: LLR demonstrates the degree of typicality for these 60 morphemes compared with the BK. Even if they are used in every conversation of the NUCC, their degree of typicality is not homogeneous. The most typical 10 morphemes relative to the BK with the highest degree of LLR and the least typical 5 with the lowest degree of LLR are shown in Table 3. The MPM indicates the number of morphemes per million.

no	Morpheme	Function	LLR	MPM
1	*un*	*Yeah, I see*	310,539	30,003
2	*ne*	TAGQ,	127,327	19,754
3	*tte*	QUO (contracted)	80,628	12,575
4	*ka*	Q	67,541	22,884
5	*teru*	PROG/PERF (contracted)	59,022	9,714
6	*sou*	*so*	51,485	11,024
7	*yo*	*I tell you*	44,561	9,790
8	*nani*	*what*	39,340	9,820
9	*keredo*	*although*	36,307	6,436
10	*a!*	INTJ	36,090	4,273
...	...	...	…	…
56	*suru*	*to do*	-2,899	14,343
57	*wa*	TOP	-4,030	25,419
58	*ni*	IO etc.	-4,301	29,498
59	*iru*	to exist, to be	-6,440	1,200
60	*wo*	ACC	-20,037	3,939

Table 3: Typical and atypical morphemes in the NUCC compared with the BK

We can easily see that interactional expressions and contracted forms are typical in face-to-face conversation. The backchannel *un* appears 30,000 times par million. This is 3% of the morphemes used in the NUCC. In contrast, the least typical 5 are indispensable grammatical morphemes in any Japanese utterance regardless of spoken or written. Negative value means that the morpheme is less used in the conversation than in books. In fact, the least typical morpheme with the lowest degree of the LLR, the accusative marker "*wo*" is often not pronounced in conversation.

3.3 NUCC compared with the IBB

We then compare the uses of these 60 morphemes in the NUCC with the IBB in order to show the difference in spoken and written interactional exchanges. These interactions are characterized by two points of view:

social closeness and physical distance between two participants of communication.

3.3.1. Typical Morphemes

The most typical 10 morphemes of the NUCC compared with the IBB are following (LLR is in bracket).

1. *un* *yeah, I see*   (324,691)
2. *da* DEC (159,975)
3. *ne* TAGQ, *you know*   (146,670)
4. *no/n* GEN, QN or INTA[4]   (108,044)
5. *ka* Q   (101,483)
6. *sou*   *so, in such a way*   (95,564)
7. *tte* QUO (contracted)   (85,429)
8. *ta* PAST   (75,684)
9. *nani*   *what*   (67,687)
10. *iu*   *to say*   (61,961)

The high frequency of *da* (declarative marker) is noteworthy. Its occurrence seems to derive from the frequent use of short turn taking in face-to-face conversation, especially the large number of casual backchannel feedback finishing with "*da*", such as "*sou-na-n-da*" (*so*-DEC-QN-DEC, *"Indeed"*), whereas this is not the case in written correspondence on the Internet. The participants are not in real-time interactions in "Questions and Answers" type exchanges, so that the frequent use of short turn taking is not common. Also the participants of the IBB do not have a close relationship between them, because in fact they do not know each other and in general the written communication does not allow them to make intimate interactions in Japanese. These are the reasons for which the informal declarative form "*da*" is typical in the NUCC, whereas the formal one "*desu*" is numerous in the IBB.

3.3.2. Verb: *To Say* in the Conversation

Among the 12 verbs in the Table 1, "*iu*" (*to say*) is the most typical one of the NUCC with LLR: 61,961, followed by *iku* (*to go,* LLR: 20,919), *yaru* (*to do,* LLR: 17,603), *suru* (*to do,* LLR: 14,343), *kuru* (*to come,* LLR: 13,558), *aru* (*to exist,* LLR: 12,403), *omou* (*to think,* LLR: 10,903), *wakaru* (*to understand,* LLR:8,613), *naru* (*to become,* LLR: 5,970), *miru* (*to see, to look at,* LLR: 5,599), *dekiru* (*to be able to,* LLR: 1,489) and *iru* (*to exist, to be,* LLR: 1,200) in descending order. This metalinguistic verb *to say* is used much more often in oral conversation than in written correspondence. It may be explained at least partially by the fact that in real-time exchanges, we talk a lot about "how to say" something. The speaker leaves traces of metalinguistic activity in his speech. For example, when we hesitate in seeking an expression, we say: *"How should I say?"*. In the example

---

[4] The occurrence of numerous "*no*" in conversation primarily comes from the frequent use of the interactional usage of this morpheme placed at the end of utterances. However there are also many "*no*" placed before the declarative "*da*" often realized "*n-da*". This frequently used bigram is often analyzed as a compound auxiliary in Japanese linguistics. This is not the case in this study, as to our morphological analyzer processes them as QN-DEC.

(1), having once used the word "*room*", the speaker corrects it with the word "*entrance*" while talking about the process of this correction: *heya-tte-iu-ka* (*Can-I say "room"?*). In this type of metalinguistic utterance, the verb: *to say* plays the main role.

(Ex.1) conversation 019

Gozenchu-wa zuutto   heya-ni
morning-TOP throughout room-LOC
heya-tte-<u>IU</u>-ka   genkan-ni   haitte-ta-n-da
room-QUO-<u>SAY</u>-Q entrance-LOC
enter-PAST-QN-DEC
 "I was in a room all morning, can-I <u>SAY</u> "room"?, in the entrance. "

    In contrast, in the activity of writing, even private texts like those found in the IBB are prepared and elaborated. That would be why there is a big gap in the use of the verb: *to say* between the IBB and the NUCC.

4.   Grammatical study: fragmentation

Finally, we will discuss how to end an utterance in Japanese conversation.

4.1   13 basic utterance-final morphemes in the NUCC compared with the BCCWJ

We analyze 13 morphemes employed at the utterance-final position in all 129 conversations of the NUCC. This position is defined by a period or a question mark in the transcription. We can consider these 13 items as the basic utterance-final morphemes in Japanese informal face-to-face exchanges. The Table 4 indicates that when compared with the BCCWJ, the most typical utterance-final morpheme of the NUCC is the interactional particle: "*ne*", while the least typical one is the auxiliary: "*ta* (Past Tense)".

    These are classified into three groups. The first includes 4 final interactional particles (Final PRT): "*ne, yo, na, ka*". The second, 3 auxiliaries (AUX): "*da, nai, ta*" and the third, 6 sentence-internal conjunctive particles (PRT): "*te, keredo(kedo), tte, kara, de, ni*" as indicates the Table 4.

    Of these three groups, the frequent use of interactional particles in conversation is entirely predictable. The normal position of these morphemes is at the end of utterances. The use of auxiliaries at the final position is also ordinary in every type of text. The most interesting phenomenon is the use of sentence-internal conjunctive particles at the utterance-final position. It is not normative in Japanese traditional grammar and absent in the written formal texts, while it is found in every conversation of the NUCC.

POS	morpheme	function	LLR
Final PRT	*ne*	TAGQ, Alignment	55,092
PRT	*te*	*and*	22,516
PRT	*keredo(kedo)*	*although*	14,129
PRT	*tte*	QUO	13,949
Final PRT	*yo*	*I tell you*	12,305
Final PRT	*na*	*I tell you, I know*	10,520
PRT	*kara*	*because*	7,526
PRT	*de*	*and*	6,583
Final PRT	*ka*	Q	6,329
PRT	*ni*	DAT, LOC, TEMP, ADVL	4,672
AUX	*da*	DEC	1,027
AUX	*nai*	NEG	270
AUX	*ta*	PAST	-7,774

Table 4: LLR of final morphemes of the NUCC compared with the BCCWJ

4.2   From sentence-internal particle to utterance-final particle or vice versa

We could say first that there are many syntactically incomplete sentences in Japanese conversation as in other languages[5] This could be due to the pragmatics of conversation: the participants of communication collaborate to finish a sentence as in example (2). The utterance of the speaker A stops at the end of the subordinate clause marked by an adversative conjunction *KEDO* (=*KEREDO* "although"). The speaker B completes A's utterance by adding the main clause.

(Ex.2) conversation 035

A: sensei-ni   mikkahodo   tomatte-morae-ba
   professor-IO several days   stay-make-if
    ii-n-desu <u>KEDO</u>.
   good-QN-DEC(formal) <u>ALTHOUGH</u>
   "Although it would be better if we could ask the professor stay here for several days."
B: A! deki-nai-n-desu-ka.
   ah   can-NEG-QN-DEC(formal)-Q
   "Ah, you can not do so."

    However in most cases, this kind of collaboration between the participants of conversation is not obvious. The particle at the end of the utterance no longer has the conjunctive function linking the subordinate and main clauses but rather has a modal function. The example 3 shows that the utterance emitted by speaker B does not adversative with that of speaker A, despite the existence of *KEDO*. The function of *KEDO* in this case is to attenuate the assertive power of the predication and to show the intention of continuing the dialogue to the interlocutor (*cf. Saegusa, 2007*).

---

[5] Syntactic fragmentation does not necessarily correspond to informational fragmentation (cf. Matsumoto 2010).

(Ex.3) conversation 092

A: dou-iu-hanashi?
   how-say story
   "what story?"
B: tabun   shi-ta-to-omou-n-da <u>KEDO</u>.
   Perhaps do-PAST-QUO-think-QN-DEC
   <u>ALTHOUGH</u>
   "Perhaps I have already spoken to you about.
   <u>KEDO</u>."
A: jaa, kika-nai-wa.
   .so ask-NEG-PRT
   "So I will not ask you."

In written normative texts, these morphemes have only one conjunctive function, while having two in conversational discourse.

This phenomenon could be viewed from a diachronic point of view. In Japanese, a SOV type language, particles are placed after their head, either conjunctives or interactionals. The resulting fragmentation can easily cause a functional and grammatical change in the role of particles. We could say first that these sentence-internal particles create new interactional functions in conversation. This is the direction from the norm to usages. However we could also point out the opposite direction: from usages to the norm in written texts. In standard written Japanese the interactional use of these particles may be put aside, while they always remain in conversation. Figure 1 indicates these two directions. This issue deserves a full review. It would be interesting to consider this question within the Macro-Syntaxe analytical framework (Blanche-Benveniste, 1990).

Subordinate+Conjunctive PRT *Nomi-tai + KEREDO (KEDO)* *I want to drink + Although*	Principal *Noma-nai* *I do not drink*

Principal *Nomi-tai* *I want to drink*	Final PRT *KEREDO* (*KEDO*) Attenuation+Continuation
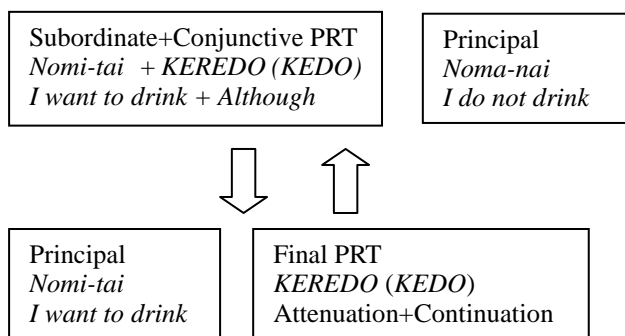
Figure 1: Linguistic change from sentence-internal PRT to utterance-final PRT or vice versa

5. Conclusion

Having compared the NUCC with the BCCWJ, several lexical and grammatical characteristics of Japanese conversation have been recognized.

1) 60 basic morphemes of spoken Japanese are identified. Personal pronouns are not included in the list. This is explained by the grammatical characteristics of the language.
2) Typical morphemes of conversation: interactional particles, interjections, markers of agreement and "*what*", reflect the involved nature of this activity, when compared with books.
3) The typical auxiliary of conversation, compared with written correspondence, is "*da* (declarative)". It may reflect the high frequency of short answers and backchannels in conversation.
4) The typical verb in conversation is "*iu (to say)*". This could come from frequent metalinguistic use of this verb in spontaneous speech, which, unlike written discourse, is not elaborated.
5) 13 basic utterance ending forms within conversation have been identified. Some of them are only used at the sentence-internal position in written texts. This is due to close and frequent exchanges between participants which cause incomplete utterances. In Japanese, because of its grammatical structure the fragmentation easily causes a functional and grammatical change in the role of particles.

Lastly, we summarize some of the features of conversational Japanese in contrast with written Japanese. It has more involved production, more metalinguistic and illocutionary traces. It also has more fragmented structures, which could cause a dynamic linguistic change. These are universal characteristics of spoken exchanges mentioned in Biber (1995), primarily due to the lack of time in real-time interactions (Biber, 2010) and secondarily to the closeness between two participants during exchanges. We also found some specific characteristics of Japanese conversation, like the absence of personal pronouns. This is explained only by the individual language structure.

6. Acknowledgements

7. References

Biber, D. (1995). Dimensions of Register Variation: A Cross-Linguistic Comparison. Cambridge: Cambridge University Press

Biber, D. (2010). Linguistic Styles Enabled by the Technology of Literacy. In M. Moneglia, A. Panunzi (Eds.), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Firenze: Firenze University Press.

Blanche-Benveniste, Cl. (1990). *Le français parlé, Études grammaticale*. Paris: Editions du CNRS.

Maekawa, K. (2007). Design of a balanced corpus of contemporary written Japanese. In *Proceedings of Symposium on Large-Scale Knowledge Resources. (LKR2007)*, pp.55--58.

Matsumoto K. (2000). Japanese intonation units and syntactic structure, *Studies in Language*, 24(3), pp.515--564.

Ogiso, T., Komachi, M., Den, Y. and Matsumoto, Y. (2012). UniDic for Early Middle Japanese: a Dictionary for Morphological Analysis of Classical Japanese. In *LREC 2012 Proceedings*. Available at: <http://www.lrec-conf.org/proceedings/lrec2012/index.html>.

Saegusa, R. (2007). Usage of GA and KEREDO in Spoken Japanese. In *Center for Student Exchange journal,*   10,   pp.11--27.   Available   at: <http://hdl.handle.net/10086/14360> (in Japanese).

In search of modality: a spontaneous speech corpus-based study

Heliana MELLO, Luciana ÁVILA, Priscila OSÓRIO, Raíssa CAETANO, Adriana RAMOS

Universidade Federal de Minas Gerais, Faculdade de Letras – UFMG

Av. Antônio Carlos, 6627 – Pampulha – Belo Horizonte, MG 31270-901 Brazil

heliana.mello@gmail.com

Abstract

Modality in speech can be taken to be a speaker's evaluation of an uttered locutive material. This paper explores the semantic notion of modality through the analysis of a Brazilian Portuguese spontaneous speech corpus. The building of the corpus took into account the utterance unit, as it is proposed in the Language into Act Theory (Cresti, 2000). This paper aims at briefly presenting modality studies developed so far within the C-ORAL-BRASIL corpus. The studies presented in this paper focus on: the identification of morpholexical modality indexes in tone units, a comparative study between modal adverbs of certainty in a sample of Brazilian and European spontaneous speech corpora and the mapping of modal adverbial constructions in Brazilian Portuguese. In all these studies, we carried a qualitative analysis, in order to describe the occurrences of the different modal indexes, such as for example: (semi-)auxiliary modal verbs, modal adverbs, verbs of propositional attitude, volitional verbs, modal adjective constructions and emerging forms.

Keywords: modality; C-ORAL-BRASIL; corpus-based research; spoken Brazilian Portuguese.

1.   What is modality?

Modality in speech can be taken to be a speaker's evaluation of an uttered locutive material following the Ballyan view that modality is the evaluation ("*Modus*") of the speaker towards his own locutionary content ("*Dictum*") (Bally, 1932). However, precisely defining this category is a difficult task, since, according to Venn (1888: 245), "[modality is] [a] variety of place upon that most thorny and repulsive of districts in the logical territory." This difficulty stems from different factors: (a) in its study tradition, modality has been the subject matter of both logical studies and natural language studies (Lyons, 1977), which implies a methodological maze not always productive for the research on its actual linguistic use; (b) this category interrelates with a number of grammatical phenomena such as time, aspect and mood (Palmer, 1986), prosody, information organization, among others; and (c) the concept of modality itself overlaps those of attitude, illocution and emotion (Mello & Raso, 2012). Therefore, for the purposes of this paper, modality in speech will be understood as the conceptualizer's evaluation of an uttered locutive material, anchored in a communicative situation.

2.   The C-ORAL-BRASIL

The investigation of modality reported in this paper was carried through the analysis of a Brazilian Portuguese Spontaneous Speech Corpus, the C-ORAL-BRASIL I (Raso & Mello, 2010, 2012). This corpus is the fifth branch of the C-ORAL-ROM project (Cresti & Moneglia, 2005), a set of corpora representative of European Portuguese, French, Italian and Spanish spontaneous speech. The C-ORAL-BRASIL follows the same architecture and technical specifications found in the C-ORAL-ROM corpora, therefore being entirely comparable to the latter.

The C-ORAL-BRASIL I is presented through a DVD in which the following files can be found: sound files (wav); metadata featuring textual, situational, participants' information; transcriptions (rtf) segmented in tone units and utterances following the Language into Act parameters (Cresti, 2000); PoS tagged transcriptions in txt and XML formats through the PALAVRAS parser (Bick, 2000), speech to text alignment in XML format through the WinPitch aligner (Martin, 2004).

The C-ORAL-BRASIL I, the informal part of the C-ORAL-BRASIL project, features a very broad diaphasic variation, that is, speech situation variation, in view of representing as accurately as possible, a range of different speech acts through actual spontaneous linguistic activity.

The corpus textual typology is branched into monologues, dialogues and conversations, which on their part, are divided into public and private.

The C-ORAL-BRASIL I also features a balanced and informationally tagged subcorpus for study purposes. The information tagging was carried following the Language into Act Theory (Cresti, 2000) and the Information Patterning Theory (Cresti & Moneglia, 2010). Searches in the subcorpus can be carried through the search interface IPIC (http://lablita.dit.unifi.it/ipic/).

3.   In search of modality

The C-ORAL-BRASIL subcorpus was used as data source for the search of modal indexes since it is balanced for textual typology and it is informationally tagged, which allows for the identification of information units that carry modal indexes. The subcorpus is composed by 20 texts of three interactional typologies: dialogic (7), monologic (7) and conversational (6), divided into private and public, in a total of approximately 30.000 words.

The procedure adopted for analysis was to manually search for modal indexes and classify them in their context of occurrence according to their typological characteristics, which are: part of speech, information unit of placement, semantic label (aletic, epistemic or deontic modality), textual typology, gender and speaker schooling level. This qualitative classification was followed by a quantitative analysis, which took into consideration

type-token ratio and a multivariate analysis supported by the R environment (http://www.r-project.org/). The semantic label assigned to each token was validated through group discussion. Cases which presented disagreements or difficulties in labeling were reassessed until reaching satisfactory classification agreement.

Among the studies that resulted from this research effort are: identification of morpholexical modality indexes in tone units (Mello *et al.*, 2010), a comparative study between modal adverbs of certainty in a sample of Brazilian and European spontaneous speech corpora (Mello *et al.*, 2011), a study about the epistemic character of conditional constructions (Ávila & Côrtes, 2011), the description of modal indexes and their pragmatic-cognitive consequences (Ávila, 2012), and the mapping of modal adverbial constructions in Brazilian Portuguese (Mello & Caetano, in progress).

The research has shown the following distribution for modal types: from 2,573 utterances examined, 250 have some kind of modal marking (9.71%). The majority of modal markings are epistemic (57.85%), with deontic marking featuring 23.57% and aletic marking exhibiting 18.57%. The modal indexes found and their morpholexical classification, along with percentage of occurrence are shown in Table 1 below.

In order to illustrate the data analyzed, some examples follow below.

(1) =$ [171] no /=PHA= thirty reals /=TOP= then I &j [/2]=SCA= I [/1]=EMP= I **suppose that** he thinks like that /=INT= Oh my goodness /=EXP_r= **maybe** at my place one **need to** go shopping and everything /=COM_r= right//=PHA=$ (*bpubmn01*)

=$ [171] não /=PHA= trinta reais /=TOP= aí eu &j [/2]=SCA= eu [/1]=EMP= eu **fico imaginando que** e' fica pensando assim /=INT= Nossa Sio' /=EXP_r= **às vezes** lá em casa **tá precisando** de fazer uma compra e tudo /=COM_r= né //=PHA=$ (*bpubmn01*)

(2) *LUC: [74] <**if** on the first time that you say a word /=SCA= > it doesn't work /=**TOP**= it never will//=**COM**= got it//=PHA=$ (*bfamcv04*)

*LUC: [74] <**se** na primeira vez que cê falou uma palavra /=SCA= não> for /=**TOP**= nunca mais **vai ser** /=**COM**= entendeu //=PHA=$ (*bfamcv04*)

(3) *PAU: [153] because **it's most likely** that I'll build a wall there //=COM=

*PAU: [153] porque **é capaz** d' eu subir uma parede lá //=COM=

As for the comparison between Brazilian and European Portuguese modal adverbs of certainty, the results indicate an overall rate of occurrence higher in EP than in BP. The explanatory hypothesis for this finding isdiscussed in Mello *et al.* (2010) and is related to social hierarchization and education level differences in the two cultures. In Table 2 below the overall token numbers are presented for both language varieties, exhibiting the higher usage of modal marking in EP vis-à-vis comparative situations in BP.

Modality morpholexical strategies	Types	Percentages
Adjectives (or nominals in adjectival function) in predicative position	(é) lógico, é provável, é importante, (é) verdade	1,42%
Adverbs and adverbial expressions	Talvez, certamente, realmente, às vezes, também, logicamente, sinceramente, com certeza, completamente, sem dúvida, possivelmente, na verdade, na realidade	6,42%
Conditionals	[if X then Y]	13,21%
Modal constructions	tem condição (de), tem chance de, o que acontece, ter que, ficar imaginando, ficar pensando, (é) para + inf., dá para + inf., ter certeza, vai saber, tem jeito	22,14%
Future	vou + inf.	1,07%
Preterit future	ia ser, ia dar, seria	3,21%
Other forms	Digamos que, de certa forma	3,57%
Verbs (indicative mood – present, perfect and imperfect; infinitive)	Dever, poder, achar, acreditar, acontecer, ver, conseguir, precisar, pensar, dar e parecer.	48,92%

Table 1: Morpholexical strategies, types and percentages

	Public EP/BP	Private EP/BP	TOTAL EP/BP
Monologues	26/5 (5.2)	23/8 (2.875)	49/13 (3.77)
Dialogues	36/25 (1.44)	11/8 (1.375)	47/33 (1.424)
Conversations	23/6 (3.83)	22/8 (2.75)	45/14 (3.214)
TOTAL	85/36 (2.36)	46/24 (1.916)	141/60 (2.35)

Table 2: Modal adverb occurrence in EP/BP

The results of a modal adverb overall study (Mello & Caetano, in progress), covering the entire C-ORAL-BRASIL I corpus, shows the following statistics: a total of 763 tokens, divided among 28 types, with a strong concentration of about 55% of occurrences being by the adverb *mesmo* 'really'. The search was carried based on PoS tagging by PALAVRAS (Bick, 2000) and was checked manually for precision and accuracy. Except for one deontic adverbial, *necessariamente* 'necessarily', all other encountered forms are epistemic. An investigation about the specificities of the usage of *mesmo* in BP is being currently carried and it aims at clarifying whether there are any skewing effects caused by specific speakers or texts in the analyzed corpus.

The study about conditional constructions and their epistemic meaning (Ávila & Côrtes, 2011) was carried based on the C-ORAL-BRASIL subcorpus previously explained. In the 6,078 utterances examined, 11 conditional constructions were found. The results indicate the following distribution of conditionals, based on textual typology and context, shown on table 3:

Textual typology	Context	Frequency
Monologue	Private	18
	Public	6
Dialogue	Private	27
	Public	13
Conversation	Private	38
	Public	9

Table 3: Conditional construction frequency

As for the frequency of protasis versus apodosis structuring the results were the following:

Syntactic structure	Frequency
Protasis- Apodosis	75
Apodosis-Protasis	12
Protasis	24

Table 4: Conditional construction typological distribution

The marking of modality in conditional constructions has evidenced epistemic values as predominant. As for the information structure organization, the most frequent structuring brings protasis in Topic and apodosis in Comment units. The cognitive value of this organization needs further study in order to determine if and how modality indexes within different informational units interact at a higher semantic level.

On a pragmatic-discursive level, especially as far as modal verbs are concerned, the major functions found in our data were: (a) mitigation of previous assertion when the modalizer occurs in Parenthetical units; (b) mark agreement or disagreement; (c) mitigation of sociocultural differences among participants in a given interaction.

4.   Provisional Conclusions

So far, our research has shown that verbs are the major modality agent in BP and epistemic modality is the most frequent semantic type found. Another interesting finding is that BP allows for multiple modal valency utterances and tone units. What that means is that the same modal index may carry different semantic values depending on the utterance and tone unit in which it is found.

The preliminary study on adverbs of certainty in a sample of BP and EP has shown an upward curve representing an increased use of modal adverbs in lower diastraty in BP if compared to higher ones, which may indicate socioculturally-based differences in the expression of politeness in the two groups. Additionally, the comparison between EP and BP indicated differences in lexical choices in these two varieties along with a much higher usage of modal markings in EP than in BP.

Modal adverbs in BP spontaneous speech have complex usage patterns. The bare modal semantic meaning of adverbials is associated with other notions such as temporality, which should be further investigated.

Additionally, we have observed a strong interface between semantics and pragmatics which we address in face of participants' roles in speech events and their stance.

Last but not least, the epistemic character of conditionals seems to indicate the different degrees of "actuality" between the protasis and the apodosis.

5.   Acknowledgements

6.   References

Ávila, L., Côrtes, P. (2011). A epistemicidade nas construções condicionais do português do Brasil: estudo baseado em corpus de fala espontânea. Paper presented at XI SILEL, 23-25 November, 2011, Uberlândia, Brazil.

Ávila, L. (2012). Pensar a modalidade: marcação epistêmica e evidencial no portuhês brasileiro falado. In *Atas do X ELC*. Belo Horizonte: FALE.

Bally, C. (1932). *Linguistique générale et linguistique*

française. Berna: Francke Verlag.

Bick, E. (2000). *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus: Aarhus University Press.

Bybee, J., Fleischmann, S. (Eds). (1995). *Modality and grammar in discourse*. Amsterdam/Philadelphia: John Benjamins.

C-ORAL-BRASIL. Available at: <http://www.c-oral-brasil.org>.

C-ORAL-ROM. Available at: <http://lablita.dit.unifi.it/coralrom>.

Cresti, E. (2000). *Corpus di italiano parlato*. Firenze: Accademia della Crusca.

Cresti, E., Moneglia, M. (2005). *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins.

Cresti, E., Moneglia, M. (2010). Informational Patterning Theory and the corpus-based description of spoken language. The compositionality issue in the Topic-Comment pattern. In M. Moneglia, A. Panunzi (Eds), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Firenze: Firenze University Press, pp. 13--46 .

Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.

Martin, P. (2004). WinPitch Corpus: A text to Speech Alignment Tool for Multimodal Corpora. Lisbon: LREC. May 2004. Available at: <http://lablita.dit.unifi.it/coralrom/papers/index.html >.

Mello, H., Carvalho, J. and Côrtes, P. (2010). Modality in Brazilian Portuguese spontaneous speech: a first mapping of morpholexical indexes. In *Revista Estudos da Linguagem*, Belo Horizonte, v. 18, n. 2, jul./dez, pp. 105--133.

Mello, H.R., Ramos, A.C., Avila, L.B.B. (2011). Probing modal adverbs in Brazilian and European Portuguese: sociocultural variability in a pluricentric language. In A.S. Silva, A. Torres and M. Gonçalves (Eds.), *Pluricentric Languages: Linguistic Variation and Sociocognitive Dimensions*. Braga: Aletheia, pp. 473-486.

Mello, H., Raso, T. (2012). Illocution, Modality, Attitude: Different names for different categories. In H. Mello, A. Panunzi and T. Raso (Eds.), *Illocution, modality, attitude, information patterning and speech annotation*. Firenze: Firenze University Press.

Mello, H., Caetano, R. (In progress). Mapeamento de advérbios modalizadores no português brasileiro falado.

Palmer, F.R. (1986). *Mood and Modality*. Cambridge: Cambridge University Press.

Raso, T., Mello, H. (2010). The C-ORAL-BRASIL corpus. In M. Moneglia, A. Panunzi (Eds.), *Bootstrapping Information from Corpora in a Cross Linguistic Perspective*. Firenze: University Press.

Raso, T., Mello, H. (Eds). (2012). *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG.

Temporal and causal uses of the connector *come* in spoken Italian

Francesca GATTA

SSLiMIT, Università di Bologna, polo scientifico didattico di Forlì

Francesca.Gatta@unibo.it

Abstract

This paper is part of a larger research project on Italian connectors. The aims is to study the contribution of connectors to the encoding of conceptual relationship between two processes. The point of view to study the relationship between encoding and inference is the conceptual framework proposed by Prandi (2004). The occurrences of *come* in spoken Italian (LIP) allow us to describe the value of the connector as proposition and conjunction. As proposition *come* has a basic modal / comparative meaning; the temporal and the causal value of *come* derives from inferences which overlays other relationship: when the contents of the connected propositions allow, the meaning of the connector may be enriched by a temporal or a causal value.

Keywords: 'Come' (conjunction); connector; encoding; inference; LIP.

1. Introduction

This paper is a small part of a larger research project on Italian connectors. The project aims to study the contribution of connectors to the encoding of conceptual relationships between two processes. The general questions we are posing are: if the relationship between two processes can be inferred, what is the function of the connector? And can the contents of the connected propositions attribute a "new" value to the connector, extending the meaning of the latter?

These are questions which concern the relationship between encoding and inference, and that between content and expression. A conceptual framework for examining such questions has been proposed by Prandi (2004, III; 2006), who argues that in some areas of language, for instance in the nucleus of the sentence, encoding is relational (roles are assigned by a grammatical relation, so the grammatical relation assigns a content), while in others, such as the more outlying parts of the sentence, coding is punctual and the conceptual content prevails over the grammatical relation. In other words, there are some cases where the grammatical relation imposes itself on the contents and is independent of them, whereas in other cases the content is independent of the linguistic expression, and the latter merely encodes a conceptual relationship which is created outside the expression as such.

We believe our findings on the temporal and the causal value of *come* in spoken Italian support this theoretical position.

2. Data

Our data is taken from corpora of spoken Italian. This first step is based only on LIP (De Mauro *et al*., 1993), but in future the analysis will be extended to CLIPS (Leoni *et al*., 2006), C-Coral ROM (Cresti & Moneglia, 2005) and PIXI (Gavioli & Mansfield, 1990). Looking only at transcripts, we lack reliable information on prosody, and it remains to be seen how far prosodic features may also influence the interpretation of connectors and of the clauses they link.

The LIP corpus (queryable online at badip.unigraz.at) contains transcripts of 469 encounters for a total of approximately 500.000 orthographic words, divided into similarly sized components from four geographical areas (Milan, Florence, Rome, Naples). The corpus is part-of-speech tagged, making for a slightly higher number of pos units than the number of orthographic words.

For each geographical area, the corpus contains five types of speech: A, B, C, are two-way encounters (face-to-face and telephone conversations, interviews, etc.: 320.331 pos units); D, E are one-way encounters (lectures, radio monologues, etc.: 203.334 pos units).

In the corpus, the forms *com'* and *come* are tagged either as prepositions (Pz) or conjunctions (C). Table 1 shows their relative frequencies in two-way and one-way encounters

	2-way		*1-way*		*Total*	
	Freq.	Freq./1000 pos units	Freq	Freq./1000 pos units	Freq.	Freq./1000 pos units
Pz	442	1.38	427	2.10	869	1.66
C	1284	4.01	631	3.10	1915	3.66
Tot.	1726	5.39	1058	5.20	2784	5.32

Table 1: Frequencies of *come/com'* in the LIP corpus

Cases where *come* is tagged as a preposition are relatively straightforward:

> Come donna ti senti realizzata o no (As a woman, do you feel realised or not?) (F B 17 61 C)
> Volevo sapere come informatica a che punto siamo noi con tutti i programmi (As a computer expert, I wanted to know where we are with all the programmes) (F A 12 5 A)
> Eh vedono vedono la loro vita come spezzata e allora ricucirla ci vuol tempo (They see they see their life as torn apart and needing time to put it together again) (F E 15 253 A)

It is more difficult to identify the value of *come* where it is tagged as a conjunction: we manually analysed the occurrences in order to identify the transphrasic relationships involved, distinguishing two-way and one-way

encounters.

Traditional Italian grammars list *come* as a conjunction in the following uses:

– introducing (a) direct interrogatives, (b) indirect interrogatives, (c) completing subordinates:

    a) ciao *come* va (R B 6 4 B)

    b) questi condoni non si sa *come* andranno a finire (we don't know *how* these new regoulations will turn out) (F A 10 82 B)

    c) il dibattito sull'opinione pubblica vediamo *come* è determinato dalla domanda se è giusto o non giusto la guerra (the debate on public opinion we will see *how* it is dominated by the question of whether the war is just or injust.) (M E 8 8 G)

– introducing adverbial clauses which are (d) comparisons or analogies (e) temporal, or (f) causal:

    d) diceva trattare l'ammalato *come se fosse la madre* come se tu infermiere o tu medico fossi sua madre e fosse lui l'unico tuo figlio (as if she was his mother) (M E 12 10 C)

    e) allora *come esce* [incomprehensible word] dal comune *come esce* lo porta su all'archivio (as soon as he walks out of the office …) (F A 5 1 A)

    f) ma come non è un ragazzo di questo (but since he's not this kind of boy) (N B 65 23 A)

Some examples, particularly those with adverbial clauses, are however ambiguous, in particular between the causal and the temporal meanings.

The temporal use of *come* is documented since Dante ("Sì tosto come il vento a noi li piega / mossi la voce …", *Inferno* V, vv. 81-82). For the dictionary GRADIT, the temporal value belongs to basic Italian ("uso fondamentale"); on the contrary, Serianni (1988) considers it typical of written and especially literary Italian. In LIP the temporal sense appears only in bidirectional encounters, supporting GRADIT's proposal that it is also a colloquial usage.

As far as concerns the causal value of *come*, GRADIT states that it is relatively infrequent ("basso uso"); similarly, Serianni claims that *come* assumes a causal value only occasionally. In LIP we found fewer causal than temporal examples, some being particularly ambiguous.

The causal interpretation appears to depend on either (a) the contents of the connected propositions; and or (b) position in the dialogue sequence. The following examples illustrate causal linking between connected propositions: in both cases there is some ambiguity between a causal interpretation and one of analogy:

Io penso che gente *come gioca alle lotterie* gioca anche al totocalcio perché insegue proprio il miraggio dl due miliardi del tre miliardi del miliardo (I think people bet on the lottery for the same reasons/in the same ways they bet on the pools) (M E 7 26 A)

Sì ma se tu me seguiti a di' sempre quando troverò *come so' passati circa sette anni* ne passeranno altri sette e io non ce sto più allora io vado a finì sotto tera o mezzo a 'n campo de patate (As about seven years have passed, another seven will) (R E 11 86 D)

The next three examples illustrate the importance of position in the dialogue sequence in suggesting a causal value (in these cases, LIP tags *come* as a preposition, while for other grammars it would be an interrogative adverb). *Come* is used to question the previous affirmation of the other speaker, in the causal sense of "why do you say that?" This is particularly clear in the second example, where speaker A explicitly confirms the causal value of his previous *come* by reformulating it with *perché* in the next utterance:

B: no tesoro non posso
A: *come* non puoi* (why can't you*)
B: tu non fossi amico di XYZ forse sì ma così non posso (M B 46 356 B)

B: e non lo vendono quella roba lì dal rivenditore grani Rapid
A: *come* non li vendono* (why don't they sell them*)
B: il grani Rapid*
A: eh non capisco *perché* non devono venderlo be' $$$ ce li ha (M B 70 15 A)

A: mo' me metto la tuta e vengo [incomprehensible word]
B: ti infili la tuta*
A: la tuta vengo in tuta
B: ma che schifo *come* vieni in tuta* (how disgusting why do you come in a tracksuit*)
A: vengo in tuta da ginnastica
B: Bleah
A: 'n ti piace*
B: no (R B 1 120 B)

3. Conclusions

To sum up, our research on a corpus of spoken Italian has provided evidence that the temporal and causal senses of *come* belong to colloquial usage as well as literary Italian. We would argue that these senses of *come* are the result of processes of inferential enrichment. From our point of view, the temporal/causal value of the connector is undercoded, and the attribution of this value derives from inferencing which overlays other relationships. If we see *come* as having a basic modal/comparative meaning, then *come* can encode this kind of relation between two clauses

without considering the contents of the propositions involved. When the contents of the connected propositions allow, however, the meaning of the connector may be enriched by a temporal or a causal value. Such enrichment is possible because – according to the theoretical viewpoint of Prandi (2004) – when we speak of adverbial clauses, we are in an area of the language in which conceptual contents are dominant with respect to grammatical relations.

4. Acknowledgements

5. References

Cresti E., Moneglia, M. (2005). *C-ORAL ROM. Integrated reference corpora for spoken Romance languages*. Amsterdam: Benjamins.

De Mauro (1999). *Grande Dizionario italiano dell'uso*. Torino: Utet.

De Mauro *et al.* (1993). *Lessico di frequenza dell'italiano parlato*. Available at: <www.badip.uni-graz.at>.

Gavioli, L., Mansfield, G. (1990). *The PIXI corpora. Bookshop encounters in English and Italian*. Bologna: Clueb.

Leoni, A., Cutugno F., Savi *et al.* (2006). *Corpora e lessici di italiano parlato e scritto*. Available at: <www.clips.unina.it>.

Mazzoleni, M. (2007). Un "come" modale, temporale e causale nell'italiano contemporaneo?. In G. Garzone, M. Mazzoleni and P. Scampa (Eds.), *"Come" et ses propositions subordonnées en italien contemporain*, «Revue Romane», 46/2, pp. 238--265.

Prandi, M. (2004). *The building Blocks of Meaning. Ideas for a Philosophical Grammar*. Amsterdam-Philadelphia: John Benjamins.

Prandi, M. (2006). *Le regole e le scelte. Introduzione alla grammatica italiana*, Torino: Utet.

Prandi, M., Gross, G. and De Santis, C. (2005). *La finalità. Strutture concettuali e forme d'espressione in italiano*. Firenze: Olschki.

Salvi, R. (Ed.). *Linguistica, linguaggi specialistici, didattica delle lingue. Studi in onore di Leo Schena*. Roma: Cisu, pp. 238-265.

Serianni, L. (1988). *Grammatica italiana. Italiano comune e italiano letterario. Suoni, forme, costrutti*. Torino: Utet.

La variazione dei verbi generali nei corpora di parlato spontaneo. L'ontologia IMAGACT

Massimo MONEGLIA, Gloria GALIARDI, Lorenzo GREGORI, Alessandro PANUNZI, Samuele PALADINI, Andrew WILLIAMS

Università di Firenze (Italia)

moneglia@unifi.it, gloria.gagliardi@unifi.it, lorenzo.gregori@unifi.it, alessandro.panunzi@unifi.it, samuele.paladini@drwolf.it, andrewwilliams.esq@gmail.com

Abstract

I verbi di azione, ad alta frequenza nel parlato, sono molto spesso "generali", perché si estendono produttivamente ad azioni che individuano oggetti ontologici diversi, ed ogni lingua presenta categorizzazioni idiosincratiche dello spazio ontologico dell'azione. Per questo motivo i verbi d'azione costituiscono un problema per la disambiguazione e per la traduzione delle lingue naturali. Questo lavoro presenta le linee di sviluppo del progetto IMAGACT, che si propone di derivare da corpora di parlato spontaneo multilingui informazioni essenziali sulla categorizzazione linguistica dell'azione, non prevedibili allo stato attuale delle conoscenze. Il progetto utilizza campioni di corpora di parlato spontaneo italiano e inglese, da cui induce l'ambito di variazione produttiva dei circa 500 verbi di azione più alti in frequenza in ciascun corpus. In IMAGACT la variazione si oggettiva in una ontologia interlinguistica le cui entrate sono costituite da scene prototipiche. L'utilizzo del linguaggio universale delle immagini evita problemi di indeterminatezza delle definizioni e facilita sia lo sviluppo, sia lo sfruttamento della base dati.

