

COMPUTERS, LITERATURE AND PHILOLOGY  
CLIP 2003

LA GESTIONE UNITARIA DELL'EREDITÀ CULTURALE  
MULTILINGUE EUROPEA E LA SUA DIFFUSIONE IN RETE



a cura di  
CARLOTA NICOLÁS MARTÍNEZ  
MASSIMO MONEGLIA

Firenze University Press

UNIVERSITÀ DEGLI STUDI DI FIRENZE  
80° ANNIVERSARIO 1924-2004

– 2 –

UNIVERSITÀ DEGLI STUDI DI FIRENZE  
80° ANNIVERSARIO 1924-2004

1. *L'Università degli Studi di Firenze 1924-2004. Atti della tavola rotonda di presentazione del volume (Firenze, 17 Dicembre 2004)*, a cura di Sandro Rogari, 2005

COMPUTERS, LITERATURE AND PHILOLOGY  
CLIP 2003

LA GESTIONE UNITARIA DELL'EREDITÀ CULTURALE  
MULTILINGUE EUROPEA E LA SUA DIFFUSIONE IN RETE

ATTI DEL CONVEGNO  
(FIRENZE, 4-5 DICEMBRE 2003)



*a cura di*

CARLOTA NICOLÁS MARTÍNEZ  
MASSIMO MONEGLIA

Firenze University Press  
2005

Computers, Literature and Philology CLiP 2003 : la gestione unitaria dell'eredità culturale multilingue europea e la sua diffusione in rete : atti del convegno : Firenze, 4-5 dicembre 2003 / a cura di Carlota Nicolás Martínez, Massimo Moneglia. – Firenze : Firenze university press, 2005. (80. anniversario 1924-2004 / Università degli Studi di Firenze, 2)

<http://digital.casalini.it/8884532957>

Stampa a richiesta disponibile su <http://epress.unifi.it>

ISBN 88-8453-295-7 (online)

ISBN 88-8453-296-5 (print)

001.30285 (ed. 20)

Discipline umanistiche-Applicazioni dell'informatica

Si ringrazia la Banca Toscana per il sostegno finanziario dato all'organizzazione del convegno e alla pubblicazione e di questi Atti.

© 2005 Firenze University Press

Università degli Studi di Firenze

Firenze University Press

Borgo Albizi, 28

50122 Firenze, Italy

<http://epress.unifi.it/>

*Printed in Italy*

## Sotto l'Alto patronato del Presidente della Repubblica Carlo Azeglio Ciampi

### Comitato d'onore

Marcello Pera - *Presidente del Senato della Repubblica*  
Pier Ferdinando Casini - *Presidente della Camera dei Deputati*  
Letizia Moratti - *Ministro dell'Istruzione, dell'Università e della Ricerca Scientifica*  
Romano Prodi - *Presidente della Commissione Europea*

### Comitato promotore

Augusto Marinelli - *Rettore Università degli Studi di Firenze*  
Claudio Martini - *Presidente della Regione Toscana*  
Michele Gesualdi - *Presidente della Provincia di Firenze*  
Leonardo Domenici - *Sindaco di Firenze*  
Gianfranco Venturi - *Presidente della Provincia di Pistoia*  
Renzo Berti - *Sindaco di Pistoia*  
Daniele Mannocci - *Presidente della Provincia di Prato*  
Fabrizio Mattei - *Sindaco di Prato*  
Vittorio Bugli - *Sindaco di Empoli*  
Alfiero Ciampolini - *Circondario Empolese Valdelsa*  
Gian Valerio Lombardi - *Prefetto di Firenze*  
Leopoldo Di Mattia - *Prefetto di Pistoia*  
Giuseppe Badalamenti - *Prefetto di Prato*  
Cesare Angotti - *Direttore Generale dell'Ufficio Scolastico Regionale per la Toscana*  
Alberto Carmi - *Presidente dell'Ente Cassa di Risparmio di Firenze*  
Aureliano Benedetti - *Presidente della Cassa di Risparmio di Firenze*  
Giuseppe Mussari - *Presidente della Fondazione Monte dei Paschi di Siena*  
Paolo Mottura - *Presidente della Banca Toscana*  
Ivano Paci - *Presidente della Fondazione Cassa di Risparmio di Pistoia e Pescia*  
Gianni Zonin - *Presidente della Cassa di Risparmio di Prato*  
Luca Mantellassi - *Presidente della Camera di Commercio Industria Artigianato Agricoltura di Firenze*  
Andrea Gualtierotti - *Presidente della Camera di Commercio Industria Artigianato Agricoltura di Pistoia*  
Luca Rinfreschi - *Presidente della Camera di Commercio Industria Artigianato Agricoltura di Prato*  
Francesco Adorno - *Presidente dell'Accademia Toscana di Scienze e Lettere "La Colombaria"*  
Francesco Sabatini - *Presidente dell'Accademia della Crusca*  
Franco Scaramuzzi - *Presidente dell'Accademia dei Geogofili*  
Giorgio Van Straten - *Sovrintendente del Teatro Comunale*

### Comitato organizzatore

Sandro Rogari - *Presidente - Delegato del Senato Accademico*  
Paolo Citti  
Luigi Lotti

## Enti sostenitori

	Regione Toscana		Ente Cassa di Risparmio di Firenze
	Provincia di Firenze		Cassa di Risparmio di Firenze
	Comune di Firenze		Fondazione Monte dei Paschi di Siena
	Provincia di Pistoia		Banca Toscana
	Provincia di Prato		Cassa di Risparmio di Prato
	Comune di Prato		Fondazione Cassa di Risparmio di Pistoia e Pescia
	Comune di Empoli		Camera di Commercio Industria Artigianato Agricoltura di Firenze
	Circondario Empolese Valdelsa		Camera di Commercio Industria Artigianato Agricoltura di Pistoia
	Università degli Studi di Firenze		Camera di Commercio Industria Artigianato Agricoltura di Prato

Si ringraziano gli Enti sostenitori delle manifestazioni per gli 80 anni dell'Università degli Studi di Firenze che contribuiscono a sostenere gli oneri finanziari delle pubblicazioni di questa collana.

# Indice

<i>M. Moneglia e M. C. Nicolás Martínez</i>	1	Introduzione
<i>M. Moneglia e M. C. Nicolás Martínez</i>	9	Introducción
<i>M. Moneglia e M. C. Nicolás Martínez</i>	17	Introduction
<i>A. H. Renear</i>	25	Text from Several Different Perspectives: The Role of Context in Markup Semantics

## APPLICAZIONI INFORMATICHE NELLA FILOLOGIA COMPUTER

### APPLICATIONS IN PHILOLOGY

<i>F. A. Marcos-Marín</i>	35	Diacronía y reconstrucción computacional
<i>S. Casciano</i>	45	Esperimento per un'edizione diplomatico-interpretativa computerizzata dei Ricordi di Francesco Guicciardini
<i>D. Fiormonte e C. Pusceddu</i>	61	Tempo e genesi del testo. Un esperimento in rete
<i>D. Silvi</i>	83	Lo stato dell'analisi testuale informatizzata: riflessioni a partire da un esperimento

## SISTEMI INFORMATICI PER LA INTEGRAZIONE, LA DIFFUSIONE E LA GESTIONE DEL PATRIMONIO CULTURALE

### COMPUTERIZED SYSTEMS FOR THE INTEGRATION, DIFFUSION AND MANAGEMENT OF CULTURAL HERITAGE

<i>M. Moneglia, A. Scarano e M. Spinu</i>	107	The C-ORAL-ROM multilingual corpus of spontaneous speech: validation of the prosodic annotation by expert transcribers
<i>M. C. Nicolás Martínez</i>	121	Plurilingüismo, traducción y diccionarios
<i>P. Cotoneschi e A. Farsetti</i>	137	L'editoria digitale della Firenze University Press
<i>R. Delmonte</i>	149	Text understanding from discourse model and internal processes

## IL MULTILINGUISMO E IL WEB

### MULTILINGUALISM AND THE WEB

<i>E. Burr</i>	201	Multilingualism and the Web
<i>L. Carrer</i>	215	Multiculturalità e rete: voci migranti in Italia
<i>P. P. Sánchez-Villalón, M. Ortega e A. Sánchez-Villalón</i>	239	Envolving to ubiquitous e-learning environments: a collaborative writing scenario
<i>E. Tosques</i>	247	About Writing in English for the Workplace



*A. Bianchi*

253 Un percorso didattico in rete per l'italiano  
lingua seconda

*R. Sitman*

259 Reflexiones en torno al uso de la prensa en  
línea para la enseñanza de lenguas desde la  
perspectiva del E/LE

# Introduzione

*Massimo Moneglia*

Università degli Studi di Firenze  
(moneglia@unifi.it)