Keywords: verbi di azione; ontologie; corpora di parlato multilingui.

1. Introduzione

I verbi di azione sono gli elementi più frequenti di strutturazione del discorso parlato e contengono l'informazione essenziale per dare senso agli enunciati (Moneglia & Panunzi, 2007). Ma i verbi d'azione sono anche i tipi linguistici meno predicibili per i dizionari bilingui e per le tecnologie di traduzione automatica (Moneglia, 2011). Questi verbi, infatti, molto spesso sono "generali", in quanto si estendono ad azioni appartenenti a differenti tipi ontologici. Per esempio in inglese ed italiano i verbi ad alta frequenza *to put* e *mettere* appartengono a questa categoria. La Tabella 1 esemplifica la varietà di atti che ricadono nella loro estensione. In 1 ad un oggetto è data locazione, in 2 un oggetto è dotato di attributi funzionali, in 3 un oggetto è modificato, in 4 una parte del corpo assume una posizione.

La diversità sostanziale tra i tipi di atti riferiti dal verbo, evidenziata dalla figura, è marcata linguisticamente dalla possibilità di identificare ciascuna azione con verbi equivalenti diversi, che si applicano in modo differenziale a ciascun tipo (*collocare*, *inserire*, *aggiungere*, *alzare*).

Malgrado una forte relazione di traduzione, *to put* e *mettere* non sono però coestensivi, dal momento che *to put* può essere esteso a 4, ma non *mettere*.

Questa differenza, individuata in seguito a lavoro su corpus, non è chiaramente identificata allo stato attuale delle conoscenze sul lessico verbale d'Azione ed è un esempio delle ragioni cruciali per cui le predicazioni del linguaggio naturale non sono idonee alla traduzione automatica: non sono identificate le entità ontologiche a cui i verbi d'azione si riferiscono nelle frasi semplici e non vi è quindi garanzia che due predicati in un dizionario bilingue selezionino la stessa entità.

Ogni lingua, con i suoi verbi generali, categorizza l'azione in un modo specifico e perciò il riferimento cross-linguistico alle attività di ogni giorno risulta scarsamente prevedibile (Moneglia & Panunzi, 2007).

ACTION TYPE	INSTANCES	EQUIVALENT VERBS
	Type 1 John puts the glass on the table John mette il bicchiere sul tavolo	*to locate* *collocare*
	Type 2 John puts the cap on the pen John mette il tappo alla penna	*to fasten* *inserire*
	Type 3 John puts water into the whisky John mette l'acqua nel whisky	*to add* *aggiungere*
	Type 4 *Mary mette su la mano Mary puts her hand up	 *to raise*

Tabella 1: Tipi azionali dei verbi *to put* e *mettere*

E' rilevante notare che tale variazione cross-linguistica non è dovuta alle fraseologie proprie di ogni lingua, ma è conseguenza del modo peculiare con cui le lingue categorizzano gli eventi, ovvero deriva da fattori semantici (Moneglia, 1998; Majid *et al.*, 2008).

Infatti l'applicazione dei verbi generali ai tipi azionali nella loro estensione è *produttiva*: in qualunque evento del tipo 1 *to put* sarà tradotto in Italiano con *mettere*, e in nessuna istanza del tipo 4 il verbo Inglese *to put*, risulterà traducibile in Italiano con *mettere*, come mostrano i seguenti esempi:

(1) John puts a glass / a pot / a dress   on the table / on the stove / on the harm chair

(1') John mette un bicchiere / la pentola / sul tavolo / sul fornello / sulla poltrona

(2) Mary puts her hand / her finger / her leg / up / aside / down

(2') *Mary mette la mano / il dito / la gamba / su / di lato / giù

Se l'applicazione di un verbo ad un tipo è produttiva, dovrebbe in linea di principio essere anche predicibile: il *range* di variazioni produttive dei verbi generali nelle diverse lingue è però, al momento, largamente sconosciuto; non è chiara, inoltre, la distinzione tra variazioni produttive e variazioni non produttive nell'estensione dei verbi generali.

Le risorse esistenti, e in particolare WordNet, che costituisce la principale e più ricca base di dati lessicale oggi disponibile (Fellbaum, 1998), non contengono informazione sufficiente a questo scopo per una varietà di ragioni (Moneglia *et al.*, 2012). Per esempio il numero di tipi (*synset*) registrati per ciascuna entrata è alto ma, non essendo la risorsa derivata da corpora, i significati periferici non sono distinti da quelli con alta probabilità di occorrenza. Inoltre, per lo stesso motivo, non esiste certezza che le variazioni principali di un verbo generale nell'uso linguistico siano censite. In aggiunta, le descrizioni date per ciascun *synset* sono vaghe e difficili da utilizzare perfino da annotatori esperti (Ng *et al.*, 1999).

Più in generale deve essere notato un problema teorico che affligge le risorse che riflettono la varietà dell'uso linguistico e rendono poco prevedibile la possibilità di traduzione, ovvero che la produttività dell'applicazione del verbo non può essere garantita da tutti i *synset* nella stessa misura. I verbi hanno infatti vari usi che si distaccano dal loro significato effettivo, ed in questi significati la relazione di traduzione non può essere predetta.

Ad esempio, tra i *synset* di WordNet del verbo *to put* è riportato il seguente:

S: (v) arrange, set up, put, order (arrange thoughts, ideas, temporal events)

In questa entrata dell'ontologia, diversamente da quanto avviene in (1) e (2), la possibilità di traduzione non corre in parallelo in tutte le istanze del tipo. Funziona in (3), ma per qualche ragione idiosincratica non in (4):

(3) I put my schedule in a certain way > Ho messo i miei impegni in un certo modo

(4) I put my life in a certain way > * Ho messo la mia vita in un certo modo

La distinzione tra tipi produttivi e tipi idiosincratici è cruciale: solo gli usi primari (come quelli nella Tabella 1) sono sicuramente produttivi, mentre gli usi fraseologici o metaforici spesso non lo sono. In altri termini, mentre la variazione in Tabella 1 identifica le variazioni in estensione su tipi di azioni diverse che un parlante nativo deve poter assentire o rifiutare sulla base della sua sola competenza linguistica, lo stesso non vale per usi marcati come in (3). Solo l'identificazione degli usi produttivi costituisce una base di conoscenza per la previsione degli ambiti di estensione dei verbi di lingue diverse nello spazio dell'azione e per rendere obiettive le relazioni di traduzione.

Il progetto IMAGACT utilizza metodologie *corpus-based* e *competence-based* per l'estrazione simultanea da risorse multilingui di parlato spontaneo di una ontologia dell'azione indipendente dal linguaggio, e permetterà la disambiguazione dei verbi di azione ad alta frequenza nel parlato rispetto ai tipi azionali in cui una applicazione produttiva può essere prevista.

Questo lavoro descrive le caratteristiche chiave del progetto. Il paragrafo 2. mostrerà la strategia *corpus-based* scelta per l'induzione delle proprietà variazionali dei verbi d'azione e presenterà in allegato le entrate verbali oggetto di analisi; il paragrafo 3. illustrerà, sulla base di un esempio concreto (la variazione di *to roll* in inglese e parallelamente la variazione di *rotolare* e *arrotolare* in italiano), la metodologia di costruzione dell'ontologia interlinguistica, specificamente basata sull'utilizzo dell'immagine.

2. Lo sfruttamento di risorse di parlato spontaneo

Le azioni specificate dai verbi usati con maggior frequenza nella comunicazione quotidiana sono anche le azioni più rilevanti per le nostre attività di ogni giorno e, in quanto tali, costituiscono l'universo di riferimento per il linguaggio. L'uso effettivo di tali verbi può pertanto essere apprezzato nella performance linguistica mediante l'osservazione delle loro occorrenze nel parlato spontaneo, in cui il riferimento all'azione è primario. I corpora di parlato spontaneo pubblicati negli ultimi due decenni sono sfruttati in IMAGACT a questo fine: la variazione di un set di predicati generali verrà infatti identificata nel corpus BNC (sezione di parlato) e, in parallelo, in una collezione di corpora italiani (C-ORAL-ROM; LABLITA, LIP, CLIPS).

IMAGACT si focalizza sui verbi ad alta probabilità di occorrenza, ovvero i 500 verbi di azione più alti in rank nelle liste di frequenza, che rappresentano il lessico

verbale di base nelle due lingue. Un'ampia selezione di questo lessico è riportata nella liste di frequenza disponibili in appendice.

Saranno annotate attraverso una infrastruttura web circa 50.000 occorrenze per lingua, derivate da un campione di 2 milioni di parole di entrambi i corpora.

Gli enunciati in cui le occorrenze compaiono nei corpora, necessariamente frammentari dal punto di vista semantico, vengono interpretati da annotatori madrelingua e ricondotti a frasi semplici nelle quali è saturata la struttura valenziale e da cui l'azione riferita risulta in modo trasparente. La presenza di una serie ampia di frasi semplici derivate dall'uso orale consente di individuare i punti essenziali della variazione d'uso di ciascun verbo e di raggrupparne in tipi gli usi produttivi.

A tal fine è adottata una metodologia specifica e una procedura di annotazione guidata dall'infrastruttura web IMAGACT a disposizione degli annotatori.

3. Formazione dell'ontologia interlinguistica dell'azione e immagine. Uno scenario "alla Wittgenstein"

Lavorando con più di una lingua, IMAGACT deve produrre un inventario di tipi *language-indipendent*. Precedenti esperienze nella costituzione di Ontologie hanno evidenziato però che il livello di consenso raggiungibile nella definizione delle entità riferite dalle espressioni linguistiche è generalmente basso, e che l'accordo nell'annotazione varia in relazione alla granularità semantica dei sensi (Brown *et al.*, 2010).

L'innovazione chiave di IMAGACT è di fornire una metodologia che sfrutti la capacità, indipendente dal linguaggio, di apprezzare somiglianze tra scene, distinguendo di fatto *l'Identificazione* dei tipi azionali dalla loro *Definizione*.

Ad esempio, la distinzione tra i tipi 1-4 nella Tabella 1 è rilevante per prevedere la variazione cross-linguistica dei concetti azionali. La differenza tra i tipi è facilmente riconosciuta dai parlanti e non richiede la definizione di un set di caratteristiche differenziali, che sono, come si diceva, radicalmente sottodeterminate.

Crucialmente solo l'identificazione, e non la definizione delle entità individuate, è richiesta per stabilire le relazioni cross-linguistiche.

In termini Wittgensteiniani: come posso spiegare a qualcuno cos'è un *gioco*? Semplicemente indicando un gioco e dicendo "Questo e simili cose sono giochi" (Wittgenstein, 1953).

Lo scenario "alla Wittgenstein" è utilizzato in IMAGACT sia per distinguere le variazioni produttive dalle variazioni non produttive all'interno dell'uso linguistico dei verbi, sia per identificare tipi azionali a livello cross-linguistico, consentendo la comparazione diretta dei tipi derivati dall'annotazione dei corpora di lingue diverse.

Per l'induzione della variazione semantica dei verbi di azione dai corpora di parlato italiano e inglese IMAGACT si sviluppa sui seguenti passi:

- distinguere gli usi primari dagli usi marcati;
- identificare in ciascun corpus di parlato i punti focali di variazione dei verbi generali su tipi di azione diversi;
- rappresentare i concetti azionali attraverso scene prototipiche a cui rapportare la variazione riscontrata nei verbi delle due lingue.

3.1 Variazione primaria vs. Variazione marcata

Il primo compito sfrutta lo scenario "alla Wittgenstein" come banco di prova della effettiva produttività dei concetti. Si deve notare, infatti, che solo gli usi che ad un parlante competente appaiono adeguati a rappresentare il significato di un predicato possono essere indicati come prototipi per l'uso del predicato stesso. In parallelo, gli usi non primari o comunque metaforici o fraseologici non possono essere indicati come istanze prototipiche di ciò che viene significato.

Si consideri ad esempio il verbo italiano *rotolare*. L'istanza (5), derivata da corpus, può essere ragionevolmente indicata come una istanza prototipica del concetto espresso dal verbo, in altri termini un parlante competente può indicare l'istanza a qualcuno che non conosce la lingua fornendo l'informazione: "questa e simili cose sono ciò che noi intendiamo con *rotolare*". Al contrario, l'istanza (6) non potrà ragionevolmente essere indicata come un'istanza di "ciò che noi intendiamo con *rotolare*".

(5) Cristina si rotola nell'erba umida
(6) Il bambino rotolò in terra dal seggiolone

Infatti, nonostante la frequenza con cui può comparire in quel contesto, in (6) il verbo è usato palesemente in senso non proprio (il bambino non *rotola*, bensì *cade*). Ciò risulta evidente ad un parlante competente. Il test consente quindi, salvo casi limite, di isolare la gran parte degli usi strettamente propri del verbo, identificando poi la loro variazione.

Lo stesso avverrà con le frasi derivate dal corpus inglese. Ad esempio, per quanto riguarda la variazione del verbo *to roll* (7), potrà essere indicata come un istanza prototipica di ciò che si intende con *to roll*, ma non (8).

(7) John rolls a cigarette
(8) John rolls the words around in his mind

Lo studio della variazione produttiva di un verbo inizia quando gli usi non produttivi sono esclusi dal campo di analisi.

3.2 Variazione verticale vs. variazione orizzontale

La variazione dei verbi generali si configura in modo simile a quanto ipotizzato originariamente da Wittgenstein, ovvero l'uso si raccoglie in una serie di famiglie, ciascuna delle quali contiene variazioni granulari rapportabili ad una istanza prototipica (Givon, 1986). Ogni concetto istanziato da un prototipo è

produttivo e distinto dagli altri dal punto di vista cognitivo, nonostante lo stesso verbo si applichi a tutte le famiglie (proprietà per cui il verbo si dice "generale"). A tale variazione si unisce poi la variazione non produttiva, non identificata nel lavoro originale del filosofo, che ovviamente non definisce entrate nell'ontologia.

L'annotazione del verbo inglese *to roll* e dei verbi italiani apparentemente in relazione di traduzione con questo, ovvero *arrotolare e rotolare*, può essere riassunta in breve nelle tabelle seguenti derivate dalla annotazione dei corpora attraverso l'infrastruttura IMAGACT. Nel corpus sono identificati una serie di tipi (variazione verticale del verbo), ognuno dei quali contiene una serie di istanze (variazione orizzontale del tipo).

TO ROLL	
Type 1	John rolls his sleeve up John rolls a cigarette The sailors roll the sail up
Type 2	The horse rolls around the field Mary rolls onto her side John rolls along the floor
Type 3	John rolls the barrel along the floor John rolls the girl onto her side John rolls the thread around
Type 4	John rolls the ball across the room John rolls the wheel into the scrapheap John rolls the apple across the table to Mary
Type 5	John rolls his ankle around John rolls his eyes John rolls his wrist around in its socket
Type 6	The car rolls into the fence The ball rolls over to the wall The car rolls into the lake
Type 7	John rolls the clay in his hands John rolls the dough into a ball John rolls the playdoh on the table

Tabella 2: Tipi azionali del verbo to roll

ARROTOLARE	
Tipo 1	Cristina arrotola il filo intorno alla ruota Cristina arrotola la benda intorno al braccio Fabio arrotola la corda intorno alla gamba
Tipo 2	Cristina arrotola una sigaretta Cristina arrotola il poster Cristina arrotola il filo

Tabella 3: Tipi azionali del verbo *arrotolare*

ROTOLARE	
Tipo 1	Matteo si rotola per terra Cristina si rotola nell'erba umida Fabio e Cristina si rotolano
Tipo 2	La ciambella di gomma rotola L'arancia rotola Il cilindro rotola

Tabella 4: Tipi azionali del verbo *rotolare*

Dopo la procedura di annotazione dei corpora, IMAGACT rilascerà un database di tipi azionali associati alla loro codifica linguistica in inglese e in italiano. L'insieme delle frasi derivate da corpora istanzieranno ogni tipo rappresentato.

3.3 Immagine e Ontologia Cross-linguistica

Sulla base dell'induzione della variazione verticale *across-types* dei verbi di azione nei corpora, IMAGACT fa uso del linguaggio universale delle immagini per riconciliare in una sola ontologia i tipi derivati dall'annotazione di corpora di diverse lingue.

Ad esempio i tipi estratti dalla annotazione di *to roll* sono rappresentati dalle scene B-H, come in Figura 1 di seguito.

La costituzione delle scene permette una rappresentazione dell'universo dell'azione valido indipendentemente dalla lingua. Per cui, a livello della costituzione dell'ontologia cross-linguistica sulla base dei dati derivati da corpus, si scoprirà che la scena B è estesa anche dal tipo 2 del verbo italiano *arrotolare*, e che i tipi 1 e 2 del verbo *rotolare* estendono rispettivamente sui tipi C e G.

Nell'insieme possiamo osservare che la variazione del verbo inglese *to roll* è più ampia rispetto alle sue controparti italiane, dato che i due verbi italiani in linea teorica corrispondenti a questo verbo inglese (*arrotolare* e *rotolare*) trovano applicazione solo in un sottoinsieme dei tipi azionali estesi da *to roll*.

Il differenziale nel significato sarà ulteriormente evidenziato nel momento in cui, dovendo identificare una scena per il tipo 1 di *arrotolare* (il tipo A di Figura 1) diventerà evidente che c'è almeno un tipo esteso da *arrotolare* che non è una possibile estensione di *to roll*. La relazione cross-linguistica risulta quindi in una intersezione tra tipi.

La corrispondenza tra tipi derivati da differenti corpora linguistici seguirà perciò dal riferimento dei tipi estratti dai corpora alla stessa galleria di scene. Questo risultato è ottenuto senza far ricorso alla comparazione tra definizioni date da differenti annotatori: identificare la corrispondenza cross-linguistica dei verbi d'azione su una ontologia *language-indipendent*, aggira la sottodeterminazione delle definizioni.

IMAGACT rilascerà una base dati di tipi azionali individuati nel riferimento linguistico alle azioni quotidiane attraverso la rappresentazione di scene prototipiche. Ogni scena sarà associata a uno o più verbi verbi italiani e inglesi che risulteranno in relazione di traduzione stretta in tutte le istanze del tipo.

IMAGACT renderà chiaro sia l'ambito di variazione dei predicati generali nelle lingue considerate, sia il differenziale semantico tra entrate lessicali a livello cross-linguistico e permetterà di basare processi di disambiguazione e traduzione su tipi ontologici produttivi oltreché rilevanti in quanto derivati da corpora rappresentativi dell'uso linguistico quotidiano.
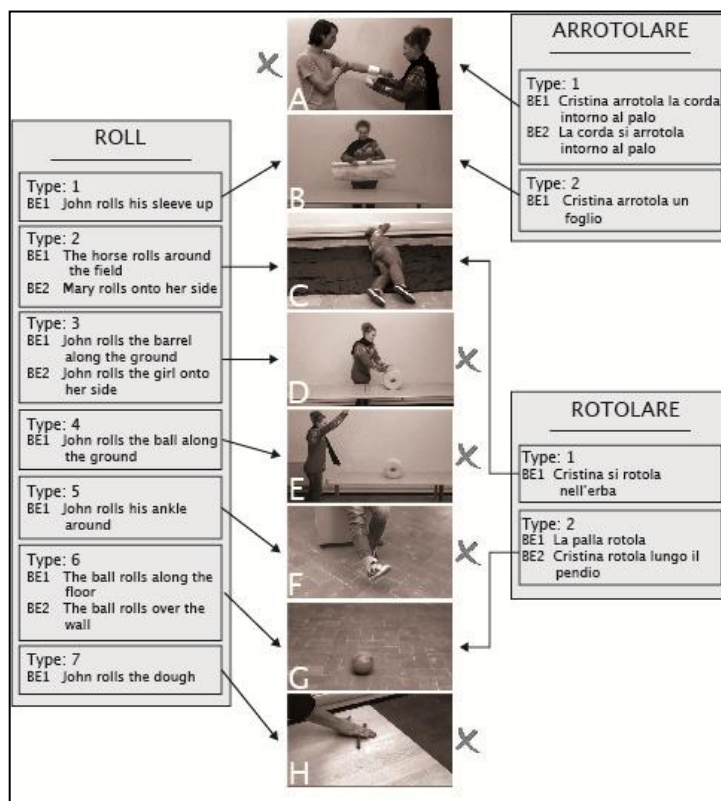
Figura 1: to roll vs. rotolare / arrotolare

4.   Referimenti

British National Corpus, version 3 (BNC XML Edition). (2007). Distributed by Oxford University Computing Services. Available at:
<http://www.natcorp.ox.ac.uk/>.

Brown, S.W., Rood, T. and Palmer, M. (2010). Number or nuance: which factors restrict reliable Word Sense Annotation? In N. Calzolari (Ed.), *Proceedings of the Seventh International Conference on Language resources and Evaluation*, pp. 3237--3243.

CLIPS Corpus. Available at: <http://www.clips.unina.it>.

CORALROM. Available at:
<http://catalog.elra.info/product_info.php?products_id=757>.

De Mauro, T., Mancini, F., Vedovelli, M., Voghera, M. (1993). *Lessico di frequenza dell'italiano parlato* (LIP). Milano: ETASLIBRI.

Fellbaum, Ch. (1998). *WordNet: An Electronic Lexical Database*. Cambridge (MA): MIT Press.

Givon, T. (1986). *Prototypes: Between Plato and Wittgenstein*. In C. Craig (Ed.), *Noun Classes and Categorization*. Amsterdam: Beniamin, pp. 77--102.

IMAGACT. Available at: <http://www.imagact.it/>.

LABLITA Corpus of Spontaneous Spoken Italian. Available at: <http://lablita.dit.unifi.it/corpora/>.

Majid, A., Boster, J.S. and Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. In *Cognition*, 109,(2), pp. 235--250.

Moneglia, M. (1998) Teoria empirica del senso e partizione semantica del lessico. *Studi di Grammatica Italiana*, XVII, pp. 363--398.

Moneglia, M. (2011). Natural Language Ontology of Action. A gap with huge consequences for Natural Language Understanding and Machine Translation. In Z. Vetulani (Ed.), *Proceedings of the 5th Language & Technology Conference* Poznań: Fundacja Uniwersytetu im. A. Mickiewicza, pp. 95--100.

Moneglia, M., Panunzi, A. (2007). Action Predicates and the Ontology of Action across Spoken Language Corpora. The Basic Issue of the SEMACT Project. In M. Alcántara, T. Declerck (Eds.), *Proceeding of the International Workshop on the Semantic Representation of Spoken Language*. Salamanca: Universidad de Salamanca, pp. 51--58.

Moneglia, M., Monachini, M., Panunzi, A., Frontini, F., Gagliardi, G.. and Russo, I. (2012). Mapping a corpus-induced ontology of action verbs on ItalWordNet. In C. Fellbaum, P. Vossen (Eds.), *Proceedings of the 6th Global WordNet Conference*, pp. 219--226.

Ng, H.T., Lim, C.Y., Foo, S.K. (1999). A Case Study on Inter-Annotator Agreement for Word Sense Disambiguation. In *Proceedings of the ACL SIGLEX Workshop on Standardizing Lexical Resources* College Park (MD), pp. 9-13.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

5. Appendice

Tabella 5: Verbi italiani di azione ad alta frequenza (entries in frequency order, read column by column)

LEMMA	FREQ.	LEMMA	FREQ.
mettere	4018	raccogliere	165
dare	3261	dividere	155
prendere	2459	rivolgere	150
parlare	2390	nascondere	147
mangiare	1662	porgere	147
portare	1509	allontanare	142
scrivere	1078	ammazzare	138
passare	942	riportare	138
lasciare	914	legare	136
tenere	852	aggiungere	131
leggere	823	fumare	130
entrare	785	spegnere	129
aprire	750	volare	129
girare	658	ballare	128
uscire	651	infilare	126
mandare	563	inserire	126
rompere	557	mantenere	126
chiudere	495	pulire	125
toccare	481	registrare	123
cadere	453	raggiungere	120
dormire	405	appoggiare	119
ridere	401	tirare - 2	119
telefonare	394	ricevere	119
pigliare	380	accompagnare	111
fermare	379	coprire	109
buttare	376	vestire	107
seguire	364	firmare	105
alzare	349	svegliare	105
partire	334	picchiare	102
porre	327	colpire	100
avvicinare	326	buttare via	98
sedere	316	tirare fuori	98
scendere	315	accendere	97
levare	300	saltare	95
togliere	282	segnare	92
bere	281	addossare	89
studiare	275	bloccare	84
rendere	269	diminuire	84
piangere	256	superare	84
camminare	255	lanciare	83
cantare	239	tendere	83
salire	234	pigiare	81
correre	233	fissare	80
tirare	227	sostenere	80
suonare	221	operare	77
cascare	204	staccare	77
attaccare	200	collegare	76
battere	198	bruciare	74
tagliare	195	guidare	72
lavare	191	ridurre	72
spostare	191	sottolineare	72
mostrare	186	eliminare	71
disegnare	185	applicare	70
costruire	182	combattere	70
rimettere	180	ritirare	68
scappare	168	aumentare	67
riprendere	166	voltare	67
spingere	166	interrompere	66
		liberare	66
		uccidere	66
gettare	65	misurare	41
curare	64	sciupare	41
sparare	64	stendere	41
comporre	63	buttare giù	40
avanzare	61	inviare	40
attraversare	61	ordinare	40
bagnare	60	rimandare	40
pesare	60	viaggiare	40
sommare	60	avviarsi	39
introdurre	59	cancellare	39
puntare	59	disporre	39
versare	59	pettinare	39
asciugare	58	premere	39
sistemare	58	rovesciare	39
bucare	56	sbattere	38
consegnare	56	scattare	38
risalire	56	agitare	37
sorridere	56	allungare	37
stringere	56	distribuire	37
procedere	55	mordere	37
tradurre	55	schiacciare	37
aggiustare	54	separare	37
baciare	54	sfuggire	37
emergere	54	assaggiare	36
giungere	54	incidere	36
tirare su	54	pescare	36
distruggere	53	provenire	35
limitare	53	ricostruire	35
proseguire	53	sciogliere	35
abbassare	52	forare	34
circondare	52	lottare	34
colorare	52	tirare via	34
condurre	52	beccare	33
cuocere	52	dondolare	33
rovinare	51	scorrere	33
sfogliare	51	addormentare	32
sporcare	50	stampare	32
unire	50	affacciare	31
allargare	49	calare	31
colare	48	respirare	31
montare	48	scavare	31
accomodare	48	trarre	31
stirare	48	trasportare	31
bussare	47	afferrare	30
piegare	47	ferire	30
cogliere	46	gocciolare	30
combinare	46	reggere	30
concentrare	46	scambiare	30
riempire	46	trattenere	30
riunire	45	volgere	30
sciare	44	accarezzare	29
servire	44	caricare	29
gridare	43	frenare	29
posare	43	incastrare	29
sollevare	43	riattaccare-2	29
strappare	43	rilasciare	29
urlare	43	scuotere	29
		segare	29
		confrontare	28
montare - 2	28	ciucciare	20
orientare	28	copiare	20
piantare	28	inseguire	20
potare	28	raffigurare	20
ricamare	28	ripassare	20
scaricare	28	spazzare	20
seccare	28	acchiappare	19
studiare-2	28	fischiare	19
buttare fuori	27	forzare	19
gesticolare	27	indirizzare	19
marcare	27	indossare	19
trasferire	27	mirare	19
abbaiare	26	percorrere	19
appiccicare	26	prelevare	19
avviare	26	scottare	19
cavare	26	soffriggere	19
conservare	26	spogliare	19
precedere	26	trascinare	19
riposare	26	volare via	19
rodere	26	annaffiare	18
sospendere	26	estrarre	18
accedere	25	freddare	18
gonfiare	25	parcheggiare	18
ridare	25	penetrare	18
ripartire	25	poggiare	18
spezzare	25	restringere	18
trascrivere	25	riaprire	18
abbattere	24	ricoprire	18
cacciare-2	24	riscaldare	18
estendere	24	appendere	17
recarsi	24	centrare	17
scattare-2	24	dipingere	17
sovrapporre	24	illuminare	17
vuotare	24	scartare-2	17
affogare	23	tossire	17
cucinare	23	trasferirsi	17
fotografare	23	annusare	16
fuggire	23	coltivare	16
numerare	23	costeggiare	16
pizzicare	23	cucire	16
rileggere	23	mescolare	16
sfilare	23	molare	16
spedire	23	pestare	16
chinare	22	proiettare	16
graffiare	22	respingere	16
parare	22	smontare	16
riscrivere	22	soffermare	16
ruotare	22	adattare	15
tappare	22	arrampicare	15
tratteggiare	22	collocare	15
assistere	21	disfare	15
associare	21	fissare-2	15
circolare	21	grattare	15
evidenziare	21	impolverare	15
menare	21	incrociare	15
passeggiare	21		
ricavare	21		
rigirare	21		
riprodurre	21		
ritirare - 2	21		
spaccare	21		
tremare	21		
investire	15	installare	11
precipitare	15	levare-2	11
pungere	15	oscillare	11
riprendere - 2	15	piagnucolare	11
sbottonare	15	rialzare	11
spuntare	15	riaccendere	11
sputare	15	rigovernare	11
succhiare	15	rincorrere	11
suddividere	15	riparare	11
verniciare	15	rivestire	11
accelerare	14	sbucciare	11
accostare	14	sorpassare	11
attrarre	14	strozzare	11
avvolgere	14	torcere	11
cacciare-3	14	affettare	10
filare	14	balzare	10
gettare via	14	bombardare	10
integrare	14	cavalcare	10
laccare	14	filtrare	10
nutrire	14	fucilare	10
richiudere	14	includere	10
rimuovere	14	incrementare	10
sciacquare	14	inginocchiare	10
sciogliere-2	14	limare	10
sdraiare	14	lucidare	10
tirare giù	14	mascherare	10
vagare	14	masticare	10
zuccherare	14	medicare	10
ampliare	13	mischiare	10
assorbire	13	moltiplicare	10
avvitare	13	piallare	10
bollire	13	poppare	10
congelare	13	rilanciare	10
fotocopiare	13	rimbalzare	10
incollare	13	scalare	10
ingrandire	13	scarnire	10
partorire	13	seppellire	10
ripigliare	13	sfiorare	10
tinteggiare	13	sottrarre	10
truccare	13	stappare	10
vociare	13		
accavallare	12		
aggirare	12		
discendere	12		
inzuppare	12		
isolare	12		
leccare	12		
riporre	12		
scaldare	12		
schermare	12		
sfondare	12		
sgonfiare	12		
spruzzare	12		
temperare	12		
apparecchiare	12		
arare	11		
cassare	11		
friggere	11		
illustrare	11		

Tabella 5: Verbi italiani di azione ad alta frequenza

Tabella 6: Verbi inglesi di azione ad alta frequenza (entries in frequency order, read column by column)

LEMMA	FREQ.	LEMMA	FREQ.
take	4006	ring	210
put	3231	lead	208
give	3030	fuck	205
work	2274	fill	199
start	1788	apply	198
keep	1282	introduce	196
leave	1272	face	195
feel	1201	prepare	195
move	1004	sign	188
bring	851	catch	187
write	795	throw	184
help	785	lie	180
run	731	present	176
show	724	kill	172
change	714	hit	164
turn	702	collect	160
speak	698	record	160
read	691	wash	155
sit	658	drop	154
stop	638	divide	151
play	615	rise	147
send	590	sleep	146
pick	524	fly	144
sort	521	push	143
carry	489	form	142
hold	463	fight	141
walk	441	serve	141
set	434	train	139
build	423	square	137
follow	420	treat	136
add	403	organize	133
eat	400	attend	132
cut	392	feed	131
listen	381	phone	129
produce	351	hurt	128
watch	349	spell	126
pass	339	lay	126
bear	333	count	125
break	307	knock	125
cover	306	clean	123
fall	305	mark	123
draw	285	point	123
round	283	press	122
receive	282	clear	117
include	280	enter	116
join	274	cross	114
wear	269	block	112
tend	267	travel	109
stick	260	reach	109
open	258	sing	109
raise	241	act	107
support	240	recording	104
shut	233	touch	103
drive	232	control	101
hang	231	tape	100
bother	223	operate	100
close	222	visit	97
pull	220	burn	96
		hand	96
		place	96
		switch	96
remove	95	fold	51
blow	94	lock	49
arrange	93	aim	49
charge	91	pack	49
print	91	demonstrate	48
contain	90	pour	48
shout	89	roll	48
copy	88	attack	47
maintain	88	destroy	47
laugh	86	direct	47
drink	85	land	47
paint	84	restrict	47
smoke	84	trade	47
replace	83	withdraw	46
cook	82	retain	45
protect	82	surround	45
deliver	81	wake	45
link	81	escape	44
shoot	81	expand	44
fix	80	circulate	44
dry	79	generate	44
measure	79	practise	44
tie	79	beg	43
gather	77	boil	43
mix	77	ride	43
jump	77	park	42
beat	76	pursue	42
cry	74	reveal	41
heat	72	rush	41
avoid	72	smash	41
alter	70	pray	40
dress	70	shit	40
spread	70	hide	39
extend	69	slip	39
chat	68	steal	39
pop	66	file	39
design	65	rub	38
indicate	65	swim	38
lift	65	plough	38
combine	63	damage	37
perform	62	colour	36
approach	61	climb	36
prevent	61	float	36
cast	61	spin	36
cough	61	squeeze	36
shed	61	tear	36
stretch	60	unite	36
tackle	60	explore	36
bowl	60	illustrate	35
bind	60	underline	35
breathe	60	warm	35
wind	60	bake	35
fire	60	secure	35
manufacture	57	bury	34
release	56	chop	34
fashion	54	ease	34
fetch	54	outline	34
knit	54	swallow	34
attach	54	tap	34
strike	53	wipe	34
light	52	smile	34
chase	52	assist	34
transfer	52	construct	33
relax	52	exercise	33
connect	51	conduct	33
step	51	score	33
repair	32	wander	24
shape	32	dot	24
rest	32	measured	24
slow	32	ruin	24
drag	32	adjust	23
distribute	31	force	23
pump	31	incline	23
rain	31	pinch	23
store	31	reverse	23
dissolve	30	search	23
emerge	30	separate	23
stuff	30	wrap	23
type	30	line	23
scream	30	creep	22
mess	29	empty	22
milk	29	hook	22
shake	29	injure	22
sink	29	kiss	22
stain	29	label	22
suck	29	access	22
taste	29	cool	22
weigh	29	cure	22
tick	29	interrupt	22
grab	29	preserve	22
crack	28	bore	21
launch	28	extract	21
sail	28	range	21
scratch	28	rip	21
struggle	28	scale	21
weave	28	seal	21
freeze	28	spring	21
absorb	27	boost	20
brush	27	chuck	20
occupy	27	divert	20
post	27	drain	20
arrest	27	erase	20
bless	27	flip	20
lodge	27	shock	20
mend	27	stamp	20
flow	26	starve	20
insert	26	trace	20
shove	26	twist	20
slide	26	pose	20
balance	26	translate	19
dictate	26	advance	19
pin	26	box	19
resist	26	capture	19
dole	26	discharge	19
bend	25	expose	19
bite	25		
clip	25		
confine	25		
nick	25		
overlap	25		
plant	25		
scrub	25		
skip	25		
sweep	25		
abandon	24		
strengthen	24		
decorate	24		
display	24		
flash	24		
free	24		
halve	24		
plug	18	transcribe	15
race	18	grasp	15
shower	18	accompany	14
stir	18	assault	14
transport	18	bash	14
canvass	18	bomb	14
model	18	constrain	14
relieve	18	lick	14
service	18	march	14
swell	18	pile	14
bang	18	rearrange	14
bolt	17	rebuild	14
bounce	17	scrape	14
collapse	17	smack	14
dispose	17	tuck	14
glaze	17	scrap	14
group	17	carve	13
guide	17	dance	13
narrow	17	explode	13
piss	17	glue	13
reinforce	17	grind	13
shine	17	hunt	13
squash	17	leak	13
strip	17	spit	13
trigger	17	tidy	13
baptize	17	water	13
restore	17	whistle	13
substitute	17	widen	13
disturb	16	wire	13
don	16	practice	13
guard	16	amalgamate	12
isolate	16	babysit	12
overtake	16	burst	12
photocopy	16	chew	12
prescribe	16	choke	12
ship	16	crucify	12
hammer	15	dash	12
load	15	demolish	12
murder	15	dip	12
project	15	exchange	12
rape	15	install	12
screw	15	leap	12
soak	15	pan	12
swap	15	refurbish	12
swing	15	rescue	12
tip	15	seat	12
mug	15	sew	12
top	15	slice	12
		spool	12
		telephone	12
		tickle	12
		toss	12
		trap	12
		unload	12
		weep	12
		deteriorate	12
		evacuate	12
		muck	12
		deposit	11
		drill	11

Tabella 6: Verbi inglesi di azione ad alta frequenza

Fictive self-quotation: quantitative and qualitative aspects of fictivity in European and Brazilian Portuguese

Luiz Fernando Matos ROCHA

UFJF

Rua José Lourenço Kelmer, s/n - Campus Universitário, 36036-900, Juiz de Fora - MG,
luiz.rocha@ufjf.edu.br

Abstract

Studies on fictivity point out that certain linguistic expressions are only indirectly related to their meant referents and that unreal scene is often presented by language users as a means of mentally accessing the real scene. By overlapping cognitive and interactional frames, the fictive self-quotation phenomenon is a discursive type of fictivity, by which its conceptualisers pose a subjectifying assessing perspective to the direct speech in the first person. The objective of this work is to analyse fictive self-quotation and its factive co-extension in oral corpora of European and Brazilian Portuguese, focusing on the construction "(I) said X-clause". As for the data, the C-ORAL-ROM Portuguese corpus (Bacelar do Nascimento *et al*., 2005), the C-ORAL Brazilian corpus (Raso & Mello, 2010, 2012), and a database from the reality show Big Brother Brasil (2002) are used, all of which subjected to electronic tools. The results point out meaningful conceptual, diatopic and diaphasic contrasts between the uses of "disse" and "falei" in the national varieties, since the verb "falar" is not often used to build a reported speech mental space in the European Portuguese and that, from a constructional standpoint, certain interactional frames seem to favour fictive self-quotation more promptly.

Keywords: cognition; fictivity; reported speech; self-quotation.

1. Introduction

Studies on fictivity (Talmy, 1996, 2000; Langacker, 1991, 1999, 2008; Pascual, 2006; Brandt, 2010) point out that certain linguistic expressions are only indirectly related to their meant referents and that unreal scene is often presented by language users as a means of mentally accessing the real scene. In the example "The fence stretches from the plateau to the valley", part of our cognition perceives the image of an object moving, following the path from the plateau to the valley. Nevertheless, another part of our cognition assesses this image as unreal, relying on the conception that nothing in the scene is actually moving. Regarding this kind of cognitive conflict, the image assessed as unreal is fictive.

By overlapping cognitive and interactional frames, the fictive self-quotation phenomenon is a discursive type of fictivity, by which its conceptualisers pose a subjectifying assessing perspective to the direct speech in the first person, differently from its factive counterpart. This is mainly due to the mismatched use between the traditional way of reporting self-speech and thought and the meaning of dicendi verbs like "dizer" and "falar", which take an exclusively epistemic status (e.g. "I said (thought) "Oh, God!"). Therefore, by means of an unreal scene of discourse reporting, the illocutionary agent reports himself to a previous and assumed speech scene, aiming at allowing mental access to the real scene of thought.

The historical methodological track followed by the studies on fictivity is analogous to the one made by Cognitive Linguistics as a whole. It begins with works which are solely based on the linguists' intuition, who developed epistemological constructs prompted by both imagery and linguistic illustrations, either made up or faked, though plausible, for postulating both psychological and cognitive state of affairs. Within this context, the main objective of this work is to describe and analyse fictive self-quotation and its factive co-extension in oral corpora of European and Brazilian Portuguese, focusing on the construction "(I) said X-clause", devoid of any directional phrases (Goldberg, 1995) or active zones (Langacker, 1991), which would unquestionably point to its factive interpretation.

As for the data, the C-ORAL-ROM Portuguese corpus (Bacelar do Nascimento *et al*., 2005) and the C-ORAL Brazilian corpus (Raso & Mello, 2010, 2012) are used, as they have similar basic architectures. A database from the reality show Big Brother Brasil (2002) is also used. They were subjected to the TextSTAT or Contextes electronic tools. On the whole, the results point out meaningful conceptual, diatopic and diaphasic contrasts between the uses of "disse" and "falei" in the national varieties, since the verb "falar" is not often used to build a reported speech mental space in the European Portuguese and that, from a constructional standpoint, certain interactional frames seem to favour fictive self-quotation more promptly, as in the case of the reality show.

However, from a discursive point of view, fictivity affects self-quotation in both varieties of the Portuguese language, mapped by clues which include monological self-report, subjectification, epistemic co-text, deictic mismatch, mental scanning, the metaphor "THINKING IS SAYING" (Rocha, 2004, 2006, 2010), speech acts such as promises, planning and appreciation. Such signs form a set of semantic and pragmatic trends extracted from the one-to-one case analysis of real interactions, making interactional and cognitive frames to converge, thus supporting the multidimensional feature of the phenomenon, basically split into epistemic and pragmatic dimensions.

This contributes to an innovative view on fictivity which, according to Talmy (2000), only refers to cognitive conflicts between discrepant (fictive and factive) ways of perceiving or conceiving the same object. On the

other hand, if we take into consideration the associative force between a given construction and a given lexical item, and if we treat it from a discursive standpoint, we conclude that a fictive cognitive frame is evoked whenever a fictive interactional frame is.

2. Fictive and Factive self-quotation

The present study investigates how discursive and prosodic aspects contribute to the recognizing of fictive self-quotation as a virtual instance of direct speech, a grammatical construction, whose features are indirectly tied with the referents, referring to the worlds, entities mentally constructed, as well as the exclusively epistemic events. Fictive self-quotation is a kind of mismatch between form and meaning. This case represents form–function mappings which are "incongruent with respect to more general patterns of correspondence in the language" (cf. Francis & Michaelis, 2003: 2). Since this construction is a non-canonical pattern, it can be a direct consequence of a grammaticalization process and mainly a product of general fictivity pattern (Talmy, 1996: 212), in which "two discrepant representations disagree with respect to some single dimension, representing opposite poles of the dimension". That is: FACTIVE AND FICTIVE SELF-QUOTATION.

We can find similar examples like these in English, as in Henry Kravis' interview:

Henry Kravis' interview (1)

FICTIVE SELFQUOTATION (FIC-SELF): My dad was reading an article in *Time* magazine about the Oxford/Cambridge of the West Coast. It's part of a group of small colleges in Claremont, along with Pomona, Scripps, and Harvey Mudd. I wanted to go to the West Coast. I'm from Oklahoma originally, but I had been in an Eastern boarding school for five years and **I said, "I want to see how the other half of the United States lives."** I tell people I went there to play competitive golf. I liked it. I used to say the first year was like a prep school with ash trays. I really went there because it was very strong in economics and political science, and those were the two areas that I wanted to focus my future on. (http://www.achievement.org/autodoc/page/kra0 int-1)

In the boldface fragment the verb "said" has an epistemic meaning, as "think" or "consider". "Said" is a *dicendi* and *sentiendi* verb at the same time. But it is not in the next example:

Henry Kravis interview (2)

FACTIVE SELFQUOTATION (FAC-SELF): After I graduated from college, that summer, I was given a job at the Madison Fund, which was a closed-end mutual fund here in New York. Ed Merkle ran it. What a terrific guy he was! After I was there for about three weeks, he said, "Kid," (they used to call me kid all the time), "I want you to go out and call on a company called Tri-State Motor Transit, in Joplin, Missouri. **And I said, "That's interesting, but who is going to go with me?"** He said, "What do you mean, who is going to go with you? You are going to go by yourself. (http://www.achievement.org/autodoc/page/kra0 int-1)

In this case, "said" is just *dicendi*. It is not an epistemic use.

There are some discursive and prosodic clues which suggest that fictive selfquotation (FIC-SELF) is abnormal in relation to canonical factive self-quotation (FAC-SELF) although FIC-SELF keeps some features inherited from this traditional pattern, as we see in the next picture. Because of it, there is a dotted arrow linking FIC-SELF and FAC-SELF as a continuum. This process involves some grammatical means of coding formal, semantic or pragmatic functional domains. In terms of argumental structure, both cases are the same (I SAID X-clause). But the last feature is different when we submitted data to PRAAT, a free scientific software program for the analysis of speech in phonetics.

Formal tendencies:

FIC-SELF   < ------ >   FAC-SELF

FICTIVE	FACTIVE
Subject + Sentiendi/dicendi verb + Speech clause (direct object)	Subject + Dicendi verb + Speech clause (direct object)
Tendecy: verb in the past tense or in historical present	Tendecy: verb in the past tense or in historical present
No complementizer (direct speech)	No complementizer (direct speech)
Prosody (1)	Prosody (2)

Table 1: Subjetive and factive

Considering the scope of tested fragments made by Professor Pablo Arantes, from Federal University of Minas Gerais (Brazil), fictive selfquotation is different from the factive one in some aspects. Such difference is provided by the comparison between five factive selfquotation occurences and four fictive self-quotation occurences. All these instances were uttered by male voices and extracted from Brazilian reality shows available on You Tube. According to the nine examples, in terms of fundamental frequency movement, which means a major acoustic manifestation of suprasegmental structures such as tone, pitch accent, and intonation, there is no outstanding differences between both selfquotations. In general, fictive and factive selfquotation show soft curves.

Even though this corpus is small, in global sense, it

shows consistent differences in terms of (i) register, a quality voice element whose purpose can make speech more expressive, and emphatic; and (ii) tessitura, a speech melody element whose melodic height variations represent cohesive function. Fictive selfquotation curves occupy low tone region (bass-pitched). Factive selfquotation curves occupy high tone region. These numbers are statistically meaningful and contribute to the fact that we have distinct vocal construals. Besides, the variability of F0 is different in both cases. In Factive selfquotation, there is more F0 curve variance than in the fictive one. As a robust and perceptual parameter, the variation range of curves in each selfquotation is too different: fictive cases (6.8 semitones); factive cases (13.8 semitones), which means there are distinct kinds of half step, as the interval between two adjacent notes in music.

The graphic below shows F0 curves of factive and fictive according to time normalization technique, whose purpose is to try to set up equivalence among sentences with different extensions and facilitate direct comparison among different points of F0 curves making them similar. Basically, on the left, this graphic presents five factive curves that occupy a large extension in terms of hertz; on the right, the four fictive cases do not. This means more tone variability in factive cases than in fictive ones.
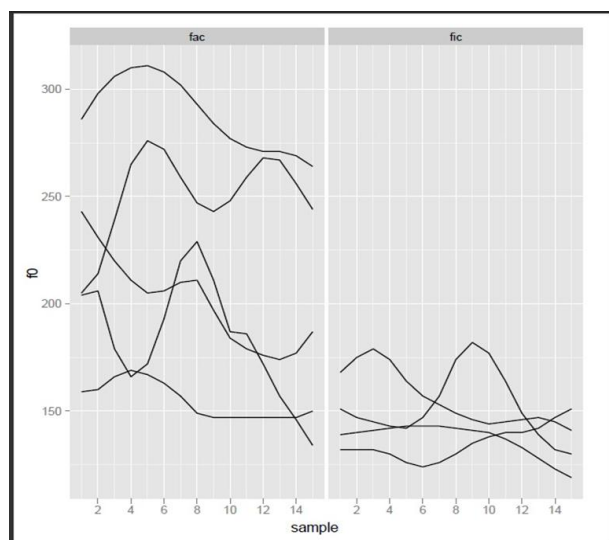
Picture 1: F0 curves of factive and fictive occurences

3.   Meaning tendencies

In this section, we have a comparison between meaning tendencies of fictive and factive self-quotation, which we have found in the corpora:

1) FIC-SELF and FAC-SELF constructions occur mainly in narrative textual types;
2) The frame of reporting scenario is monologic in FIC-SELF; in FAC-SELF, dialogic;
3) There is previous co-textual information before fictive selfquotation, like other epistemic verbs; in FAC-SELF, there is none;
4) In FIC-SELF, there is an epistemic space-builder

whose semantic value is sentiendi and dicendi at the same time in the sense of "think" or "consider"; but in the factive case, this value is only dicendi; in FIC-SELF, there is the metaphor THINKING IS SAYING and the metonymy SAYING FOR THINKING.

5) The first one evokes an assessing frame and the second one a speech communication frame;
6) Fictive selquotation tends to present speech acts in terms of promising, planning, evaluation, and concluding; factive tends to present speech acts in terms of requests, advice, suggestion, instruction, and asserting;
7) Considering all the scenario around the verb "falei" or "disse" in corpora, there is a strong tendency: fictive self-quotation is pairing with a fellowship face. On the other hand, factive selfquotation is pairing with competence face;
8) In fictive self-quotation, addressee in reported narrative is the speaker himself; but in factive, it is another character;
9) In fictive, vocative is a generic entity, for example, "Deus" (God), "gente" (folks), but in factive, we commonly have a person's name;
10) Even though we do not find such clues, deixis phenomena in the embedded clause can help us to distinguish both constructions. Let us see an example:

BRAZILIAN PORTUGUESE:
JUL: <teve um dia que alguém me falou assim / Nossa / cê tá velha / hein / sua menina tá com dez anos / **eu falei / velha é ela** //
   (C-ORAL Brasil - RASO & MELLO, 2012)

TRANSLATION:
JUL: someday someone **told** me: "You're old! Your daughter is ten!". **I said: she is old!**

The exchange of "you", second person, as "she", third person, in the X-clause (VELHA É ELA = SHE IS OLD, not YOU ARE OLD) becomes the direct speech (I said: she is old) a fictive self-quotation, although we have a previous direct speech frame: "someday someone **told** me: 'You're old! Your daughter is ten!'." The third-person deixis 'she' is inconsistent with that scenario marked by past tense verbs "told" and "said". Besides, if it would be a case of factive self-quotation, in the reported interaction, the speaker JUL would have to use YOU and to say: **YOU ARE OLD!**, as the character "someone" does. It means we have just one clue to read all the self-quotation as fictive, which is discrepant with respect to a single deitic dimension.

4.   Quantitative analysis

For the quantitative analysis from those corpora, I have searched the pattern (EU) DISSE/FALEI X-ORACIONAL (I SAID X-clause, in English) to find self-quotations in first person, using TextSTAT

concordance software and Contexts concordance from C-ORAL-ROM project.

In European Portuguese, the verb "falar" (to say), in general, does not profile dicendi substructure. In this sense, it is similar to the verb "speak", in English. In European Portuguese, this function belongs to the verb "dizer" (to say). In Brazilian Portuguese, the verbs "dizer" and "falar" can profile dicendi substructure. In relation to selfquotation, all these numbers that we will see signalize important contrasts between national varieties of Portuguese, for example, the preference for "dizer" instead "falar" as a dicendi verb in European Portuguese than in Brazilian one. The former profiles a punctual process of demonstration by word of beliefs and convinced attitudes. The latter profiles a general process of verbalization, which refers to skills and abilities of speech production.

In European Portuguese Corpus, we have found 50 types of the pattern (EU) DISSE X-ORACIONAL, being 44 FAC-SELFS and only six FIC-SELFs. The mainly reason for that is the specificity of this pattern, which is semi-instantiated. On the other hand, in the same corpus, we have found just 21 occurences of "falei" associated with prepositional phrase in general, which for us means that there is no dicendi function. This is a kind of counter-evidence of FIC-SELF.

In terms of Brazilian Corpus, the word form "disse" (I said) occurs two times, being two cases of FAC-SELF and there are no FIC-SELF cases with this form. But the word form "falei" (I said) occurs 351 times, being 153 instances integrated to FAC-SELF scenarios and 68 to a FIC-SELF ones.

Considering other data, a Brazilian Reality Show (2002), in four and half hours of continuous recording, we found 69 occurrences of direct speech in first person with verb "falei" ("I said", in English); 43 are cases of fictive self-quotation and 26 are cases of factive self-quotation. These numbers can't be understood as a mere generalization. It signals that we use it a lot, depending on the interactional frame. Note that in a reality show, reported speech frame is a powerful and pervasive construction as "war" strategy. In this sense, fictive selfquotation justifies the reporting thoughts through an epistemic and discrepant use of "say" ("falei") with the purpose of profiling more action and confidence than the mere use of "think" or "consider".

5.   Conclusion

It is important to highlight that the abundance of virtual computational architectures to study linguistics has a single purpose: to gain more precise access to language and to what is psychologically real in processing it. In other words, the fictivity of the proper linguistics investigation seems to be the current point of no return in the history of linguistics. In the case of this work, PRAAT and Corpus Linguistics instruments have permitted that fictive selfquotation is understood as a phenomenon which depends on its integrated features to be mapped. With PRAAT, we can say that fictivity has an specific

melody when we can constrast it with factivity occurences through similar constructional patterns. With Corpus Linguistics, we can see the integration of grammatical constructions with discourse more clearly; and show in more details how it happens; and verify how the conceptualizer sets up alternatives forms of construal for the same referent or situation, conventionalizing language changes. The comparison between European Portuguese Corpus and Brazilian Portuguese Corpus has revealed that both national varieties have their proper way of profiling fictive selfquotation. As their corpora architectures are the same (both under C-ORAL-ROM project), the numbers of fictive self-quotation occurences are not very different proportionally, but when we compare these corpora with another one (reality show), we can see how fictivity depends on the interactional frame to be more or less productive. Cognitive frames of fictivity are strongly in action when interactional frames of fictivity are in action, too.

6.   Acknowledgements

7.   References

Bacelar do Nascimento, M.F. *et al*. The Portuguese Corpus. (2005). In E. Cresti, M. Moneglia (Eds.), *C-ORAL-ROM*, *Integrated Reference Corpora of Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins, pp. 163--207.

Brandt, L. (2010). *Language and enunciation* – A cognitive inquiry with special focus on conceptual integration in semiotic meaning construction. Faculty of the Humanities: Aarhus University.

Francis, E.J., Michaelis, L.A. (Eds.). (2003). *Mismatch*: *form-function incongruity and the architecture of grammar*. Stanford: CSLI Publications.

Goldberg, A. (1995). *Constructions*: *a construction grammar approach to argument structure*. Chicago: The University of Chicago Press.

Langacker, R.W. (1991). *Concept, Image, and Symbol: The Cognitive Basis of Grammar.* Berlin and New York: Mouton de Gruyter.

Langacker, R.W. (1999). Virtual reality.*Studies in the*

Linguistics Sciences. V. 29, n. 2.

Langacker, R.W.   (2008). *Cognitive Grammar*: a basic introduction. New York: Oxford University Press.

Pascual, E. (2006). Fictive interaction within the sentence: a communicative type of fictivity in grammar. *Cognitive Linguistics* 17–2, 245–267.

Raso, T., Mello, H.R. (2010). The C-ORAL-BRASIL corpus. In M. Moneglia, A. Panunzi (Eds.) *Bootstrapping information from corpora in a crosslinguistic perspective*. 1ed. Firenze: Firenze University Press, 2010, v. 1, pp. 193--213.

Raso, T., Mello, H.R. (2012). *C-ORAL-BRASIL 1*: *corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG.

Rocha, L.F.M. (2004). A construção da mímesis no reality show: uma abordagem sociocognitivista para o discurso reportado. Tese (Doutorado em Linguística). Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Rocha, L.F.M. (2006). A fala silenciosa reportada: metáfora, metonímia e mesclagem. In *Linguística*, Rio de Janeiro, v. 2, n. 1, pp. 23--38.

Rocha, L.F.M. (2010). Fictivity and Factivity in Brasilian Portuguese: prosodic and pragmatic aspects in a socio-cognitive approach to reported speech. In *Programme and abstracts of International Conference Pluricentric Languages: linguistics variation and sociocognitive dimensions*. Braga: Universidade Católica Portuguesa 2010, pp. 152--154.

Talmy, L. (1996). Fictive motion in Language and "Ception". In P. Bloom, M.A. Peterson and M.F. Garrett (Eds.), *Language and Space*. MIT Press: Cambridge, pp. 211--276.

Talmy, L. (2000). *Toward a cognitive semantics.* 2 volumes. Cambridge, MA: Massachusetts Institute of Technology.

Modeling the grammar of ASSESSMENT of casual conversation in Brazilian Portuguese: the design of a corpus to investigate language probabilities functioning in context

Giacomo Patrocinio FIGUEIREDO, Adriana Silvina PAGANO

Federal University of Ouro Preto; Federal University of Minas Gerais

Rua do Seminário, s/n, Mariana, MG, Brazil 35420-000

giacomojakob@yahoo.ca, apagano@ufmg.br

Abstract

This paper presents a corpus-based model for the interpersonal system of ASSESSMENT in the clause grammar of casual conversation (Eggins & Slade, 1997) in Brazilian Portuguese. More specifically, it examines Modal Particle use. Data were obtained from a sample of casual conversation retrieved from CALIBRA, a monolingual corpus of Brazilian Portuguese designed following a context-based typology of texts. The texts were analyzed according to systemic functional theory categories (Halliday & Matthiessen, 2004; Figueredo, 2011) and semi-automatically annotated for grammar categories with the software CorpusTools (O'Donnell, 2008). Variation in the patterns of Particle use was found for the whole corpus and those in the subsection where casual conversation is located. Results pointed to a more frequent use of Modal Particles for Assent, Understand, Confirm and Conclude and therefore to a more intense contribution of those categories to the process of negotiation among interactants in casual conversation. On the other hand, Modal Particles related to the systems of PERSUASION and PROSODY were observed to contribute less to the variation found in casual conversation.

Keywords: casual conversation; assessment; modal particles; monolingual corpus; Brazilian Portuguese.

1. Introduction

Drawing on the notion of *probabilistic grammar* (Halliday, 1991), this paper presents a corpus-based model for the interpersonal system of ASSESSMENT in the clause grammar of casual conversation (Eggins & Slade, 1997) in Brazilian Portuguese. 'Modeling', as construed here, can be defined as the description of grammar features and statement of their probabilities of instantiation for the text type under investigation. Halliday (1978) conceives of language as a naturally evolved semiotic system, its main purpose being to offer a reservoir of meaning-making resources for humans to interpret and organize both our natural world and our social relations. Grammar is, in turn, the stratum of language responsible for creating meaning. Since meaning is, in fact, the contrast of paradigmatic features (Saussure, 2006), for any given language subsystem, the job done by the grammar is to change (responding to the pressure of new contextual demands) the systemic (paradigmatic) organization of features in order to create meaning. This process of specialization leads to language variation. As a result, language is modeled in terms of (Halliday, 1991): (i) its relations to the context of culture – the "environment" in which it takes place, in which it is meaningful; and (ii) the process through which language as a reservoir of meaning-making potential (the system) becomes, via grammar operations, language in context (the text). Consequently, the modeling of "actual" grammar – the grammar that creates meaning functioning in the context of situation – needs to account for (a) the way context is materialized in language (examining the systemic dimension of realization) and (b) the probabilities for a potential grammatical feature to be instantiated as text (the dimension of instantiation) (Halliday, 1991). Thus, to model the contextual pressure that ultimately causes language variation – in other words,

to model any given text type – including casual conversation, it is necessary to account for the dimensions of realization and instantiation. Following Halliday's (1978) conceptualization of language, a great number of studies have explored grammar from a realizational point of view (cf. Martin, 1992; Caffarel, Martin & Matthiessen, 2004, among others and Eggins & Slade, 1997, specifically for casual conversation). A smaller number of studies have explored the instantiational process (cf. Matthiessen, 2001; Martin, 2008, among others). There are fewer studies still drawing on the realization-instantiation complementarity (Matthiessen, 2004). To a large extent, this is due to the fact that the process of instantiation leading to the modeling of specific text types is not fully understood (Martin, 2008). By presenting a modeling of casual conversation interpersonal grammar systems, this paper aims at exploring the complementarity of realization and instantiation, as well as contributing to the understanding of probabilities in the constitution of text types. More specifically, it presents a study of the interpersonal grammatical system of ASSESSMENT in Brazilian Portuguese, including its distribution across text types and relates that to the distribution of ASSESSMENT functions in casual conversation. Such relation can ultimately lead to the modeling of casual conversation in Brazilian Portuguese and contribute to consolidating corpus-based investigation as a necessary step towards the understanding of the instantiation process.

2. Theoretical underpinnings

2.1 The design of a corpus to investigate language probabilities functioning in context

Drawing on the concept of text as "language functioning in context" (Halliday & Hasan, 1976), Matthiessen, Teruya & Wu (2008) propose a typology based on the

contextual variables of field (type of social action), tenor (role relationships between speaker and listener) and mode (role played by language). More specifically, they model their typology on specific parameters of field and mode, namely the field parameter of socio-semiotic process and the mode parameters of medium and turn. By socio-semiotic process, they mean the uses to which language is put in order to fulfill a social activity. These are eight: doing (using language in an ancillary form to perform a social activity); exploring (comparing different positions and arguing for one of them); expounding (taxonomizing and explaining phenomena); reporting (chronicling phenomena); recreating (recounting and narrating activities in other socio-semiotic processes); sharing (negotiating and calibrating interpersonal relations); recommending (advising on a course of action), and enabling (instructing and regulating behavior). Each process has a particular configuration of tenor and mode. This has to do with whether language use involves specialization (specialized/non-specialized), with the role of language in situation (ancillary/constitutive), the mode of production (written/spoken) and the turns in interaction (monologue/dialogue). Table 1 displays the main parameters of a context-based typology and provides examples of prototypical text types for each variety.

LANGUAGE USE		SOCIO-SEMIOTIC PROCESS	MODE			
			WRITTEN		SPOKEN	
			DIALOGUE	MONOLOGUE	MONOLOGUE	DIALOGUE
Specialized	**reflection**	expounding	*letter* *exam*	*textbook* *research article*	*lecture* *plenary*	*debate* *tutorial*
		exploring	*letter to editor*	*review* *editorial*	*speech*	*panel discussion*
		recommending	*agony aunt letter* *promotional letter*	*ad* *blurb*	*prayer*	*consultation*
		enabling	*regulation, law procedures*	*open letter*	*sermon*	*demonstration*
Non-specialized		reporting	*questionnaire*	*news report* *recount* *biography*	*statement*	*media interview*
		recreating	*cartoons*	*novel* *short story*	*anecdote*	*theatre play*
		sharing	*e-mail*	*blog* *diary*	*reminiscence*	*gossip, chat*
	action	doing	*business letter* *invitation*	*shopping list*	*ceremony*	*service encounter*

Table 1: Context-based typology

A corpus design based on the typology above allows for the study of language frequencies of grammatical systems, both globally in the language system as a whole, and "broken down" according to typological features of language in the context of culture (Halliday, 1992). CALIBRA, which stands for Catálogo da Língua Brasileira, is one such corpus, designed on the basis of the language typology proposed in Matthiessen, Teruya & Wu (2008). CALIBRA is a monolingual corpus of Brazilian Portuguese, which compiles language produced in a natural communicative setting and representative with respect to each of the socio-semiotic processes mentioned above. It is a raw corpus with minimal header annotation and encoding in UTF-8. Texts compiled in CALIBRA were produced within the 1990-2010 decades. As regards the spoken mode, texts were recorded from spontaneous speech and subsequently transcribed to be incorporated. The corpus design allows for mapping a particular language variety. For the purposes of the present study, which targets casual conversation, texts can be located in the typology as non-specialized, spoken, dialogic texts within the sharing process. A detailed account of this variety is provided in the following sections.

2.2 Casual conversation

As a species, human beings are part of the animal world. This means that our biological constitution needs food and shelter; safety and companionship. No human can live their whole life alone apart from other humans. It is also part of our species programing to be able to keep track and record of time towards the past by building and storing personal and collective memories and to the future by predicting, planning and realizing projects, such as finding food, building shelter, or maintaining relationships. As a result, our biology determines only partly what humans are, since it is embedded in our social world and in our history – not only individual histories of each human being, but the history of our social world

(Malinowski, 1935).

The shaping of biology [by society [shaped by history] lies at the core of a functional theory of culture. The process is called *symbolic modeling* and ultimately explains why 'mating' becomes 'marriage', a 'pack' becomes a 'family', and 'feeding' becomes a 'dinner party'. Culture, then, is a symbolic system of conditioning for human beings, turning the specimens into people with a place in society for a given period of history.

Language has a crucial part to play in symbolic modeling. It is through language that culture conditions human beings. Education, the law, religion and all institutions responsible for passing on a means of survival, a code of values and so on to the next generation are all fully dependent on language. Malinowski (1935) states that language creates the symbols of a social group, it organizes institutions by developing particular discourses and stores knowledge in the texts that are taught and shared among its members.

Casual conversation, thus, assumes a special status in this process, given that it responds for creating and passing on knowledge and values efficiently in a very specific context – that of people who are closest to each other. Eggins and Slade (1997) state that casual conversation is a resource frequently deployed in negotiating our social identity and establishing our "social geography" – the people (along with their values and social relations) who are close or distant from us. The reiteration and multiplication of such texts through a period of time contribute to social stratification and distribution of power among people in a social group.

2.2.1. The grammar of casual conversation

Language can serve as the most resourceful tool in symbolic modeling because it has a grammar (Halliday, 1978). Semiotic systems are bi-stratal, in which a symbol is characterized by the univocal correspondence between its content plane ("semantics") and its expression plane ("phonetics") (Saussure, 2006). Language, however, has evolved to formally organize the content (Hjelmslev, 1969). The content is, as it were, divided into two: the substance of content (semantics) and the form of content (grammar). Grammar, then, is defined as the formal organization of language content plane. Consequently, the meaning of a linguistic symbol is not conveyed by the univocal correspondence between content and expression; rather, the understanding of content can change depending on its formal organization. Since meaning is, in fact, a paradigmatic contrast between symbols, grammar operates altering the organization of systems in order to create new meanings. Whenever there is need for a reshaping of some aspect of human life – different aspects of symbolic modeling – there is also a contextual pressure for new meanings and new texts. Grammar reorganizes features of systems, changing both their paradigmatic contrast and their probability, thus creating new meanings through variation of text types.

The grammar of casual conversation is one example of such process. Responding to the contextual pressure of negotiating social identity and drawing social geography maps, the grammar of casual conversation has created meanings to materialize such contexts (cf. Eggins & Slade, 1997). For example, interpersonal systems (MOOD, MODALITY, ASSESSMENT and POLARITY) are deployed to establish a "sympathy relation" towards the speaker's values and positions. Ideational systems (TRANSITIVITY and EPITHESIS) help building the narrative underlying casual conversation as well as passing judgment and ascribing voice and thought to other people. Textual systems (THEME and INFORMATION) help staging phases of casual conversation, as well as giving prominence to interpersonal and ideational systems relevant to the construction of typical features of casual conversation such as sympathy, narrative and judgment (cf. Eggins & Slade, 1997).

From the point of view of instantiation, the grammar changes the typical, non-prominent, ratio of feature instantiation, due to contextual pressure. Although fewer features in relation to the whole of the system are deployed, these are relatively more frequent in casual conversation. One such case is found in the interpersonal system of ASSESSMENT in Brazilian Portuguese.

2.2.2. The system of ASSESSMENT

Any interaction between people can be viewed as the negotiated process of converting interactants' [personal] opinions into [interpersonal] shared knowledge. The amount of opinion converted into shared knowledge is likely to determine the social proximity/distance among interactants for a given interaction. By the same token it is likely to indicate distribution of power, knowledge, expertise, authority, etc., contributing to determine their social identity. In general, a speaker tends deploy resources (from systems such as MODALITY, ASSESSMENT, INTONATION, etc.) which may increase the chances of his/her opinions being accepted. In this sense, the concept of 'valid or not-valid' is a very important feature of social relations, since it is an outcome of negotiation (cf. Halliday & Matthiessen, 2004). The interpersonal grammar deploys a set of sub-systems precisely to negotiate positioning, power, values and "social geography". These are collectively responsible for exchanging evaluation and can be characterized by two features: (i) extension of evaluation – the speaker marks his/her position towards what s/he is saying; (ii) orientation of evaluation – the speaker marks his/her position towards his/her own role as speaker, or demands an assessment from the listener to do so. The interpersonal systems mostly associated with (i) are MODALITY, POLARITY and partially MODAL ADJUNCTS (mood and comment). The interpersonal systems associated with (ii) are partially MODAL ADJUNCTS (mood and comment) and ASSESSMENT.

Martin and White (2005: 95) describe the semantic region of engagement among interactants as:

> "when speakers/writers announce their own attitudinal positions they not only

self-expressively 'speak their own mind', but simultaneously invite others to endorse and to share with them the feelings, tastes or normative assessments they are announcing. Thus declarations of attitude are dialogically directed towards aligning the addressee into a community of shared value and belief".

Engagement meanings, in turn, are grammaticalized by the system of ASSESSMENT, defined by Halliday and McDonald (2004) as: "*a grammatical system ... whereby the speaker signals attitude to, and degree of involvement in, the proposition or proposal of the clause* (p. 341)."

In Brazilian Portuguese, the system of ASSESSMENT is realized by Modal Particles (Lam, Figueredo & Espíndola, 2010) as displayed in Figure 1.
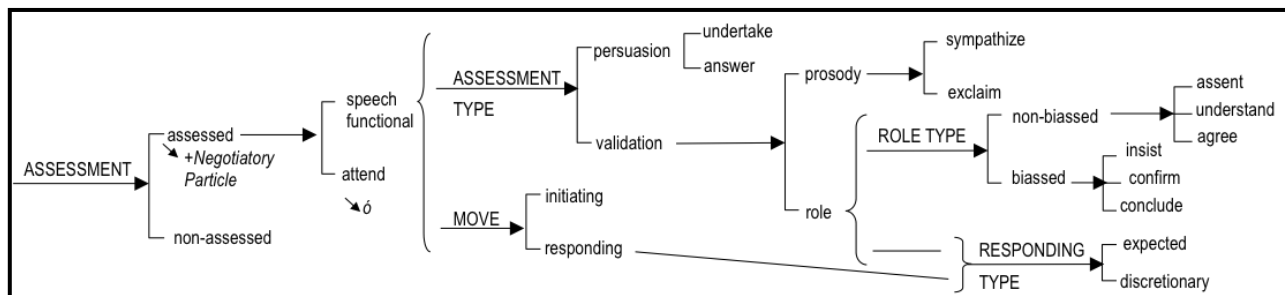
Figure 1: The system of ASSESSMENT in Brazilian Portuguese

Particles function adding further options to MOOD selection, shaping statements, questions, commands and offers according to the speaker's need for their interlocutor's assessment of a move, such as exhorting, agreeing, concluding, etc. By using Modal Particles in Brazilian Portuguese a speaker can not only assess what is being said, but also invite the listener to assess the speaker's own role as speaker [the one who assess what is being said]. Modal Particles can be more strongly associated with propositions – the exchange of information – realized by Indicative Mood; and those associated with proposals – the exchange of goods-and-services – realized by Imperative Mood. Modal Particles carry two complementary interpersonal functions in the clause: they indicate how the clause should be valued in terms of agreement, assent, exhortation, etc.; and they are picked up by the listener as a means of propelling dialogue. Examples of ASSESSMENT functions in Brazilian Portuguese retrieves from CALIBRA can be found below, a gloss and a free translation being provided for each of them.

Ó   João   você   toma   conta   deles
ATTEND   João   you   take   care   of.they
*"**Listen to me** João, you take good care of them."*

Todo mundo lá   gostava dele   **né**
All   world there liked   of.he ASSENT
*"Everybody liked him, **don't you think so too**?"*

S1 Não deve de ser para ligar para elas
   Not must of be   to   call   for   them

"We are not supposed to call them."
S2 Eu acho que é sim **sô**.

   I   think that be yes INSIST

*"**But I do** think we are."*
Vocês não voltam pra lá   **viu**

You   not return to   there   UNDERSTAND

*"You should never go back there, **is it clear**?"*
Você grava   as   minhas aulas   **é**
You   record the my   lectures CONFIRM
*"So you tape my lectures, **do you**?"*

E   eu   tava animado **sô**
And   I   was   excited   SYMPATHIZE
*"And I was **extremely** excited."*

Ôxe   quem tá ligando pra isso?
EXCLAIM   who   is caring   to   this
*"Why **on earth** would anyone care about it?!"*

Fala   **aí**   o   quê que você faz
Speak ATTENUATE   the what that you do
*"**Please**, tell me what you do."*

Então o   que   é   **tchê**?
So   the what is   EXHORT
*"**Just say** what it is."*

3.   Methodology

To model casual conversation, the following methodology was adopted. A spoken language corpus of 10,000 tokens (10 texts of 1,000 tokens) of casual conversation was compiled from CALIBRA (Catalogue of the Language of Brazil). As previously mentioned, casual conversation texts are located in the typology as non-specialized, spoken, dialogue texts within the sharing process. For CALIBRA, spoken texts, including casual conversation texts used in this research, are recorded from spontaneous conversations and subsequently transcribed. Few features are inserted into the transcriptions, including basically clause/information unit separation: '.' falling tone; '?' rising tone; '…' level tone; ',' short pause; '--' hesitation or turn-taking; '[' more than one speaker speaking simultaneously.

After compilation, texts in the casual conversation

MODELING THE GRAMMAR OF ASSESSMENT OF CASUAL CONVERSATION IN BRAZILIAN PORTUGUESE: THE DESIGN OF A CORPUS TO INVESTIGATE LANGUAGE PROBABILITIES FUNCTIONING IN CONTEXT

421

corpus were analyzed according to systemic functional theory categories (Halliday & Matthiessen, 2004; Figueredo, 2011) and semi-automatically annotated for grammar categories with the software CorpusTools (O'Donnell, 2008). This software allows researchers to annotate texts with categories of interest and retrieve their frequency along the corpus tested for statistical significance.

Drawing on Halliday (1991b), who states that counting frequencies in a text is, in fact, stating instantiation probabilities in the grammar, the frequencies obtained were analyzed in order to reach a probabilistic grammatical profile of Particles based on the generalization of frequencies found in the corpus.

4. Modeling ASSESSMENT for casual conversation in Brazilian Portuguese

The concept of 'modeling' implies that the results of a study carried out for a sample allow us to make estimates for the whole of the population. When performing a modelling of a grammar feature, an account of the functional distribution of a particular resource (realization), together with its variation across text types (instantiation), is needed. The sample in this case is defined by two complementary steps. First, a grammar description is needed, so the "strings of sounds" found in the corpus can be converted into grammar features. As a result, the corpus under investigation – the "true" corpus – is a sample of grammar patterns. When querying CALIBRA for the categories in the system of ASSESSMENT, the patterns in Table 2 were found:

Particle function	Occurrence No.	Relative frequency
assent	498	60,8%
conclude	73	8,9%
attend	61	7,4%
exclaim	60	7,3%
agree	48	5,9%
attenuate	41	5,0%
understand	12	1,5%
exhort (ans.)	10	1,2%
confirm	6	0,7%
exhort (und.)	5	0,6%
challenge	4	0,5%
sympathize	1	0,1%
TOTAL	819	100%

Table 2: Global model for ASSESSMENT

Secondly, a distribution of grammar features across text types is needed, so variation patterns can be observed. The results obtained from CALIBRA are shown in Table 3.

Based on these complementary distributions, it is possible to see if there is significant variation between the patterns in Particle use found for the language and those in casual conversation, evidenced by texts located in CALIBRA within the sharing process sharing. This variation is seen in Table 4.

Process Function	Expo	Rep	Rec	Sha	Do	Recom	Ena	Expl	TOTAL
attend	2	3	13	6	1	1	6	29	61
exhort (und.)	0	0	2	1	0	0	1	1	5
attenuate	5	1	13	5	14	0	0	3	41
exhort (ans.)	0	0	4	2	0	3	0	1	10
challenge	0	0	1	1	0	1	0	1	4
exclaim	0	3	11	18	10	2	2	14	60
sympathize	0	0	1	0	0	0	0	0	1
assent	15	110	25	142	16	33	45	112	498
understand	2	0	3	5		1	1	0	12
agree	5	3	3	2	10	0	11	14	48
confirm	0	0	1	3	1	1	0	0	6
conclude	13	0	15	24	15	0	0	6	73
TOTAL	42	120	92	209	67	42	66	181	819

Table 3: Typological variation for ASSESSMENT

Legend: EXPO= Expounding; REP = Reporting; REC = Recreating; SHA = Sharing; DO = Doing; RECOM = Recommending; ENA. Enabling; EXPL = exploring

Particle function	Sharing process (casual conversation)		Global	
	Occurrence no.	Relative frequency	Occurrence no.	Relative frequency
attend	6	2,87%	61	7,45%
exhort (und.)	1	0,48%	5	0,61%
attenuate	5	2,39%	41	5,01%
exhort (ans.)	2	0,96%	10	1,22%
challenge	1	0,48%	4	0,49%
exclaim	18	8,61%	60	7,33%
sympathize	0	0,00%	1	0,12%
assent	142	67,94%	498	60,81%
understand	5	2,39%	12	1,47%
agree	2	0,96%	48	5,86%
confirm	3	1,44%	6	0,73%
conclude	24	11,48%	73	8,91%
TOTAL	209	100,00%	819	100,00%

Table 4: ASSESSMENT model for casual conversation

The data presented in Table 4 show how casual conversation departs from the expected ratio for the whole language. In terms of ASSESSMENT more specifically, it is possible to see skewing in the use of Modal Particles, there being more frequency for Assent, Understand, Confirm and Conclude, all belonging to the sub-system of ROLE TYPE (see Figure 2, above). As a result, it is possible to estimate that this region of ASSESSMENT is contributing more intensely to the process of negotiation among interactants in casual conversation. On the other hand, there is skewing to a less frequent use for the other Modal Particles, suggesting that the sub-systems of PERSUASION and PROSODY contribute less to the variation found in casual conversation.

5. Conclusion

The results obtained for Modal Particle use in casual conversation texts drawing on a corpus sample validate the methodology used to model the interpersonal system

of ASSESSMENT and can be further applied in order to describe other grammar features in Brazilian Portuguese and state their probabilities of instantiation in a particular text type. The idea of text variation – including the modeling of grammar systems – needs to account for the small perturbation in the average feature choices for any given text. The results presented in this paper can show how ASSESSMENT is deployed in such fashion, as to point how feature choices are skewed to vary the system towards casual conversation.

6. References

Caffarel, A., Martin, J., Matthiessen, C. (Eds.) (2004). *Language typology*: *a functional perspective*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Eggins, S., Slade, D. (1997). *Analyzing casual conversation.* London: Cassell.

Figueredo, G. (2011) *Introdução ao perfil metafuncional do português brasileiro*. Federal University of Minas Gerais. (unpublished PhD dissertation).

Halliday, M.A.K. (1978). *Language as social semiotic: the social interpretation of language and meaning*. London & Baltimore: Edward Arnold & University Park Press.

Halliday, M.A.K. (1991). Corpus studies and probabilistic grammar. In K. Aijmer, B. Altenberg. (Eds), *English Corpus Linguistics: Studies in Honour of Jan Svartvik.* London: Longman.

Halliday, M. A. K., Hasan, R. (1976). *Cohesion in English.* London: Arnold.

Halliday, M. A. K., Matthiessen, C. (2004). *An introduction to functional grammar*. 3rd ed. London: Edward Arnold.

Halliday, M. A. K., McDonald, E. (2004). Metafuncional profile of the grammar of Chinese. In A. Caffarel, J. Martin, and C. Matthiessen. (Eds.), *Language typology*: a functional perspective. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hjelmslev, L. (1969). *Prolegomena to a Theory of Language* (trans F. J. Whitfield), revised English edition. Madison, WI: The University of Wisconsin Press.

Lam, M., Figueredo, G. and Espíndola, E. (2010). An SFL Contrastive Analysis of Interpersonal Particles in Cantonese and Portuguese. *36th International Systemic Functional Congress*. Tsinghua Printing House: Beijing.

Martin, J. R. (1992). *English text: system and structure*. Philadelphia and Amsterdam: John Benjamins Publishing Company.

Malinowski, B. (1935). *Coral gardens and their magic: a study of the methods of tilling the soil and of agricultural rites in the Trobriand Islands*. v.2. New York: American Book Company.