*María Carlota Nicolás Martínez*

Università degli Studi di Firenze  
(carlota.nicolas@unifi.it)

## 1. RINGRAZIAMENTI E NOTIZIE

CLiP (*Computers Literature and Philology*) è una delle più note sedi di discussione nell'informatica umanistica, per il suo livello scientifico e il suo valore internazionale. La Conferenza CLiP si riunisce annualmente dal 1998, ogni anno in una diversa università europea. Partecipano principalmente scienziati europei che lavorano nel campo dell'*Humanities Computing*, ma anche studiosi non europei, e in special modo americani. Il Comitato Scientifico di CLiP è composto dai seguenti membri:

Lou Burnard (Oxford University )

Elisabeth Burr (Universität Bremen)

Domenico Fiormonte (Università di Roma III)

Francisco Marcos-Marín (Universidad Autónoma de Madrid e Università di Roma "La Sapienza")

María Carlota Nicolás Martínez (Università di Firenze)

Concha Sanz Miguel (Universidad de Castilla-La Mancha)

Il primo incontro, Edimburgo 1998, è stato, convocato su iniziativa di Domenico Fiormonte e Jonathan Usher ai quali si sono uniti poi Lou Burnard, Oxford University; Elisabeth Burr, Universität Bremen; Francisco Marcos-Marín, Universidad Autónoma de Madrid e Antonio Zampolli, Università degli Studi di Pisa, consapevoli del significato delle nuove tecnologie per la filologia e le altre scienze umanistiche e dell'importanza dello sfruttamento dei contenuti umanistici in formato digitale e in rete. Gli incontri successivi seguono questa impostazione nell'evoluzione del quadro culturale e tecnologico europeo e si sono tenuti nelle seguenti sedi

1999 - Roma, Università di Roma "La Sapienza"

2000 - Alicante, Universidad de Alicante e Instituto Cervantes

<<http://www.cervantesvirtual.com/CLiP2000/index.htm>>

2001 - Duisburg, Gerhard Mercator University

<<http://www.uni-duisburg.de/FB3/CLiP2001/index.htm>>

2002 - Albacete, Gestión del Conocimiento y Nuevas Tecnologías Universidad de Castilla-La Mancha

<<http://www.uclm.es/gcynt/clip2002/index.htm>>

Gli atti delle precedenti conferenze CLiP sono pubblicati rispettivamente in:

Fiormonte, Domenico / Usher, Jonathan (eds.) (2001): *New Media and the Humanities: Research and Applications. Proceedings of the first seminar Computers, Literature and Philology*, Edinburgh 7-9 September 1998. Oxford: Humanities Computing Unit, University of Oxford.

<<http://www.hcu.ox.ac.uk/publications/clip.html>>

Fiormonte, Domenico (a cura di) (2003): *Informatica umanistica: dalla ricerca all'insegnamento. Atti dei convegni CLiP 1999 (Roma "La Sapienza") e 2000 (Universidad de Alicante)*. Roma: Bulzoni.

<[http://www.selc.ed.ac.uk/italian/digitalvariants/events/Clip99\\_2000.htm](http://www.selc.ed.ac.uk/italian/digitalvariants/events/Clip99_2000.htm)>

Proceedings of the CLiP conference N° 3 Albacete, 2002, Monograph in *La Musa Digital*

<[http://www.uclm.es/lamusa/lista\\_articulos\\_mono.asp?numero=3&clengua=en](http://www.uclm.es/lamusa/lista_articulos_mono.asp?numero=3&clengua=en)>

CLiP 2003 si è celebrato nell'ambito delle manifestazioni degli *80 anni dell'Università degli Studi di Firenze* tenutesi sotto l'alto patrocinio del Presidente della Repubblica Italiana Carlo Azeglio Ciampi e ha ricevuto il patrocinio e il sostegno di Regione Toscana, CSIAF (Centro Servizi Informatici dell'Ateneo Fiorentino), FUP (Firenze University Press); Facoltà di Lettere, Corso di Laurea Specialistica in Informatica per le discipline Umanistiche, Dipartimento di Lingue e letterature Neolatine, Dipartimento di Italianistica (Laboratorio Linguistico-LABLITA).

Il seminario è stato coordinato da María Carlota Nicolás Martínez e si è svolto presso il Dipartimento di Italianistica dell'Università di Firenze nei giorni 4 e 5 Dicembre 2003.

Comitato organizzatore:

María Carlota Nicolás Martínez (Dipartimento di Lingue e Letterature Neolatine)

Anna Dolfi (Dipartimento di Italianistica)

Giovanna Angeli (Dipartimento di Lingue e Letterature Neolatine)

Massimo Moneglia (Dipartimento di Italianistica)

Eugenio Picchi (ILC-CNR, Pisa)

Il lavoro di CLiP 2003 sono stati seguiti da un Post-conference workshop didattico offerto dall'*Association for Literacy and Linguistic Computing (ALLC)*, dedicato ai linguaggi di *markup* e allo standard di codifica dei testi TEI. Al *Workshop*, diretto da Harold Short, Direttore del *Centre for Computing in the Humanities*, Kings College di Londra, hanno preso parte giovani del corso di laurea specialistica in Informatica per le discipline umanistiche dell'Università di Firenze e molti intervenuti al seminario. Harold Short, Alejandro Bia e Elena Pierazzo, che hanno tenuto le lezioni meritano una menzione e un ringraziamento.

A nome del comitato scientifico di CLiP ringraziamo in particolare il comitato organizzatore degli *80 anni dell'Ateneo Fiorentino*, nella persona del Prof. Sandro Rogari,

per aver accolto questa conferenza tra le manifestazioni celebrative. Un ringraziamento particolare va al Dipartimento di Italianistica e al suo direttore Prof. Anna Dolfi, per l'ospitalità data al convegno e un caldo grazie ai collaboratori del Laboratorio LABLITA per aver sostenuto l'organizzazione logistica e informatica del convegno. Ringraziamo inoltre Ida Tucci per aver impaginato questi Atti.

A nome di tutti i partecipanti alla conferenza vogliamo esprimere ancora una volta l'unanime cordoglio per la ancora recente scomparsa di Antonio Zampolli, la personalità che più di ogni altra ha contribuito alla nascita e allo sviluppo dell'informatica umanistica in Europa e che ha onorato questa conferenza prendendo parte al suo comitato scientifico

## 2. CLiP 2003

L'obiettivo della conferenza internazionale CLiP è l'integrazione delle discipline umanistiche e delle tecnologie dell'informazione e della comunicazione (TLC). Letteratura e Filologia sono termini da intendere in senso ampio. Per Letteratura si intendono i testi di ogni tipo (parlato, scritto, ipertesto) nei quali possono essere contenute immagini, suoni, grafici ecc.; la Filologia si occupa dell'analisi scientifica di tali testi da diverse prospettive (linguistica, culturale, critica testuale, ecc.).

Il CLiP 2003 è orientato alle tematiche relative al multilinguismo nei contenuti umanistici e ha come obiettivo principale di riunire studiosi e operatori che operano nell'informatica umanistica intorno al tema: *La gestione unitaria dell'eredità culturale multilingue europea e la sua diffusione in rete.*

CLiP affronta le tematiche del multilinguismo relativamente al contesto dell'Unione Europea e s'interessa soprattutto alla collaborazione fra insegnamento e ricerca. Il multilinguismo dell'unione e l'educazione alla cultura multilingue dei cittadini europei sono infatti due condizioni di fondo per la gestione e la fruizione della comune eredità culturale.

“Le risorse culturali e scientifiche europee sono un'unica risorsa pubblica che forma e sviluppa la memoria delle nostre diverse società e assicura una base solida per lo sviluppo della nostra industria dei contenuti digitali in una società della conoscenza sostenibile.” (Conclusioni dell'Expert-meeting degli Stati Membri dell'Unione Europea, Lund, aprile 2001).

I lavori attuali nel settore dell'eredità culturale condividono tale assunzione espressa nei Lund Principles, ma perché le diverse risorse culturali nazionali possano essere sentite come un bene comune deve essere affrontato l'ostacolo della natura multilingue dell'eredità culturale europea e dalla persistente mancanza di sistemi robusti che garantiscano la possibilità di un accesso multilingue in rete ai cittadini dell'Unione.

CLiP 2003 affronta il nesso tra multilinguismo e eredità culturale anche in relazione ai nuovi strumenti di integrazione offerti dal VI Programma Quadro dell'Unione

(Network of Excellence and Integrated Projects) e focalizza sulle necessità, gli strumenti e le opportunità offerte dal collegamento tra formazione e ricerca scientifica e tecnologica nel campo dell' Humanities Computing.