Martin. J.R. (2008). Tenderness: Realisation and instantiation in a Botswanan Town. In N. Nørgaard. (Ed.), *Systemic Functional Linguistics in Use.* Odense: Working Papers in Language and Communication.

Martin, J., White, P. (2005). *The language of evaluation*: appraisal in English. London: Palgrave.

Matthiessen, C. (2001). The environments of translation. In E. Steiner, C. Yallop. (Eds.), *Exploring translation and multilingual text production, beyond content*. Berlin & New York: Mouton de Gruyer, pp. 41--124.

Matthiessen, C. (2004). *Descriptive motifis and generalizations*. In A. Caffarel, J. Martin and C. Matthiessen. (Eds.), *Language typology*: *a functional perspective*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Matthiessen, C., Teruya, K. and Wu, C. (2008) Multilingual studies as a multi-dimensional space of interconnected language studies. In J. Webster. (Ed.), *Meaning in Context*: *implementing intelligent applications of language studies*. London and New York: Continuum.

O'Donnell, M. (2008). *The UAM CorpusTool:* software for corpus annotation and exploration (2008). In C.M. Bretones Callejas *et al.* (Eds.), *Applied Linguistics Now*: understanding language and mind / la linguística aplicada hoy: comprendiendo el lenguaje y la mente. Almería: Universidad de Almería. pp. 1433--1447.

Saussure, F. (2006). *Curso de Linguística Geral.* São Paulo: Cultrix, 2006.

Fragments as instantiation of syntactic slots: complexity of the interface between lexicon, grammar and discourse in spoken French

Marie-Noëlle ROUBAUD, José DEULOFEU

EA 4671 ADEF; Aix Marseille Université; ENS de Lyon, IFE; 13248, Marseille, France
LIF-CNRS UMR 7279; Aix Marseille Université
mnroubaud@yahoo.fr, jose.deulofeu@gmail.com

Abstract

A particular feature of spontaneous speech syntax is the abundance of "fragments" (non clausal text units) which are generally analysed as independent syntactic text units. The main purpose of the paper is to show that many of them are in fact licensed by a verb of a preceding text unit, directly or by means of complex constructions, one of which will be discussed in detail : PSS. We will show that by reconsidering the syntax lexicon interface. We assume following Blanche-Benveniste *et al*. (1984) that this interface is highly complex. There are many ways in which a syntactic slot can be filled: null, pronominal, simple lexical, list of lexical items. And finally by means of a "discourse grafting" device (Deulofeu, 2010). One subcase, PSS, already investigated (Roubaud, 2000; Blanche-Benveniste, 1986, 2010), displays a combination of two fillers: a pronoun or a "light" lexical unit followed by a second one bringing a progressive semantic specification. In those patterns, the second clause does not obligatorily meet the subcategorisation requirements of the main verb. Such patterns pose the question of the limits between syntax and discourse. And also between structural and "online" syntax. Finally, we will show that PSS is combined with higher level discourse patterns in order to overcome processing problems.

Keywords: spoken French; corpus; syntax; progressive semantic specification.

1. Introduction

The main purpose of our paper is to revisit the way "fragments" and more generally syntactic structures are linked to the linguistic context through a corpus based study. This study is a piece of a more general project aiming to develop a competence grammar compatible with descriptive generalizations captured through spontaneous speech analysis This amounts to specifying the interface between grammar, lexicon and discourse.

Our empirical domain can be defined as "extended fragments". "Fragments" are non sentential utterances syntactically autonomous but linked to a host construction by means of syntax-semantics interface rules (Culicover & Jackendoff, 2005):

L1 who came yesterday L2 Bo

We further define "extended fragments" as lexical items or constructs linked to a syntactic slot of a construction within « discourse patterns ». We look at defining the nature of that link.

2. Framework

We rely on the theoretical framework of *Approche Pronominale* (Blanche-Benveniste *et al.*, 1984) revisited with *Basic Linguistic Theory* (Dixon, 2009). This framework, which can be compared with the one presented in chapter 8 of Biber *et al*. (1999); has been applied to spoken language analysis in numerous studies (Blanche-Benveniste & Jeanjean, 1987; Blanche-Benveniste *et al.*, 1990; Blanche-Benveniste, 1986, 1997, 2010-b; Deulofeu, 2010).

The main *Approche Pronominale* (AP) assumptions are the following:

AP stands as a lexicalist approach of syntactic structures: lexical items licence syntactic slots: *manger* (eat) [P0, P1];

Pronouns and not full lexical items or phrases are default fillers of syntactic slots: *je le mange* ;

The paradigms of pronouns which can be built in the slots determine their grammatical features.

Lexical heads (constructeurs) with their underspecified syntactic slots are the basic components of syntactic constructions (skeletons). Syntactic skeletons slots are filled with lexical features to give full-fledged constructions

Lexicalization can be "direct " lexical items fill directly the slots or "indirect", involving additional grammatical devices (dispositifs).

As for the interface with performance, we assume, departing from the view that fragments are self contained syntactic units that an abstract syntactic construction (competence) can be uttered at once or in several times by the same speaker or several ones, which can result in a concatenation of fragments.

This lexicalization strategy can be linked to various competence performance interaction studies (Apotheloz, 2008; Auer, 2005; Blanche-Benveniste, 1990; Deulofeu, 2011).

More specifically, as particular structures are concerned, we propose to include lexicalization within the "performance" patterns identified by Iwasaki & Ono (2002): "to eyes used to the constructed data in linguistic literature, sentences in Japanese conversation look rather chaotic… though these types of utterances have been traditionally regarded as performance errors, careful examination reveals several clearly identifiable patterns, which we call "on line mechanisms"... We think these patterns are systematic enough to deserve a place in grammar…: phenomena of interpolation, incrementation, reformulation, local management and bridging… furthermore it is our hope that continuing analysis of spoken data in different languages will allow us to

construct a typological and universal model for a grammar of human language."

We will further rely on a research on pseudo-clefts in spoken French (Roubaud, 2000) following Higgings (1973), Peters & Bach (1968, 1971) for English.

The empirical basis from witch our examples are taken consists of various spoken French corpora: GARS (Groupe Aixois de Recherches en Syntaxe), CRFP (Corpus de Référence du FrançaisPparlé: 400 000 words), CERF (Corpus Evolutif de Référence du Français: 10 million including 1million words in spoken).

3.   Indirect filling

3.1   The case of lists of fillers

Consider the syntactic skeleton: *faire* P0 [-pers], P1 [-pers, -verbal]. The lexicalization of this abstract pattern can be Direct or Indirect.

A Direct pronominal filling will give the following construct: *ça fait ceci*. In the same way, a Direct lexical filling will give: *son truc faisait une minerve* [his stuff was (like) a neck brace]

Various types of "Indirect" lexical filling are possible as the utterance is processed: double filling, list filling, zero filling (contextual inferences).

Example:

Indirect lexical filling of a syntactic skeleton *ça fait ceci* by a « list » of lexical items with two speakers:

(1) L1: ça fait un + un + comment on dit je sais plus
      + une chose + là
   L2: une écharpe + un col roulé + un
   L1: mais non + le truc blanc là + qu'ils ont ceux
      qui se sont cassé la + le
   L2: ah oui
   L1: la + la + la + la cheva-
   L2: la minerve
   L1: voilà + la minerve (oral, privé)

According to our assumptions all the NPs which look like independent fragments are to be linked as indirect lexicalizations to the object syntactic slot of *faire*. This results in a fragmented filling (Deulofeu, 2011) of a syntactic slot. The link of the structural skeleton and the on line processing can be visualized by means of a graphic device: a « grid » as defined by Blanche-Benveniste & Jeanjean (1987). The structure SVO can be read horizontally whereas one can see vertically how the" Indirect filling" is processed:

(1) L1 ça fait un
            un
            comment on dit
            je sais plus
            une chose   là
      L2   une écharpe
            un col roulé
            un

L1 mais non le truc blanc   là
            qu'ils ont ceux qui se sont cassé la
                                        le
L2   ah oui
L1   la
            la
            la
            la cheva-
L2   la minerve
L1 voilà   la minerve

What is interesting to notice is that if the syntactic status of the fragments is the same (object of *faire*), their semantico pragmatic status is different. The material which is added to the NP - disfluences, metalinguistic remarks (*comme on dit*, *je ne sais plus*) discourse markers (*viola*) - helps the participants to evaluate the information status of the fragments - approximation, invalid lexical search (*non*), successful filling (*oui, voilà*). This material is not to part of the abstract syntactic structure. It comments on the process of lexicalization which belongs to the utterance building level.

3.2   The pseudo-cleft case

In the former example the lexicalization process involves paradigmatic listing of one grammatical category (NP) with added items not integrated in the grammatical structure (oui, voilà…).

In other cases Indirect lexicalization involves a grammatical device: the combination by means of the pseudo-cleft construction of two possible fillers of a syntactic slot between which stands a semantic relationship of "progressive specification" (Roubaud, 2000). In the following examples the two possible fillers of the object of *faire* are c*e que (what)* and the NP *le saut en extension*:

(2) ce que je sais faire c'est le saut en extension
      (oral, privé)
   [what I can do is the extension jump]

Part 1 (what I can do) is semantically underspecified
what = [gr. function : P1], [-pers], [-verbal]
I can do = head verb and other dependants
Part 2 (is the extension jump) is semantically specified
the extension jump = lexical features : [movement of body] [+ extension of body]

Notice that when a full pseudo cleft pattern is used semantic progressive specification must obey grammatical constraints as both fillers must meet subcategorisation rules coming from the lexical structure of the "main" verb as well as lexical restrictions; with the verb *say* the filler introduced by *c'est* must be something that can be said :

(3) ce que je peux dire c'est que nous ne sommes pas

inutiles (oral, TV)
[what I can say is that we are not useless]

This pattern of progressive semantic specification (PSS) acts further as a repair device for clausal filling. This has many advantages. For example, it facilitates the processing of a clausal subject, which is almost excluded in spontaneous speech as filler in Direct filling:

(4) ce qui rendait les choses particulièrement difficiles c'est que la variation est double (oral, public)
     * que la variation est double rendait les choses particulièrement difficiles
     [what made the things particularly difficult was that this variation was double]

(5) ce qui me choquait un petit peu c'est qu'il s'agissait toujours d'orgie (oral, privé)
     ? qu'il s'agissait /s'agît/ s'agisse toujours d'orgie me choquait un petit peu
     [what shocked me a little bit was that it was always the case of an orgy]

As a consequence of his discourse nature, PSS allows lexicalization even to go beyond grammatical constraints of subcategorisation:

- the specifying part of the utterance may contain direct discourse

(6) ce qui m'a paru bizarre c'est que quand je lui ai dit je vous mets à l'ordre quel ordre Monsieur il m'a dit non non non laissez laissez j'ai l'habitude je le ferais moi-même (oral, privé)
     [what looked strange to me was that when I said to him I put (on this check) payable to payable to whom sir he answered no no don't bother I know how to manage I will do this by myself]

or a kind of rhetorical self addressed question

(7) ce qui est embêtant c'est que c'est que quelle est l'opération la plus simple en général c'est l'addition (oral, privé)
     [what is annoying is (that) is (that) what is generally the simplest operation : it is addition]

- or allows freer contrastive patterns than in direct licensing:

(8) et maintenant dans l'imprimerie ce qu'on demande à un imprimeur c'est non pas + d'être un artiste c'est d'être un gestionnaire (oral, professionnel)
     [now in printing what you require from a printer is not to be an artist is to be a manager]

In direct lexicalization, *mais* (*but*) is needed:

(8') on demande à un imprimeur non pas d'être un artiste mais d'être un gestionnaire

- allows category mismatch in lists

(9) moi ce que je proposerais au comité de quartier + c'est que nous fassions une commission malgré euh ce qu'on a pu nous dire que il n'était + le projet était pas encore bouclé de s'emparer des des données que l'on a déjà + et de voir nous en tant qu'habitants + ce qu'on souhaiterait qui + enfin ce qui nous inquiète et que le le cabinet qui est en train de donc de plancher sur le projet on lui amène nous aussi des éléments de réflexion + (oral, public)
     [as for me, what I would propose to the district assembly is that we set up a committee - in spite of the fact that they said that the project was not completed – in order to consider the data that we already have and to see as neighbors what we would like well what bothers us and the consulting office who is working on the project to bring him elements to think about]

In direct lexicalization, complementizers preferably match:

(9') ? je proposerais… que nous fassions une commission …
            de s'emparer des … données …
            et de voir … ce qu'on souhaiterait …
            et que … on lui amène … des éléments …

- allowing filling by paratactic constructions

(10) ce que je peux rajouter même mieux que ça c'est qu'en fait + elle était la première soliste à l'orchestre moi j'étais le second flûtiste + (oral, public)
     [what I can add better than this is that in fact she was the first soloist of the orchestra (and) me I was second flautist]

In direct lexicalization, complementizer *que* is needed:

(10') je peux rajouter même mieux que ça qu'en fait + elle était la première soliste à l'orchestre et que moi j'étais le second flûtiste

The discursive and on line nature of PSS can even result in specific strategies based on paratactic syntactic patterns without *c'est,* in which "the semantic underspecification of the first member let the hearer expect the second" (Blanche-Benveniste, 2010-a):

(11) ce qui m'est arrivé au début + j'ai décollé dans le vent un peu trop fort (oral, privé)
[what happened to me in the beginning + I landed off against the wind somewhat too hard]

(12) il y avait une chose chez maman euh elle était illettrée (oral, privé)
[there was one thing with ma well she was illiterate]

Claire Blanche Benveniste noticed that all this patterns can be ordered in a cline, such that : « La cohésion la plus forte est fournie par le modèle canonique de pseudo-clivée qui réunit un faisceau de propriétés grammaticales favorisant la cohésion. D'autres modèles n'utilisent qu'une partie de ce faisceau de propriétés, la cohésion la moins forte étant celle des organisations par parataxe. » (Blanche-Benveniste, (2010-a)

4. From processing repairs to discourse patterns

PSS has to face processing constraints, due to what can be called the "efficient communication paradox". On one side, the indirect lexical specification by extended pseudo clefts allows the speaker to accurately make his point in spite of lack of "right word" by means of "periphrasis". But a "long" lexical specification puts the main verb licensing the lexical part out of short time memory and even introduces irrelevant grammatical material blurring coherent transition with following discourse units. There seems to be a way out of the paradox: a "reformulation" step, using constructions with *c'est*, clitic "doubling", etc. For example, an indirect lexicalization in which the speaker wants to explain what bothers him (*ce qui me gêne un peu*) and which evolves to a long piece of speech (In square brackets below) becoming more and more autonomous is "recapitulated" by the word *choses* allowing to reintroduce through the verb *inquiéter* (synonym of *gêner*) at the end of the discourse unit the semantic role of the lexicalization (source of bothering for the speaker):

(13) enfin moi ce qui *ce qui me gêne un peu* c'est [aujourd'hui on a + on a un projet hein vous l'avez l- vous l'avez lu comme moi j'ai entendu des choses qui m'ont quand même beaucoup inquiété moi quand ici en réunion publique on m'a dit deux fois une voie que j'en sois j'entends parler deux fois deux voies après j'entends + au niveau logement quand je fais et et tout tout est acté hein puisqu'il y a euh phase un il y a euh les logements qui vont être construits par exemple cette école maternelle qu'on nous dit qu'elle sera pas euh construite tout de suite elle est phasée en phase deux + c'est-à-dire qu'elle est phasée elle est euh + c'est phasé le budget est là tout tout est là hein euh je sais pas si vous l'avez lu comme moi si vous pouvez confirmer je pense (…) hein donc

euh alors qu'on nous a dit qu'elle serait pas euh réalisée tout de suite parce que effectivement les étoles les écoles qu'il y avait étaient pas encore à saturation]+ moi ça ç- je dirais qu'il y a d- dans le projet il y a des *choses* qui *m'inquiètent* beaucoup (oral, public)

For Blanche-Benveniste (2010-a), the reformulation appears as the conclusion of a discourse unit. But it is not always the case. The reformulation can be a specific move within larger discourse patterns and be a step for further clarification. In the following example of discourse pattern, we can see this scheme: explanation, **summary**, *synthetic reformulation* and clarification

(14) L2: ah oui ah oui + fidéliser le le client c'est important + surtout les gens âgés ils aiment bien qu'on s'occupe d'eux + ils arrivent ici faut faut même si ils doivent se servir ils aiment bien que qu'on les serve quand même + ils prétexteront toujours quelque chose pour qu'on qu'on aille se s- les aider et + voilà + des fois les X il faut les ramener chez eux parce qu'ils ont pris trop de marchandise(s) + donc il faut les ramener chez eux parce qu'ils sont ils en ont trop ils peuvent pas marcher + quand il y a trop de vent quand il pleut + c'est vraiment à part + c'est vraiment à part en grande(s) surface(s) c'est sûr qu'on leur fait pas + ça ils arrivent ils se débrouillent et + ils rentrent par leurs propres moyens
L1 : donc là vous pouvez faire la différence
L2 : ouais + ***c'est ce qu'ils recherchent*** + *les gens très âgés qui peuvent pas se déplacer ce qu'ils recherchent c'est la proximité + puis le la façon de +* les petits commerçants c'est vrai on a le temps de s'occuper de des gens + en grande surface ils ont pas le temps + les employés sont pas là pour ça de toute façon + (oral, professionnel)

The item *proximité* synthesizes/ summarizes the former long explanation, opening an opportunity for further clarification (the superiorty of small shops over supermarkets for attending old costumers).

As puts Apotheloz (2008: 91), "we can conclude from these observations that the identificative constructions [our PSS] are a central device for the sequential organization of some discursive patterns. From this point of view they appear as building the interface between grammar and discourse".

It is nevertheless to notice that the reformulation devices play a complementary part in maintaining discourse cohesion and coherence when PSS is used. In the last example the speaker wants to emphasize that it is important to point out the scandalous attitude of some

occupational doctors:

(15) enfin il y a quand même quelque chose à signaler qui est important c'est que la médecine du travail + lui a demandé de ne jamais parler de son diabète à son employeur + *donc ça c'est quand même quelque chose d'assez grave + qu'il faut encore noter* parce qu'on rentre en l'an deux mille quand même hein (oral, privé)

[well there is something to be pointed out which is important it is that occupational medicine staff asked him not to mention his diabetes to their boss *so this is something quite serious that has to be noticed* because we are now on the way to the years 2000 anyway, aren't we]

The clause introduced by *parce que* qualifies as even more scandalous their attitude as we are entering the years 2000. The reformulation part in italics appears as a necessary part of the whole pattern.

Indeed a coherence gap appears if we erase the reformulation step. In the following example, "parce que" has default scope on the preceding clause and not on the main clause of the PSS:

(15') enfin il y a quand même quelque chose à signaler qui est important c'est que la médecine du travail + lui a demandé de ne jamais parler de son diabète à son employeur parce qu'on rentre en l'an deux mille quand même hein

5. Conclusion

Beyond PSS patterns, the progressive specification semantic relationship plays a crucial part both in structuring the interface between syntax and semantics and, as a "projection device", in smoothening "online" building and processing of utterances.

Our next step is a corpus based typology of discourse patterns involving fragments and progressive specification with more registers and comparison with similar facts in other languages (Cresti & Moneglia, 2005).

On methodological grounds, this study shows that it is important to take in consideration larger contexts than one sentence even complex to investigate properly the links between grammar, lexicon and discourse. Such wide scope "useful contexts" (Blanche-Benveniste, 1988) further allow us to sort out in what kind of discourse contexts such complex constructions appear, beyond the argumentative ones pointed out for PSS by Roubaud (2000) and Apothéloz (2008).

6. References

Apothéloz, D. (2008). A l'interface du système linguistique et du discours: l'exemple des constructions identificatives (e.g. pseudo-clivées). In O. Bertrand, S. Prévost, M. Charolles, J. François and C. Schnedecker (Eds.), *Discours, diachronie, stylistique du français. Études en hommage à Bernard Combettes*. Berne: Peter Lang, pp. 75--92.

Auer, P. (2005). Projection in Interaction and Projection in Grammar, In *Text*, 25.1, pp. 7--36.

Biber, D., Johannson, S., Leech, G, Conrad, S. and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. New-York: Longman.

Blanche-Benveniste, C. (1986). Une chose dans la syntaxe verbale. In *Recherches sur le français parlé*, 7, pp. 141--168.

Blanche-Benveniste, C. (1988). La notion de contexte dans l'analyse syntaxique des productions orales : exemples des verbes actifs et passifs. In *Recherches sur le français parlé*, 8, pp. 39--57.

Blanche-Benveniste, C. (1990). Grammaire première et grammaire seconde : l'exemple de *en*. In *Recherches sur le français parlé*, 10, pp. 51--73.

Blanche-Benveniste, C. (1997). *Approches de la langue parlée en françai*s. Paris: Ophrys.

Blanche-Benveniste, C. (2010-a). Les pseudos-clivées et l'effet deux points. In M.-J. Béguelin, M. Avanzi and G. Corminboeuf (Eds.), *La Parataxe*, t. 2, *Structures, marquages, exploitations discursives*. Bruxelles: Peter Lang, pp. 185--217.

Blanche-Benveniste, C. (2010-b). *Le français. Usages de la langue parlée*. Leuven-Paris: Peeters.

Blanche-Benveniste, C., Deulofeu, J., Stéfanini, J. and van den Eynde, K. (1984). *Pronom et syntaxe. L'approche pronominale et son application au français*. Paris: SELAF.

Blanche-Benveniste, C. and Jeanjean, C. (1987). *Le français parlé. Transcription et édition*. Paris : Didier Erudition.

Blanche-Benveniste, C., Bilger, M., Rouget, C. and van den Eynde, K. (1990). *Le français parlé: Études grammaticales*. Paris : CNRS.

Cresti, E., Moneglia, M. (Eds.). (2005). *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Culicover, P., Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.

Deulofeu, J. (2010). La greffe d'un énoncé sur une construction : une combinaison originale de parataxe et de rection. In M.-J. Béguelin, M. Avanzi and G. Corminboeuf (Eds.), *La Parataxe* : *Structures, marquages, exploitations discursives*. Bruxelles: Peter Lang, vol 2, pp. 111--135.

Deulofeu, J. (2011). De l'inutilité de la notion de « fragment » pour la description des énoncés « fragmentés »'. Available at : <http://lingo.stanford.edu/sag/papers/procpap-final>.

Dixon, R.M.W. (2009). *Basic Linguistic Theory*. Oxford: Oxford University Press.

Higgins, F.H. (1973). *The Pseudo-cleft construction in English*. M.I.T. Dissertation.

Iwasaki, S., Ono, T. (2002). Sentence in spontaneous spoken Japanese discourse. In J. Bybee and M. Noonan

(Eds.), *Complex sentences in grammar and discourse, Essays in honor of Sandra A. Thompson*, Amsterdam: John Benjamins, 2001, pp. 176--202.

Peters, S., Bach, E. (1968). Pseudo-cleft Sentences. Preliminary version, unpublished ms, Austin: Texas.

Peters, S., Bach, E. (1971). Pseudo-cleft Sentences. Unrevised version. In Report to *NSF: On the Theory of Transformational Grammar*, GS-2468, Austin: Department of linguistics, University of Texas at Austin, pp. 171--190.

Roubaud, M.-N. (2000). *Les constructions pseudo-clivées en français contemporain*. Paris: Champion.

"Subordinate" clauses and syntactic annotation of spoken French

Frédéric SABIO

Aix-Marseille University, Laboratoire *Parole et Langage*
5 avenue Pasteur, 13100 Aix-en-Provence, France
frederic.sabio@orange.fr

Abstract

The grammatical analysis of clauses introduced by a "subordinating conjunction" has always been a challenge for linguists because, on the one hand, spontaneous spoken data exhibits highly variable syntactic and discursive organizations which have never been properly described through the sentence-based framework of traditional grammar; and on the other hand because continuous reference to the notion of "subordination" tends to unify in an artificial way several types of syntactical configurations that it would be advisable to distinguish carefully. Within the *Rhapsodie* project (directed by Anne Lacheret, Univ. of Paris Ouest), which is devoted to the syntactic and prosodic tagging of spoken French, we have been directly confronted to such difficulties, and we have had to make some methodological choices which will be the theme of our paper. The tagging system which has been developed both annotates the microsyntactic dependences and the macrosyntactic groupings. Taking those two levels of analysis into account allows us to describe most of the attested uses of conjunctions, including the most problematic ones. The annotation system will be illustrated with a selection of corpus-drawn utterances.

Keywords: syntax; spoken language; French; tagging; subordination; macrosyntax.

1. Introduction

This study has been conducted within the *Rhapsodie* project (headed by Anne Lacheret, Univ. Paris-Ouest) which is a four-year program (2008-2012) which aimed at annotating a 36.000 words spoken French corpus on both syntactic and prosodic grounds (*cf.* http://www. projet-rhapsodie.fr.). The ultimate goal of the project was to model the interface between syntax and prosody and to identify the existing correlations between prosodic and syntactic boundaries. The present paper is not meant to give a detailed account of the *Rhapsodie* framework; it will not even address the diverse aspects of the syntactic annotation system (see Benzitoun *et al.*, 2009, 2010); it will merely illustrate some specific issues regarding the analysis and annotation of "subordinate" clauses. Spontaneous speech seems to be a particularly valuable type of data for the description of sequences which are introduced by so called "subordinating conjunctions", since it offers a large and somewhat puzzling variety of forms which would not be properly described by the sentence-based framework of traditional grammar.

Before we introduce our annotation system, we will say a few words about the drawbacks of the traditional concept of subordination.

2. Subordination as syntactic dependency

Grammatical tradition quite commonly assumes that any clause which is introduced by a conjunction such as *when*, *because*, *since* or other morphemes of the same kind be automatically regarded as a "subordinate clause" (Riegel *et al.*, 1994). In our view, continuous reference to the notion of subordination tends to unify in an artificial way several types of syntactical and discursive configurations that it would be advisable to distinguish carefully. If we wish to make a reasonable - and somewhat more restricted - use of the concept of subordination, we must stop considering that the conjunctional status of the initial morpheme is a reliable syntactic criterium *per se*, and keep the notion to sequences that share a real dependency relation to the verb of the construction (Debaisieux, 2006a; Deulofeu, 2011).

Obviously, what can be regarded to be a "real dependency relationship" is no simple matter and crucially depends on some theoretical choices. We will refer to the theoretical frame of "Pronominal approach" (Blanche-Benveniste, 1980; Blanche-Benveniste *et al.*, 1984, 1990; Deulofeu, 1991) which postulates that syntactic dependency ("rectional relations") must necessarily correlate with a set of paradigmatic properties, such as the equivalence with a pronoun, the possibility to be cleft, and a few other features that will be detailed below. The application of these criteria is useful since it enables us to distinguish between clearly *dependent* sequences, that pertain to the strict domain of syntax and can readily be analyzed as subordinate clauses; and other configurations that do not possess any paradigmatic property, and thus appear to only be *linked* to the neighboring constructions, sharing with them mere "association", or paratactic, relations.

The following example will serve to characterize *dependant* subordinate clauses:

il viendra   [quand on lui demandera]
he will come [when he will be asked to]

Here is a set of criteria that show that the *When*-clause is syntactically dependant on (or governed by) the verb *venir* (*to come*), and could therefore be considered as a genuine subordinate clause. The temporal sequence:

(a) could be replaced by a pronominal form such as the interrogative pronoun *when* or a quasi-pronominal expression like *at that moment*:

quand est-ce qu'il viendra? *[when will he come?]*
il viendra *à ce moment-là [he will come at that moment]*

(b) can occupy a focus position within some sentence-types like cleft constructions, among others:

c'est *quand on lui demandera* qu'il viendra
[it is when he will be asked to that he will come]

(c) is liable to develop a contrast between positive and negative modality:

il viendra *quand on lui demandera* et pas quand il le decidera
[he will come when he will be asked to, and not when he will decide to]
il viendra non pas quand il le decidera mais *quand on lui demandera*
[he will not come when he will decide to but when he will be asked to]

(d) can be modified by a paradigmatic adverbial like *seulement, uniquement, surtout* (*only*, *mostly*):

il viendra *seulement quand on lui demandera* [he will come *only when he will be asked to*]

Here are three corpus-drawn oral utterances in which the clause between brackets is syntactically dependent on the main verb of the construction:

le métier de fleuriste était pas drôle [*parce que* il fallait avoir les mains dans l'eau]
lit: working as a florist wasn't funny [*because* you always had to keep your hands in the water]
nous avons vu une euh euh un crépuscule euh [*alors que* nous étions d- au au sommet de la mosquée]
lit: we saw a er – er a twilight [*while* er we were i- at the at the top of the mosque]
il chantait à Saint Laurent à la cathédrale [*quand* il y avait des fêtes]
lit: he used to sing at Saint-Laurent in the cathedral [*when* there were parties]

In contrast with such canonical examples, the following conjunctional clauses (in brackets) would react in a negative way to the paradigmatic criteria listed above: they have no equivalence with a pronoun, cannot be cleft, and so on.

vos clients euh pourront euh à cet endroit admirer la vue sur le lac et le barrage - [parce que n'oubliez pas que le le Muséoscope surplombe le lac de Serre Ponçon hein]
lit. your customers er can er in this place admire the sight on the lake and the dam - [because don't forget that the Muséoscope overhangs the lake of Serre Ponçon]
ici par exemple c'est du corail qu'elle va porter dans sa corne d'abondance - [alors que là-bas ça sera des fruits]

lit. here for example it is coral that she is going to carry in her horn of plenty - [while over there that will be fruits]
[quand je vois les les les les les élèves qui descendent dans la rue et tout] moi je les soutiens
lit. [when I see the the the the the pupils who go down in the street and stuff] me I support them

In the *Rhapsodie* project, it was essential to make a clear distinction between the syntactically dependent conjunctional sequences, and those that have a non-dependent status. But of course, other aspects had to be taken into account, such as distributional and prosodic properties. We have chosen to study such phenomena in the theoretical frame of macrosyntax (Blanche-Benveniste *et al.*, 1990; Deulofeu, 2003; Sabio, 2012).

3.   Macrosyntactic patterning

3.1   Presentation

To put it simply, macrosyntax relates to a level of organization which allows the description of sequences which could not be analyzed on the sole basis of their microsyntactic properties, since they share a somewhat discursive relationship with the surrounding context. At the macrosyntactic level, the utterances can be seen as sequences of successive units making up the following pattern:

Utterance: [Pre-Nucleus – *Nucleus* – Post-Nucleus]

What distinguishes those three units has to do with the modality that they are liable to express, certain prosodic properties, and their linear position:

- *The Nucleus* is the basic macrosyntactic unit. It bears an illocutionary value which can be interpreted as a speech act (declarative, question, exclamation), and is liable to form an autonomous utterance. Prosodically, it is associated with a choice of terminal contours that make up a paradigm of prosodic forms, each of them being related to an illocutionary value.

- *The "ad-Nucleus"* (pre- and post-Nucleus) bear no illocutionary value: they seem to be "deactivated" (Verstraete, 2007) as to their capacity to convey any kind of illocutionary content. As a consequence, they cannot constitute an independent communicative unit. Pre- and post-Nucleus are respectively placed before and after the Nucleus unit.

Regarding the way in which micro- and macrosyntax articulate to form utterances, it must be pointed out that in our approach both levels are largely autonomous one from another. This means that two units sharing the same microsyntactic status (that is, with the same syntactic function) may well be realized as two different

macrosyntactic units. Inversely, two elements which have the same macrosyntactical status can fulfill different syntactic functions. Let us consider the following utterances:

> il n'a pas vu Paul *à Paris* (mais à Londres)
> [he didn't see Paul *in Paris* (but in London)]
>
> *à Paris*, il n'a pas vu Paul
> [*in Paris*, he didn't see Paul]

The Prepositional Phrase *à Paris* works in both cases as a syntactic adjunct to the verb *voir* (*to see*). But their macrosyntactic integration within the utterance is different: in the first case, the locative sequence is part of the Nucleus; in the second utterance, it forms the pre-Nucleus unit.

Here is a second example: the two following utterances share the same macrosyntactic pattern: a Nucleus unit followed by a post-Nucleus unit.

> 10 ans il avait (en réponse à "il avait quel âge ?")
> [10 years old he was (as an answer to "how old was he?")]
>
> il est trop jeune je trouve
> [he is too young I think]

10 ans	il avait
Il est trop jeune	je trouve
Nucleus	*Post-Nucleus*

Table 1: examples of a Nucleus unit followed by a post-Nucleus unit

But their microsyntactic organization is quite different: *10 ans* is an object to the verb *avait*; whereas there is no direct dependency relationship between *il est trop jeune* and *je trouve*.

3.2 Macrosyntactic annotation

The *Rhapsodie* annotation system is organized on several levels. This paper will only mention the first level, which is mainly concerned by major grammatical groupings (such as macrosyntactic ones). The following labels are used, which will be illustrated in 4 below:

//	marks the end of a macrosyntactic utterance
<	marks the frontier between pre-Nucleus and Nucleus
>	marks the frontier between Nucleus and post-Nucleus
+	Indicates a (microsyntactic) dependency relationship between two successive macrosyntactic units

Table 2: Labels used in *Rhapsodie*

4. "Subordinate" clauses: a typology

The micro- and macrosyntactic frame which has briefly been introduced above leads us to distinguish between 5 different configurations involving sequences introduced by a "subordinating conjunction". This typology constitutes an exhaustive classification of all the types of subordinate and "pseudo-subordinate" clauses that we have found in written or spoken French corpora.

The three following features are needed to distinguish between our 5 types:

(a) The conjunctional sequence *is / is not* dependent on the verb, on a strictly syntactic base (*cf.* section 2 above).
(b) The conjunctional sequence *constitutes / doesn't constitute* an autonomous macrosyntactic utterance (*cf.* section 3 above).
(c) The conjunctional sequence *is / is not located* in the same macrosyntactic unit as the main verb (that is: in the same Nucleus, or the same pre-Nucleus Unit, or the same post-Nucleus Unit).

The following table indicates the three features *a-b-c* on the X-axis, and the 5 syntactic types on the Y-axis:

Conjunctional sequence / Type	a) Dependent on the verb (microsyntactic level)	b) Forms an autonomous macrosyntactic utterance	b) Located in the same macrosyntactic unit as the main verb
❶ Dependent sequences inside a macrosyntactic unit	+		+
❷ Dependent sequences forming a macrosyntactic unit inside the utterance	+	-	-
❸ Dependent sequences forming a macrosyntactic utterance	+	+	
❹ Non-dependent sequences inside a macrosyntactic utterance	-	-	-
❺ Non-dependent sequences forming an macrosyntactic utterance	-	+	

Table 3: The 5 syntactic configurations

Each type is tagged in the following way ["CS" for "conjunctional sequence"]:

Type 1	No tagging	
Type 2	// CS + <	[if the CS is a pre-Nucleus]
	// CS + >	[if the CS is a Nucleus]
Type 3	// + CS //	
Type 4	// CS <	[if the CS is a pre-Nucleus]
	> CS //	[if the CS is a Nucleus]
Type 5	// CS //	

Table 4: The tags used in *Rhapsodie*

We will now illustrate each of those grammatical types. Due to lack of space, we will not go in much detail but will only present an overview of our typology.

4.1 Type 1: *dependent sequences inside a macrosyntactic unit*

With this first type, the conjunctional sequences appear to be grammatically integrated both in terms of microsyntax (since they are dependent on the verb) and in terms of macrosyntax (since they are realized into the same unit as the verb itself, showing no detachment of any kind). For example, the conjunctional sequence and the rest of the construction can be placed in the same Nucleus Unit, as in:

//il est parti plus tôt que prévu *parce qu'il avait un rendez-vous* //
[// he went away earlier than expected *because he had an appointment*//]
// il ne viendra que *si cela est nécessaire* //
[//he will come only *if this is necessary*//]

But the whole of the construction can be realized in another macrosyntactic Unit, such as a pre-Nucleus Unit:

// si Pierre a l'intention d'arriver *quand la réunion sera terminée* < autant qu'il reste chez lui //
[//if Pierre intends to turn up *once the meeting is over* < he'd better stay home//]

Those subordinate clauses are obviously the most canonical and easy to describe and annotate, since the micro- and macro-syntactic levels strictly overlap. At the first level of our annotation system, we do not feel the need to specify that the adjunct has been realized as a conjunctional sequence (rather than a Prepositional Phrase or any other category). This is why we only annotate the beginning and end of the macrosyntactic Unit with no internal delimitation.

4.2 Type 2: *dependent sequences forming a macrosyntactical unit inside the utterance*

This type deals with the conjunctional phrases that are dependent on the verb (as in type 1 above, or type 3 below), but are realized as a specific macrosyntactic unit placed at the initial position of the construction. The conjunctional phrase can either be a pre-Nucleus Unit or a Nucleus Unit.

Here is an example in which the subordinate clause constitutes a pre-Nucleus unit:

// quand ils vont rentrer dans la vie active + < ça va être dur pour eux // [oral, corpaix]
[*lit.* // when they will enter the labour market + < it will be hard for them //]

Notice that two labels are used for the annotation: a) the left angle bracket "<", which signals the frontier between pre-Nucleus and Nucleus Units ; b) the "+" sign, which indicates that there is a dependency relationship between the initial temporal clause and the verb located into the Nucleus Unit.
Here is an example where the subordinate clause has the value of a Nucleus Unit:

Loc.1: // vous allez aller vous promener ? //
Loc.2: // seulement s'il fait beau +> on ira //
[Speaker 1: // will you be going for a walk?//]
[Speaker 2: //only if the weather is fine + > we will be going //]

The last mentioned examples are quite frequent in everyday conversation (Sabio, 2006). The initial clause constitutes the Nucleus Unit, that is, the macrosyntactic element which bears the illocutionary value of an assertion.
As in the preceding example, 2 labels will be useful here: a) the right angle braket ">", signaling the limit between the Nucleus and the post-Nucleus. 2) the "+" sign indicating that the two macrosyntactic units are linked by a dependency relationship.

4.3 Type 3: *dependent sequences forming a macrosyntactic utterance*

Here, the conjunctional clause is once again syntactically dependent on a verb, but it appears to be completely detached from the rest of the construction, in such a way that it forms a completely independent macrosyntactic utterance; thus the construction appears to be realized as a sequence of two successive utterances. Such examples have sometimes been analyzed as "delayed complements" or "supplements" (Debaisieux, 2006b); it appears that the subordinate clause can be detached from the preceding sequence in several different ways:

- In dialogues, it can take the form of a "supplement" which is given by one of the speakers:

Baga: Et si je ne faisais que dormir comme toi < qui est-ce qui lèverait les impôts ? // Tu dépenses tout pour bouffer. //
Le roi: //+ Parce que je n'ai rien d'autre à faire.//
(Architruc, R. Pinget, Ed. Minuit, 16-17)

Baga: // If I spent my time sleeping as you do < who would levy the taxes? // You spend all the money to buy food. //
The king: //+ Because I have nothing else to do.//

In the *Rhapsodie* annotation system, these detached sequences are isolated between a double-slash symbol, based on the fact that they are utterances on their own. In addition, the *"+"* sign is here to indicate that there is a rectional link between the verb of the first utterance and the subordinate clause of the second utterance.

- Prosodic or graphic cues indicate that such a "detachment effect" can be found in monologues as well:

// quand je sors de la consultation + < je suis euphorique //+ parce que j'ai aimé être avec les gens //

// when the medical examinations are over + < I am thrilled //+ because I like being with people//

With that kind of delayed clauses, the conjunction appears to be frequently preceded by a variety of elements; like:

(a) A connective morpheme like *et* (*and*) or *mais* (*but*):

moi < je préfère une édition originale //+ *mais* pas parce qu'elle est originale //
me < I prefer an original version //+ *but* not because it is original//

(b) A negation mark:

// c'est un métier pénible d'accord //+ mais *pas* parce que c'est un métier privé ou parce que c'est un métier public //
// it is a hard work indeed //+ but *not* because it is in the private sector or in the public sector//

(c) A paradigmatic adverbial like *seulement* (*only*) or *surtout* (*mostly*, *especially*):

// j'aimais pas du tout les cours de français // + *surtout* quand on faisait des dictées //
[//I didn't like French classes //+ *especially* when we made dictations //]

les jeunes en Angleterre < euh quand ils parlent < c'est fou // faut s'accrocher pour comprendre //+ *surtout* quand tu es pas anglais //
[the young people in England < er when they speak < it's amazing // it is necessary to hang on to understand //+ *especially* when you are not English //]

(d) The conjunction can be preceded by a pre-Nucleus Unit like *pour moi* (*for me*) or *à mon avis* (*in my view*):

// il y allait souvent //+ mais *d'après ce qu'on m'a dit* < beaucoup moins régulièrement quand l'hiver arrivait // [invented ex.]

[//he went there often //+ but *as far as I know* < much less regularly when wintertime came//]

A very special pre-Nucleus type we can find in those specific configurations is the expression *et ce* or *et cela* (lit. *and this*), for example:

// il répondait par l'affirmative, //+ *et ce* parce qu'il en avait toujours été ainsi.//   [written ex.]
[*lit.* // he gave a positive answer,//+ *and this* because he had always done so.//]

Let us point out once again that that we consider the delayed clause as a syntactically dependant clause. That position is easy to justify on the basis of two examples like:

il a parlé // mais pas *à* Paul
[he spoke // but not *to Paul*]
il a accepté de se désister // mais pas *en faveur de Paul*
[he accepted to withdraw // but not *in favor of Paul*]

The prepositions *à* (*to*) and *en faveur de* (*in favor of*) clearly show that the delayed sequence has the grammatical form of a canonical complement.

4.4   Type 4: *non dependent sequences inside an autonomous macrosyntactic utterance*

We will give very few illustrations for this type:

// *vu que ça se transmet par les moustiques* < c'est quand même relativement dangereux //
[// since it is a mosquito-borne disease < it is quite dangerous //]

// *comme on le sait* < il y a pas eu d'effusion de sang //
[// as we know < there has been no bloodshed //]

Here, the underlined clauses have the status of a pre-Nucleus Unit. But in contrast with the second type described above, there is absolutely no dependency relationship between that initial sequence and the verb of the following construction (see section 2 above). C. Blanche-Benveniste (1980) describes the link between such clauses and the following verbal construction as a mere "association" relationship.

One hint to the absence of dependency is the impossibility to develop the conjunctional sequence as a cleft:

- it is since it is a mosquito-borne disease that it is quite dangerous;
- it is *as we know* that there has been no bloodshed.

In our annotation system, the angle bracket indicates the end of the pre-Nucleus sequence, but (in contrast with type 2, we use no *"+"* sign, in order to show the absence of any syntactic dependency.

We would adopt the same tagging for sequences placed after the Nucleus Unit, with a right bracket instead of a left one, in order to indicate that the structure is organized as a succession of a Nucleus Unit and a post-Nucleus Unit, as in:

// il y a de la bière dans le frigo > si tu as soif //
[// there is beer in the fridge > if you are thirsty //]

4.5   Type 5: *non-dependent sequences forming a macrosyntactic utterance*

The last configuration we would like to mention is found in examples like:

// ce film n'a pas du tout fonctionné en France tout du moins // parce que en Amérique + < beaucoup de gens sont allés le regarder // [ex. Debaisieux]
[// that film had no success at all in France anyway // because en America +< many people went to see it //]

// généralement < les mâles sont aussi plus beaux et plus colorés dans la plupart des espèces // bien que chez les poissons comme les Trichogaster leeri < ils sont exactement pareils // [ex. Debaisieux]
[// usually < males are more beautiful and more colorful in most species // although with fishes like Trichogaster leeri < they look exactly the same //]

// vos clients euh pourront euh à cet endroit admirer la vue sur le lac et le barrage // parce que n'oubliez pas que le le Muséoscope surplombe le lac de Serre Ponçon hein //
[// your customers er can er in this place admire the sight on the lake and the dam // because don't forget that the Muséoscope overhangs the lake of Serre Ponçon //]

In such examples, the conjunctional sequences (*because…*, *although,…*) are totally distinct from what precedes them both regarding microsyntax, since no dependency relationship can be postulated between the successive sequences, and macrosyntax, since they form utterances bearing their own illocutionary force.

The last example is particularly striking since it shows that the successive constructions are liable to be associated to two different modality values, that is, a declarative in the first one ("your customers can admire the sight on the lake"), and a command in the second utterance ("don't forget that the Muséoscope overhangs the lake of Serre Ponçon"). Just to give another example, the following sequence presents a declarative in the first utterance, and a question in the second (which is in fact some kind of a "rhetorical" question):

// on est influençable par rapport à l'anglais > finalement // parce que pourquoi emprunter des mots euh à l'anglais et pas à l'espagnol ou à l'allemand //

[// we are influenced by English > in fact // because why should we borrow words from English instead of Spanish or German //]

In our view, it would be extremely misleading to describe those conjunctional sequences as "subordinates": all things being equal, the conjunctions seem to behave like connective markers that operate at the discursive level.

The only kind of "independence" they lack is discursive independence, not grammatical one: just like a construction starting with *but*, *therefore* or *anyway* could not be considered as "independent" at the discursive level, the structures illustrated here have to be placed after an utterance on the basis of which they can be interpreted.

5.   Conclusion

Spoken data shows that French conjunctions seem to be used in two very different ways: as a syntactic tool liable to achieve microsyntactic integration; and as a discursive marker devoted to macrosyntactic organization. In the past, most of the studies have mainly focused on the microsyntactic structures, which appear to be more canonical and easier to deal with. But the description of spoken data makes it urgent to go into the detail of macrosyntactic aspects of the problem. In the Rhapsodie frame, we have adopted a range of 4 labels (<, >, +, //) which make it possible to annotate both the dependency relations of the conjunctional phrases, and some major macrosyntactic characteristics (such as the fact that conjunctional phrases are liable to form utterances on their own, or the fact that they can be used as an "ad-Nucleus", bearing no illocutionary value).

6.   References

Bally, C. (1932). *Linguistique française et linguistique générale*. Paris, Leroux.

Benzitoun C., Dister A., Gerdes K., Kahane S., Marlet R. (2009). Annoter du des textes tu te demandes si c'est syntaxique tu vois. In *Arena Romanistica*, 4, pp. 16--27.

Benzitoun, C., Dister, A., Gerdes, K., Kahane, S., Pietrandrea, P. and Sabio, F. (2010). Tu veux couper là faut dire pourquoi. Propositions pour une segmentation syntaxique du français parlé. In *Actes du Congrès Mondial de Linguistique Française* (CMLF 2010). La Nouvelle Orléans, pp. 2075--2090.

Blanche-Benveniste C. (1980). La complémentation verbale: valence, rection, associé. In *Recherches sur le français parlé*, 3, pp. 57--98.

Blanche-Benveniste, C., Deulofeu, J., Stefanini, J. and Van Den Eynde, K. (1984). *L'Approche pronominale et son application au français*. Paris: SELAF.

Blanche-Benveniste, C., Bilger, M., Rouget, C., Van Den Eynde, K. and Mertens, P. (1990). *Le français parlé, études grammaticales*. Paris, Éditions du CNRS.

Debaisieux, J.-M. (2006a). La distinction entre dépendance grammaticale et dépendance macrosyntaxique comme moyen de résoudre les paradoxes de la subordination . In *Faits de Langues*, 28,

pp.119--132.

Debaisieux, J.-M. (2006b). Un cas peu étudié de détachement: les éléments régis en épexégèse. In *Colloque international Les linguistiques du détachement*. Nancy, non publié.

Debaisieux, J.-M., Deulofeu, J. (2009). When a construction constructs the context. In A. Bergs, G. Diewald (Eds.), *Context and constructions*, Amsterdam, John Benjamins, pp.43--62.

Deulofeu, J. (1991). La notion de dépendance syntaxique dans l'approche pronominale. In *L'information grammaticale*, 50, pp.19--24.

Deulofeu, J. (2003). L'approche macrosyntaxique en syntaxe: un nouveau modèle de rasoir d'Occam contre les notions inutiles. In *Scolia*, 16, pp. 77--95.

Deulofeu J. (2011). L'approche Macrosyntaxique en syntaxe: un outil pour traiter le problème des constructions improprement appelées subordonnées. In JJ. Bustos Tovar, R. Cano Aguilar, E. Mendez Garcia de Paredes and A. Lopez Serena (Eds), *Syntaxis y analisis del discurso hablado en Espanol. Honenaje a Antonio Norbona*. Publicaciones Universidad de Sévilla, Sevilla, pp. 731--747

Miller, J., Weinert, R. (1998). *Spontaneous spoken language. Syntax and Discourse*. Clarendon Press, Oxford.

Mithum M. (2005). On the assumption of the sentence as the basic unit of syntactic structure, Linguistic diversity and language theories. Z. Frayzingier, A. Hodges and D.S. Rood (Eds.), *Studies in language*, Companion series. Amsterdam, John Benjamins, pp. 169--183

Narbona Jiménez A. (1990). *Las subordinadas adverbiales impropias en español (II). Causales y finales, comparativas y consecutivas, condicionales y concesivas*, Málaga, Librería Ágora.

Riegel, M., Pellat, J.-C. and Rioul, R. (1994). *Grammaire méthodique du français*, Paris, PUF.

Sabio F. (2006). L'antéposition des complements dans le français contemporain: l'exemple des objets directs. *Lingvisticae Investigationes*, 29:1. Fascicule spécial: Ordre des mots et topologie de la phrase française. Kim Gerdes and Claude Muller (Eds.), John Benjamins Publishing Compagny, pp. 173--182.

Sabio, F. (2012). *Syntaxe et organisation des énoncés, observations sur la grammaire du français parlé*, Mémoire de HDR, non publié.

Sandfeld, K.R. (1943). *Syntaxe du français contemporain – II Les propositions subordonnées*, Paris, Droz.

Verstraete, J.-C. (2007). *Rethinking the coordinate subordinate dichotomy*. Berlin, Mouton.

SPEECH AND SOCIOLINGUISTICS

Nominal agreement in the speech of students from urban areas of Sao Tome

Silvia Figueiredo BRANDÃO

Universidade Federal do Rio de Janeiro/CNPq
Rio de Janeiro, Brazil
silvia.brandao@terra.com.br

Abstract

In this study, performed according to the theoretical and methodological assumptions of variational sociolinguistics, we take up the question of non-implementation of number agreement mark in Noum Phrase (NP) in the speech of Sao Tome, considering individuals from 10 to 18 years in various stages of schooling. It has been designed to test, in speaking of these individuals, the role of variables that were salient for not applying the number mark in the noun phrase (SN). Non-implementation of the nominal plural mark in the speech of students of Sao Tome will depend, among other factors, on the domain or partial knowledge of another language(s) spoken in the region, more interaction with speakers of these languages and on the lower level of education. In the urban variety of Sao Tome, level of education is a variable of primary importance to the distribution of polarized variant patterns of agreement. We discuss the claim of Hagemeijer (2009: 19-20) that, given the linguistic situation of Sao Tome and Principe, which is probably the only country in the Portuguese-speaking Africa where the majority of the population now has Portuguese as first language, there would be conditions for the emergence of a new variety.

Keywords: number agreement; Noun Phrase; Portuguese of Sao Tome; urban variety.

1. Introduction

Questions concerning the loss of inflectional morphology and rules of agreement are important parameters for defining the status of varieties emerging from the contact between linguistically and culturally distinct populations. In this sense, studies about nominal and verbal agreement have served as the basis for the formulation of different interpretations about the emergence and development of varieties of Portuguese, as well as to characterize the Portuguese-based creoles.

Unlike what occurs in relation to the Portuguese of Brazil (PB), there are few studies carried under variational sociolinguistic perspective that focus the nominal agreement in African countries where Portuguese is the official language. In general, studies have been focusing on the Portuguese-based Creole and on cases classified as restructured Portuguese that are observed in rural areas (Baxter, 2009; Figueiredo, 2010). Only recently was awarded the speech of individuals who have Portuguese as L1 and live in urban areas, as found in Brandão (2011a, 2011b), who dealt with this variable in the capital of Sao Tome and Prince, national state which has marked multilingualism,

Brandão (2011a) argues that, among educated speakers, the agreement rule is rated semicategorical, approaching what is seen in European Portuguese, while among those with high school and/or fundamental education, it has variable character, conditioned by linguistic and social factors.

2. Goals

In the current study, we take up the question of non-implementation of number agreement mark in Noum Phrase (NP) in the speech of urban areas of Sao Tome, this time also considering individuals from 10 to 19 years in various stages of schooling. It has been designed to test, in the speech of these students, the role of variables that were salient for not applying the number mark in the noun phrase (NP) according to Brandão (2011b). It starts with the hypothesis that non-implementation of the nominal plural mark in the speech of students of Sao Tome will depend, among other factors, on the domain or partial knowledge of another language(s) spoken in the region, on more interaction with speakers of these languages, on the level of education and particularly on the socio-economic conditions of individuals.

3. The linguistic situation of Sao Tome

In the archipelago of Sao Tome and Prince, located in the Gulf of Guinea, several languages coexist due to a series of historical contingencies related to its colonization process: the Forro (or Santome) and the angolar on the island of Sao Tome, the Lung'ie on the island of Prince, as well as the Creole of Cape Verde, the Portuguese of Tonga and remnants languages from the Bantu group Bantu -, these ones used by a smaller contingent of the population.

Figure 1: Map of Sao Tome and Prince

In this set, stand out the Portuguese and the Forro, which, according to data from the 2001 census, are spoken respectively by 98.9% and 72.4% of individuals over five years (Hagemeijer, 2009: 18), which in general speak two or more of the said languages.

4. Theoretical framework, methodology and brief profile of the informants

The study was conducted according to the theoretical and methodological assumptions of Variationist Sociolinguistics, based on sample selected of nine of the recordings made by Tjerk Hagemeijer on the island of Sao Tome in 2009 and supported in the program Goldvarb-X. Surveys, of the type DID and with 15 to 30 minutes, deal with aspects of life of the informant and his community. Twelve variables were controlled: four extralinguistic, and eight structural.

All the nine informants are only students. Natural of Sao Tome, they live, from birth, in its urban area and have Portuguese as their mother tongue (L1). Family members of some of them live in rural areas, the so-called "roças".

5. Data analysis

The total of 633 constituents of 312 NPs were analyzed. In only 31 cases (4.9%) the number marker was not implemented, as is displayed in Figure 2. The overall index is lower than that obtained by Brandão (2011b) in the speech analysis of 22 individuals from primary and secondary levels of education (12.8%) and different age groups (18-75 years) that have already ended their schooling process.
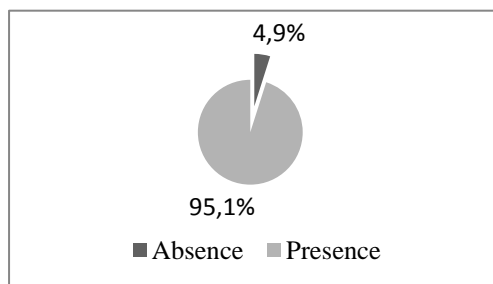

Figure 2: Number marker in NPs

The variationist analysis indicated that the input of the absence of plural mark is very low (.05) and is subject to contraints relating to the performance of the individual (Table 1) and the linear and relative position of the constituent in the the NP (Table 2).

INDIVIDUAL PERFORMANCE							
Informant/ Number of NPs	N	%	R.W.	Informant/ Number of NPs	N	%	R.W.
ST-E1-E6m (10 NPs)		8/19	42.1	.91	ST-E6-FDh (36 NPs)	8/88	9.1	.48
ST-E2E6m (15 NPs)		0/28	0	---	ST-E7-FDh (17 NPs)	3/38	7.9	.38
ST-E3-F8h (36 NPs)		0/62	0	---	ST-E8-FDh (41 NPs)	3/83	3.6	.26
ST-E4-F8m (98 NPs)		0/261	0	---	ST-E9-FDm (44 NPs)	0/91	0	---
ST-E5-F8m (15 NPs)		9/26	34.6	.91				
Input: .05				Significance: .000			

Table 1: Individual Performance

LINEAR AND RELATIVE POSITION OF THE CONSTITUENT IN THE NP		N	%	R.W.
Pre-nuclear	1st position	4/262	1.5%	.25
	2nd/3rd positions	2/26	7.7%	.77
Nuclear	1st position	0/12	0%	---
	2nd position	16/268	6%	.62
	3rd/4th positions	3/35	8,6%	.79
Post-Nuclear	2nd/3rd/4th positions	6/30	20%	.90
Input: .05				Significance: .000

Table 2: Linear and relative position of the constituent in the NP

Of the nine informants, four categorically applied the rule of canonical agreement. Among the five informants to which the rule is variable, two girls showed a greater tendency not to apply the rule: one of the 6th, another of the 9th grade (R.W. .91 in both cases). The remaining three, all male and attending the 10th or 11th

grade, remained below the rate of .50.

Despite the low input of the rule and the small number of data, this analysis confirmed what has been observed in other studies on nominal agreement in both the Brazilian Portuguese and the Portuguese of Sao Tome: linear and relative position of the constituent in the NP is the most relevant linguistic variable to the presence or absence of number marker So, as shown in Table 2, (a) the marks are concentrated (W. R. .25 and .77) in the area to the left, the pre-nuclear area; (b) in the nucleus and from there marks will be less frequent: (i) the nucleus in the second position: R. W. 62, in the third or fourth, R.W. .79; (ii) constituents on the right, R.W. .90.

All nuclei in the first position (located therefore far left) presented plural mark, a trend also observed in the aforementioned analyses. It is, however, one observation on the behavior of the pre-nuclear constituent in second or third position: the R. W. obtained for the non-implementation of the plural mark is often far above the reported rate, usually not more than 20 points higher than that observed in the first position.

6.   Final remarks

Although we have not done a classical variacionist analysis, since it was based on the speech of a small number of informants and not filling with the same number of informants all social cells, the indication of individual performance as the most important variable for the absence/presence of the plural marker in NP suggests that the agreement, in Sao Tome society, has strong socio-economic-cultural implications. Regardless of the level they are in school, while, in the speech of four students, the rule is categorical, in five others, has variable character in a greater or lesser degree. This, of course, is linked to aspects not controlled in this study and which relate to their family environment, to their greater or lesser exposure to cultural goods, to languages spoken in the region, and to the type of school they attend. It is worth noting the remarks of two of the students who use categorically the rule: one claimed that his father gives him all the means for his intellectual development, and another said that their parents prefer her to study at the Portuguese School because they think that in this school the teachers are better prepared, which, consequently, would provide a better quality of teaching.

[+ marks]						[- marks]	
Pre -nucleus		Nucleus			Post-nucleus		
Pos.	Pos.	Pos.	Pos.	Pos.	Pos.	Pos.	Pos.
1	2/3	1	2	3/4	2	3	4/5

Figure 3: Continuum of marking plurality in the NP constituents in non-European varieties of Portuguese

In the speech of the students who apply variably the rule of agreement, the main restrictions governing the marking of plurality, as has been observed also in the PB, are related to the linear and relative position of constituents in NP, which obeys the scale represented in Figure 3 and shows that the marks are concentrated to the left of the nucleus or in the nucleus in first position, decreasing in constituents in the right.

This study, as well as those mentioned here, that is based on *corpora* of spontaneous speech, and that focus nominal agreement in Portuguese of Sao Tome, have confirmed the observations of Hagemeijer (op. cit) regarding the existence of different "registers" (or standards) dependent on the actuation of socio-economic and cultural factors.

This confirms also the tendencies indicated by Brandão (2011a, 2011b), which outlined, for the urban area, a framework of strong sociolinguistic polarization, despite the low overall rate of absence of plural mark in constituents of the NP.

7.   References

Brandão, S.F. (2011a) Concordância nominal em duas variedades do português: convergências e divergências. In *Veredas - Revista de Estudos Linguísticos*, 15 (1), pp. 164--178.

Brandão, S.F. (2011b). O cancelamento da marca de número nominal na variedade urbana não *standard* do Português de São Tomé. In: XVI Congreso Internacional de la ALFAL, 2011, Alcalá de Henares. In *Documentos para el Congreso Internacional de la ALFAL*. Alcalá de Henares: ALFAL/Universidad de Alcalá.

Baxter, A. (2009). A concordância de número. In D. Lucchesi, A. Baxter and I. Ribeiro (Eds.). *O português afro-brasileiro*. Salvador, EDUFBA, pp. 269--318.

Figueiredo, C.F.G. (2010). A concordância plural variável no sintagma nominal do português reestruturado da comunidade de Almoxarife, São Tomé (Desenvolvimento das regras de concordância variáveis no processo de transmissão-aquisição geracional), Tese de Doutorado, Universidade de Macau, Macau.

Hagermeijer, T. (2009). As línguas de São Tomé e Príncipe. In *Revista de Crioulos de Base Lexical Portuguesa e Espanhola* 1 (1), pp. 1--27.

Spoken *corpora* and variation: case-studies

Dinah CALLOU[1,2], **Carolina SERRA**[1], **Erica ALMEIDA**[1,2]

[1]Universidade Federal do Rio de Janeiro, [2]CNPq

dcallou@gmail.com, carolserraufrj@gmail.com, ericadesousa@ig.com.br

Abstract

This paper focuses on four linguistic processes in Brazilian Portuguese: (i) the use of subjunctive *versus* indicative mood in embedded clauses; (ii) the replacement of morphological simple future by periphrastic future; (iii) ***R***-deletion and (iv) vowel harmony. The data are extracted from a *corpus* of informal interviews with university graduates (standard dialect), stratified for age groups (25-35; 36-55; 56 on), gender and geographical region. The analysis makes use of sociolinguistic methodology (Labov, 1994) and the theory of prosodic hierarchy (Selkirk, 1984; Nespor & Vogel, 1986). We conclude that (i) the use of subjunctive in embedded clauses is related to the semantic/lexical component of the main clause and not all verbs license variable use; (ii) in spoken language the morphological simple future has been replaced by periphrastic forms and the hypothesis is that children incorporate the simple morphological future only in school; (iii) there is a gradual process of ***R***-deletion and even the IP and PhP boundaries no longer inhibit deletion of the segment; (iv) vowel harmony process shows stability in Brazilian Portuguese and similar behaviour in all cities. In order to have a clear picture of all processes it is necessary to understand the interplay of grammatical, prosodic and social constraints.

Keywords: variation; subjunctive mood; periphrastic future; *R*-deletion; vowel harmony.

1. Introduction

The aim of this paper is to discuss four variable linguistic processes in standard dialects of Brazilian Portuguese: (i) the use of subjunctive *versus* indicative mood in embedded clauses (eu **não** *acho que seja/é* 'I do not think that it be/is'); (ii) the ongoing replacement of the morphological simple future by the periphrastic future (*cantarei* 'I will/shall sing' ~ *vou cantar* 'I am going to sing'); (iii) ***R***-deletion (*canta∅* ~ *canta**r*** 'to sing') and (iv) vowel harmony (*p**i**rigo* ~ *p**e**rigo* 'danger').

All analyses are based on spoken corpora -- informal interviews --, collected in the 70´s and in the 90´s, with University graduates (standard dialects), in urban centers of Brazil, Salvador, Recife (Northeastern region), Rio de Janeiro, São Paulo (Southeastern region), and Porto Alegre (Southern region). The samples are stratified for age (1= 25; 2 = 36-55; 3 = 56 on) and gender. These speech samples have been built within the Project *"Estudo da norma lingüística urbana culta (NURC)"* and more than 1500 hours of standard dialect are available for research. The analysis makes use of sociolinguistic methodology (Labov, 1994) and VARBRUL/GOLDVARB computational programs.

2. Subjunctive *versus* indicative

The usual explanation for the variable use of subjunctive *versus* indicative mood in Brazilian Portuguese is that there is a difference in meaning between the two constructions: the indicative mood expresses factual reality and the subjunctive mood -- considered by traditional grammar the prototypical mood of subordination -- expresses eventuality and potentiality (the *irrealis* hypothesis).

This variable use is not restricted to Portuguese and has been also attested in other Romance languages such as French (Poplack, 1992) and Spanish (Rivero, 1971; Bosque & Demonte, 1999). Mattos e Silva (1989:

741) points out that this alternation has been in use since the 13th century.

The subjunctive/indicative mood variation occurs not only in adverbial (1), but also in embedded clauses (2), although with different rates.

(1) *Embora o homem **diga/*diz** que está pobre*
Although the man says that (he) is poor

(2) *A mãe de Maria não quer que ela **vá/*vai***
Mary's mother does not want that she go(es)

The use of subjunctive in embedded clauses -- around 20% -- is related to the semantic/lexical component of the main clause (the matrix verb). Not all verbs present variable use of the subjunctive.

Verbs of 'opinion'	Oco/total	% Subj.	% Ind.
Acreditar/crer (believe)	34/50	68%	32%
Supor (suppose)	04/04	100%	0%
Achar (think)	123/1046	12%	88%
Pensar (think)	05/16	31%	69%
Parecer (seem)	01/54	4%	96%

Table 1: Frequency of subjunctive/indicative mood, according to each verb

Comparing dialects (Figure 1 below), we can see that there is a more significant difference of use between the three cities with two verbs: 'acreditar' (believe) and 'pensar' (think).
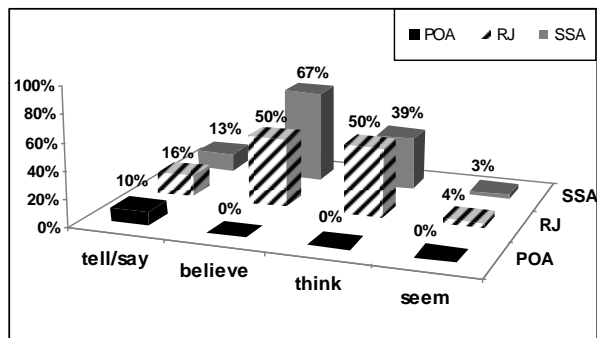
Figure 1: Frequency of use of each verb in each city

Three significant factor groups were pointed out in all dialects. The subjunctive mood (23% - *input* .24) is more frequent when the verb is in the first person rather than in the third person; there is a negative particle in the matrix clause; and the matrix verb is in the past tense, as in example (3), from Callou & Almeida (2009).

Person	Oco / total	%	P.R.
First person	44 / 110	40%	.76
Third person	13 / 135	10%	.28

Table 2: Person of the matrix verb

(3) *eu **pensei** que **fosse** alguma coisa que ele tivesse roubado ...*

I thought that it was something that he had stolen …

Negation effect	Oco / total	%	P.R.
Negative	14 / 19	74%	.92
Assertive	43 / 226	19%	.45

Table 3: Negation effect

(4) *eu **não** acho que casar e ter filhos **seja** uma coisa natural, da vida*
I do not think that getting married and having children be a natural thing, of life

The embedded clause analysis reveals age-group differentiation, when the verb believe 'acreditar' is pointed out (Figure 2): older -- rather than younger -- speakers use the subjunctive more often. Regional and time variables also play a role in mood choice: the use of subjunctive forms is less frequent in Rio than in Salvador (Figure 3), once more, with the verb 'acreditar' (*believe*); from the 70's to the 90's, the use of subjunctive mood is related to the lexical item (Figure 4).
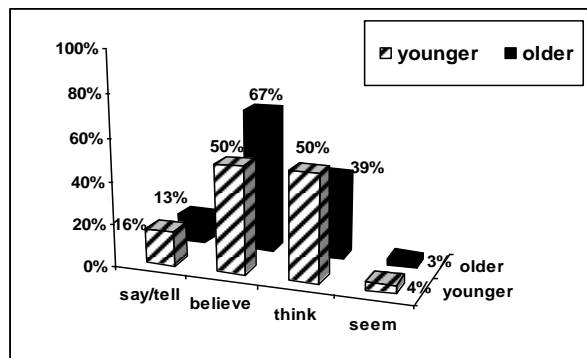

Figure 2: The use of subjunctive with each verb according to age
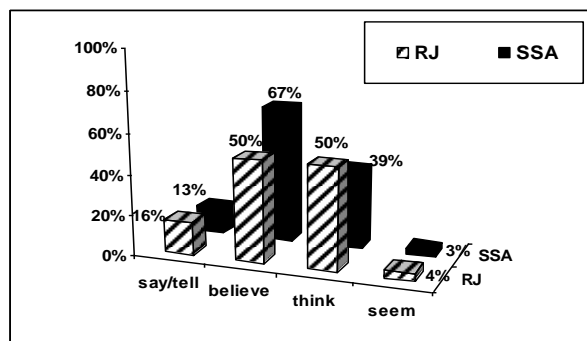
Figure 3: The use of subjunctive with each verb in each city
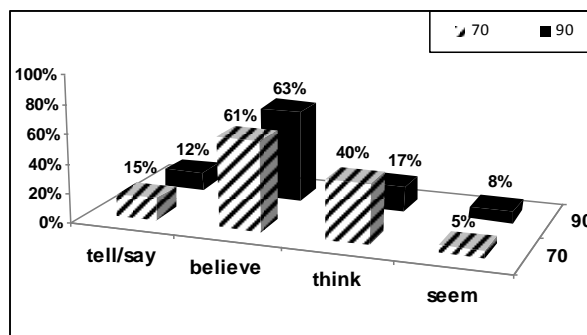
Figure 4: The use of subjunctive with each verb according to decade

3. Periphrastic future *versus* simple morphological future

In Portuguese, future tense is mainly expressed by two simple forms (morphological simple future; simple present tense + obligatory time marker) or by periphrastic forms (present/future tense of modal auxiliary verb *ir* ('to go') + main verb infinitive). In contemporary spoken Brazilian Portuguese the morphological simple future has been replaced by periphrastic forms, except when the auxiliary and the main verbs are the same, as in example (5) below.

(5) eu *vou ir* ao cinema
'I *will go* to the movies'

Nowadays, the use of *haver+de+infinitive* is very rare and put emphasis on the action.

(6) *Hei de trazer* o livro amanhã
'I *will bring* the book tomorrow *for sure*'

spoken language	
morphological simple future	7%
periphrastic form (*ir*+inf.)	77%
simple present tense	16%
Tokens	393

Table 4: Future constructions in contemporary Brazilian Portuguese

Nevertheless, the grammaticization process in Portuguese is still in progress, and a complete merger of adjacent elements has not yet occurred (Oliveira, 2006) and the two elements maintain a certain degree of independence, allowing insertion of adverbs between the auxiliary and the main verb:

(7) ela *vai* simplesmente *escrever…*/ * she will simply write…).

We conclude that variation between simple and periphrastic forms is a reflection of competition between two grammars, following Kroch's proposal (1994), the same way as variation of *ter/haver-existential* constructions. Language acquisition researches have shown that children incorporate the simple morphological future to their lexical inventory only on exposure to a wider range of written language in school.

4. R deletion

Regarding ***R***, our hypothesis is that, besides linguistic and social factors, such as morphological class – non-verbs (ma(***r***) '*sea*') *versus* verbs (canta(***r***) 'to sing') -- age group and region, the prosodic structure also plays a role in the loss of the segment in final coda position. We postulate that the domain of deletion is not the syllable but rather a prosodic boundary, i.e., this phenomenon is also prosodically motivated.

Similar to other segmental phenomena, as external *sândi,* for instance, which takes into consideration prosodic constituent boundaries (Bisol, 1996, 2002; Tenani, 2002), the hypothesis is that ***R***-deletion is also conditioned by the position of the syllable as regards the edge of the prosodic domain:

prosodic word (***Pw***) -- A prosodic word has one and only one primary accent and a PW ^{max} has one and only one prominent element (Vigário, 2003).

A prosodic word is, for instance, the domain of dactylic lowering and neutralization in the direction of a high vowel in Brazilian Portuguese (Battisti & Vieira, 1996);

phonological phrase (***PhP***) -- A phonological phrase should contain more material than one prosodic word (Frota, 2000; Tenani, 2002) and the domain of ɸ-formation is defined by the configuration [… Lex XP …]$_{Lex}$^{max} (where *Lex* stands for the head of a lexical category, and *Lex^{max}* for the maximal projection of a lexical category). In Brazilian Portuguese, ɸ caracterizes itself by regular occurrence of a pitch accent in its more prominent element (Frota & Vigário, 2000; Tenani, 2002; Fernandes, 2007); or

intonational phrase (***IP***) -- The domain of ***IP*** may consist of all the ɸs in a string that is not structurally attached to the sentence tree or any remaining sequence of adjacent ɸ in a root sentence (Nespor & Vogel, 1986). Long phrases (in number of syllables and/or prosodic words) tend to be divided in the same way as small phrases tend to form a unique ***IP*** with an adjacent ***IP***, i.e, balanced phrases are preferred (Frota, 2000; Serra, 2009). In Brazilian Portuguese, the domain of ***IP*** is indicated by a nuclear contour (pitch accent + boundary tone) and a potential pause in its right boundary. There is also a preferential occurrence of L+H* associated to the first stressed syllable of ***IP***, no matter this syllable is the most prominent of ɸ (Tenani, 2002; Moraes, 2007; Serra, 2009; Silva, 2011).

Taking into consideration these three domains, ***R*** deletion would be more frequent at lower levels rather than at higher levels, as we can see in example (8):

(8) [[(pra *sair*)$_{pw}$]$_{php}$]**IP** [[(*teØ*)$_{pw}$]$_{php}$ [(que *ficaØ*)$_{pw}$ (quietinho)$_{pw}$]$_{php}$]IP / to go out (to) have to keep quiet

Data from Votre (1978) and from Gomes (2006) – adult and child speech, respectively, have shown that the presence of a pause -- durational trace frequently associated with the right edge of IP – licenses *R* realization. This reasoning represents another argument in favor of our hypothesis.

In recent research about coda acquisition, in European Portuguese, Jordão (2009) asseverates that the final position of IP clearly favors not only the reconstruction strategies but also the realization of coda.

Moreover, this interpretation could be able to explain the higher frequency of deletion in final coda position (46%) and lower frequency in internal coda position (3%) – Callou *et al.*, 1998.

This analysis is restricted to age group from 25 to 35 years old, male and female, confronting Rio de Janeiro and Salvador data, in order to explain the

trajectory of the phenomenon from initiation to completion, as far as **R**-deletion was strongly concentrated on speakers of this age group (72%), at least, at the beginning of the process. We make use of sociolinguistic methodology (Labov, 1994) and the theory of prosodic hierarchy (Selkirk, 1984; Nespor & Vogel, 1986).

In Rio de Janeiro, **R**-deletion may be considered a midrange change, and in Salvador a change nearing completion, affecting almost every word in which the given sound appears, no matter whether a *verb* (97%) or *non-verb* (78%), as we can see in Figure 5.
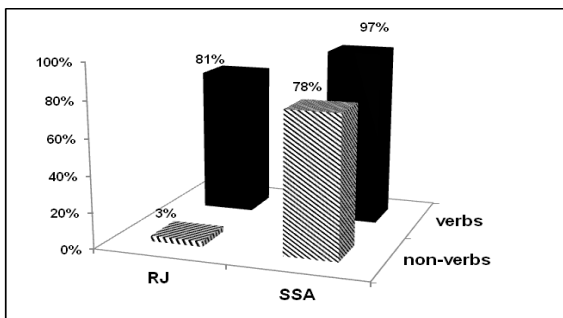

Figure 5: *R* deletion in final coda position, in standard dialect, in Rio de Janeiro and Salvador, in the 70's, according to morphological class

This analysis confirms previous studies with several different samples which have always pointed to morphological class (verbs / non-verbs) as the predominant conditioning factor of this sound change: **R**-deletion is much more frequent in verbs, although it conveys semantically relevant information, for it is a marker of the infinitive and of the subjunctive mood (quere**r** 'to want'; se eu quise**r** 'if I want').

If we compare Rio de Janeiro dialect in real time, in the 70's and in the 90's, we will be able to say that **R**-deletion has continued to advance (Figure 6) and is always conditioned by morphological class.
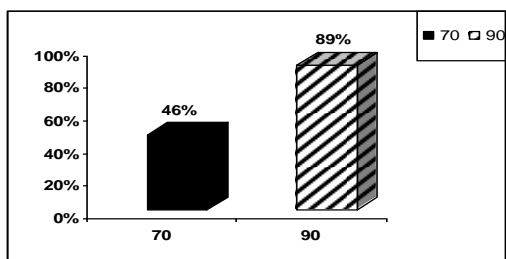

Figure 6: *R* deletion in final coda position, in standard Rio de Janeiro dialect, in the two decades

In Salvador, it is possible to affirm that among young speakers, in the 90's, R-deletion process is completed, no matter the word in which the segment is inserted is a *verb* (100%) or a *non-verb* (99%).

According to the hierarchy prosodic hypothesis, **R** deletion would be more frequent at lower levels rather than at higher levels. The multivariate analysis of 232 tokens allows to conclude that IP and PhP boundaries favor the preservation of the segment while PW favors **R**-deletion, in the 70´s.

The opposition between *verbs* and *non-verbs* remains significant and must be taken into consideration, since it is only if we analyze each boundary separately that it is possible to have a wider vision of the process. At least, at the 70's, in Rio de Janeiro dialect, **R**-deletion in non-verbs is restricted to word boundary (PW).

There is a gradual process of deletion and from the 1970's to the 1990's even the IP and PhP boundaries no longer inhibit deletion of the segment (Figure 7).
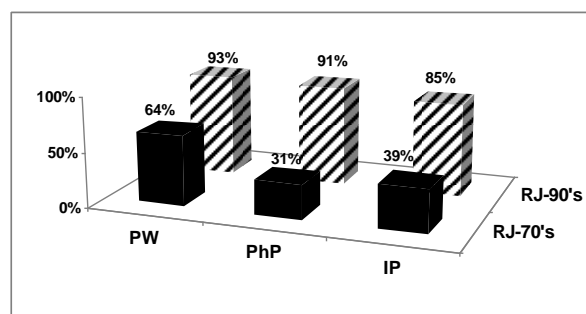

Figure 7: *R* deletion in final coda position, in standard dialect, Rio de Janeiro dialect, in the two decades, according to prosodic boundary

To sum up, we are still trying to understand the interplay of grammatical, prosodic and social constraints which governs **R**-deletion in Brazilian Portuguese.

5. Vowel harmony

Traditionally, vowel harmony is defined as the raising of pre-stressed mid vowels **e** and **o** due to high vowels **i** or **u** in the stressed syllable (p**e**rigo → pirigo '*danger*'; c**o**ruja→c**u**ruja '*owl*'). It can also apply to the lowering of pre-stressed mid vowels in the environment of a low vowel in the stressed syllable, as in *bolota ~ b[⊡]'l[⊡]ta* "ball"; *Pelé~ P[⊡]'l[⊡]* 'Brazilian soccer player' .

Vowel harmony process shows stability in Brazilian Portuguese, although it is a process almost completed in European Portuguese since the 15th century. The analysis has shown that the target vowels */ e /* and */ o /* behave differently in Brazilian Portuguese. We observe that vowel harmony is a split phenomena as far as raising of pre-stress mid vowels can be obtained either by the quality of adjacent syllable high vowel or due to the articulatory or acoustic assimilation of neighboring adjacent consonants: *moqueca* ➔*[m][u]queca* "kind of food"; *boneca* ➔*[b][u]neca* "doll";*pomada* ➔*[p][u][m]ada* "cream"; *colher* ➔ *[k][u]lher "spoon"* .

The comparison of mid vowel raising in five Brazilian cities -- São Paulo (SP), Rio de Janeiro (RJ), Salvador (SSA) and Recife (RE) -- shows a similar

behavior: almost the same general input and conditioning environments, as related above.
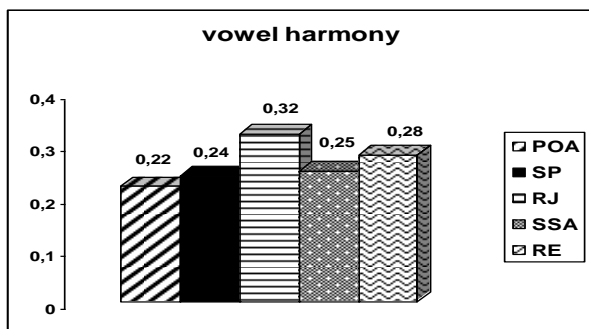

Figure 8 - Comparing dialects (input)

The trapezoid form of the mouth cavity allows a larger vertical space for the production of front vowels than the vertical space for the production of back vowels. Within this hypothesis [i] is higher than [u] (Bisol 1989) and this would explain why [i] is a better trigger than [u]. Bisol's results are based on Porto Alegre data.