Nel convegno sono stati presentati e discussi vari tipi di strumenti informatici oggi disponibili per l'accesso, la gestione e lo studio del patrimonio culturale, in particolare quella parte del patrimonio a cui si interessano tipicamente i filologi e i linguisti, ovvero il testo nelle sue molte manifestazioni. Sono state considerate le opportunità offerte dalla rete sia per la rappresentazione delle varie espressioni culturali di una società multilingue quale quella europea, sia per il potenziamento dell'educazione linguistica dei cittadini che, come si diceva, costituisce una delle premesse per l'apprezzamento del patrimonio culturale comune dell'Unione. Sono state considerate inoltre, in modo particolare, le potenzialità offerte dell'editoria digitale per la salvaguardia delle lingue minoritarie, ma anche per la diffusione e la tutela delle realizzazioni scientifiche prodotte nelle diverse lingue di cultura tradizionali.

Per facilitare il lettore nella identificazione dei molti argomenti trattati, questi atti, così come era stato il convegno, sono divisi in aree tematiche.

Gli atti si aprono con il lavoro di Allen H. Renear che aveva introdotto i lavori della conferenza con la sua lettura magistrale, *Text from Several Different Perspectives: The Role of Context in Markup Semantics*. Renear propone l'argomento oggi universalmente al centro dei lavori dell'informatica umanistica, ovvero il concetto di *markup* testuale in XML. Il lavoro evidenzia i problemi teorici che sottostanno la pratica del *markup* dal punto di vista della teoria dell'atto linguistico, ovvero il quadro teorico sviluppato per primo da Austin nel suo libro *How to do things with words*.

Il *markup* può essere ambiguo pragmaticamente, per cui Renear nota che il significato reale del *markup* di un documento non è pienamente determinato dalla sua sintassi e della sua semantica, ma che un ruolo fondamentale è giocato dalla forza illocutiva (il tipo di atto compiuto attraverso l'operazione di *markup*). In altri termini il *markup* ha una forza illocutoria che varia con il contesto. L'ambiguità relativa ai diversi valori illocutori porta a vari problemi che sono stati notati nella pratica di annotazione del testo e possono essere risolti in modo unitario prendendo in considerazione la teoria dell'atto linguistico.

Nella prima sezione si discutono specificamente le applicazioni informatiche e le strategie in uso nel campo della filologia per lo sfruttamento delle potenzialità della cultura digitale. Francisco A. Marcos-Marín (Universidad Autónoma de Madrid e Università di Roma "La Sapienza") nel contributo *Diacronía y reconstrucción computacional* propone una sintesi delle principali questioni ed esperienze realizzate riguardo ai metodi computazionali per la ricostruzione linguistica, sia nella Linguistica Spagnola sia nella Filologia Classica.

Simona Casciano (Università di Roma "Tor Vergata") nel lavoro *Esperimento per un'edizione diplomatico-interpretativa computerizzata dei Ricordi di Francesco Guicciardini* propone l'applicazione delle tecniche computazionali e del linguaggio di *markup* XML ad un classico problema di filologia italiana: L'edizione dei ricordi di Francesco

Guicciardini. L'articolo propone una edizione elettronica basata sul linguaggio XML che permette l'accesso simultaneo ai manoscritti e alle edizioni esistenti. L'edizione elettronica consente sia uno studio del testo da più prospettive sia la formulazione di ipotesi ricostruttive delle sorgenti manoscritte perdute.

Domenico Fiorimonte (Università di Roma III) e Cinzia Pusceddu (Università di Edinburgo) nel loro lavoro *Tempo e genesi del testo. Un esperimento in rete*, ripercorrono le radici storiche e lo sviluppo del nuovo paradigma testuale che sottostà alle attuali forme digitali di scrittura (che può essere fatto risalire ai movimenti letterari degli anni '30). L'articolo presenta un esperimento *on line* di rappresentazione digitale della genesi testuale di alcune poesie dello scrittore italiano Valerio Magrelli.

Daniele Silvi (Università degli Studi di Roma Tor Vergata) *Lo stato dell'analisi testuale informatizzata: riflessioni a partire da un esperimento* passa in rassegna le principali tecnologie esistenti per l'analisi del testo e presenta l'utilizzo di uno dei principali standard *de facto* per lo studio delle varianti testuali, ovvero il programma TACT.

Nella seconda sezione sono trattati vari tipi di sistemi che facilitano la gestione e la localizzazione e la diffusione dell'informazione contenuta nel patrimonio culturale.

Uno dei temi più attuali della tecnologia digitale è la raccolta di grandi basi di dati multilingui dedicate allo sviluppo di tecnologie del linguaggio che facilitino il multilinguismo. C-ORAL-ROM è una risorsa multilingue di parlato spontaneo per le principali lingue romanze, che raccoglie quattro *corpora* comparabili di parlato spontaneo Italiano, Francese, Spagnolo e Portoghese. Tale risorsa è stata prodotta in un progetto IST coordinato dall'Università di Firenze in stretto contatto con istituzioni che partecipano a CLiP, come l'Istituto Cervantes e la Universidad Autónoma de Madrid. Il lavoro *The multilingual corpus of spontaneous speech C-ORAL-ROM: validation of the prosodic annotation by expert transcribers* proposto da Massimo Moneglia, Antonietta Scarano e Mario Spinu (Università di Firenze), presenta i risultati del progetto. Il lavoro mostra in particolare che l'evidenza prosodica può essere assunta a livello cross-linguistico come parametro per la sincronizzazione dei dati linguistici nelle basi dati multimediali di parlato.

Le tendenze attuali verso il multilinguismo lasciano oggi la loro impronta in diversi ambiti delle nostre culture: nel web, nella letteratura, nella traduzione letteraria e nei dizionari. Il contributo di María Carlota Nicolás Martínez (Università degli Studi di Firenze) *Plurilingüismo, traducción y diccionarios* presenta alcuni risultati specifici raccolti in questi ambiti che mettono in rilievo la relazione tra due lingue: lo spagnolo e l'italiano.

Tra le strategie più promettenti per la gestione, la localizzazione e la diffusione del patrimonio culturale dell'Unione *l'e-publishing* ha certamente un ruolo centrale. *L'e-publishing* sta in questo momento aumentando la propria rilevanza come sistema di pubblicazione che consente sia una garanzia di serietà scientifica sia una facile diffusione, in particolare per le pubblicazioni appartenenti al dominio della ricerca accademica. Il lavoro di Patrizia Cotoneschi and Antonella Farsetti (Firenze University Press) *L'editoria digitale e della Firenze University Press* presenta le esperienze realizzate presso l'Università di Firenze nel campo dell'*e-publishing* e nell'area *books on demand*. Il lavoro mette in

rilievo sia i vantaggi degli archivi di metadati in XML (in via di realizzazione presso la FUP), sia i vantaggi offerti all'utente dai sistemi di recupero dell'informazione *full text*, che consente la ricerca per lemma e le liste di frequenza nei testi in archivio.

L'articolo *Text Understanding from discourse model and inferential process* di Rodolfo Delmonte (Università "Ca' Foscari"), presenta invece una tecnologia del linguaggio sofisticata che può trovare applicazione come tecnologia abilitante in molti problemi di trattamento automatico del linguaggio (sommario, estrazione di parole chiave, archiviazione automatica, ecc.). Il *Text understanding* è un processo basato su regole linguistiche che attiva, nei casi strettamente necessari, le conoscenze del mondo necessarie per specificare formalmente il contenuto del testo. Tale conoscenza è stimolata dall'attivazione di processi inferenziali. Il funzionamento di questa tecnologia del linguaggio sviluppata all'Università di Venezia è illustrato in dettaglio e provato sulla base di casi test (storie per bambini e testi di giornali).

I contributi di Elisabeth Burr e Luisa Carrer nella sezione seguente sono dedicati alle possibilità offerte dal web per lo sviluppo di una società multilingue.

Elisabeth Burr (Università di Bremen), nel lavoro *Multilingualism and the Web*, mette in rilievo il fatto che la rete è divenuta un ambiente multilingue che favorisce il multilinguismo fornendo uno spazio per l'effettivo utilizzo di lingue minoritarie e dialetti locali. Burr evidenzia che il processo che porta dal monolinguisimo al multilinguismo, fino all'ibridazione, è una caratteristica importante della cultura digitale.

Luisa Carrer (University of Edinburgh) nel contributo *Multiculturalità e rete: voci migranti in Italia* illustra le potenzialità del web, specificamente per la comunicazione pluriculturale dei migranti in Italia e per la loro educazione. Attraverso lo spoglio delle principali risorse *on-line* disponibili ne evidenzia l'attuale utilizzo, in particolare per l'informazione, la comunicazione letteraria e la didattica dell'italiano come L2.

Nel VI Programma Quadro dell'Unione Europea è stato lanciato un importante programma di *e-learning*. Tale programma riguarda in modo particolare l'apprendimento delle lingue intese come preparazione ad un uso attivo del linguaggio nella comunicazione e nell'interazione sociale. Gli atti del convegno si concludono con una serie di interventi dedicati in particolare all'utilizzo degli strumenti di *e-learning* per l'insegnamento della seconda lingua, con particolare riguardo all'acquisizione dell'abilità di scrittura.

Pedro P. Sánchez-Villalón, Manuel Ortega, Asunción Sánchez-Villalón (Universidad de Castilla La Mancha) nel loro lavoro *Evolving to Ubiquitous eLearning Environments* propongono uno scenario tecnologicamente avanzato per l'acquisizione del linguaggio nel quale un ambiente ricco e motivante, in associazione con tecnologie di *e-learning*, permette di praticare la scrittura dell'inglese da parte di apprendenti.

L'acquisizione della capacità di scrittura è essenziale per gli studenti nell'Europa contemporanea e in un contesto mondiale globalizzato, ma contrasta con la tendenza generale a privilegiare la comunicazione orale e per immagini (Tosques). L'uso delle tecnologie dell'informazione per sviluppare le abilità linguistiche di scrittura è anche l'argomento del contributo di Tosques (Università degli Studi di Firenze) *About Writing in English for the Workplace*. Tosques presenta un *corpus* di materiali per l'acquisizione

assistita dell'abilità di scrivere in inglese. I materiali multimediali interattivi sembrano essere in effetti un veicolo privilegiato per l'acquisizione della capacità di scrittura.

Anna Bianchi (Istituto Tecnologie Didattiche del C.N.R., Genova) nel suo lavoro *A presentation of a teaching plan for teachers of Italian as a foreign language, implemented within a didactic archive on the net, inside the project "Netform2"* presenta un progetto sviluppato in una collaborazione tra una scuola secondaria e l'Istituto Tecnologie Didattiche del C.N.R. di Genova. Il progetto vuol creare e testare una griglia finalizzata a mettere in atto qualsiasi tipo di piano di insegnamento. La griglia è valida per diversi argomenti e per più livelli scolastici, tra i quali il lavoro mette in evidenza l'insegnamento dell'italiano agli stranieri. Il progetto intende creare un archivio on line che sarà disponibile agli insegnanti.

Nell'ultimo intervento viene messo in rilievo l'utilizzo della rete come fonte di contenuti validi per l'insegnamento della seconda lingua. Tutti i giornali in lingua spagnola sono attualmente presenti real time in rete. Rosalie Sitman (Università di Tel Aviv) nel suo lavoro *Reflexiones en torno al uso de la prensa en linea para la ensenya de lenguas desde la perspectiva de E/LE* nota che questo fatto, ha incoraggiato gli insegnanti di lingua in tutto il mondo ad usare tale sorgente per l'insegnamento dello spagnolo. Considerando il ruolo della stampa nella società l'articolo propone l'esempio prototipico dell'insegnamento dello spagnolo, che può essere significativo in generale per l'insegnamento della lingua seconda.

In conclusione ci sia concessa una precisazione sulla forma linguistica multilingue adottata in questi atti. La conferenza CLiP, in corrispondenza agli argomenti e alle tradizioni culturali dei contenuti trattati, accetta contributi in tutte le lingue dell'Unione, ma tradizionalmente le lingue del CLiP sono l'inglese, lo spagnolo e l'italiano. Conseguentemente questi atti appaiono in forma multilingue e ospitano contributi scritti sia in italiano sia in inglese sia in spagnolo. La parte introduttiva del volume è replicata in ognuna di queste lingue. Al fine di permettere una maggiore diffusione dei risultati ogni contributo è preceduto da un *abstract* e da parole chiave in lingua inglese.





# Introducción

*Massimo Moneglia*

Università degli Studi di Firenze  
(moneglia@unifi.it)

*María Carlota Nicolás Martínez*

Università degli Studi di Firenze  
(carlota.nicolas@unifi.it)

## 1. AGRADECIMIENTOS E INFORMACIONES

CLiP (*Computers Literature and Philology*) es uno de los foros de discusión con mayor renombre en círculos especializados debido a su nivel científico y a su valor supranacional. El CLiP se convoca todos los años desde 1998 en una universidad europea. Participan en este foro de discusión investigadores de países europeos que trabajan en el campo de la Informática humanística, y también estudiosos de países no europeos, particularmente americanos. El Comité Científico está constituido por:

Lou Burnard (Oxford University)

Elisabeth Burr (Universität Bremen)

Domenico Fiormonte (Università di Roma III)

Francisco Marcos-Marín (Universidad Autónoma de Madrid / Università di Roma “La Sapienza”)

María Carlota Nicolás Martínez (Università di Firenze)

Concha Sanz Miguel (Universidad de Castilla-La Mancha)

El primer encuentro, en Edimburgo 1998, surgió por iniciativa de Domenico Fiormonte y Jonathan Usher a los cuales se unieron posteriormente Lou Burnard, Oxford University; Elisabeth Burr, Universität Bremen; Francisco Marcos-Marín, Universidad Autónoma de Madrid y Antonio Zampolli, Università degli Studi di Pisa; conscientes de la validez de las nuevas tecnologías para la Filología y el resto de las ciencias humanísticas, y de la importancia de sus contribuciones para alcanzar una cultura digital general y para la construcción de Internet. Los sucesivos encuentros respetan este planteamiento inicial adaptándose a la evolución del ambiente cultural y tecnológico europeo y se han realizado en las siguientes sedes:

1999 - Roma, Università di Roma “La Sapienza”

2000 - Alicante, La Universidad de Alicante y el Instituto Cervantes  
<<http://www.cervantesvirtual.com/CLiP2000/index.htm>>

2001 - Duisburg, Gerhard Mercator University

<<http://www.uni-duisburg.de/FB3/CLiP2001/index.htm>>

2002 - Albacete, Gestión del Conocimiento y Nuevas Tecnologías Universidad de Castilla La Mancha  
<<http://www.uclm.es/gcynt/clip2002/index.htm>>

Las actas de las precedentes conferencias CLiP han sido publicadas en:

Fiormonte, Domenico / Usher, Jonathan (eds.) (2001): *New Media and the Humanities: Research and Applications. Proceedings of the first seminar Computers, Literature and Philology*, Edinburgh 7-9 September 1998. Oxford: Humanities Computing Unit, University of Oxford.

<<http://www.hcu.ox.ac.uk/publications/clip.html>>

Fiormonte, Domenico (a cura di) (2003): *Informatica umanistica: dalla ricerca all'insegnamento. Atti dei convegni CLiP 1999 (Roma "La Sapienza") e 2000 (Università de Alicante)*. Roma: Bulzoni.

<[http://www.selc.ed.ac.uk/italian/digitalvariants/events/Clip99\\_2000.htm](http://www.selc.ed.ac.uk/italian/digitalvariants/events/Clip99_2000.htm)>

Proceedings of the CLiP conference N° 3 Albacete, 2002, Monograph in *La Musa Digital*.

<[http://www.uclm.es/lamusa/lista\\_articulos\\_mono.asp?numero=3&lengua=en](http://www.uclm.es/lamusa/lista_articulos_mono.asp?numero=3&lengua=en)>

Clip 2003 se ha celebrado como parte de los actos conmemorativos de los 80 años de la Università degli Studi di Firenze bajo el patrocinio del Presidente de la Repubblica Italiana Carlo Azelio Ciampi y ha recibido el patrocinio y el apoyo de la Regione Toscana; el CSIAF (Centro Servizi Informatici dell'Ateneo Fiorentino); la FUP (Firenze University Press); la Facoltà di Lettere, Corso di Laurea Specialistica in Informatica per le discipline Umanistiche, Dipartimento di Lingue e Letterature Neolatine y Dipartimento di Italianistica (Laboratorio Linguistico-LABLITA).

El Congreso ha sido coordinado por María Carlota Nicolás Martínez y ha tenido lugar en el Dipartimento di Italianistica de la Università degli Studi di Firenze los días 4 y 5 de diciembre de 2003.