Acoustic studies of Brazilian stressed vowels (Moraes, Callou & Leite, 1996) shows, however, that the articulatory explanation does not work in all Brazilian dialects. In Recife, Salvador, São Paulo for instance [i] and [u] have the same F1 value. So F1, related to vowel height, can not be the explanation for the asymmetric behavior of *i / u*.

An alternative hypothesis is that the distinctive feature for back vowels is not degree of openness but degree of labialization (lip rounding). Figure 1 shows that the acoustic space of [o] and [u], based on F1 and F2 plotation, is practically the same, reinforcing this hypothesis. If it is rounding that is the distinctive feature for back vowels, Brazilian vowel system is asymmetrical, as far as for front vowels the distinctive feature is height while for back vowels it is roundness.
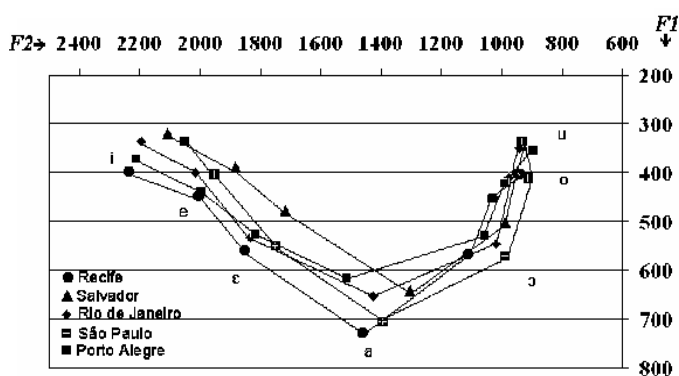

Figure 9: Acoustic space of the stressed BP vowel system of each city

6. References

Battisti, E., Vieira, M.J.B. (1996). O sistema vocálico do português. In *Introdução a estudos de fonologia do português brasileiro*. L. Bisol (Ed.), Porto Alegre: EDIPUCRS, pp. 159--194.

Bisol, L. (1996). O sândi e a ressilabação. In *Letras de Hoje*, v. 31, n. 2, pp. 159--168.

Bisol, L. (2002). A degeminação e a elisão no VARSUL. In L. Bisol, C. Brescancini (Eds.), *Fonologia e variação: recortes do português brasileiro*. Porto Alegre: EDIPUCRS, pp. 231--250.

Bosque, I., Demonte, V. (1999). Gramática descriptiva de la Lengua Española 2. *Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales*. Real Academia Española. Colección Nebrija y Bello. Espasa Calpe, S. A., Madrid.

Callou, D., Almeida, E. (2009). Mudanças em curso no português brasileiro: contrastando duas comunidades. In: *Textos selecionados*. Braga 2008. XXIV Encontro Nacional da Associação Portuguesa de Lingüística. Lisboa/APL, pp. 161--168.

Callou, D., Leite, Y. and Moraes, J. (1998). O sistema pretônico do português do Brasil: regra de harmonia vocálica. In: *Atas do XXI Congresso Internazionale de Lingüística e Filologia Romanza: Sezione 5*.Tübingen, Max Niemeyer Verlag, pp. 95--100.

Callou, D., Serra, C. (2011). Variação do rótico e estrutura prosódica. *Revista do GELNE*.

Fernandes, F.R. (2007). Ordem, focalização e preenchimento em português: sintaxe e prosódia. Tese de Doutorado em Lingüística. Campinas: LEL/UNICAMP.

Frota, S. (2000). *Prosody and focus in European Portuguese. Phonological phrasing and intonation*. New York: Garland Publishing.

Frota, S., Vigário, M. (2000). Aspectos de prosódia comparada: ritmo e entoação no PE e no PB. In R.V. Castro, P. Barbosa (Eds.), *Actas do XV Encontro Nacional da Associação Portuguesa de Lingüística,* v.1. Coimbra: APL, pp. 533--555.

Gomes, C.A. (2006). Aquisição do tipo silábico CV(r) no português brasileiro. In Scripta: Belo Horizontev. 9. N. 18, pp. 11--28.

Jordão, R.M. (2009). A estrutura prosósica e a emergência de segmentos em coda no PE: um estudo de caso. Dissertação de Mestrado em Linguística Portuguesa. Lisboa: Universidade de Lisboa/FLUL.

Labov, W. (1994). *Principles of linguistic change*. Internal factors. Cambridge, Blackwell.

Matos e Silva, R.V. (1989). *Estruturas trecentistas*. Lisboa, Imprensa nacional: Casa da moeda.

Moraes, J.A. (2007). Nuclear and pre-nuclear contours in Brazilian Portuguese intonation. Available at: <www.fl.ul.pt/dlgr/SonseMelodias/PaPI2007ToBIw orkshop>.

Moraes, J., Callou, D.M.I. and Leite, Y. (1996). Neutralização e Realização Fonética: A Harmonia Vocálica no Português do Brasil. In *Anais do Congresso Internacional sobre o Português*, Lisboa: Editora da APL, pp. 395--404.

Nespor, M., Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Oliveira, J. (2006). O futuro da língua portuguesa ontem e hoje: variação e mudança. Tese de Doutorado. Faculdade de Letras, UFRJ.

Poplack, S. (1992). The inherent variability of the French subjunctive. In C. Laeufer, T. Morgan (Eds), *Theoretical Analyses in Romance Linguistics*. John Benjamins, Amsterdam/Philadelphia, pp. 235--263.

Tenani, (2002). Domínios prosódicos do português do Brasil: implicações para a prosódia e para a aplicação de processos fonológicos. Tese de Doutorado em Lingüística. Campinas: LEL/UNICAMP.

Rivero, M.L. (1971). Mood and Presupposition in Spanish. In *Foundations of Language 7*, pp. 305--336.

Selkirk, E. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge: M.I.T. Press.

Serra, C.R. (2009). Realização e percepção de fronteiras prosódicas no português do Brasil: fala espontânea e leitura. Tese de Doutorado em Língua Portuguesa. Rio de Janeiro: UFRJ/Faculdade de Letras.

Silva, J.C.B. (2011). Caracterização prosódica dos falares brasileiros: as orações interrogativas totais. Dissertação de Mestrado em Língua Portuguesa. Rio de Janeiro: UFRJ/Faculdade de Letras.

Vigário, M. (2003). *The prosodic word in European Portuguese*. Berlin/New York: Mouton de Gruyter.

Votre, S. (1978). Variação fonológica no Rio de Janeiro. Tese de Doutorado. Rio de Janeiro/PUC.

Banco de dados sociolinguísticos do Norte do Brasil

Regina CRUZ[1,2], **Carlos NEDSON**[1], **Raquel COSTA**[1], **Josivane SOUSA**[3], **Socorro CAMPOS**[1], **Orlando CASSIQUE**[1], **Doriedson RODRIGUES**[1], **Mara COSTA**[2,4]

[1]UFPA; [2]CNPq; [3]PMPA; [4]Bolsista PIBIC

Av. Augusto Correa, s/n – Campus do Guamá – Belem (PA) – 66075-900

E-mail: regina@ufpa.br, nedson@ufpa.br, raqmaria@ufpa.br, josivane@yahoo.com.br, bmscampos@ufpa.br, cassique@ufpa.br, doriedson@ufpa.br, mara.costa@il.com.br

Abstract

This paper presents how they formed corpora for study of the unstressed mid vowel of the linguistic varieties of Brazilian Portuguese (PB) spoken in Amazon are being organized, processed and annotated. The *NORTE VOGAIS* Project aims to verify the variations of unstressed mid vowel in Amazon PB to provide a sociolinguistic configuration of the phenomena like vocalic harmony or rising in Pará state, for example. So far the formed corpora are from the following cities: Belém (Sousa, 2010; Cruz *et al*., 2008); Cametá (Rodrigues & Araujo, 2007; Rodrigues & Reis, 2012; Costa, 2010); Mocajuba (Campos, 2008); Breves (Cassique *et al*., 2009; Dias *et al*., 2007) and Breu Branco (Marques, 2008). The NORTE Vogais project's team has been investigating three vowel processes in variation: a) unstressed (pretonic) vowel mid rising; (Cruz, 2012, 2010; Sousa, 2010; Rodrigues & Araujo, 2007; Campos, 2008; Cassique *et al*. 2009; Dias *et al*., 2007; Marques, 2008); neutralization of non-final post-tonic vowel (Costa, 2010) and allophonic nasalization (Rodrigues & Reis, 2012). The *NORTE VOGAIS* project has speech samples of 342 PB speakers from Amazon in its database and it is linked to PROBRAVO team.

Key words: sociolinguistic corpora; Amazon Brazilian Portuguese; PROBRAVO project; pretonic mid vowel; linguistic variation.

1. Introdução

Desde 2007, quando passou a integrar o grupo PROBRAVO, o projeto Norte Vogais já efetuou estudos do processo de variação das vogais médias pretônicas do português falado em cinco localidades do Estado do Pará, a saber: i) Cametá (Rodrigues & Araújo, 2007; Rodrigues & Reis, 2012; Costa, 2010); ii) Mocajuba (Campos, 2008); iii) Breves (Cassique *et al,*. 2009; Dias *et al.*, 2007); iv) Belém (Sousa, 2010; Cruz *et al*., 2008) e; v) Breu Branco (Marques, 2008; Coelho, 2008; Campelo, 2008). Todas são descrições sociolinguísticas de cunho variacionista e apresentam um tratamento quantitativo dos dados, que possibilitam uma comparação de seus resultados quanto ao fenômeno estudado, no caso as vogais átonas. São justamente estes procedimentos que passaremos a detalhar no presente trabalho.

2. Projeto Norte Vogais

O projeto Norte Vogais está diretamente ligado ao Diretorio nacional de pesquisa do CNPq PROBRAVO[1], coordenado por Dr. Marco Antônio de Oliveira (PUCMG) e Dr. Seung-Hwa Lee (UFMG). O grupo de investigadores do PROBRAVO realiza uma investigação multidisciplinar – sócio-histórica e linguística – para descrever as realizações fonéticas das vogais nos dialetos do Sul ao Norte do Brasil. Até o presente momento cinco regiões foram investigadas no Estado do Pará: Belém, Breves, Cametá, Mocajuba e Breu Branco, tanto nas suas zonas rurais quanto urbanas.

De maneira geral, a equipe da UFPA pretende ao mesmo tempo caracterizar o sistema vocálico átono e suas variantes, com base em amostra estratificada e em termos variacionistas, assim como analisar e explicar o processo de variação das vogais médias pretônicas e postônicas não-finais no português falado no Norte do Brasil interna e qualitativamente.

3. Fenômenos investigados

As descrições sociolinguísticas empreendidas pela equipe da UFPA priorizam a investigação de três aspectos fonéticos em particular: a) a variação das vogais médias pretônicas; b) a variação das vogais médias postônicas mediais e; c) a nasalidade alofônica, cujos detalhes são fornecidos nesta secção.

3.1 Vogais médias pretônicas

Muitos estudos já foram realizados sobre as vogais médias em posição pretônica no Brasil. Elencamos aqui, a partir de uma sucessão temporal, aqueles realizados na região Norte: Rodrigues (2005) sobre o alteamento /o/> [u] no português falado em Cametá (PA); Dias *et al*. (2007) sobre a alteamento na fala rural de Breves (PA); Oliveira (2007) sobre a harmonização vocálica no português urbano de Breves (PA); Araújo & Rodrigues (2007) sobre as vogais médias /e/ e /o/ no português falado no município de Cametá (PA); Cruz *et al*. (2008) sobre a harmonização das médias pretônicas no português falado nas ilhas de Belém (PA); Campos (2008) sobre o alteamento vocálico em posição pretônica no português falado no município de Mocajuba (PA); Marques (2008) sobre o alteamento das vogais médias pretônicas no português falado no município de Breu Branco (PA) e; Sousa (2010) sobre a variação das vogais médias pretônicas no português falado na área urbana do município de Belém (PA).

---

[1] A equipe do PROBRAVO é responsavel pelo projeto nacional Descrição Sócio-Histórica das Vogais do Português (do Brasil) e pode ser melhor conhecida pelo site http://www.geocities.com/probravo/.

Em sua maioria as descrições sociolinguísticas realizadas pelo projeto Norte Vogais investigaram as vogais médias pretônicas na perspectiva do alteamento (Rodrigues & Araújo, 2007; Oliveira, 2007; Campos, 2008; Marques, 2008; Cassique *et al.,* 2009; Sousa, 2010). Apenas Dias *et al.* (2007) e Cruz *et al.* (2008) analisaram o fenômeno de variação das médias pretônicas na óptica da harmonização vocálica. De forma generalizada, os dados demonstraram uma tendência ao não alteamento nos dialetos paraenses. Os resultados sobre o alteamento confirmaram a afirmativa de Bisol (1981) de serem as vogais altas na sílaba seguinte um contexto altamente favorecedor (Rodrigues & Araújo, 2007; Dias *et al.,* 2007; Campos, 2008; Cruz *et al.,* 2008; Cassique *et al.,* 2009). Outro resultado convergente diz respeito ao fato de os dados de fala de informantes de mais baixa escolaridade e de maior faixa etária apresentarem maior probabilidade de alteamento.

Como se pode constatar avançou-se bastante nas descrições sociolinguísticas das vogais médias pretônicas no português falado na Amazônia Paraense, os procedimentos metodológicos adotados foram comuns, principalmente no que diz respeito a formação dos *corpora* e tratamento dos dados.

3.2   Vogais postônicas não-finais

O único trabalho sobre postônicas mediais realizados no seio do PROBRAVO pela equipe da UFPA é o de Costa (2010). A autora verifica o comportamento das vogais médias /e/ e /o/ em posição postônica não-final de itens lexicais no português falado nas áreas urbana e rural do município de Cametá. O *corpus* foi constituído com amostras de fala de 96 informantes estratificados em sexo, faixa etária, nível de escolaridade e procedência. A coleta dos dados foi realizada através de dois tipos de entrevista: a *livre* (48 informantes); e o *teste* ou nomeação de figuras (48 informantes).

O *corpus* apresenta **2.177** dados, sobre o qual se observou a partir de uma análise estatística, no programa computacional *Varbrul*, considerando variáveis linguísticas e não linguísticas, que o fenômeno de alteamento com peso relativo de **.46** apresenta probabilidade menor de ocorrência do que o não alteamento com peso relativo de **.54**.

Este trabalho apresenta igualmente uma análise qualitativa do comportamento das vogais médias - /e/ e /o/ - postônicas não-finais, as quais apresentam quatro variantes possíveis: manutenção [e]/[o], alteamento [i]/[u], apagamento [ø] e abaixamento [E]/[O].

Costa (2010) procede igualmente a uma descrição fonológica das vogais médias postônicas - /e/ e /o/ - não-finais, cujo objetivo é verificar como o ambiente fonético é determinante no comportamento das quatro variantes identificadas, a saber: manutenção (abób[o]ra / velocíp[e]de), alteamento (abób[u]ra / velocíp[i]de), abaixamento (abób[O]ra / cér[E]bro) e apagamento (abób[ø]ra / velocíp[ø]i).

3.3   Nasalidade alofônica

Outro estudo sobre vogais átonas no escopo do projeto PROBRAVO foi o de Rodrigues & Reis (2012) sobre a nasalidade alofônica na variedade do português falada em Cametá (PA). De acordo com os resultados de Rodrigues & Reis (2012) há maior probabilidade de ocorrer a nasalização vocálica pretônica, decorrente da assimilação do traço nasal da consoante da sílaba seguinte, em detrimento da não nasalização vocálica pretônica.

O outro trabalho sobre o fenômeno da nasalidade alofônica é o de Cassique (2002) que estudou o português falado na zona urbana de Breves, na ilha do Marajó. Cassique (2002) detectou de 2013 ocorrências de nasalidade alofônica na variedade do português falada em Breves, que 1070 são manifestações para a variante nasalizada, 53%, e 943 dados atestando a variante não-nasalizada, 47%. Comparando-se os resultados de Cametá e Breves com os das cinco capitais brasileiras, presentes em Abaurre & Pagotto (2002), obteve-se o seguinte quadro de tendência de nasalidade do português brasileiro, como visualizado no Gráfico 1 abaixo.

Gráfico 1: Tendência da nasalidade alofônica do norte ao sul do Brasil. Fonte: Cruz (2010: 253)

Constata-se, portanto, que há um declínio da nasalidade do norte ao sul do Brasil. O índice baixo da variedade de Breves parece não contrariar tal tendência, uma vez que Breves tem indícios de apresentar uma situação sociolinguística particular que será comentada na secção 6.

4.   Procedimentos metodológicos adotados por projetos

Os dados foram coletados em trabalho de campo, com gravações em áudio. Para a coleta destes, priorizaram-se as narrativas de experiência pessoal nos moldes da teoria da variação (Tarallo, 1988). Utilizou-se para cada variedade investigada uma amostra estratificada em sexo, faixa etária (15 a 25 anos; 26 a 45 anos e acima de 46 anos) e escolaridade (analfabeto, fundamental, médio e superior).

Uma vez as gravações concluídas, os dados obtidos foram transcritos grafematicamente observando os parâmetros da Análise da Conversação (Castilho, 2003).

Um arquivo contendo a triagem dos dados, tomando como unidade de análise o grupo de força

como estabelecido por Câmara Jr. (1969), foi criado, por informante. Uma cópia do mesmo foi feita, para nela se proceder à transcrição fonética do vocábulo contendo o fenômeno estudado. Utilizou-se para a transcrição fonética o alfabeto SAMPA[2].

Uma vez a transcrição fonética concluída, procedeu-se à codificação dos dados. Para os estudos sobre vogais médias pretônicas, utilizou-se o mesmo arquivo de especificação do PROBRAVO, de autoria de Orlando Cassique e Doriedson Rodrigues. Costa (2010) e Rodrigues & Reis (2012), por conta da especificidade de seus estudos, utilizaram arquivos de especificação mais adequados a seus objetos de estudo.

De maneira geral, os arquivos de especificação contém fatores de diversas naturezas: a) fonéticos b) morfológicos; c) sintático entre outros, além dos fatores sociais. Por último, realizou-se o tratamento estatístico dos dados pelo programa VARBRUL.

5. Caracterização dos *corpora* formados

Os *corpora* do projeto Norte Vogais possui um número total de informantes variando de 24 (vinte e quatro) a 72 (setenta e dois), como podemos visualizar no Quadro 1 abaixo.

Localidade	Total de informantes	Fonte
Breves (urbano)	42	Oliveira (2007)
Breves (rural)	36	Dias *et al* (2007)
Belém (urbano)	48	Sousa (2010)
Belém (rural)	24	Cruz *et al* (2008)
Cametá	48	Costa (2010)
	72	Rodrigues (2005)
Mocajuba	48	Campos (2008)
Breu Branco	24	Marques (2008), Campelo (2008) e Coelho (2008)

Quadro 1: Número total de informantes do Projeto Norte Vogais por variedade investigada com a indicação da fonte de cada estudo realizado. Fonte: Atualizado de Cruz (2012: 200)

Localidade	Total de informantes	Duração total das gravações	Fonte
Breves	42	10 h 35 min	Oliveira (2007) Dias *et al* (2007)
Belém	48	15 h 28 min	Sousa (2010) Cruz *et al* (2008)
Cametá	120	45 h 21 min	Costa (2009) Rodrigues (2005)
Mocajuba	48	24 h 21 min	Campus (2008)
Breu Branco	24	4 h 24 min	Marques (2008) Campelo (2008) Coelho (2008)

Quadro 2: Tamanho do corpus do Projeto Norte Vogais em horas de gravação

O Projeto Norte Vogais do Brasil conta com um banco de dados de amostra de fala de 342 (trezentos e quarenta e dois) informantes nativos da Amazônia Paraense, originários de cinco variedades locais: Belém, Cametá, Breves, Breu Branco e Mocajuba, em suas zonas rural e urbana.

Além das transcrições, o *corpus* contém o áudio das gravações realizadas em trabalho de campo. O Quadro 2 contem uma descrição do *corpus* em horas gravadas.

6. Tendência do Português da Amazônia Paraense

De forma geral, as descrições sociolinguísticas realizadas sobre o português falado na Amazônia Paraense tem demonstrado uma tendência à não aplicação da regra de alteamento das vogais médias em posição pretônica, como podemos constatar no Quadro 3 abaixo.

Dialeto	Não aplicação da regra	Aplicação da regra	Fonte
Breves (urbano)	81	19	Oliveira (2007)
Breves (rural)	57	43	Dias et al. (2007)
Breves (geral)	67	33	Cassique et al. (2009)
Cametá	60	40	Rodrigues & Araújo (2007)
Belém (urbano)	64	36	Sousa (2010)
Belém (rural)	53	47	Cruz et al. (2008)
Mocajuba	51	49	Campos (2008)
Breu Branco	76	24	Marques (2008)

Quadro 3: Percentual de alteamento nas variedades linguísticas investigadas pelo Projeto Norte Vogais. Fonte: Atualizado de Cruz (2012: 202)
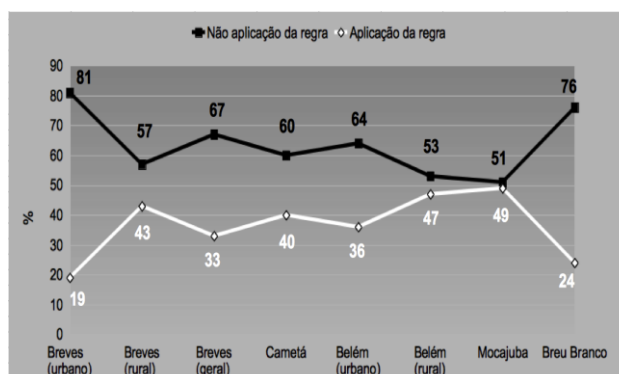

Gráfico 2: Tendência ao não alteamento das vogais médias pretônicas no Português da Amazônia Paraense, de acordo com os resultados dos trabalhos realizados pela Equipe do Projeto Norte Vogais da UFPA. Fonte: Atualizado de Cruz (2012: 203)

Outro resultado relevante compreende a inexpressiva ocorrência de vogais médias baixas nas

---

[2] http://www.phon.ucl.ac.uk/home/sampa/index.html

posições átonas. Tais resultados contrariam de um lado a divisão dialetal de Nascente que caracteriza os dialetos do Norte do Brasil como apresentando uma tendência à realização das vogais médias abertas nas posições átonas, em oposição aos dialetos do Sul do Brasil que prefeririam as vogais médias fechadas. Por outro lado os resultados reforçam a hipótese de Silva Neto (1957) de que o Pará compreenderia uma ilha dialetal na classificação de Antenor Nascente entre os dialetos do Norte do Brasil. Silva (1989) menciona nos seus resultados, uma predominância das vogais baixas no seu *corpus* formado com amostras de fala do dialeto alvo – o de Salvador -, que fora confrontado com amostras de fala de 50 pontos do território baiano e de uma localidade do estado de Sergipe emprestadas, respectivamente, do Atlas Prévio dos Falares Baiano e de Mota (1979).

Os resultados dos estudos empreendidos pela equipe do Projeto Vozes da Amazônia têm buscado prioritariamente caracterizar o português regional paraense. Nesse sentido, os resultados sobre as vogais médias pretônicas têm demonstrado uma tendência ao uso de suas variantes com probabilidade de maior ocorrência de manutenção das médias pretônicas em decorrência do alteamento das mesmas, inclusive com índices percentuais muito próximos de ocorrência da manutenção das médias pretônicas entre as variedades investigadas (Breves (rural), Belém, Cametá e Mocajuba). Duas, das variedades investigas (Breves (urbano) e Breu Branco) confirmam a tendência à manutenção, mas apresentam percentuais muito destoantes das quatro outras variedades comparadas.

Os resultados do estudo da variação das médias pretônicas no português da Amazônia Paraense mostram que os percentuais de alteamento são muito baixos de modo geral nas zonas dialetais do Pará.

Os índices mais destoantes de Breves (33%) e de Breu Branco (24%), por indicarem a necessidade de uma investigação mais aprofundada sobre a situação sociolinguística destes dois municípios em particular, levaram a equipe da UFPA vinculada ao PROBRAVO a lançar uma nova edição do **Vozes da Amazônia** destinada a investigar o português falado nas zonas de migração do Pará3. Breves e Breu Branco apresentam em comum o fato de terem sido justamente regiões que receberam um fluxo migratório considerável em decorrência de projetos econômicos da região.

O município de Breves sozinho apresenta um terço da população de todo arquipélago marajoara. O inchaço populacional sofrido por Breves se deu no segundo ciclo da borracha, durante a segunda guerra mundial, quando o governo apostando em um crescimento econômico oriundo da borracha, fez vir nordestinos para trabalharem na exploração da borracha na Amazônia, os ditos soldados da borracha. Uma vez terminada a guerra e o declínio do segundo ciclo da borracha, os imigrantes nordestinos não tiveram como voltar para a sua terra de

origem e fixaram residência obrigatoriamente na Amazônia, uma boa parte deles ficou justamente na cidade de Breves.

Breu Branco é um dos municípios de criação recente no Pará, seus moradores, em sua maioria, são brasileiros originários de diferentes regiões do Brasil – mineiros, paulistas, gaúchos, paranaenses, maranhense, cearense, piauiense, tocantinenses – que migraram para o Pará para trabalhar na construção da hidrelétrica de Tucuruí na década de oitenta. Com a conclusão da primeira etapa dos trabalhos de implantação da Hidrelétrica de Tucuruí, a maioria desses trabalhadores fixou residência nos municípios da região. Desta forma a população atual de Breu Branco se assemelha a de Brasília. Breu Branco, portanto, apresenta a mesma situação linguística atestada em Brasília (DF) e no sul do Pará onde por questões econômicas – no caso de Breu Branco (PA) tal situação foi ocasionada pela construção da hidrelétrica de Tucuruí – vários dialetos do português brasileiro convivem em uma mesma localidade, ocasionando de tal contato dialetal uma nova norma linguística.

Os resultados dos estudos sobre as vogais médias das variedades da Amazônia Paraense demonstraram que estas duas variedades investigadas fogem completamente a uma característica comum das variedades da Amazônia paraense que é a quase neutralização da variação entre as médias. As variedades de Breu Branco (próximo a Tucuruí) e da zona urbana de Breves (no Marajó) têm como pontos em comum o fato de serem localidades que receberam uma forte migração de falantes do português de outras regiões do Brasil por conta de projetos econômicos. Essas regiões não possuem marcas de identidades (e aí em todos os sentidos) com a Amazônia paraense, e tudo indica inclusive na variedade linguística.

Nossa hipótese é a de que os fatores externos são relevantes no condicionamento da realização das variantes das médias pretônicas e fazem com que tais variedades sejam muito diferentes das demais da Amazônia Paraense. Para comprovar tal hipótese procederemos a uma nova coleta de dados, controlando como principal fator a origem ou ascendência do falante, como fez Bortoni-Ricardo (1985). Acreditamos ser talvez o fator que esteja controlando a realização dessas variantes. Verificaremos também além da variável origem do falante, o fator faixa etária, em especial a fala dos mais jovens, a fim de se verificar se se trata de uma variação estável ou mudança em progresso.

Como última hipótese, acreditamos que nas regiões em questão – Breu Branco e Breves - ainda não se cristalizou uma nova norma resultado do contato intervariedades, como ocorrido em Brasília, e o fato desta nova norma ainda não ter sido estabelecida resulta em contraste muito acentuados da realização das variantes atestadas.

Os resultados sobre a nasalidade vêm justamente fortalecer nossa hipótese de sustentação de uma

---

3 Trata-se do Projeto de Pesquisa *Vozes da Amazônia*, (Portaria Nº 075/2009 ILC).

investigação diferenciada para o português falado nas zonas de migração, uma vez que os dados de Breves (Cassique, 2002) contrariam a tendência da nasalidade do português falado no Norte que seria de ocorrência de alto índice de nasalidade.

7. Conclusão

O presente trabalho apresenta os corpora formados pela equipe do Projeto Norte Vogais vinculado ao PROBRAVO que estuda prioritariamente o vocalismo átono no Norte do Brasil, mais especificamente na Amazônia Paraense.

O projeto conta com corpora formados da variedade do português falada nas localidades de: Cametá, Mocajuba, Breves, Breu Branco e de Belém.

Ao todo o banco de dados do referido projeto contém amostras de fala de 342 informantes nativos do Pará e um total de mais de 100 horas de gravação.

Este banco de dados já subsidiou a investigação de três fenômenos relacionados diretamente ao vocalismo átono: o alteamento das vogais médias pretônicas; a neutralização das vogais postônicas mediais e a nasalização alofônica.

8. Referências

Abaurre, B., Pagotto, E. (2002). Nasalização no português falado no Brasil. In D. Hora (Ed.), *Gramática do português falado,* 2. ed. rev. v. 6. São Paulo: Editora da Unicamp, pp. 491--515.

Bisol, L. (1981). Harmonia Vocálica: uma regra variável. Tese (Doutorado em Linguística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Bortoni-Ricardo, S.M. (1985). *The urbanization of rural dialect speakers: a sociolinguistic study in Brazil*. Cambridge: Cambridge University Press, 265p.

Câmara Jr, J. (1969). *Estrutura da Língua Portuguesa*. Petrópolis: Vozes.

Campelo, M. (2008). A Variação das Vogais Médias Anteriores Pretônicas no Português Falado no Município de Breu Branco(PA): uma Abordagem Variacionista. Trabalho de Conclusão (Graduação em Letras) – Universidade Federal do Pará, Belém.

Campos, B. (2008). Alteamento vocálico em posição pretônica no português falado no Município de Mocajuba-Pará. Dissertação (Mestrado em Letras) – Universidade Federal do Pará, Belém.

Cassique, O. (2002). Minina bunita... olhos esverdeados (um estudo variacionista da nasalização vocálica pretônica no Português falado na Cidade de Breves/PA). Dissertação (Mestrado em Letras) – Universidade Federal do Pará, Belém.

Cassique, O. *et al*. (2009). Variação das Vogais Médias Pré-tônicas no português falado em Breves (PA). In D. Hora (Ed.), *Vogais no ponto mais oriental das américas.* João Pessoa (PB): Ideia, pp. 163--184.

Castilho, A. (2003). *A língua falada no ensino do português*, 5ª. Edição. São Paulo: Contexto.

Coelho, M.L. (2008). A Variação das Vogais Médias Posteriores Pretônicas no Português Falado no Município de Breu Branco(PA): uma Abordagem Variacionista. Trabalho de Conclusão (Graduação em Letras) – Universidade Federal do Pará, Belém.

Costa, R. (2010). Descrição sociolinguística das vogais médias postônicas não-finais /o/ e /e/ no português falado no município de Cametá-PA. Dissertação (Mestrado em Letras). Universidade Federal do Pará, Belém.

Cruz, R. (2012). Alteamento vocálico das médias pretônicas no português falado na Amazônia Paraense. In S.H. Lee (Ed.), *Vogais além de Belo Horizonte*. Belo Horizonte: Faculdade de Letras da UFMG, pp. 194--220.

Cruz, R. *et al*. (2008). As Vogais Médias Pretônicas no Português Falado nas Ilhas de Belém (PA). In M.S. Aragão (Ed.), *Estudos em fonética e fonologia no Brasil*. João Pessoa: GT-Fonética e Fonologia / ANPOLL.

Cruz, R. (2010). Panorama Sociolinguístico do Português Falado na Amazônia Paraense. In M. do S. Simões (Ed.). *Navegando entre o Rio e a Floresta por vias do Marajó: com vista a ensino, pesquisa e extensao*. Belém, pp. 243--261.

Dias, M. *et al*. (2007). O alteamento das vogais pré-tônicas no português falado na área rural do município de Breves (PA): uma abordagem variacionista. In *Revista Virtual de Estudos da Linguagem (REVEL)*. Porto Alegre, n. 9, vol. 5, jul. Disponível em: <http://www.revel.inf.br/site2007/_pdf/9/artigos>.

Marques, L. (2008). Alteamento das Vogais Médias Pré-tônicas no Português Falado no Município de Breu Branco (PA): uma Abordagem Variacionista. Trabalho de Conclusão (Graduação em Letras) – Universidade Federal do Pará, Belém.

Mota, J. (1979). Vogais antes do acento em Ribeirópolis (SE). Dissertação (Mestrado em Língua Portuguesa) - Universidade Federal da Bahia, Salvador.

Oliveira, D. (2007). *Harmonização vocálica no português falado na área urbana do município de Breves/PA: uma abordagem variacionista*. Belém: UFPA. (Plano PIBIC/CNPq).

Rodrigues, D. (2005). Da zona urbana à rural/entre a tônica e a pretônica: alteamento /o/ > [u] no português falado no município de Cametá/Ne paraense: uma abordagem variacionista. Dissertação (Mestrado em Letras) – Universidade Federal do Pará, Belém.

Rodrigues, D., Araújo, M. (2007). As vogais médias pretônicas / e / e / o / no português falado no município de Cametá/PA – a harmonização vocálica numa abordagem variacionista. In L. Bisol, C. Brescancini (Eds.), *Cadernos de Pesquisa em Linguística, Variação no Português Brasileir*, volume 3, Porto Alegre, novembro, pp. 104--126.

Rodrigues, D., Reis, G. (2012). Variação da Nasalização Vocálica Pretônica Seguida de Consoante Nasal na

Sílaba Seguinte no Português Falado no Município de Cametá – Pará. In S.H. Lee (Org.), *Vogais além de Belo Horizonte*. Belo Horizonte: UFMG, pP. 322--348.

Silva, M.B. (1989). As pretônicas no falar baiano: a variedade culta de Salvador. Tese (Doutorado em Língua Portuguesa) – Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Silva Neto, S. (1957). *Introdução ao Estudo da Língua Portuguesa no Brasil*. 4 ed. Rio de Janeiro, Presença.

Tarallo, F. (1988). *A pesquisa sociolinguística*. São Paulo: Ática. (Série Princípios).

Sousa, J. (2010). *A Variação das Vogais Médias Pretônicas no Português Falado na Área Urbana do Município de Belém/PA*. Dissertação (Mestrado em Letras) – Universidade Federal do Pará, Belém: UFPA.

And I'd say "This week, we're not going to clean the windows": direct reported speech within a domestic labor workplace context

Kellie GONÇALVES

University of Bern, English Department
Länggassstrasse 49, 3000 Bern 9, Switzerland
goncalves@ens.unibe.ch

Abstract

This study investigates the meta-discursive accounts of successful and unsuccessful communication within a domestic labor workplace context of a multilingual cleaning company in New Jersey, USA. 41 semi-structured interviews were carried out with Portuguese-speaking domestics, language brokers and their Anglophone clients in order to understand how meaning is negotiated within this particular language contact situation. The analysis indicates that the main linguistic feature employed by participants was that of direct reported speech (DRS). Using DRS functioned to dramatize the effect of their speech events, represented the development of their accounts among interlocutors at the time of the actual conversation as well as claiming authenticity about their actual language practices within their daily interactions. The specific linguistic features investigated include personal, spatial and temporal deictic markers, marked changes in prosody, and speech verbs.

Keywords: reported speech; deictic markers; domestic labor workplace; discourse analysis.

1. Introduction

This study is about a specific language contact situation among Portuguese-speaking domestics and English-speaking clients in New Jersey, USA. It is part of a larger project on communication among domestics and their Anglophone clients, where meta-discursive strategies and the significance of dense, tightly-knit social networks (Milroy, 1980; Milroy & Milroy, 1992; Wei, 1993; Stoessel, 2002) are investigated as well as the linguistic landscapes of the neighborhood in which domestics reside. Preliminary results indicate that domestics' use of English in the workplace consists of meta-linguistic strategies such as 'basic' English, gestures, as well as communicating through 'language brokers' (Tse, 1996; Weisskirch & Alva, 2002; Weisskirch, 2005; Del Torto, 2008)[1]. As a result of living in a Portuguese-speaking community, most of these women do not require English on a daily basis since most of their interactions can be carried out in Portuguese only. In meta-discursively reconstructing their interactions with one another, direct reported speech (DRS) (Volosinov, 1971; Bakhtin, 1981; Goffman, 1981; Coulmas, 1986; Li, 1986; Tannen, 1986, 1989; Clark & Gerrig, 1990; Buttny, 1997; Biber *et al.*, 1999; Holt, 1996, 2000, 2009; Myers, 1999; Carter & McCarthy, 2006; Sams, 2007, 2010) is employed, which functions to convey authenticity of the actual speech event (Coulmas, 1986; Li, 1986; Mayes, 1990; Holt, 1996, 2000, 2009), as well as representing the development of the conversation between parties and the interlocutors' respective stances (Holt, 1996; Niemelä, 2005). Moreover, the use of DRS within this context functions to depict the story's climax (Drew, 1998; Clift, 2000; Golato, 2000) and dramatize (Mayes, 1990; Myers, 1999) the effect of achieving both successful and unsuccessful communication within the reported interaction between

domestics, clients and language brokers. The features of DRS that are scrutinized in this study include personal, spatial and temporal deictic markers, marked changes in prosody, and speech verbs (Holt, 1996). More specifically, the personal pronouns investigated include (I, you, she, we, they) while the spatial and temporal markers include those tense (present, continuous, past, etc.) and time adverbials (then, now), while the speech verbs consists of the reporting clause, namely a pronoun or name followed by a reported verb such as "said" or the quotative "like". For Carter and McCarthy indexical markers or deictic words "are especially common in situations where joint actions are undertaken and where people and things referred to can be seen by the participants" (2006: 178). Deictic markers index the various ways individuals orient themselves and their interlocutors in interaction and function to make reference to physical, psychological and emotional closeness and distance as well as expressing contrast and difference (ibid.). A discourse analytic approach is employed within this study in order to reveal how the use of DRS within the context of spoken discourse functions and deems communication among Portuguese-speaking domestics and their Anglophone clients as successful or unsuccessful. The research questions driving this study are:

1) What linguistic strategies are used by participants to meta-discursively describe communication in their workplace?
2) What linguistic features are employed in their descriptions and what functions do they serve?

2. Data Collection

Obtaining data for a project among domestics and their employers can be extremely challenging and has been well documented by several researchers (Rollins, 1985; Anderson, 2000; Chang, 2000; Parreñas, 2001; Romero, 2002; Lan, 2006 and Parreñas, 2008). While Romero

---

[1] A language broker functions as an intermediary between individuals coming from two different L1 backgrounds.

(2002) worked as a domestic herself, Rollins (1985: 9) "worked for a month as a domestic to submerge [herself] in the situation prior to designing the research in order to sensitize [herself] to the experience of domestic work and of relating to a female employer". I was fortunate that I had direct access to a cleaning company in New Jersey through familial ties and was able to conduct interviews with both employees and clients.

The data for this study consists of 41 semi-structured interviews, 18 with domestics, 19 with clients and 4 with language brokers. The interviews were recorded and lasted between 16 minutes – 1 hour and 30 minutes producing a total of 21.5 hours of recordings. Due to the data-driven nature of this study, hypotheses were not addressed in an *a priori* fashion. Rather, several thematic categories emerged from the transcripts and corpus, which are indicated in table 1.0

Categories	Domestics	Clients
*Language use & practices at work	X	X
Language attitudes	X	X
English skills among domestics	X	X
Social networks	X	

Table 1: Thematic categories

For the purposes of this study, I looked at language use and practices at work among domestics and clients. Below I scrutinize three excerpts, one from a Luso-Brazilian Portuguese-speaking domestic, one from an Anglophone language broker and the last one from an Anglophone client. In investigating how communication is achieved in the workplace context, I analyze how meaning is negotiated by interviewees' evaluations and the DRS employed to reconstruct their conversations, which are deemed successful or not.

In extract 1 below, Livia, a Brazilian domestic, who has been residing in the U.S. for seven years discusses her difficulties of speaking English, but describes her ability to understand English at work when it is in written form. In order to exemplify what she means, Livia employs DRS to reconstruct a telephone conversation she had with Dona Magda, the company owner and language broker, concerning the content of a note left for Livia by an English-speaking client:

Extract 1) A domestic's interpretation

1. L: mas olha eu não consigo soltar a língua (.)
2.   não sei se é vergonha também (.) sabe (.)
3.   não sei
4. K: e com os clientes?=
5. L: =ãh?=
6. K: =e com os clients (.) por exemplo?
7. L: entendo que é XXX (.) igual quando elas
8.   escreve alguma coisa eu sempre entendo (.)

9.   eu sempre ligo pra dona magda e falo
10.   "dona magda olha eu (1.0) <u>tá assim assim</u>
11.   <u>assado</u>" "<u>ah</u> (.) <u>mas é isso</u>?" "tá <u>ok</u>" é o que
12.   eu falei era aquilo mesmo (.) ela falou (.)
13.   "não (.) tá tudo certo"

Livia begins this extract by explaining her challenges of speaking English when she employs the metaphor "soltar a língua" (line 1). She continues and states that she is not sure why, but confesses that it could be her embarrassment "vergonha também" (line 2) at actually speaking. When asked about her communication with clients, Livia states that she always understands when they write her notes "quando elas escreve alguma coisa eu sempre entendo" (line 8). Her use of the adverb of frequency always "sempre" is repeated in line 6 when she claims to always call her boss in order to confirm that she has understood the client's note of instruction through written text. Livia reconstructs this conversation by using several features of DRS such as personal and temporal deixis markers, reported verbs as well as a shift in prosody. First, Livia uses the personal pronouns I "eu" and she "ela" to refer to herself and Dona Magda (lines 9 & 12) as the speakers of the conversation. Second, Livia employs the reported verb say in "falo" (line 9) to introduce her reported utterance and the pronoun-plus-speech-verb "ela falou" (line 12) to reintroduce Dona Magda into the conversation. This reintroduction of Dona Magda occurs in line 10 subsequent to the adjacency pair of a question and answer sequence that has been exchanged by Livia and her boss through the changes in prosody, represented in the extract by the underlined words, to mark both speakers (lines 10 & 11). Finally, Livia's use of the verb tenses within this conversation are the present tense of the verb to be in "tá assim", "é isso" and "tá ok" and are considered "appropriate to the reported speaker/context rather than the current one" (Holt 1996: 222). The exchange between Livia and Dona Magda presented in this extract is one that occurs on a regular basis in order to confirm Livia's comprehension of the English instructions left for her by her English-speaking client. The DRS within this exchange indexes Livia as somebody who understands English well, but may be just embarrassed to speak it while simultaneously depicting Dona Magda as the language broker who provides encouragement and confirmation of Livia's English comprehension skills "tá tudo certo"(line 7). As a result, this sequence depicts the communication between Livia, Dona Magda and the client as successful.

In the next extract, Janet, the English-speaking driver, who also functions as the main language broker when the company owner is unavailable, discusses and assesses Bella's (a Portuguese domestic) English skills. Janet claims that because of Bella's language insecurity, communication is stymied, which has previously led to prolonged and unnecessary problems:

Extract 2) A language broker's view

```
1. Janet  bella's problem is (.) is her inse:cu:rity
2.        about her english and i tell her that (.) i
3.        said (.) "bella (.) I understand everything
4.        you::'re sa::ying to me" and you know like
5.        over christmas (.) one of her insecurities (.)
6.        i felt (1.0) if she wouldn´t have felt so
7.        insecure (.) we could´ve resolved some
8.        problems faster
```

In this extract, Janet reconstructs the conversation she had with Bella by using DRS, which functions to replicate the actual conversation as well as dramatize the hardships concerning their communication. This is done through Janet's use of the speech verb "I said" (line 3) as well as the personal pronouns "I", "you" and "me". The personal pronouns "I" and "me" are co-referential with Janet who is doing the reporting. Similar to the co-referential functions of the pronouns used, are the temporal references of the present tense and present continuous tense of the verb forms in "I understand" and "you're saying" (lines 3 & 4). The shift in prosody used within the reported utterance (underlined segment in lines 3 & 4) functions to dramatize the speech event and emphasize Janet's comprehension and Bella's intelligible English-speaking skills. The main problem of communication between Bella and Janet, however, lies in Bella's apparent insecurity of speaking English (lines 1 & 5), which has led to delays of problem solving among domestics and clients. As a result, the utterance analyzed using DRS functions to dramatize communication between one particular domestic and language broker as often unsuccessful due to Bella's linguistic insecurity.

In the final extract, Mrs. Malloy, an English-speaking client, discusses how she communicates with Patricia, her Portuguese-speaking domestic, by using both verbal communication as well as gestures. In exemplifying a typical situation, Mrs. Malloy uses DRS to offer evidence for the reported speech event as it actually happened:

Extract 3) A client's perspective

```
1.  M:   i'd say erm (.) "patricia this week we're
2.        not going to clean the windows" and i'll
3.        point to the window and I'll say (.) "i have
4.        had them a:ll cleaned they're fine (.) you
5.        don't need to touch them (.) so they're a:ll
6.        fine" like @@@ and we do hand signals
7.        so and i say (.) "do you under- ok?" and
8.        she's like (.) "ok" and i don't know if that
9.        means "yes (.) I understand you" or "ok,
10.       (.) you've said something" you know? i
11.       (1.0) that (.) there is no like (.) there is no
12.       real verbal communication back
```

In this extract Mrs. Malloy begins with the reported verb "say" and then continues her account of the conversation by addressing Patricia directly (line 1), which functions to convey that these were the actual words uttered during the initiation of the conversation. Second, she uses the inclusive personal pronoun "we", the present continuous verb tense "going", as well as the spatial deictic marker this week (line 1), all of which function to signal Mrs. Malloy's point of view at that particular time. Her next DRS utterance (line 3) includes features such as temporal reference in the past perfect tense "I have had them all cleaned" as well as the present tense and personal pronoun "you don't need to touch them" (lines 4 & 5), which function to indicate the time of speaking during the actual conversation with her interlocutor. Her claim of pointing to the window and their joint use of hand signals (line 6) suggest that Mrs. Malloy and Patricia use both linguistic and non-linguistic strategies in order for communication to be achieved which prove to work for both Mrs. Malloy and Patricia. In order to confirm Patricia's understanding of Mrs. Malloy's instructions, however, she inquires directly. This is seen in (line 7) when Mrs. Malloy uses the reported verb "I say", which precedes the direct question "do you under-, ok?". What is interesting about this question is Mrs. Malloy's initial report about comprehension. She begins her utterance by asking if Patricia understands her instructions, but then resorts to simplifying her request by asking "ok?", which is marked by a shift in prosody and rising intonation. In this context, Mrs. Malloy employs basic English skills in order for the communication between her and Patricia to be regarded as successful. Mrs. Malloy further states that Patricia confirms her request by her response when Mrs. Malloy makes use of the quotative in "she's like "ok" (line 8). She then employs DRS to report a hypothetical account of her thought process and how the exchange developed (Sams 2007; 2010). This is done when Mrs. Malloy confesses to not knowing how she should socio-pragmatically understand Patricia's use of "ok" by giving two possible options of its potential meaning. The first meaning could be a preferred response in positively responding back to Mrs. Malloy's question while the second option "ok, you said something" (line 10), acknowledges Mrs. Malloy's utterance. Despite the fact that Mrs. Malloy employs DRS to reconstruct this conversation and hypothetical thought process, which has the effect of dramatizing her account, she states that "no real communication" has taken place because the socio-pragmatic meaning of Patricia's "ok" in response to Mrs. Malloy's question remains ambiguous. Nevertheless, the reconstructed conversation reveals that the communicative event of giving directions between Mrs. Malloy and Patricia using gestures and basic English is ultimately deemed successful.

3.   Conclusion

According to Coulmas (1986: 2) the use of DRS "evokes the original speech situation and conveys, or claims to convey, the exact words of the original speaker" in the interaction. The effect of employing DRS within storytelling or narratives also functions to dramatize the

unfolding events of interlocutors' interactions at the time and place of the actual speech event. In my analysis, I showed how the use of DRS among domestics, language brokers and clients was employed as a prominent linguistic strategy, which functioned to convey authenticity of the actual speech event between domestic and language broker or domestic and client. This was shown in all three extracts analyzed above. The second function DRS had within the analysis was to represent the development of the conversation between interlocutors' as well as their particular stances concerning their joint communication of the speech event. The final function that DRS had within this study was to depict the story's climax and dramatize the effect of achieving both successful or unsuccessful communication within a specific language contact situation within a domestic labor and workplace context. In presenting the analysis, I focused on typical DRS features, which included personal pronouns, spatial and temporal markers, shifts in prosody as well as speech verbs. In her work on workplace discourse, Holmes states that "few researchers have ventured into blue collar worksites; they tend to be noisy and dirty and often rather uncomfortable places for academics undertaking research" but asserts that "this is undoubtedly another direction in which it is important to expand workplace discourse research (forthcoming: 15). The aim of this study was to "venture" into an area of research that is not always easily accessible to researchers and as a result, a dearth of linguistic studies exists within the context of domestic labor. The intention of my study was to expand the direction of workplace studies in general and thus shed light on how meaning is negotiated between Portuguese-speaking domestics and their Anglophone clients. Research on workplace studies outside of white-collar contexts is indeed challenging yet, I hope to have shown that communicative strategies within a domestic labor context yields fruitful insight into how meaning is achieved and reported on between interlocutors of different language backgrounds.

4. References

Anderson, B. (2000). *Doing the Dirty Work? The Global Politics of Domestic Labour*. London: Zed Books.

Bakhtin, M.M. (1981). *The Dialogic Imagination*. University of Texas Press.

Biber, D., Leech, G. N. and Conrad, S. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Buttny, R. (1997). Reported speech in talking race on campus. *Human Communication Research* 23, pp. 477--06.

Carter, R., McCarthy, M. (Eds.). (2006). *Cambridge Grammar of English*. Cambridge and New York: Cambridge University Press.

Chang, G. (2000). *Disposable Domestics: Immigrant Women Workers in the Global Economy*. Cambridge, MA: South End Press.

Clift, R. (2000). Stance-taking in reported speech. In *Essex Research Papers in Linguistics*, vol. 32. University of Essex.

Coulmas, F. (1986). Reported speech: Some general issues. In F Coulmas (Ed.), *Direct and Indirect Speech*. Berlin: Mounton de Gruyter, pp. 1--28.

Clark, H.H., Gerrig, R.J. (1990). Quotations as Demonstrations. In *Language* 66, pp. 764--805.

Del Torto, L.M. (2008). Once a broker, always a broker: Non-professional interpreting as identity accomplishment in multigenerational Italian-English bilingual family interaction. In *Multilingua* 27, pp. 77--97.

Drew, P. (1998). Complainable matters: the use of idiomatic expressions in marking complaints. In *Social Problems*, 35, pp. 398--417.

Ehrenreich, B., Hochschild, A. (2004). *Global Woman: Nannies, Maids, and Sex Workers in the New Economy*. New York: Henry Holt and Co.

Goffman, E. (1981). *Footing: Forms of Talk*. Philadelphia: University of Pennsylvania Press, pp. 124--59.

Golato, A. (2000). An innovative German quotative for reporting embodied actions: Und ich so / und er so 'and I'm like / and he's like'. In *Journal of Pragmatics* 32, pp. 29--54.

Holmes, J. (forthcoming). Discourse in the workplace. In K. Hyland, B. Paltridge (Eds.), *Continuum Companion to Discourse Analysis*. London: Continuum.

Holt, E. (1996). Reporting on Talk: The Use of Direct Reported Speech in Conversation. In *Research on Language and Social Interaction* 29(3), pp. 219--245.

Holt, E. (2000). Reporting and Reacting: Concurrent Responses to Reported Speech. In *Research on Language and Social Interaction* 33(4), pp. 425-454.

Holt, E. (2009). Reported speech. In S. D'hondt, J.O. Ostman and J. Verschueren (Eds.), *Pragmatics of Interaction. Handbook of Pragmatics Highlights*, vol.4. John Benjamins Publishing Company, pp. 190--205.

Lan, P.C. (2006). *Global Cinderellas: Migrant Domestics and Newly Rich Employers in Taiwan*. Durham and London: Duke University Press.

Li, C. (1986). Direct and indirect speech: A functional study. In F. Coulmas (Ed.), *Direct and indirect speech*. Berlin: Mouton de Gruyter, pp. 29--45.

Mayes, P. (1990). Quotation in spoken English. Studies in Language 14, pp. 325--363.

Milroy, L. (1987 [1980]). *Language and Social Networks*, 2nd ed. Oxford: Blackwell Publishing.

Milroy, L. and Milroy, J. (1992). Social network and social class: Toward an integrated sociolinguistic model. In *Language in Society 21*. pp. 1--26.

Myers, G. (1999). Unspoken speech: Hypothetical reported discourse and the rhetoric of everyday talk. In *Text* 19(4), pp. 571--590.

Niemelä, M. (2005). Voiced Direct Reported Speech in Conversational Storytelling: Sequential Patterns of Stance Taking. In *SKY Journal of Linguistics* 18, pp. 197--221.

Parreñas, R.S. (2001). *Servants of Globalization. Women Migration, and Domestic Work*. Stanford, CA: Stanford

University Press.

Parreñas, R.S. (2008). *The Force of Domesticity: Filipina Migrants and Globalization*. New York: New York Press.

Rollins, J. (1985). *Between women: Domestics and their employers*. Philadelphia: Temple University Press.

Romero, M. (2002).   *Maid in the U.S.A*. London: Routledge.

Sams, J. (2007). Quoting the Unspoken: An analysis of quotations in spoken discourse. In *Colorado Research in Linguistics*, 20, pp.1--16.

Sams, J. (2010). Quoting the unspoken: An analysis of quotations in spoken discourse. In *Journal of Pragmatics*, 42, pp. 3147--3160.

Stoessel, S. (2002). Investigating the role of social networks in language maintenance and shift. In *International Journal of the Sociology of Language* 153, pp. 93--131.

Tannen, D. (1986). Introducing constructed dialogue in Greek and American conversational and literary narrative. In F. Coulmas (Ed.), *Direct and indirect speech*. Berlin: Mouton de Gruyter, pp. 311--332.

Tannen, D. (1989). *Talking voices*. Cambridge, England: Cambridge University Press.

Tse, L. (1996). Language brokering in linguistic minority communities: The case of Chinese-and Vietnamese-American students. In *The Bilingual Research Journal* 20 (3-4), pp. 485--498.

Volosinov, V.N. (1971). Reported Speech. In Matejka, L. and K. Pomorska (Eds.), *Readings in Russian Poetics: Formalist and Structuralist Views*: MIT Press.

Wei, L. (1993). Mother Tongue Maintenance in a Chinese Community School in Newcastle Upon Tyne: Developing a Social Network Perspective. In E. Murphy (Ed.), *ESL: A Handbook for Teachers and Administrators in International Schools*. Clevedon: Multilingual Matters Ltd, pp. 199--215.

Weisskirch, R. S. (2005). The relationship of language brokering to ethnic identity for Latino early adolescents. *Hispanic Journal of Behavioral Sciences* 24(3), pp. 286--299.

Weisskirch, R. S. and Alatorre A. S. (2002). Language brokering and the acculturation of Latino children. *Hispanic Journal of Behavioral Sciences* 24(3), pp. 369--378.

5.   Appendix

Transcription Conventions:

@@	= signals laughter
wo::rd	= perceptible lengthening
(.)	= pause shorter than one second
(1.0)	= pause lengths in seconds
?	= rising intonation, often signals questions
=	= latched talk
___	= underlined text is marked for changes in prosody

Mapping Paulistano Portuguese: the SP2010 project

Ronald Beline MENDES, Livia OUSHIRO

University of Sao Paulo

Av. Prof. Luciano Gualberto, 403 - Sala 16, Cidade Universitária, 05508-010 - São Paulo - SP

rbeline@usp.br, livia.oushiro@usp.br

Abstract

This paper reports on the objectives, methods, and results from the project SP-2010 (Mendes, 2011), currently under the execution by the *Grupo de Estudos e Pesquisa em Sociolinguística* (GESOL-USP). Its main objectives are (i) to build a contemporary and representative sample of São Paulo Portuguese; (ii) to develop studies of sociolinguistic variation in the city, an understudied speech community (Mendes, 2009; Rodrigues, 2009); and (iii) to make the corpus of recordings and transcripts available online for a wider group of researchers. The first phase of the project aims at collecting 60 sociolinguistic interviews with speakers stratified by sex/gender, age, and level of education by 2013. In view of the highly heterogeneous sociodemographic make-up of the city of São Paulo, fieldworkers also observe distinctions in informants' social class, family generation in the city, and area of residence. Interview recordings follow Variationist Sociolinguistics premises (Labov, 1984, 2006; Tagliamonte, 2006) and data transcription norms are designed as to facilitate automatic data handling in softwares such as R.

Keywords: spoken corpus; Paulistano Portuguese; variationist sociolinguistics; data collection; transcripts.

1. Introduction

Although São Paulo Portuguese has already been documented and analyzed through broad and significant research projects such as Projeto NURC-SP (Castilho & Preti, 1986, 1987; Preti & Urbano, 1988, 1990) and Projeto Para a História do Português Paulista (Castilho 2007), most works within these projects aim at analyzing "Brazilian Portuguese," either in contrast with European Portuguese (e.g., studies on parametric variation), or in relation to its internal processes of change (e.g., studies on grammaticalization).

Among the very few works about Paulistano Portuguese in its social context, Rodrigues (1987) analyzed variable subject-verb agreement (e.g., *nós vamos* vs. *nós vai* 'we go') in the speech of 40 (semi-)illiterate speakers in two favelas, and Coelho (2006) analyzed the variable use of 1PP pronouns (*nós* vs. *a gente* 'we') in the speech of 24 speakers living in a working class community. Yet, to date, little is known about the linguistic production and perception of many other (supposedly) typical Paulistano variants (e.g., the realization of coda /r/ as a tap in words such as *porta* 'door,' the diphthongization of nasal /e/ in words such as *fazenda* 'farm') and other variants in the city, as well as their social distribution and evaluation in the speech community at large.

This may be due to the difficulties of building a representative speech corpus of a heterogeneous and multicultural city with more than 11 million people, highly diverse in terms of their geographical origin, socioeconomic class, and cultural background. According to a recent survey by the Instituto de Pesquisa Econômica Aplicada (IPEA, 2011), 46% of the adult working population (between 30 and 60 years old) living the the São Paulo Metropolitan Area were not born in the state of São Paulo (see Figure 1). Although the survey does not refer exclusively to the city itself, it gives an idea of the intense presence of non-native inhabitants in this region. One can consider that the number of non-Paulistanos living in the city may be even greater, since the 54% of Paulistas include all people born in the state of São Paulo and not only the capital city.
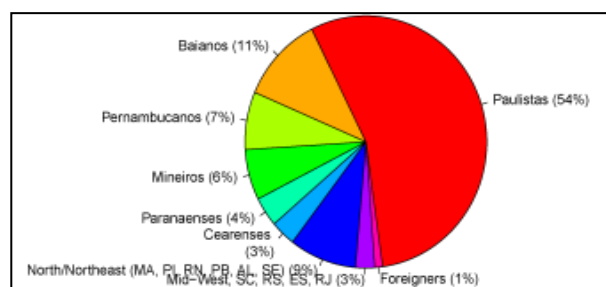

Figure 1: Adult population living in the São Paulo Metropolitan Area. Source: IPEA 2011

This fact raises a number of questions: which social parameters are most relevant for linguistic differentiation and stratification and how to reach speakers of varied social networks? How to gather detailed ethnographic information from each informant (Poplack, 1989), acknowledging a persistent point made by the "third-wave" of sociolinguistic studies (Eckert, 2005) on the importance of observing individuals' social practices? Which methodologies are best for handling a large amount of spoken linguistic data?

In this paper, we report on the objectives, methods, and results from the Project SP-2010 (Mendes, 2011), currently under execution by the *Grupo de Estudos e Pesquisa em Sociolinguística da USP* (GESOL-USP),[1] which aims at: (i) building a contemporary and representative sample of Paulistano Portuguese; (ii) fostering the development of sociolinguistic studies in the city; and (iii) making the corpus of recordings and transcripts available online for a wider group of researchers.

---

[1] http://linguistica.fflch.usp.br/gesol.

2.   Methods and Results

In 2009-2010, GESOL-USP collected 82 sociolinguistic interviews with residents of the city of São Paulo, native or not to the city, of both sexes and different sexual orientations, from 15 to 89 years of age, with different levels of education, of varied socioeconomic statuses, living in 59 different neighborhoods in the city. In view of São Paulo's great sociodemographic complexity, these exploratory recordings had the objective of defining the most relevant social variables for the sociolinguistic description of Paulistano Portuguese; elaborating an interview schedule; developing best practices in approaching possible informants; identifying possible technical and methodological problems that may occur during the recordings (e.g. avoiding noise, making the informant comfortable) and coming up with solutions for them; and elaborating criteria for transcribing the interviews.

From this experience, we observed that certain sociolinguistic profiles are hard to locate – for instance, younger native Paulistanos who have not concluded at least high school, especially women living in more central areas, or people over 70 who were actually born in the city, especially in more suburban areas. In addition, in spite of our initial aim of locating prototypical speakers from certain neighborhoods (e.g. Mooca, Bexiga, Pinheiros), geographic and socioeconomic mobility seems to be characteristic of the city and its inhabitants, many of whom prefer not to settle in a single place for life. Further, a technical but not to be ignored challenge is the presence of noise (traffic, constructions, people), even in residential areas of the city. The methods designed for this project try to address some of these issues.

In the present phase, to be concluded by 2013, the social parameters for constituting the sample are sex/gender (men and women), three age groups (20-34 y.o.; 35-59 y.o.; 60+ y.o.), and two levels of education (up to high school; college). As our focus is on the *social meaning of variation* (Chambers, 1995), these variables have been chosen primarily because of their potential to shed light on the relationship between variable linguistic uses and social identities, as well as to enable cross-comparisons with other linguistic corpora of Brazilian Portuguese – e.g. VARSUL (Bisol *et al.*, s/d), VALPB (Hora, 2004), PEUL (Paiva & Scherre, 1999), ALIP (Gonçalves, 2003).

Sex/Gender and Age have been broadly analyzed in sociolinguistic studies and have been shown to be correlated with variables whose variants are differently evaluated in terms of prestige: a number of works have have observed that the prestigious forms in the community tend to be employed by women (Chambers, 1995; Labov, 2001; Cheshire, 2004), and that unprestigious forms tend to be avoided by speakers in the intermediary age group, who mostly suffer pressures of the linguistic market (Bourdieu, 1991; Labov, 2001). Correlations with Age can also point to possible changes in progress in the linguistic system through *apparent time analyses* (Labov, 2001). The three age groups are mostly

based on their relative position in the job market, but also take into account each group's general lifestyles in a big city. The younger speakers, those between 20 and 34 years old, comprise young adults who tend to be relatively less stable than people in the other two age groups; in São Paulo, it is not rare to find people up to 34 years old who are not married, who do not own their own place, who go to college or who lead life more similarly to people in their early 20s. The group aged between 35 and 59 years old, in turn, is intended to comprise people more fully inserted in the job market and relatively more stable. Finally, the group over 60 years old refers to people in or close to retirement.

Level of education is also directly associated with stigmatization and prestige. The general hypothesis is that more educated speakers will tend to avoid unprestigious forms in the community, or otherwise that the forms they employ will be considered more "correct." In Brazilian sociolinguistic studies, the division between "educated" and "uneducated" speakers is normally taken as an index of socioeconomic status (Rodrigues, 2009: 151). This situation seems to be changing in São Paulo as well as in many other urban centers through extensive public policies of improved access to primary, secondary, and higher education (for instance, *Progressão Continuada* in the state of São Paulo and *ProUni* in a national scope); the division between only two levels of education is a consequence of these changes. However, general increase in average levels of education is not always followed by a direct ascension in individual socioeconomic status, which means that the equation between level of education and social class should not be overestimated. We suggest that level of education should be treated as constitutive of speakers' social class, but not as its substitute.

The combination of these social parameters yields 12 sociolinguistic profiles (e.g. men between 20-34 y.o. without a college degree), each of which is to be filled by 5 speakers, in a total of 60 sociolinguistic interviews. Each of these 5 speakers per cell should reside in a different zone of the city (North, South, East, West, Central), and each cell should contain at least one speaker of three city areas (Downtown, Extended Central Area, Suburbs), as a way to ensure a broad coverage of the city. The speakers' place of residence is defined as the place where he/she has lived for the most part in the past 10 years.

In a second stage, we will focus on social class, a social factor generally overlooked in Brazilian sociolinguistic studies due to lack of reliable criteria for categorizing speakers in different socioeconomic groups (Rodrigues, 2009; Mendes, 2011). In the city of São Paulo, speakers' socioeconomic status possibly should take into account, in addition to their income and level of education, their type of residence, occupation, and access to cultural goods. The corpus will also be stratified according to speakers' generation in the city, in order to examine the contribution of different groups of migrants and immigrants in the community, and speakers' area of residence, which is also an index of socioeconomic status.

During this first phase of the project, information on these variables is collected through the sociolinguistic interview and post-recording questionnaires, which will enable preliminary analyses of their role in the sociolinguistic stratification in São Paulo.

Speakers to be recorded have been contacted through the "friend of a friend" method (Milroy, 2004). Our experience has shown that speakers in the city are very resistant to talking to a "stranger" (the researcher); however, when introduced by a common acquaintance, speakers tend to be much more receptive and solicitous, a fact that also has consequences for naturalness of speech. After a speaker has been recorded, the researcher asks her/him to suggest another speaker. As a means to ensure that informants do not belong to the same or few social networks, the new suggested speaker can only be recorded if he or she is not acquainted with the person who indicated the current informant. For instance, in the example in Figure 2, B has indicated two new speakers, C and D, but only the latter can be selected as a new informant.
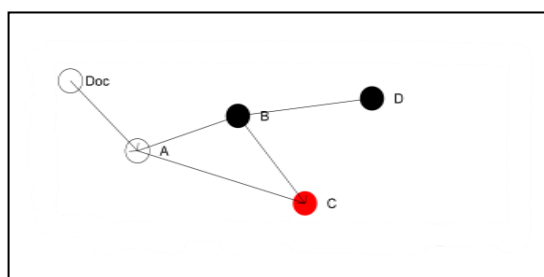

Figure 2: Selection of informants

The interview schedule has the twofold objective of obtaining samples of spontaneous speech by Paulistanos of varied sociolinguistic profiles and more information on these speakers' living conditions, sociolinguistic evaluations and perceptions (Labov, 2006). It is divided into two parts. The first one is more personal and covers topics such as the speakers' neighborhood, childhood, parents and family, education, current occupation, social network, and leisure activities. It aims at obtaining narratives in the past (e.g. "What was your childhood like in neighborhood X?"), in the present (e.g. "In your leisure time, what do you and your family like to do?") and in the future (e.g. "What would you do if you won the lottery?"), as well as opinion accounts (e.g. "What do you think of the new law for gay marriage?"). The second part contains more specific questions about the speakers' relation to the city and their perceptions on Paulistano identities (e.g. "When you were in (another city), did people recognize you as a Paulistano? If so, how?"). In the last part of the interview, speakers are asked to read a word list, a news report, and a 'statement' (a text with strong marks of oral language). Although the interview schedule is divided into two parts, it enables easy transition between topics and has yielded natural sounding conversations.

After the interview is recorded, the fieldworker fills out a form with detailed speaker's sociolinguistic information (date of birth, occupation, family's place origin and first generation that migrated to São Paulo, schools, place(s) of residence etc.), and makes note of any relevant additional information in the fieldwork journal. The informant is also asked to fill out a socioeconomic form, if he/she feels comfortable to do so, containing seven multiple-choice questions about their monthly income and living conditions. Our experience has shown that the multiple-choice form greatly improves the chance of obtaining these data (instead of having the informant orally answer these questions directly to the fieldworker).

Each sociolinguistic interview is about 60-70 minute long and has been stored in .wav (stereo, 44,100 Hz) format. The recordings have been made with TASCAM DR-100 recorders and two Sennheiser HMD26 microphones (one for the fieldworker and one for the informant). Although it could be argued that the presence of these technical paraphernalia possibly enhances the Observer's Paradox (Labov, 2006), we find that speakers' occasional uneasiness tends to decrease considerably after some 15 minutes of recording and, more importantly, that the improved audio quality is worth the trouble, especially in a city as noisy as São Paulo.

All interviews are then evaluated by four members of the research group not involved in the field recordings, according to the speakers' fitness to the sociolinguistic profile, audio quality, naturalness of conversation, and conformity to the interview schedule. The 82 previously collected interviews during the pilot experience have also been evaluated according to the same parameters, and some of them may be included in the final corpus to be made available online, in addition to the 60 recordings of the present data collection phase, as long as they meet the high-quality requirements.

The criteria for transcribing the recordings follow a simplified semiortographic approach in order to make the material more easily available in a written medium. The following criteria aim at facilitating the manipulation of text files in softwares such as R (Gries, 2009; Hornik, 2011) to automatically identify and extract tokens of a variable into a spreadsheet program (Oushiro, 2012).

Transcripts do not contain any special formatting such as boldface, italics, tab stops, columns, and are saved in plain text (.txt) with UTF-8 encoding. Orthographic rules of Brazilian Portuguese are followed in every case, even if speakers produce variants that differ from the written standard. The idea here is that a transcriber is unable to pay attention to all variable phenomena simultaneously – e.g. monophthongization of /ow, ej/, diphthongization of nasal /e/, postvocalic /r/ deletion, nasal assimilation of /ndo/, vowel raising of unstressed /e,o/, to name a few. In addition to creating unintelligible texts, this would probably cause transcripts to be unstandardized; further, the fact that the recordings will be made available lessens the need for a highly detailed transcript. On the other hand, grammatical variables should not be "corrected" by the transcriber (e.g. lack of

nominal agreement). Punctuation is limited to ellipses (to signal pauses), and question and exclamation marks (to indicate intonation of certain phrases). Capital letters are only employed in proper names (e.g. cities, institutions), abbreviations (e.g. USP, and identifying speakers (e.g. S1, D1).

GESOL-USP has also been developing parallel data collection projects, in addition to gathering a sample from the community at large. These parallel projects and studies are centered on specific groups of speakers and/or social variables within the city: residents of the upper class neighborhood Itaim Bibi (Ciancio, 2012); social class (Faria, 2012); gay men and gender (Soriano, 2012); different groups of migrants – Paraibanos (Mendes, forth) and Alagoanos (Silva, 2012). These studies aim at describing and contrasting general sociolinguistic patterns of the community and their uses within certain social groups residing in the city.

Based on the corpus collected so far, the research group has been developing studies of sociolinguistic variation in Paulistano Portuguese: the variable realization of coda (-r) as a tap or a retroflex, in words such as *porta* 'door' and *mulher* 'woman' (Mendes, 2009, 2010; Mendes & Oushiro forth); variable nasal (e) as a monophthong or a diphthong, in words such as *fazenda* 'farm' (Mendes, 2010; Oushiro, 2011); verbal negative structures (e.g. *Não vou* vs. *Não vou não* 'I won't go') (Rocha, 2012); nominal and verbal agreement (Silva, 2012; Oushiro, 2011).

3.   Conclusion

The SP-2010 Project has been collecting a contemporary corpus of Paulistano Portuguese and fostering the development of sociolinguistic studies focusing on the correlations between variable linguistic uses and social identities. By 2013, more than 60 sociolinguistic interviews (audio and transcriptions) will be made available online to the linguistic community. Parallel to this data collection project, a number of studies have also been analyzing specific social networks and communities of practice in the city, in contrast with larger community variational patterns, as to provide a broader and more detailed description of linguistic uses in São Paulo.

4.   Acknowledgements

5.   References

Bisol, L., Menon, O.P.S. and Tasca, M. (s/d). VARSUL, um banco de dados. Available at: <http://www.varsul.org.br/?modulo=secao&id=9>.

Bourdieu, P. (1991) *Language and symbolic power*. Cambridge: Polity Press.

Castilho, A.T., Preti, D. (Eds.). (1986). *A linguagem falada culta na cidade de São Paulo: materiais para seu estudo,* vol. I – Elocuções Formais. São Paulo: T.A. Queiroz.

Castilho, A.T., Preti, D.   (Eds.). (1987). *A linguagem falada culta na cidade de São Paulo: materiais para seu estudo*, vol. II – Diálogos entre dois informantes. São Paulo: T.A. Queiroz/FAPESP.

Castilho, A.T. (2007). Projeto para a História do Português Paulista. Projeto Temático de Equipe (Proc. FAPESP n. 06/55944-0).

Chambers, J.K. (1995). *Sociolinguistic theory: linguistic variation and its social significance*. Oxford: Blackwell.

Cheshire, J. (2004). Sex and gender in variationist research. In J.K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change*. Oxford: Blackwell.

Ciancio, R. (2012). Estudo sociolinguístico da fala paulistana por falantes do Itaim Bibi. Research Project.

Coelho, R.F. (2006). É nóis na fita! Duas variáveis linguísticas numa vizinhança da periferia paulistana. O pronome de primeira pessoa do plural e a marcação de plural no verbo. Master's dissertation.

Eckert, P. (2005). Three waves of variation study: the emergence of meaning in the study of variation. Manuscript, s/d. Available at: <www.stanford.edu/~eckert/PDF/ThreeWavesofVariation.pdf>.

Faria, C.B. de (2012). Para a inclusão de "classe social" nos estudos sociolinguísticos em São Paulo. Research Project.

Gonçalves, S.C.L. (2003). O português falado na região de São José do Rio Preto: constituição de um banco de dados anotado para o seu estudo. Research project. Available at: <www.iboruna.ibilce.unesp.br/historico/Projeto>.

Gries, S.Th. (2009). *Quantitative Corpus Linguistics with R*. New York: Routledge.

Hora, D. da (Ed.). (2004). *Estudos Sociolinguísticos*: perfil de uma comunidade. Santa Maria: Palotti.

Hornik, K. (2011). R FAQ. Available at : <http://cran.r-project.org/doc/FAQ/R-FAQ.html>.

IPEA (2011). Comunicados do IPEA no. 115 – Perfil dos migrantes em São Paulo, 2011. Available at: <www.ipea.gov.br/portal/images/stories/PDFs/comunicado/111006_comunicadoipea115.pdf>.

Labov, W. (1984). Field methods of the Project on Linguistic Change and Variation. In J. Baugh, J. Sherzer (Eds.), *Language in Use: Readings in Sociolinguistics*. Englewood Cliffs: Prentice Hall, pp. 28--54.

Labov, W. (2001). *Principles of linguistic change*: social factors. Oxford & Cambrige: Blackwell.

Labov, W. (2006). *The Social Stratification of English in New York City*. Cambridge: Cambridge University Press.

Mendes, R.B. (2009). Who sounds /r/-ful? The pronunciation of coda /-r/ in the city of São Paulo. *Paper presented at NWAV38*. University of Ottawa.

Mendes, R.B. (2010). Sounding Paulistano: variation and correlation in São Paulo. *Paper presented at New Ways of Analyzing Variation - NWAV39*. University of Texas at San Antonio.

Mendes, R.B. (2011). SP-2010 – Construção de uma amostra da fala paulistana. Projeto de Pesquisa.

Mendes, R.B. (forth). A pronúncia retroflexa do /-r/ na fala paulistana. In D. Hora, E.V. Negrão (Eds.), *Estudos da Linguagem. Casamento entre temas e perspectiva*. João Pessoa: Ideia/Editora Universitária, pp. 282--299.

Mendes, R.B., Oushiro, L. (forth). Percepções sociolinguísticas sobre as variantes tepe e retroflexa na cidade de São Paulo. In D. Hora, E.V. Negrão (Eds.), *Estudos da Linguagem. Casamento entre temas e perspectiva*. João Pessoa: Ideia/Editora Universitária, pp. 262--281.

Milroy, L. (2004). Social networks. In J.K. Chambers, P. Trudgill and N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change*. Oxford: Blackwell.

Oushiro, L. (2011). Identidade na pluralidade: produção e percepção linguística na cidade de São Paulo. Research project.

Oushiro, L. (2012). Analyzing (-r) with R. Paper presented at the 2012 GSCP International Conference. Available at: <www.letras.ufmg.br/gscp2012-eng/data1/arquivos/84.pdf>.

Paiva, M.C., Scherre, M.M.P. (1999). Retrospectiva sociolinguística: contribuições do PEUL. In *DELTA* (online), vol. 15, n.spe., pp. 201--232.

Poplack, S. (1989). The care and handling of a mega-corpus: the Ottawa-Hull French Project. In R. Fasold, D. Schiffrin (Eds.), *Language Change and Variation*. Amsterdam: Benjamins, pp. 411--451.

Preti, D., Urbano, H. (Eds.). (1988). *A linguagem falada culta na cidade de São Paulo*: *materiais para seu estudo*, vol. III – Entrevistas. São Paulo: T.A. Queiroz/FAPESP.

Preti, D., Urbano, H. (Eds.). (1990). *A linguagem falada culta na cidade de São Paulo: materiais para seu estudo*, vol. IV – Estudos. São Paulo: T.A. Queiroz/FAPESP.

Rocha, R.S. (2011). Negação pós-verbal no português paulistano: definição do envelope de variação. In *Seminário do GEL, 59, Programação*, Bauru: GEL, 2011. Available at: <http://gel.org.br/detalheResumo.php?trabalho=7564>.

Rodrigues, A.C.S. (1987). A concordância verbal no português popular em São Paulo. PhD Thesis. FFLCH-USP.

Rodrigues, A.C.S. (2009). Fotografia sociolinguística do português do Brasil: o português popular em São Paulo. In A.T. Castilho (Ed.), *História do Português Paulista*. Campinas: Instituto de Estudos da Linguagem/UNICAMP.

Silva, F.G. (2012). Concordância nominal: um contraste dentro da cidade de São Paulo. Research project.

Soriano, L. (2012). Estudo sociolinguístico de gays paulistanos em diferentes situações de fala. Research project.

Tagliamonte, S. (2006) *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Documentação da Língua Indígena Brasileira Yaathe (Fulni-ô)

Januacele da COSTA, Miguel OLIVEIRA Jr., Fábia SILVA

Universidade Federal de Alagoas

Maceió – AL, Brasil

januafc@bol.com.br, miguel@fale.ufal.br, fabia_fulnio@hotmail.com

Resumo

Este artigo tem por objetivo descrever o Projeto de Documentação Linguística da Língua Indígena Brasileira Yaathe, falada pelo povo Fulni-ô. O povo Fulni-ô, que vive no município de Águas Belas, interior de Pernambuco, é o único povo indígena do nordeste brasileiro que preservou sua língua depois do processo colonizador. A despeito do uso sistemático que os Fulni-ô fazem de sua língua, sobretudo em situações privadas, ela tem sido considerada por órgãos internacionais uma língua em extremo risco de extinção. Justifica-se, desta maneira, a urgência de um projeto de documentação como o que se descreve a seguir. O artigo apresenta um breve histórico do povo Fulni-ô, situando-o socio-historicamente, descreve a situação atual de sua língua, lista os objetivos do projeto a ser desenvolvido, justificando a sua relevância, e detalha metodologia específica a ser adotada na coleta e no tratamento de dados, metodologia que segue padrões hoje adotado por bancos de dados de línguas em perigo de extinção.

Palavras-chave: Yaathe; Fulni-ô; documentação linguística.

1. Introdução

A língua Yaathe, pertencente ao tronco Macro-jê (Rodrigues, 1986), é ainda falada pela maior parte da população Fulni-ô. Em um estudo sociolinguístico para definir o perfil linguístico da comunidade (Costa, 1993), ficou demonstrado que 91,5% dos índios são falantes ativos ou passivos da língua original do grupo. A designação *Yaathê* significa literalmente "nossa fala", de [ya] "possessivo, 1ª pessoa do plural" e ['jat^{h}e] "fala".

Os índios Fulni-ô vivem no município de Águas Belas, no oeste-sudoeste de Pernambuco, a cerca de 300 quilômetros de Recife, a capital do Estado de Pernambuco. A reserva indígena Fulni-ô está localizada a pouca distância da margem esquerda do Rio Ipanema, um dos afluentes, também da margem esquerda, do Rio São Francisco.

Um dos aspectos mais interessantes da situação dos índios Fulni-ô é a sobrevivência da língua, uma vez que todas as outras línguas indígenas faladas nessa parte do país já desapareceram. Embora se possa afirmar a vitalidade da língua neste momento, divergências internas e outros problemas, como o empobrecimento cada vez mais crescente da região e o descaso das autoridades regionais, poderiam vir a mudar esse quadro em poucos anos. As pessoas mais jovens da comunidade foram encorajadas, por um período de cerca de 40 anos, a não falar sua língua ou viver de acordo com os costumes de seu povo. Esse direcionamento e as atitudes dele decorrentes, vêm, todavia, mudando nas últimas décadas. Atualmente, o grande desejo dos Fulni-ô é a manutenção da sua língua e da sua cultura.