Organización:

María Carlota Nicolás Martínez (Dipartimento di Lingue e Letterature Neolatine)

Anna Dolfi (Dipartimento di Italianistica)

Giovanna Angeli (Dipartimento di Lingue e Letterature Neolatine)

Emanuela Cresti (Dipartimento di Italianistica)

Massimo Moneglia (Dipartimento di Italianistica)

Eugenio Picchi (ILC-CNR, Pisa)

Antes de la conclusión de CLiP 2003 ha tenido lugar un Seminario propuesto y financiado por la *Association for Literacy and Linguistic Computing (ALLC)* dedicado a los lenguajes de anotación y a los estándar de codificación de los textos TEI. Al Seminario, dirigido por Harold Short, Director del *Centre for Computing in the Humanities*, Kings College de Londres, han tomado parte jóvenes del Corso di Laurea Specialistica in Informatica per le Discipline Umanistiche de la Università di Firenze y muchos de los

participantes al Congreso. Queremos mostrar nuestro reconocimiento y agradecimiento a Harold Short, Alejandro Bia y Elena Pierazzo, que han dirigido el Seminario.

En nombre del Comité Científico de CLiP queremos dar las gracias al Comité Organizador de los *80 anni del Ateneo Fiorentino*, personalmente al Prof. Sandro Rogari, por haber incluido nuestro Congreso entre los eventos de las celebraciones. Un agradecimiento especial al Dipartimento di Italianistica y a su director Prof. Anna Dolfi, por su hospitalidad y un afectuoso gracias a los colaboradores del Laboratorio LABLITA por haber contribuido a la organización logística e informática del seminario. Agradecemos también a Ida Tucci el enmaquetado de estas Actas.

En nombre de todos los participantes al Congreso queremos expresar, una vez más, nuestro unánime pésame por la reciente desaparición de Antonio Zampolli, la personalidad que más que ninguna otra ha contribuido al nacimiento y al desarrollo de la Informática Humanística en Europa y que ha honrado los Seminarios CLiP formando parte de su comité científico.

## 2. CLiP 2003

El objetivo general de los Seminarios Internacionales *Computers, Literature and Philology* (CLIP) es la integración de las materias humanísticas y las tecnologías de la información y de la comunicación (TIC). *Literature* y *Philology* deben entenderse en un sentido amplio. *Literature* abarca textos de todo tipo (hablados, escritos, hipertextos) en los que también pueden incluirse imágenes, sonidos, gráficos, etc., mientras que la *Philology* abarca el análisis científico de estos textos desde diversas perspectivas (lingüística, cultural, de crítica textual, etc.).

El Congreso CLiP 2003 incluye temas relacionados con el plurilingüismo en ámbito humanista; su objetivo principal es el de reunir a estudiosos y operadores que trabajan en la informática humanista en relación con el tema: *La gestión unitaria de la herencia cultural multilingüe europea y su difusión en la red*.

En el CLiP 2003 se discute específicamente sobre proyectos vinculados al ámbito del plurilingüismo en relación con la Unión Europea y está interesado sobre todo en la colaboración entre enseñanza e investigación.

El plurilingüismo de la Unión y la educación de los ciudadanos europeos en una cultura plurilingüe son dos premisas básicas para la gestión y para la utilización de la herencia cultural compartida.

“Los recursos culturales y científicos europeos son un valor público único que forma y desarrolla la memoria de nuestras diversas sociedades y asegura una base sólida para el desarrollo de nuestra industria de contenidos digitales en una sociedad de conocimiento sostenible” (Conclusiones del Expert-meeting de los Estados Miembros de la UE, Lund, abril de 2001).

Los trabajos actuales en el sector de la herencia cultural comparten las afirmaciones expuestas en Lund, pero para que los múltiples recursos culturales nacionales

puedan ser reconocidos como un bien común se plantea el problema de la naturaleza plurilingüe de la herencia cultural europea y de la pertinaz carencia de sistemas capaces de garantizar la posibilidad de acceso plurilingüe, a través de la red, a todos los ciudadanos de la Unión.

En el CLiP 2003 se trata del nexo entre plurilingüismo y herencia cultural en relación con los nuevos instrumentos de integración ofrecidos en el VI Programa Marco de la Unión Europea (Network of Excellence and Integrated Projects). El CLiP 2003 se centra en la necesidad, en los instrumentos y en las oportunidades que ofrecen la relación entre formación e investigación científica y tecnológica en el campo de la Informática Humanística.

Durante el Congreso se han presentado y se ha debatido sobre varios tipos de instrumentos informáticos actualmente en uso para la gestión y el estudio del patrimonio cultural, en especial todo lo relacionado con el patrimonio filológico y lingüístico, es decir, el texto visto desde varias perspectivas. Se ha tratado de las oportunidades que ofrece la red tanto para la representación de las múltiples expresiones culturales de una sociedad plurilingüe europea, como para potenciar la educación lingüística de los ciudadanos, que constituye, como se ha dicho, una de las premisas para la valorización del patrimonio cultural compartido de la Unión. Además se han mostrado, en concreto, la potencialidad que ofrece la edición digital para salvaguardar las lenguas minoritarias, y también para la difusión y la tutela de las obras científicas producidas en las varias lenguas de cultura tradicionales.

Para que el lector pueda consultar con mayor rapidez los argumentos tratados se ha organizado este volumen en secciones que corresponden con los temas tratados, de la misma manera que se hizo durante el Congreso.

Este volumen empieza con el trabajo de Allen H. Renear que abrió el Congreso con su conferencia: *Text from Several Different Perspectives: The Role of Context in Markup Semantics*. Renear propone un tema que actualmente se encuentra en un lugar central en los trabajos de informática humanística, es decir, el concepto de anotador textual en xml. Su trabajo pone en evidencia los problemas teóricos que subyacen a la práctica concreta de anotación de textos desde el punto de vista de la teoría del acto lingüístico, el cuadro teórico desarrollado por Austin en su libro *How to do things with words*.

La anotación puede ser ambigua pragmáticamente, por lo que Renear nota que el significado real de la anotación de un documento no está totalmente determinado por la sintaxis y por la semántica, sino que el papel fundamental lo juega la fuerza ilocutiva (el tipo de acto realizado mediante la operación de anotación). En otras palabras la anotación tiene una fuerza ilocutiva que varía con el contexto.

La ambigüedad relativa a los varios valores ilocutivos lleva a varios problemas que han surgido en la práctica de las anotaciones del texto y que pueden ser resueltos de manera unitaria teniendo en consideración la teoría del acto lingüístico.

En la primera sesión se han discutido específicamente las aplicaciones informáticas y las estrategias que se utilizan actualmente en el campo de la filología para aprovechar la potencialidad de la cultura digital.

Francisco A. Marcos-Marín (Universidad Autónoma de Madrid y Università di Roma “La Sapienza”) en su intervención *Diacronía y reconstrucción computacional* propone una síntesis de las principales cuestiones y experiencias realizadas respecto a los métodos computacionales para la reconstrucción lingüística, en la Lingüística Española y en la Filología Clásica.

Simona Casciano (Università di Roma “Tor Vergata”) en su trabajo *Esperimento per un'edizione diplomatico-interpretativa computerizzata dei Ricordi di Francesco Guicciardini* propone la aplicación de las técnicas computacionales y del lenguaje de anotación xml en un clásico problema de filología italiana: La edición de los recuerdos de Francesco Guicciardini. El artículo propone una edición electrónica basada en el lenguaje xml que permite el acceso simultáneo a los manuscritos y a las ediciones existentes. La edición electrónica ofrece la posibilidad de hacer tanto un estudio del texto desde varias perspectivas como la formulación de hipótesis de reconstrucción de las fuentes perdidas del manuscrito.

Domenico Fiorimonte (Università di Roma III) y Cinzia Pusceddu (University of Edinburgh) en su trabajo *Tempo e genesi del testo. Un esperimento in rete* hacen un recorrido sobre el texto que va de las raíces históricas al desarrollo del nuevo paradigma textual que es la base de las actuales formas digitales de escritura (que remonta a los movimientos literarios de los años 30). El artículo presenta un experimento en línea de representación digital de la génesis textual de algunas poesías del escritor italiano Valerio Magrelli.

Daniele Silvi (Università degli Studi di Roma “Tor Vergata”) *Lo stato dell'analisi testuale informatizzata: riflessioni a partire da un esperimento* revisa las tecnologías más importantes existentes para el análisis del texto y presenta la utilización de uno de los principales estándares de facto para el estudio de las variantes textuales, es decir, el programa TACT.

En la segunda sección se han tratado varios tipos de sistemas que facilitan la gestión, la localización y la difusión de la información contenida en el patrimonio cultural.

La tecnología digital, actualmente, se afronta con mucha frecuencia la creación de grandes bases de datos plurilingües dedicadas al desarrollo de tecnologías del lenguaje que faciliten el plurilingüismo. C-ORAL-ROM es un recurso plurilingüe de habla espontánea para las principales lenguas románicas, que recoge cuatro corpus comparables de habla: italiano, francés, español y portugués. Este recurso ha sido producido en un proyecto IST coordinado por la Università degli Studi di Firenze en estrecho contacto con instituciones que participan en el CLIP, como el Instituto Cervantes y la Universidad Autónoma de Madrid. El trabajo *The multilingual corpus of spontaneous speech C-ORAL-ROM: validation of the prosodic annotation by expert transcribers* propuesto por Massimo Moneglia, Antonietta Scarano y Mario Spinu (Università degli Studi di Firenze), presenta los resultados del proyecto y muestra concretamente que la evidencia prosódica puede ser asumida desde el punto de vista lingüístico transversal como parámetro para la sincronización de los datos lingüísticos en las bases de datos multimedia de habla.

Las tendencias actuales del plurilingüismo dejan hoy su huella en muchos ámbitos propios de nuestras culturas: en la web, en la literatura, en la traducción literaria y en los diccionarios. El artículo de María Carlota Nicolás Martínez (Università degli Studi di Firenze) *Plurilingüismo, traducción y diccionarios* presenta algunos resultados específicos extraídos de estos ámbitos que resaltan la relación entre dos lenguas: el español y el italiano.

Entre las estrategias con más futuro para la gestión, la localización y la difusión del patrimonio cultural de la Unión tiene sin duda un papel importante la edición electrónica que está en este momento aumentando su propia relevancia como sistema de publicación que permite una garantía de seriedad científica y una fácil difusión, en especial para las publicaciones que pertenecen al dominio de la investigación académica. El trabajo de Patrizia Cotoneschi y Antonella Farsetti (Firenze University Press) *L'editoria digitale e della Firenze University Press* presenta las experiencias obtenidas en la Università degli Studi di Firenze en el campo de la edición electrónica en el campo de *books on demand*. Este trabajo pone de relieve las ventajas de los archivos de metadatos en xml (en proceso de realización en la FUP). Ventajas que ofrecen al usuario la recuperación de la información del texto completo, y le permiten las búsquedas por lema y la creación de listas de frecuencias de los textos archivados.

El artículo *Text Understanding from discourse model and inferential process* de Rodolfo Delmonte (Università di Venezia Ca' Foscari) presenta una tecnología del lenguaje sofisticada cuya aplicación permite resolver muchos problemas de tratamiento automático del lenguaje (realización de sumarios, extracción de palabras claves, creación de archivos automática, etc.). El *Text understanding* es un proceso basado en reglas lingüísticas que activa, cuando es estrictamente necesario, los conocimientos del mundo necesarios para especificar formalmente el contenido del texto. Tal conocimiento se pone en marcha mediante la activación de procesos de inferencia. El funcionamiento de esta tecnología del lenguaje desarrollada en la Università di Venezia es ilustrada detalladamente y probada basándose en un ejemplo test (historias de niños y textos de periódicos).

Los artículos de Elisabeth Burr y Luisa Carrer en la sección sucesiva están dedicados a las posibilidades ofrecidas en la red para el desarrollo de una sociedad plurilingüe.

Elisabeth Burr (Universität Bremen), en su trabajo *Multilingualism and the Web*, pone de relieve el hecho de que la red ha llegado a ser un ambiente plurilingüe que favorece el pluringüismo ofreciendo un espacio para la concreta utilización de lenguas minoritarias y dialectos locales. Burr subraya que el proceso que lleva del monolingüismo al plurilingüismo, hasta llegar a un punto de hibridez, es una característica importante de la cultura digital.

Luisa Carrer (University of Edinburgh) en el artículo *Multiculturalità e rete: voci migranti in Italia* ilustra las potencialidades de la web, en concreto para la comunicación pluricultural de los emigrantes en Italia y para su educación. Mediante la extracción de los principales recursos en uso de la red pone en evidencia su actual utilización, especialmente para la información, la comunicación literaria y la didáctica de la lengua italiana como L2.

El VI Programa Marco de la Unión Europea ha lanzado un importante programa de enseñanza con ordenador, dicho programa se refiere específicamente al aprendizaje de lenguas vistas como preparación para un uso activo del lenguaje en la comunicación y en la interacción social. Las actas de este congreso concluyen con una serie de artículos dedicados a la utilización de los instrumentos de enseñanza asistida con ordenador para el aprendizaje de segundas lenguas, especialmente referidos a la adquisición de la destreza escrita.

Pedro P. Sánchez-Villalón, Manuel Ortega, Asunción Sánchez-Villalón (Universidad de Castilla-La Mancha) en su trabajo *Evolving to Ubiquitous eLearning Environments* proponen un escenario tecnológicamente avanzado para la adquisición del lenguaje, un ambiente rico y motivador asociado con tecnologías de enseñanza con ordenador, que permite la práctica de la escritura del inglés por parte de los dicentes.

La adquisición de la competencia escrita es esencial para los estudiantes en la Europa contemporánea en un contexto mundial globalizado, pero contrasta con la tendencia general que lleva a privilegiar la comunicación oral y mediante imágenes (Tosques). El uso de las tecnologías de la información para el desarrollo de la competencia escrita es también el tema del artículo de Tosques (Università degli Studi di Firenze) *About Writing in English for the Workplace*. Tosques presenta un corpus de materiales para la adquisición asistida de la competencia escrita del inglés. Los materiales multimedia interactivos parecen ser de hecho un vehículo privilegiado para la adquisición de dicha competencia.

Anna Bianchi (Istituto Tecnologie Didattiche del C.N.R., Genova) en su trabajo *A presentation of a teaching plan for teachers of Italian as a foreign language, implemented within a didactic archive on the net, inside the project "Netform2"* presenta un proyecto desarrollado mediante la colaboración de un Instituto de Enseñanza Superior y del Istituto di Tecnologie Didattiche del C.N.R. di Genova. El proyecto pretende crear y comprobar la validez de un esquema único cuya finalidad es poner en marcha cualquier tipo de plan de enseñanza. El cuadro es válido para varios temas y para más de un nivel escolar, en el trabajo se muestra la enseñanza de italiano a extranjeros. El proyecto pretende crear un archivo en red que estará a disposición de los profesores.

En la última intervención se subraya el uso de la red como fuente de contenidos válidos para la enseñanza de la segunda lengua. Todos los periódicos en lengua española están actualmente en la red. Rosalie Sitman (Universidad de Tel Aviv) en su trabajo *Reflexiones en torno al uso de la prensa en línea para la enseñanza de lenguas desde la perspectiva de E/LE* nota que este hecho ha animado a los profesores de lengua de todo el mundo a usar esta fuente de información para la enseñanza del español. Considerando el papel de la prensa en la sociedad, el artículo propone como ejemplo la enseñanza del español, que puede ser significativo en general para la enseñanza de la segunda lengua.

Para concluir queremos precisar algo sobre la forma lingüística adoptada en estas Actas. El Seminario CLiP, en sintonía con los temas y las tradiciones culturales de los contenidos tratados, acepta artículos en todas las lenguas de la Unión, pero tradicionalmente las lenguas del CLiP son el inglés, el español y el italiano. Como consecuencia estas actas aparecen en una forma multilingüe puesto que en ellas se recogen



artículos escritos en italiano, en inglés y en español. La parte introductiva de este volumen se presente en cada una de estas lenguas. Para que haya una mayor difusión de los resultados cada uno de los artículos empieza con un resumen e incluye palabras claves en lengua inglesa.

# Introduction

*Massimo Moneglia*

Università degli Studi di Firenze  
(moneglia@unifi.it)

*María Carlota Nicolás Martínez*

Università degli Studi di Firenze  
(carlota.nicolas@unifi.it)

## 1. ACKNOWLEDGE

The CLiP seminars are considered by specialists to be a very prestigious forum for discussion of outstanding scientific quality and supra-regional importance. To the congress participate scholars working in the field of Humanities Computing from all Europe but also from non-European countries, in particular, Americans. To the Scientific Committee participate the following scholars:

Lou Burnard (Oxford University)

Elisabeth Burr (Universität Bremen)

Domenico Fiormonte (Università di Roma III)

Francisco Marcos-Marín (Universidad Autónoma de Madrid and Università di Roma “La Sapienza”)

María Carlota Nicolás Martínez (Università degli Studi di Firenze)

Concha Sanz Miguel (Universidad de Castilla-La Mancha)

The first meeting, Edinburgh 1998, was organized under the initiative of Domenico Fiormonte and Jonathan Usher later joined by Lou Burnard, Oxford University, Elisabeth Burr, Universität Bremen; Francisco Marcos-Marin, Universidad Autónoma de Madrid and Antonio Zampolli, Università degli Studi di Pisa. These scholars were aware of the significance of new technology in philology and other humanistic sciences and the importance of utilizing humanistic content in a digital form and via the Internet. The following meetings are along the same line concerning European cultural and technological frame and have been held in the following locations:

1999 - Roma, Università di Roma “La Sapienza”.