Este artigo descreve um projeto de pesquisa ora em curso, financiado pelo CNPq (Edital MCT/CNPq N. 014/2010 – Universal, Processo Nº 475763/2010-6), cujo objetivo é a documentação da língua *Yaathê*, em formato digitalizado, para disponibilização à comunidade científica. Objetivos mais específicos, relacionados aos interesses do grupo de pesquisa que se propõe desenvolvê-lo são, além da formação de um banco de dados, a elaboração de uma gramática descritiva, passível de ser utilizada no ensino-aprendizagem, ou, no mínimo, fornecer subsídios para a elaboração de materiais didáticos e a produção de artigos sobre aspectos da língua em todos os níveis de análise, bem como de dissertações e teses visando à formação de novos pesquisadores para o estudo de línguas indígenas.

Na região Nordeste, os grupos indígenas existentes quando do descobrimento foram rapidamente atropelados pelo processo colonizador que, partindo do ciclo da cana-de-açúcar, no litoral, empurrou as nações indígenas que não foram dizimadas para o sertão interior. Mais tarde, o ciclo do gado cumpriria a sua parte na extinção dos nativos, ora dizimando populações inteiras, sobretudo as que ocupavam as margens dos rios, como o São Francisco e seus afluentes, principalmente, para ocupar as terras com o criatório do gado, ora aniquilando a cultura por desfazer grupos inteiros, espalhando-os para longe de suas aldeias, obrigando-os, assim, a viverem isolados e fazendo parte de uma população sertaneja anômica.

Parte das populações indígenas que sobreviveram ao massacre, tanto étnico quanto físico, graças à ação dos missionários franciscanos e capuchinhos, que os agruparam em missões, perderam elementos importantes do seu equipamento cultural, o que os diferenciava das populações não-índias vizinhas e entre si. Entre as perdas de marcas de identidade, a mais marcante foi a perda da língua nativa. Atualmente, das cerca de 23 nações que vivem no Nordeste, das quais a maior parte teve sua identidade étnica reconhecida e suas terras legitimadas apenas na segunda metade do século passado, só os Fulni-ô, no sul do Estado de Pernambuco, preservaram a sua língua nativa, o Yaathe. Sendo a língua um fator determinante da identidade étnica, só por esse motivo a documentação com o objetivo de preservação já se faria importante. Entretanto, além disso, uma documentação fundamentada da língua, visando a diferentes objetivos e diferentes análises, é, seguramente, de grande importância para a ciência linguística.

2. Justificativa

Recentemente, a UNESCO divulgou um relatório sobre línguas em risco de extinção e, de acordo com os critérios

utilizados pela pesquisa, o Yaathe é uma língua que se encontra em "extremo perigo de extinção".[1]

Apesar de os números indicarem uma alta porcentagem de falantes de Yaathe entre os Fulni-ô (cerca de 3.000 pessoas, o que corresponde a mais de 90% da população total), o uso da língua está restrito a situações bastante específicas. Raramente os Fulni-ô fazem uso de sua língua nativa em situações públicas; há, no entanto, evidências de que quase todos eles a utilizam em situações privadas. Nas famílias, por exemplo, os pais, em geral, dão ordens ou fazem perguntas aos filhos em Yaathe, a despeito de estes invariavelmente responderem em português. Estudos recentes indicam que crianças muito pequenas dominam aspectos particulares do uso da língua, como, por exemplo, a caracterização de gênero.

A despeito do uso sistemático que os Fulni-ô fazem de sua língua em situações privadas, e do esforço que o povo tem demonstrado em manter vivas a sua língua e a sua cultura, através de iniciativas educacionais, há ainda muito pouco registro do Yaathe, o que dificulta bastante quaisquer atividades relacionadas à preservação de suas manifestações linguísticas e culturais.

Atualmente, o material usado nas escolas como recurso de ensino-aprendizagem da língua na reserva indígena Fulni-ô é bastante escasso e de qualidade questionável. [2] Os professores fazem o que podem: escrevem seus próprios textos, preparam aulas e planos de aula, conforme exigido pelas instâncias oficiais, falam sobre cultura e religião, incentivam o uso da língua e o respeito pela cultura como um todo, tudo feito de maneira muito pouco sistemática e sem amparo em usos reais, documentados, da língua. Além de uma cartilha elaborada nos anos 90 do século passado, não há outro material oficial para o ensino da língua.[3] Há, por outro lado, muito material criado e produzido pelos professores, e um esforço cada vez mais constante no sentido de se padronizar a escrita de modo a ser aceita pela comunidade. [4] Parece evidente que o acesso a um banco de dados da língua será de vital importância para a elaboração de materiais didáticos mais adequados, bem como para auxiliar no processo de sistematização da grafia da língua.

Há, sobre o Yaathe, alguns trabalhos acadêmicos de descrição e análise linguística. Entre os mais importantes, citam-se Meland (1968), Meland e Meland (1967), Lapenda (1968) e Barbosa (1991). Meland e Meland (1967) é uma descrição da fonologia, elaborada sob o modelo tagmêmico, bem como Meland (1968). Lapenda (1968) descreve a estrutura da língua de um ponto de vista mais tradicional e Barbosa (1991) é uma descrição fonética e

fonológica, também apoiada no modelo tagmêmico.

Mais recentemente, três trabalhos foram efetuados sobre a língua. Costa (1993) procurou investigar a atual situação lingüística dos Fulni-ô, dada a sua peculiaridade de última língua nativa no Nordeste do Brasil, a fim de verificar tendências à substituição ou ao deslocamento em relação ao Português. Esta investigação serviu como pano de fundo para a observação de fenômenos de atitudes linguísticas de professores não-índios, face à variedade de Português falada pelas crianças índias que chegam à escola da cidade, e de interferências de uma língua na outra, mais precisamente da influência do Yaathe – que consideramos língua materna – sobre o Português – segunda língua. Neste caso, tratava-se da variedade de Português falada pelas crianças índias. Os resultados de tal trabalho podem, por um lado, ajudar a clarear e a melhorar a compreensão dos professores de língua portuguesa das variedades linguísticas que são utilizadas pelos alunos de procedências diversas. Por outro lado, devem contribuir para o conhecimento e o autoconhecimento das nações indígenas. Costa (1999) detém-se sobre a estrutura do Yaathe, procurando descrever e explicar o sistema (fonologia e gramática) e o seu funcionamento. Cabral (2009) enfocou o sistema prosódico da língua, buscando descrever o acento no nível da palavra, experimentalmente.

Atualmente, há estudos em andamento dentro do projeto *Gramática descritiva (de usos) do Yaathe (Fulni-ô)*, desenvolvido no PPGLL/UFAL, sendo duas monografias de iniciação científica (uma sobre gênero e outra sobre nasalidade em Yaathe) e uma dissertação de mestrado (sobre a estrutura da sílaba em Yaathe). A disponibilização de um banco de dados etiquetado, transcrito e devidamente anotado em muito auxiliará a boa execução destes e de futuros estudos acerca da língua.

3.   Objetivos

Em vista do trabalho que vem sendo efetuado há algum tempo na aldeia e com a língua, já se dispõe de uma quantidade razoável de material coletado – listas de palavras, textos variados (letras de músicas, narrativas, cânticos religiosos) e respostas a questionários variados. Parte desse material foi gravado em formato digital. Entretanto, é preciso que se faça um tratamento mais consistente em termos de digitalização e organização para armazenamento e disponibilização pública, de modo a que esse material possa efetivamente vir a constituir um banco de dados da língua.

O objetivo central deste projeto é compor um banco de dados o mais abrangente possível acerca da língua Yaathe, constituído de materiais já coletados e de materiais por coletar. O banco de dados seguirá os modelos hoje adotados por bancos de dados de línguas em perigo de extinção [5], contendo materiais transcritos, anotados e acessíveis à comunidade. Os dados já coletados serão organizados, etiquetados, transcritos e anotados.

Também o projeto tem por objetivo coletar materiais complementares para constituir o banco de dados. Assim, e de acordo com as necessidades estabelecidas a partir da sistematização dos dados já existentes, objetiva-se fazer

---

[1] http://www.unesco.org/culture/ich/index.php?pg=00139.

[2] A escola da aldeia oferece educação básica, do maternal ao ensino médio, incluindo educação de adultos, recebendo, aproximadamente, 1.000 alunos em condições precárias.

[3] Neste ano de 2010, a língua foi incluída na matriz curricular da escola regular da aldeia, sendo assim uma das poucas línguas indígenas brasileiras a ser oficialmente incluída no ensino regular, reconhecida pelo MEC e pela Secretaria de Educação do Estado de Pernambuco.

[4] Cabe observar que a equipe que se propõe a desenvolver este projeto participa deste movimento, apoiando-o, fornecendo assessoria linguística e propondo descrições mais minuciosasde as pectos da língua, que contribuirá para a elaboração de materiais didáticos mais adequados.

[5] Utilizaremos, para este fim, as recomendações feitas pela E-MELD School of Best Practice (http://www.emeld.org/school/).

coleta de dados acústicos de alta qualidade, contendo não apenas material proveniente de listas (como as clássicas Swadesh, Lingua Descriptive Quesionnaire, e aquelas propostas por Healey, em seu Manual de trabalho de campo), mas, sobretudo, exemplares discursivos, entre os quais narrativas de experiência pessoal, mitos, narrativas procedimentais e conversas espontâneas. Muito desse material também será gravado em vídeo, uma vez que informações visuais têm sabidamente importância fundamental para a compreensão de determinados fenômenos linguísticos.

Esse banco de dados é, como já se apontou, o produto principal deste projeto. Entretanto, espera-se que a constituição do banco de dados sirva como ponto de partida para novas pesquisas acerca da língua, para a implementação de estudos já em andamento, para o aprofundamento das discussões acerca de um sistema gráfico aprovado pela comunidade e para a elaboração de materiais didáticos para o ensino da língua. O projeto que aqui se propõe tem por objetivo envolver e formar pesquisadores em diferentes níveis – da IC ao doutorado – e professores pesquisadores, na tarefa de descrição e estudo dos diferentes aspectos da estrutura do Yaathe.

4.   Metodologia

O material já existente será selecionado, levando-se em conta a qualidade da gravação e a potencial utilidade do mesmo. Os exemplares escolhidos serão tratados (digitalizados e editados, em alguns casos), etiquetados e organizados dentro de uma estrutura computacional hierárquica a ser definida.

Uma vez que se tenha uma ideia do material aproveitável dentro do corpus não-catalogado já existente, uma coleta de dados em campo será organizada, tendo como objetivo complementar o material já disponível para compor o banco de dados.

Entre os dados que se planeja coletar incluem-se listas de palavras e frases, tendo como modelo as já clássicas listas Swadesh (Swadesh, 1955), LDQ (Comrie & Smith, 1977), e aquelas propostas por Healey (1975), em seu Manual de trabalho de campo, e uma série de exemplares discursivos, entre os quais narrativas de experiência pessoal, mitos, narrativas procedimentais e conversas espontâneas. Um dos objetivos principais desta coleta de dados é incluir dados de vídeo, uma vez que informações visuais têm reconhecida importância para a compreensão de determinados fenômenos linguísticos. Portanto, objetiva-se gravar também em vídeo a maior parte das sessões de coleta de dados em campo.

Os dados de áudio e vídeo serão gravados e arquivados respeitando todas as medidas e indicações propostas pela *E-MELD School of Best Practice*[6], que vem sendo adotadas em projetos de documentação de línguas indígenas internacionalmente, pelo *Open Archival Information System* (OAIS)[7], que é um modelo de referência, com padrão ISO (14721:2003), adotado pelos bancos de dados linguísticos mais recentes, e anotados seguindo os preceitos do Metadata Encoding and Transmission Standard (METS)[8], também adotados por bancos de dados internacionais.

Após essa fase de organização e coleta de dados, proceder-se-á à etapa seguinte: a transcrição, tradução e anotação dos dados. Essa é uma fase que costuma demandar um tempo considerável de trabalho, pelo que estima-se que apenas um percentual do material será transcrito e anotado. Por conta disso, uma cuidadosa seleção será feita do material a ser transcrito e anotado, levando-se em conta a representatividade e potencial utilidade do mesmo.

A transcrição e tradução serão feitas com o auxílio dos professores de Yaathe, o que resultará em um produto mais acurado e proporcionará uma discussão acerca de um modelo adequado de grafia a ser adotado, com aprovação da comunidade.[9] As transcrições serão feitas no programa Praat (Boersma & Weenik, 2007), uma vez que este programa dá acesso a detalhes acústicos dos dados, o que não apenas facilita a transcrição, nos mais diferente níveis, mas também auxilia a feitura de estudos acústicos os mais diversos. É importante ressaltar que um dos objetivos deste projeto é elaborar um banco de dados que seja disponibilizado para a comunidade acadêmica, com o objetivo de propiciar estudos linguísticos os mais diversos. Portanto, é preciso levar em consideração o arcabouço tecnológico utilizado na construção do banco de dados. Os aplicativos computacionais que se pretende utilizar na execução do presente projeto têm sido sistematicamente utilizados por diversos projetos internacionais de documentação de línguas, por possuírem uma estrutura de fonte aberta, por funcionarem em diversas plataformas operacionais e por estarem em constante desenvolvimento.

Os dados transcritos em Praat serão exportados para o programa ELAN (Hellwig & Uytvanck, 2007), que permite uma maior liberdade de uso para anotação, possibilitando inclusive o alinhamento da transcrição e anotação com arquivos de vídeo. As tecnologias tanto do Praat quanto do ELAN possibilitam que os dados transcritos sejam disponibilizados online para consulta, através do programa open source *Spock*[10], que permite efetuar buscas no corpus transcrito devolvendo transcrição e som correspondente.

Além de disponibilizar os dados localmente, nos servidores da Universidade Federal de Alagoas, para livre consulta pela comunidade, os dados serão depositados em bancos internacionais, tais como o do LAT (Language Archiving Technology[11], garantindo assim a sua preservação.

5.   Considerações Finais

Entende-se, de acordo com Himmelmann (2006), que documentação de línguas é um campo de investigação e de prática linguística cujas preocupações básicas são a

---

[6]   E-MELD   School   of   Best   Practice (http://www.emeld.org/school/).

[7]   Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1 Blue Book January 2002 (Washington, DC: CCSDS   Secretariat,   2002).   Disponível   online: http://public.ccsds.org/publications/archive/650x0b1.pdf.

[8]   Library of Congress, "METS: Metadata Encoding & Transmission   Standard"   (2007), http://www.loc.gov/standards/mets/.

[9]   Cumpre notar que o projeto conta com a participação de uma falante nativa do Yaathe, Fábia Pereira da Silva.

[10]   Spock   -   a   Spoken   Corpus   Client: http://www.iltec.pt/spock/?page=main-pt.

[11]   http://corpus1.mpi.nl.

compilação e a preservação de dados linguísticos primários e interfaces entre esses dados e vários tipos de análises neles baseadas. Além disso, embora preocupação com línguas em risco de extinção seja uma boa razão para que se desenvolvam projetos de documentação de línguas, não é a única. Documentações de línguas fornecem subsídios para as bases empíricas da linguística e de disciplinas afins, tais como tipologia linguística, antropologia cognitiva, etc., que dependem muito de dados de comunidades de fala pouco conhecidas para verificação das suas hipóteses, economizando, assim, recursos de pesquisas.

A principal contribuição do presente projeto de pesquisa é, assim, auxiliar a preservação de uma língua nativa brasileira em estado de iminente extinção, oferecendo uma documentação linguística abrangente e representativa, que poderá ser utilizada não apenas para estudos acadêmicos, mas também para a elaboração de materiais didáticos utilizados no ensino da língua na comunidade indígena.

É importante salientar que o esforço para a preservação de línguas em estado de extinção tem sido considerável, por meio, sobretudo, de agências de fomento internacionais (como a UNESCO e a VolkswagenStiftung, por exemplo). O Yaathe não está incluído em nenhum desses programas, o que torna o financiamento deste projeto ainda mais urgente e relevante. Como apontado acima, o Yaathe é a única língua indígena brasileira ainda sobrevivente no Nordeste do Brasil, o que torna qualquer esforço no sentido de sua preservação extremamente importante, no sentido de valorizar e preservar a identidade da cultura nativa dessa região do país.

6. References

Barbosa, E.A. (1991). Aspectos fonológicos da língua Yatê. (Dissertação de Mestrado). Brasília: UnB.

Boersma, P., Weenik, D. Praat. Disponível em: <www.praat.org>.

Cabral, D.F. Descrição fonético-experimental do acento em Yaathe. (Dissertação de Mestrado). Maceió: UFAL, 2009.

Comrie, B., Smith, N. (1977). *Lingua Descriptive Studies: Questionnaire (= Lingua 42.1)*. Amsterdam: North-Holland. 72 pp.

Costa, J.F. (1993). Bilinguismo e atitudes linguísticas interétnicas: aspectos do contato Português-Yaathe. (Dissertação de Mestrado). Recife: UFPE.

Costa, J.F. (1999). Ya:thê, a última língua nativa do nordeste do Brasil: aspectos morfofonológicos e morfossintáticos. (Tese de Doutorado). Recife: UFPE.

Costa, J.F., Silva, F.P. (2009). Dêixis de gênero em Yaathe. In *Leitura. Revista do Programa de Pós-Graduação em Letras e Linguística – UFAL*. Maceió, nº 43/44, jan./jun. 2009, pp. 123--138.

Foley, W.A. (1997). *Anthropological linguistics. An introduction*. Londres: Blackwell.

Healey, A. (Ed.) (1975). *Language Learner's Field Guide*. Ukarumpa, EHD, Papua New Guinea: Summer Institute of Linguistics Printing Department.

Hellwig, B., Uytvanck, D. van. (2007). Manual do EUDICO Linguistic Annotator (ELAN). Disponível em: <http://www.lat-mpi.eu/tools/elan/>.

Himmelmann, N.P. (2006). Language documentation: What is it and what is it good for? In N.P. Himmelmann, U. Mosel. (Eds.), *Essentials of language documentation*. Berlin/New York: Mouton de Gruyter.

Lapenda, G. (1968). *Estrutura de lingual Yatê*. Recife: UFPE.

Meland, D., Meland, D. (1968). *Word and morpheme list of the Fulni-ô Indian language*. Dallas, Texas: Summer Institute os Linguistics.

Meland, D., Meland, D. (1967). Fulni-ô (Yahthe) phonology statement. In *Arquivo lingüístico* n. 025. Brasília: Summer Institute os Linguistics.

Meland, D. (1968). Fulni-ô grammar. In *Arquivo lingüístico* n. 026. Brasília: Summer Institute os Linguistics.

NWO Advisory Committee on Endangered Language Research. (2000). Endangered language research in the Netherlands. Amsterdam.

Ribeiro, D. (1996). *Os índios e a civilização. A integração das populações indígenas no Brasil moderno*. São Paulo: Companhia das Letras.

Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. In *International Journal of American Linguistics*, 21, pp. 121--137.

Silva, F.P. (2008). Revisão da fonologia do Yaathe para uma proposta de uniformização da escrita na língua. (Trabalho de Conclusão de Curso de Graduação). Maceió: UFAL.

Silva, F.P. (2011). A sílaba em Yaathe. (Dissertação de Mestrado) Maceió: UFAL.

Tupinambá Nheenga: considerações sobre um dicionário escolar do Tupinambá de Olivença, BA

Clara Carolina SANTOS, Consuelo COSTA
Universidade Estadual do Sudoeste da Bahia
Estrada do Bem Querer, km 3, Vitória da Conquista, Bahia
claracarolina@gmail.com

Abstract

A intenção é elaborar um vocabulário bilíngue que compreenda um acervo lexical representativo da língua Tupinambá com informações fonéticas correspondentes a cada entrada. Este vocabulário deverá ser de utilidade nas atividades escolares voltadas para o ensino e fortalecimento da língua Tupinambá e pode constituir-se como uma importante referência da língua e de aspectos da cultura Tupinambá. Os resultados deste estudo deverão servir como material de apoio à escola e nucleadas Tupinambá, mas também para o ensino do português, pois atualmente os Tupinambá buscam uma aprendizagem escolar nas duas línguas. O vocabulário escolar bilíngüe Tupinambá – Português terá a inovação, em relação aos dicionários escolares em línguas indígenas em Tupinambá, de apresentar a transcrição fonética dos verbetes o que - em conjunto com as *oficinas de fonética e fonologia* oferecidas aos professores indígenas - proporcionará um suporte material que auxiliará de modo seguro o uso da língua na escola e sua retomada pela comunidade. Além disso, este vocabulário diferenciar-se-á dos demais dicionários do Tupi Antigo (língua da qual o Tupinambá é uma variedade) por considerar a convenção ortográfica dos índios de Olivença.

Keywords: Tupinambá; línguas indígenas; fonologia.

1. Paper

Quando é impressa em 1595 uma *Gramática* de José Anchieta[1] para uso na Companhia de Jesus à variedade de língua ali descrita não é atribuído nenhum nome (Rodrigues, 2010: 28). É apenas no decurso da empresa lusitana que a *língua mais usada na costa do Brasil* é denominada *língua brasílica* ou *língua do Brasil*[2]. Nos primeiros livros sobre o Brasil, língua da costa, língua brasílica ou apenas língua é alusão à língua nativa das nações habitantes da quase totalidade da costa brasileira (Rodrigues, 1994), foi uma variedade empregada na missão jesuítica nos séculos XVI e XVII (Câmara Jr., 1979: 99) e, a partir do século XIX, é considerada uma língua das origens do Brasil (Dietrich, 2010: 10). Em estudos mais recentes, a delimitação da língua da costa é referida como "uma realidade linguística complexa (Dietrich, 2010: 9)".

Para ilustrar esta diversidade, o tupinambá corresponde a uma variedade de língua da família tupi-guarani (Rodrigues, 1996: 57, a*pud* Dietrich, 2010: 9)*,*

"em que se baseiam as línguas gerais da época colonial, a língua brasílica, a língua geral paulista e a língua geral amazônica (Dietrich, 2010: 9)". Para Dietrich & Noll (2010) esta *variedade "se falava entre casais de portugueses com mulheres indígenas e seus filhos mestiços* (Dietrich & Noll, 2010: 81)*"* na costa brasileira e, tendo servido aos fins catequizadores da Companhia de Jesus, com possíveis empréstimos do tupinambá no português, os jesuítas passam a denominar esta variedade de língua brasílica ou língua do Brasil (Rodrigues, 2010 a*pud* Dietrich & Noll, 2010). Do contato entre uma variedade linguística da costa e a língua portuguesa, surge a língua geral *que "do ponto de vista linguístico, já não designava o tupi genuíno, mas uma forma modificada desta língua* (Dietrich & Noll, 2010: 81)", mais simplificado, *"sobretudo na sua fonética e na morfossintaxe* (Dietrich & Noll, 2010: 81). Neste exemplo, três variedades de língua são descritas nos livros que servem de referência a este estudo. A primeira é a língua representada na gramática de Anchieta, contada nas cartas e nos relatórios ultramarinos; a segunda é possivelmente oriunda do contato entre portugueses e suas mulheres e filhos, como nos ensina Dietrich & Noll (2010) e a terceira começa a ser delineada a partir do século XVIII e

"se referia inicialmente à língua dos índios tupinambás (do Pará), para diferenciar a forma genuína do seu tupi da língua geral amazônica que se formou no curso da expansão portuguesa na bacia do rio Amazonas nos séculos XVII e XVIII (Dietrich & Noll, 2010: 81-82)".

No caso específico do Tupinambá, julga-se que esta variedade tenha se espalhado "por causa das migrações contínuas dos Tupinambás (Dietrich, 2010: 12)" entre Santa Catarina, Bahia, Maranhão e Amazônia. Neste texto, faremos referência ao uso da variedade Tupinambá

---

[1] Anchieta (1595)**.**

[2] Sobre o desenvolvimento dos modos de nomear a língua mais usada na costa, Rodrigues (2010) cita como exemplo relatórios da Companhia de Jesus. No texto, eles seguem a ordem cronológica de impressão no período seiscentista. É interessante esta sequência, pois demonstra no curso do tempo como palavras como "língua" e "brasílica" são paulatinamente associadas à "língua da costa". São estes os documentos enumerados: "*(...) Nomes das partes do corpo humano, pella língua do Brasil pelo Padre Pero de Castinho (manuscrito datado de 1613 publicado por Ayrosa, 1937); Catecismo na lingoa brasilica (edição do padre Antonio d'Araujo, 1618), Arte da língua brasilica pelo padre Luis Figueira (1621), Vocabulario na língua brasilica (manuscrito anônimo datado de 1622, publicado por Ayrosa, 1938), Catecismo brasilico da doutrina christaã, pelo P. Antonio de Araújo, emendado nesta segunda impressão pelo P. Bertholomeu de Leam (1685), Arte de grammatica da língua brasilica do P. Luis Figueira.* (p.28)*"*

entre os indígenas em Olivença, Ba. Para este estudo, as línguas da família tupi-guarani formam "um grupo com outras línguas mais distantes na sua diferenciação histórica, mas que, elas também, apresentam correspondências regulares de sons, de palavras e de formas gramaticais (Dietrich, 2010: 10)". De um modo geral, escolhemos denominar a língua por Tupinambá, pois este é o uso corrente entre os indígenas em Olivença, embora saibamos que, em seu estudo na escola e uso primeiro, a língua alvo é o Tupi Antigo.

Contrastando diferentes registros seiscentistas da língua falada na costa, considerando algumas condições de impressão e escrita destes textos, Rodrigues (2010) constata que há "alguma diversidade (...) entre a fala dos tupis e a dos demais falantes da língua brasílica, diversidade que aparece também nos textos em língua indígena escritos por Anchieta nos primeiros dez anos em que esteve atuando entre os tupis (Rodrigues, 2010: 28[3])". Isso não é dado novo.

No contato com as nações da costa brasileira é possível que os jesuítas tenham esbarrado nas cerca de 79 línguas descritas ou meramente referidas na narrativa extensa de Fernão Cardim (1925)[4]. Curiosamente, esta diversidade foi ignorada em seu uso primeiro pois aos jesuítas importava tratar aquelas línguas não travadas, isto é, ignorava-se aquelas línguas "muito difíceis de pronunciar, línguas consideradas anômalas dentro do

egocentrismo (Câmara Jr, 1979: 99)" europeu. Estudos contemporâneos reafirmam a idéia de que o registro das variedades do tupi é basicamente vinculado a relações amigáveis entre portugueses e índios no litoral de São Vicente e, "serra acima, na região de Piratininga e do Alto do Rio Tietê (no atual estado de São Paulo) (Rodrigues, 2010: 28)". Neste contexto de "disciplinização da língua Tupi (Câmara Jr., 1979: 102)", duas variedades de língua concorrem nos textos basilares seiscentistas, referências para o estudo que, agora, apresentamos. Conforme Rodrigues (2010: 28):

> "Embora Anchieta tivesse elaborado uma primeira versão de sua gramática já antes de 1560, enquanto ainda estava entre os tupis de São Vicente, a versão publicada dessa obra foi revista e adaptada às características da língua falada ao longo da costa do Rio de Janeiro e para o norte, tendo sido completada ou na Bahia ou no Espírito Santo, portanto ao norte do Rio de Janeiro, fato este que determinou escrever, na versão publicada, que os tupis são além dos tamoyos do Rio de Janeiro".

Além destas variedades não podemos esquecer da apropriação dos textos seiscentistas a partir dos tupinólogos novecentistas. Parte daquilo que o senso comum compreende como "língua indígena" é esse imaginário romântico que associa o nome *tupi* à construção da nacionalidade brasileira (RODRIGUES, 2010: 29). No século XIX, o Tupi e as línguas do seu tronco "passaram a ser consideradas o protótipo das nossas línguas indígena*s* (Câmara Jr, 1979: 99)" e, embora os estudos novecentistas almejem esta pureza numa língua originária, eles partem de registros já com uma ampla difusão da língua e, por isso, "já não designava o tupi genuíno, mas uma forma modificada desta língua (Dietrich & Noll, 2010: 81)" de modo que, em alguns registros, confunde-se com a língua geral, com o próprio tupi (Silva Neto, 1986: 30-51 *apud* Dietrich & Noll, 2010: 81) e, em alguns casos, com um "construto dos jesuítas (Dietrich & Noll, 2010: 81)". Sobre este assunto, Aryon Rodrigues (2010) diz que o tupi é "reativado entre os intelectuais, sobretudo na primeira metade do século XIX, logo após a independência do país, quando se buscava uma identidade nacional (p. 29)". Rodrigues (2010) lembra o estudo de Eldeweiss (1947), para quem esta reativação é fruto de publicações em catálogos espanhóis do final do século XVIII sobre a língua tupi em território brasileiro (Eldeweiss, 1947, *apud* Rodrigues, 2010)[5].

---

[3]   Para esclarecer o lapso na citação, a variação que esta citação faz referência é a pronúncia dos verbos acabados em consoantes, descritos no *Vocabulário da Lingua Brasílica*, além de diferenças morfológicas na forma indicativa dos verbos transitivos iniciados por *m* que não recebem o prefixo relacional – i após o prefixo do sujeito, tendo nulo em seu lugar (cf. Rodrigues, 2010: 28-29).

[4]   Entre diversas nações, sobre os Tupinambás, cuja variedade é foco neste estudo, assim diz este registro: "Outros há a que chamam Tupinabas: estes habitam do Rio Real até junto dos Ilhéus; estes entre si eram também contrários, os da Bahia com os do Camamu e Tinharê.Por uma corda do Rio de São Francisco vivia outra nação a que chamavam Caaété, e também havia contrários entre estes e os de Pernambuco. Dos Ilhéus, Porto Seguro até Espírito Santo habitava outra nação, que chamavam Tupinaquim; estes procederam dos de Pernambuco e se espalharam por uma corda do sertão, multiplicando grandemente, mas já são poucos; estes foram sempre muito inimigos das cousas de Deus, endurecidos em seus erros, porque eram vingativos e queriam vingar-se comendo seus contrários e por serem amigos de muitas mulheres. Já destes há muitos cristãos e são firmes na fé". (Cardim, F., 1925). O percurso deste livro é curioso. Embora tenha sido recuperado no movimento modernista como um registro fidedigno da "realidade da nação brasileira" sabe-se que a sua primeira impressão é realizada em terras inglesas em 1625, pois o navio de seu autor naufragou e, assim, seus escólios e sobreviventes do naufrágio são capturados pelo capitão James Cook. Escrito entre as décadas de 1580 e 1625, data da primeira publicação do Tratado, este livro é reimpresso pelos lusitanos apenas no século XVIII a mando de D. Manuel, para divulgar a história portuguesa, ilustrando, assim, o seu império. Não sei bem, por isso, se este livro pode ser atualizado como referência aos escritos jesuíticos da Companhia de Jesus. Por outro lado, sua atualização no século XX é bastante proveitosa para o conhecimento da diversidade de línguas indígenas dos seiscentos brasileiro e, neste texto, serve a este fim.

[5]   A importância do Tupi é divulgado em terras não brasileiras por meio da circulação de livros, em especial, de relatos de viajantes. Conforme Rodrigues (2010): "Um dos primeiros escritores brasileiros a destacar o nome tupi foi o poeta e pesquisador Gonçalves Dias, em sua poesia romântica de grande ressonância. O naturalista Martius (1863-67), no primeiro ensaio de classificação dos povos indígenas do Brasil, distinguiu nove grupos étnicos, ao primeiro dos quais deu o nome de tupis e guaranis; essa classificação foi reorganizada pelo etnólogo von den Steinen (1886), que distinguiu oito grupos e chamou o

Conforme Rodrigues, se por um lado a partir da rememoração Tupi no século XIX como a língua originária brasileira esta variedade ganha destaque entre os estudos, por outro o Tupinambá "foi caindo em desuso com o quase total extermínio" dos tupinambás na Bahia e a "progressiva catequização e assimilação" (Rodrigues, 2010: 30) dos tupinambás no Maranhão. Esta repercussão pode ser sentida tanto no desenvolvimento de estudos contemporâneos quanto na apropriação das línguas em contato com jesuítas das expedições ultramarinas ao fixarem a gramática da língua indígena.

Há uma controvérsia bastante conhecida sobre a delimitação da língua Tupi Antigo em oposição à Tupinambá e, diz-se, se partirmos do preceito de que estas línguas devem ser comparadas em sua variação histórica, mesmo estudiosos como Aryon Dall'Igna Rodrigues teriam "confundido" os termos Tupinambá e Tupi Antigo, embora tenha levado a termo um trabalho magistral na língua a que nos referimos neste trabalho. Contradições à parte, recusamos esta delimitação arbitrária, bem como os discursos que a amparam, pois que a noção de tempo histórico vinculada a este tipo discussão é aquele progressista, acumulador, no qual exemplos passados podem servir para atualizações presentes. Outro motivo para desconsiderarmos esta discussão histórica e formalista (e talvez o mais contundente) é porque a nós importa a atualização da língua em seu contexto contemporâneo, de revitalização e constituição identitária para as comunidades indígenas em Olivença. Sendo um estudo para revitalização da língua Tupi Antigo como língua estrangeira na comunidade Tupinambá de Olivença os processos linguísticos devem ser respeitados em seu uso contemporâneo.

O efeito desta história é bem conhecido entre os Tupinambás de Olivença e, mesmo lá, em uma comunidade que teve sua língua violentamente apagada, predomina-se uma "noção geral de que o modelo, o verdadeiro exemplo típico das línguas indígenas do Brasil são os dialetos Tupi da costa" (Câmara Jr., 1979: 100), argumento que Eduardo de Almeida Navarro não se cansa de lançar mão em seu *Curso Moderno de Tupi Antigo*, chegando ao extremo de escolher como verbo para "chegar" um verbete citado *apenas uma única vez* na Gramática de Figueira (o *iepotar*). Chegaram os Portugueses e l*a nave va*[6]...

Anterior à Assessoria Linguística do *Projeto Tupinambá*, um *Curso de Tupi* era ministrado na comunidade pelos próprios professores das escolas. O livro de referência para este estudo era o *Curso Moderno de Tupi Antigo*, de Eduardo Navarro (2005), e, por isso, a primeira lição do livro, "Chegaram os portugueses", foi estudada durante as oficinas oferecidas em 2011 na escola sede. Este manual, no entanto, é a) destinado a professores que já estejam familiarizados com algum estudo gramatical de alguma língua, o que não é o caso para todos os professores indígenas da escola e b) não cumpre o fim pedagógico de ensinar às crianças da escola estruturas da língua Tupinambá. Espera-se que, com o desenvolvimento de oficinas nas escolas, novos textos dos professores e dos alunos, bem como cantigas e mitos da comunidade, sejam integrados ao ensino da língua Tupinambá nas escolas[7].

A permanência das guerras aos indígenas por meios aparentemente pacíficos é história que, infelizmente, conta com grande documentação na historiografia brasileira. Isso não significa, entretanto, que os Tupinambás não tenham resistido (como é comum esta nação ser referida nas histórias desde os seiscentos). Uma das tentativas de revitalização de sua cultura e da língua dos seus ascendentes partiu da própria comunidade indígena que, tendo participado do encontro *C-Indy* na Universidade Estadual da Bahia, organizado pela professora Consuelo Costa, requisitaram um Curso de Tupi, a princípio na escola Sapucaeira, em Olivença, na intenção de implantar uma escola bilíngue.

2. References

Anchieta, J. (1595). *Arte de Gramática da língua mais usada na costa do Brasil feita pelo padre Joseph de Anchieta de Cõpanhia de IESU*. Coimbra, por Antonio de Mariz.

Ayrosa, P. (Ed.). (1938). *Vocabulário na Língua Brasílica: manuscrito português-tupi do século XVII*. Volume XX da Coleção Departamento de Cultura. São Paulo.

Barbosa, Pe. (1956). *Curso de Tupi Antigo: Gramática, Exercícios, Textos*. Rio de Janeiro: Livraria São José. Disponível em: <http://biblio.etnolinguistica.org/barbosa_1956_curso>.

Caldas, R.B.C. Dicionários bilíngues: uma reflexão acerca do tratamento lexical em línguas Tupi. In *Línguas e Culturas Tupi*, Volume 2. Brasília: Ed. Curt Nimuendajú, pp. 105--115.

Câmara Jr, J.M. (1979). *Introdução às Línguas Indígenas Brasileiras*. Rio de Janeiro: Ao Livro Técnico.

Câmara Jr, J.M. (2003) *Introdução às línguas indígenas brasileiras*. 3 ed. – Rio de Janeiro: Ao livro técnico.

Cardim, F. (1925). *Tratados da Terra e Gente do Brasil*. Introdução e notas de Batista Caetano, Capistrano de Abreu e Rodolpho Garcia. Rio de Janeiro: Ed. J.

---

primeiro deles simplesmente tupis. Já dez anos antes Couto de Magalhães, autor brasileiro de grande prestígio, tinha publicado, sob patrocínio do governo imperial, o seu curso de língua geral amazônica..." (p.30).

[6] Ao leitor atento que se interesse pelas questões de variação e sobredeterminações acerca a língua Tupinambá, Tupi Antigo, possíveis divergências entre os modos de nomear as línguas deixamos como sugestão a bibliografia seguinte: Freire, J.R.B. & Rosa, M.C. (2003); Câmara Jr, J.M. (2003).

---

[7] Para aqueles curiosos, é interessante compreender que este estudo do Tupi na escola indígena de Olivença é amparado por um conjunto de leis da Bahia, a saber, a Lei no. 18.629/2010 (que institui o plano de carreira para o professor indígena na Bahia); pelo Decreto n. 8.741 de 12 de março de 2013 que cria a categoria de escola indígena baiana e pela resolução CEE no. 106/2004 que estabelece diretrizes e procedimentos para a organização e oferta da Educação escolar indígena no sistema Estadual de Ensino da Bahia.

Leite.

Dias, A.G. (1959). *Dicionário da Língua Tupy chamada Língua Geral dos Indígenas do Brasil por A. Gonçalves Dias*. Lipsia: F. A. Brockhaus, Livreiro de S. M. o Imperador do Brasil.

Dietrich, W. O tronco tupi e as suas famílias de línguas. Classificação e esboço tipológico. In V. Noll, W. Dietrich. (Eds.), *O português e o tupi no Brasil*. São Paulo: Contexto.

Dietrich, W., Noll, V. (2010). O papel do tupi na formação do português brasileiro. In V. Noll, W. Dietrich (Eds.), *O português e o tupi no Brasi*l. São Paulo: Contexto.

Fargetti, C.M. (2010). *Cultura Material indígena: questões lexicográficas*. Línguas e Culturas Tupi, Volume 2. Brasília: Ed. Curt Nimuendajú. pp. 117--129.

Figueira, L. (1880). *Arte de Gramática da Língua Brasílica do Padre Luiz Figueira, Theólogo da Companhia de Jesus*. Lisboa, na Oficina de Miguel deslundes, na rua da Figueira, 1657. Nova edição anotada por Emílio Allain, Rio de Janeiro, Tipografia e litografia de Lombacta, Ourives, no. 7.

Freire, J.R.B., Rosa, M.C. (Eds.). (2003). *Línguas Gerais: Política Linguística e Catequese na América do Sul no Período Colonial*. Rio de Janeiro: Ed. UERJ.

Lee, K. (2005). *Conversing in colony: The Brasílica and the vulgar in Portuguese America, 1500-1759*. Maryland: The John Hopkins University.

Lemle, M. (1971). *Internal classification of the Tupi – Guarani Linguistic family*. Tupi Studies I, Summer Institute of Linguistics and Related Fields, No. 29, Ed. Benjamin F. Elson, University of Oklahoma, Norman, USA.

Rodrigues, Aryon. (1994). *Línguas Brasileiras: Para o conhecimento das línguas indígenas*. São Paulo: Edições Loyola.

Rodrigues, A. (2010). Tupi, tupinambá, línguas gerais e português no Brasil. In V. Noll, W. Dietrich (Eds.), *O português e o tupi no Brasil*. São Paulo: Contexto.