2000 - Alicante, Universidad de Alicante and Instituto Cervantes  
<<http://www.cervantesvirtual.com/CLiP2000/index.htm>>

2001 - Duisburg, Gerhard Mercator University  
<<http://www.uni-duisburg.de/FB3/CLiP2001/index.htm>>

2002 - Albacete, Gestión del Conocimiento y Nuevas Tecnologías. Universidad de Castilla-La Mancha  
<<http://www.uclm.es/gcynt/clip2002/index.htm>>

The proceedings of the congress are available in:

Fiormonte, Domenico / Usher, Jonathan (eds.) (2001): *New Media and the Humanities: Research and Applications. Proceedings of the first seminar Computers, Literature and Philology*, Edinburgh 7-9 September 1998. Oxford: Humanities Computing Unit, University of Oxford.

<<http://www.hcu.ox.ac.uk/publications/clip.html>>

Fiormonte, Domenico (ed.) (2003): *Informatica umanistica: dalla ricerca all'insegnamento. Atti dei convegni CLiP 1999 (Roma "La Sapienza") e 2000 (Universidad de Alicante)*. Roma: Bulzoni.

<[http://www.selc.ed.ac.uk/italian/digitalvariants/events/Clip99\\_2000.htm](http://www.selc.ed.ac.uk/italian/digitalvariants/events/Clip99_2000.htm)>

Proceedings of the CLiP congress N° 3 Albacete, 2002, Monograph in *La Musa Digital*.

<[http://www.uclm.es/lamusa/lista\\_articulos\\_mono.asp?numero=3&lengua=e](http://www.uclm.es/lamusa/lista_articulos_mono.asp?numero=3&lengua=e)>

CLiP 2003 took place amidst the *80th anniversary celebrations of the University of Florence*, which was held under the auspices of the President of the Italian Republic Carlo Azelio Ciampi and received the support and encouragement of the Region of Tuscany, CSIAF (the Computer Service Centre of the University of Florence), FUP (Firenze University Press); the Literature Faculty, the Specialization Degree Program in Computer Science for the Humanities disciplines, the Romance Language and Literature Department, and the Italian Department (Language Laboratory-LABLITA).

The seminar was coordinated by María Carlota Nicolás Martínez and took place in the Italian Department of the University of Florence on December 4th and 5th, 2003.

Organizing Committee:

María Carlota Nicolás Martínez (Dipartimento di Lingue e Letterature Neolatine)

Anna Dolfi (Dipartimento di Italianistica)

Giovanna Angeli (Dipartimento di Lingue e Letterature Neolatine)

Massimo Moneglia (Dipartimento di Italianistica)

Eugenio Picchi (ILC-CNR, Pisa)

The CLiP 2003 activities were followed by a Post-congress teaching workshop offered by the *Association for Literacy and Linguistic Computing (ALLC)* devoted to the mark-up and standard languages of TEI text codes. Participating in the workshop, conducted by Harold Short, Director of the *Centre for Computing in the Humanities*, Kings College of London, were young students of the Specialization Degree Course in *Computer Science for the Humanities Disciplines* of the University of Florence. Harold Short, Alejandro Bia e Elena Pierazzo, who taught the lessons, deserve mention and thanks.

On behalf of the CLiP supervisory committee we wish to give special thanks to the organizing committee for the *80<sup>th</sup> anniversary of the University of Florence*, in the person of Prof. Sandro Rogari, for having welcomed this congress among the other celebrative events. Special thanks go to Prof. Anna Dolfi, for the hospitality given to

the congress and warm thanks to the collaborators of the LABLITA lab for having provided the logistical and computer know-how for the congress. We also thank Ida tucci for the paging of these Acts.

In the name of all the participants in the congress we wish to express yet again the unanimous condolences for the recent decease of Antonio Zampolli, the person who more than any other contributed to the birth and development of computer science applications to humanities studies in Europe and who honoured this congress by taking part in the supervisory committee.

## 2. CLiP 2003

The goal of the international congress on *Computers, Literature and Philology* (CLiP) is the integration of the humanistic disciplines with information technology and communication (TLC). Literature and Philology should be understood in a broad sense; by literature we mean text of all types (spoken, written and Internet), in which the content can also include images, sounds, graphics, etc. Philology considers the scientific analysis of different texts from different perspectives (linguistic, cultural, text analysis, etc.).

CLiP 2003 is centred around themes that are relative to multilingualism in a humanistic content and has a principal objective of bringing together scholars and technicians working in the field of Humanities Computing focusing on the theme of: *The unitary management of the multilingual European cultural heritage and its exploitation on the net.*

CLiP examines the themes of multilingualism in the context of the European Union and is, above all, interested in collaboration between teaching and research.

The multilingual nature of the Union and the education for a multilingual culture of European citizens are two basic premises for developing the common cultural heritage.

“Europe’s cultural and scientific knowledge resources are a unique public asset forming the collective and evolving memory of our diverse societies and providing a solid basis for the development of our digital content industries in a sustainable knowledge society.” (Conclusion of Expert-meeting of States Members of Europe Union, Lund, April 2001).

The present work in the sector of cultural heritage shares the assumptions expressed in the Lund Principals. However, because of diverse national and cultural resources in the European Union, felt to be a common resource, obstacles must be faced due to the nature of a multilingual cultural heritage and the persistent lack of a strong system that guarantees multilingual access to European Union citizens over the Internet.

CLIP 2003 will also examine the connection between multilingualism and cultural heredity in relationship to new tools for integration offered by the European Union’s VI Framework Programme (Network of Excellence and Integrated Projects) and focus

on the needs, the tools and the opportunities offered by connecting training, scientific research and technology in the field of Humanities Computing.

During the congress there were presentations and discussions of the various kinds of computer tools available today for accessing, managing and studying cultural heritage materials, in particular the area which is of special interest to philologists and linguists, namely texts in their multiple manifestations. Consideration was given to the opportunities offered by the web both for representing the various cultural expressions of a multilingual society like Europe's and for strengthening language learning among its citizens, who, as it was affirmed, constitute one of the premises for an appreciation of the cultural heritage of the European Union. Special consideration was also given to the potential offered by digital publishing for safeguarding the minor languages, as well as for broadcasting and protecting the cultural achievements produced among the various traditional languages of culture.

The proceedings open with the paper by Allen H. Renear, *Text from Several Different Perspectives: The Role of Context in Markup Semantics*, that also opened the congress in plenary. The paper deals with the main theoretical and practical issue that humanities computing is universally facing in the present days, that is text mark up in XML.

The paper deals with important theoretical problems underlining the practices of mark up from the point of view of speech act theory, that is the framework of language studies early developed in Austin's *How to do things with words*. Crucially markup can be pragmatically ambiguous and therefore the full meaning of document markup is not fully determined by syntax and semantics, but pragmatics and, specifically illocutionary force play a role as well (the act accomplished through the mark up). Renear argues that markup has an illocutionary force, which can vary with context. He shows how mark up ambiguity due to the different values is responsible for several puzzles that can be solved in a unified manner taking into account speech act theory.

In the first section the tools and strategies used in Philology to exploit the potential of digital culture are presented and discussed Francisco A. Marcos-Marín (Universidad Autónoma de Madrid and Università di Roma "La Sapienza") in the paper *Diacronía y reconstrucción computacional* makes a synthesis of the main issues and experiences concerning the use of computational methods for language reconstruction within Spanish Linguistics and Classic Philology

Simona Casciano (Università di Roma "Tor Vergata") with her *Esperimento per un'edizione diplomatico-interpretativa computerizzata dei Ricordi di Francesco Guicciardini* propose the application of computational techniques and XML mark-up language to an important philological problem of Italian philology (The *Ricordi* by Francesco Guicciardini). The paper offers an XML based electronic edition which allows a study of the text from several viewpoints regarding simultaneous access to existing manuscripts and editions. The edition offers a multi-perspective study of the text and facilitates reconstruction hypothesis.

Domenico Fiormonte (Università di Roma III) and Cinzia Pusceddu (University of Edinburgh) in their paper *Tempo e genesi del testo. Un esperimento in rete*, investi-

gate the historical roots and development of the new textual paradigm embodied in present digital forms of writing (that can be traced back to literary movements of the 1930s). They present an online experiment for the representation of the textual genesis of several poems of the Italian writer Valerio Magrelli.

Daniele Silvi (Università degli Studi di Roma "Tor Vergata") *Lo stato dell'analisi testuale informatizzata: riflessioni a partire da un esperimento* summarizes the state of the art of present technologies for textual analysis and presents the use of one of the main de facto standard in the field (the TACT program) for the study of textual variants.

One of the main topics of digital technologies is presently the collection of large multilingual language repositories. This kind of cultural heritage digital archives provides essential data for research in linguistics, human language technologies and human language technologies dedicated to facilitate multilingualism. C-ORAL-ROM is a multilingual corpus of spontaneous speech for the main romance languages gathering four comparable collections of Italian, French, Portuguese and Spanish spoken texts. C-ORAL-ROM have been delivered by and IST project co-ordinated by the University of Florence in strict connection with institutions that are involved in the CLiP Seminar like Instituto Cervantes and Universidad Autónoma de Madrid.

The paper presented by Massimo Moneglia, Antonietta Scarano e Marius Spinu (Università degli Studi di Firenze) *The multilingual corpus of spontaneous speech C-ORAL-ROM: validation of the prosodic annotation by expert transcribers* present the main results of the Project. More specifically, shows that the prosodic evidence is a cross-linguistic parameter that allows a significant synchronization of the linguistic data of spoken language data in multimedia repositories.

Present day multilingual tendencies are testified at many levels: in the web, in traditional literature, in literary translation. The contribution of María Carlota Nicolás Martínez (Università degli Studi di Firenze) *Plurilingüismo, traducción y diccionarios* present some finding about the relation between Spanish and Italian languages.

The present trend toward multilingualism leave their mark today in several areas of our cultures: on the web, in literature, in literary translation and in dictionaries. The contribution of María Carlota Nicolás Martínez (Università degli Studi di Firenze) *Plurilingüismo, traducción y diccionarios* presents certain specific results gathered in these areas which highlights the relationship between two languages: Spanish and Italian.

Among the more promising strategies for management, localization and diffusion of the cultural heritage in the union a specific role is played by e-publishing is now increasing its importance as an easy and authoritative publication system suitable for research publication in the all academic domains. The paper by Patrizia Cotoneschi and Antonella Farsetti (Firenze University Press) *L'editoria digitale della Firenze University Press* presents the experience currently developed at the University of Florence in e-publishing and book on demand area. The paper focus on the xml based metadata structure of the archive, which is presently under construction, and on the facilities offered to the reader for the localization of content. to a full text retrieval system that allows the research of lemma and frequency lists within the archive.

The article *Text Understanding from discourse model and inferential process* by Rodolfo Delmonte (Università di Venezia “Ca’ Foscari”) present a sophisticated language technology that can be applications as enabling technology for many issues concerning the automatic treatment of language data (summarization, keyword extraction, etc.)

*Text Understanding* is a linguistically-based process which activates the Knowledge of the World needed for specifying formally the content of a text. The knowledge of the world is activated, only when needed, by triggering inferential processes. The functioning of this language technology, developed at the University of Venice is explained in detail and proved on test cases (children story and texts taken from newspapers).

The contributions of Elisabeth Burr and Luisa Carrer in the subsequent section are dedicated to the possibilities offered by the web for the development of a multilingual society

Elisabeth Burr (Universität Bremen) in the paper *Multilingualism and the Web* highlights that the web has become a multilingual environment that favours multilingualism. The web provides space for minority languages and local dialects. She argues that the process from monolingualism to multilingualism and to hybridity is an important ingredient of digital culture.

Luisa Carrer (University of Edinburgh) in her paper *Multiculturalità e rete: voci migranti in Italia* consider the potential of e-publishing for both multi-cultural communication and educational application. Present on line resources are recorded and the main fields of application are highlighted: news; multi-cultural representation of migration, literary expression, educational field, specifically for teaching Italian.

In The VIth framework program the European Union has launched an eLearning Programme which regards language learning, intended as a preparation for the active use of the language for communication, as social interaction. The proceedings of the congress are concluded with a set of papers that are specifically dedicated to e-learning experiences for second language acquisition and that focus on the acquisition of writing skills

Pedro P. Sánchez-Villalón, Manuel Ortega, Asunción Sánchez-Villalón (Universidad de Castilla-La Mancha) in their paper *Evolving to Ubiquitous eLearning Environments* propose a Technology-Enhanced Language Learning Scenario, in which a reach and motivating interactions environment, in association with e-learning technologies, enable the practice of writing English as a Foreign Language

The acquisition of written skills is essential to students in today’s Europe and in globalized professional environment, but contrast with the general tendency to assign a privilege to spoken and image communication (Tosques). The use of information technologies to enhance writing ability in foreign languages is also the topic of the contribution by Edward Tosques (Università degli Studi di Firenze) *About Writing in English for the Workplace*. Tosques presents a corpus of computer-assisted learning materials for the acquisition of writing in English. He claims that interactive multimedia tools appear a particularly appropriate vehicle for the acquisition of writing skills.

Anna Bianchi (Istituto Tecnologie Didattiche del C.N.R., Genova) in the paper *A presentation of a teaching plan for teachers of Italian as a foreign language, implemented within a didactic archive on the net, inside the project “Netform2”* presents a project

accomplished at in Genova in a collaboration between a secondary school and the Istituto Tecnologie Didattiche del C.N.R. di Genova. The project is aimed to create, and to test, a grid suitable for putting in effect whatever kind of teaching plans, in different subjects and for different school levels. The paper focus focus on teaching Italian to foreign pupils.

In the last contribution the use of the net as a source for contents to be use for second language acquisition goals is highlighted. All news papers issued in Spanish are now real time available in internet. Rosalie Sitman (Universidad de Tel Aviv) in the paper *Reflexiones en torno al uso de la prensa en línea para la enseñanza de lenguas desde la perspectiva de E/LE* notices that this fact have encouraged language teachers all around the world to use o news papers for teaching Spanish. Starting form the role played by press in the society the paper propose the prototypical example E/EL that may turn out significant for teaching second language in general.

To conclude, we would like to make a clarification regarding the multilingual form adopted in these acts. The CLiP congress, in accordance with the cultural topics and traditions of the contents treated, accepts contributions in all the languages of the Union, but traditionally the languages favoured by CLiP are English, Spanish and Italian. Consequently, these acts appear in multilingual form and welcome contributions written in Italian, English or Spanish. The introductory part of the volume is repeated in each of these languages. For the purpose of favouring a wider diffusion of the results each contribution is preceded by an abstract and by key words in English.





# Text from Several Different Perspectives: The Role of Context in Markup Semantics

*Allen H. Renear*

University of Illinois at Urbana-Champaign  
(renear@uiuc.edu)

*Recent work in markup studies suggests a profound role for context in understanding critical semantic distinctions in the use of XML markup. Specifically we argue that markup has illocutionary force, which can vary with context even while the syntax and propositional content of a markup utterance remains constant, and that the development of this insight can generalize and unify existing solutions to previously identified problems as well as contribute to an improved understanding of textual communication in general.*

*Keywords:* speech acts, XML, markup, pragmatics, context

## 1. INTRODUCTION

Defenders of realist approaches to text generally reject the claim that “what text really is” depends upon such things as the context of creation and reception. However recent work in markup studies demonstrates that independently of the realist/anti-realist debate context can be seen to play a profound though poorly understood role in document markup semantics. Specifically, we argue that the full meaning of document markup is determined not just by syntax and (propositional) semantics, but by pragmatics as well, and, specifically, by *illocutionary force*. We then show that it is the *context of use* that establishes illocutionary force: independent of context of use markup is pragmatically ambiguous even when precisely characterized as to syntax and propositional semantics. We believe that this ambiguity is responsible for several puzzles in markup studies. Here we show how the solutions that have been offered for two of these puzzles are in fact each special applications of the general insight that illocutionary force is constitutive of markup meaning and determined by context. These results suggest that a general theory of context and markup pragmatics, exploiting the accomplishments of speech act theory, could provide unifying explanations of various phenomena in textual communication.

## 2. PUZZLES INVOLVING MOOD AMBIGUITY IN MARKUP

In what follows we will see that the traditional distinction between descriptive and procedural markup is flawed; that the distinction conflates two different featu-

res – mood and domain – which in fact can vary independently. Typical markup, identified syntactically or through its narrow prepositional semantics, is actually ambiguous, varying in its full meaning with context of use. An adequate markup taxonomy must, among other things, recognize speech act distinctions in order to represent the full extent of markup meaning, and, in addition, will require an account of how *context* of use functions to disambiguate mood and determine the relevant illocutionary force of the utterance<sup>1</sup>.

### *2.1 The descriptive/procedural distinction*

Early discussions of text markup in the 1970s and 1980s distinguished markup from text and then further categorized markup into two or more kinds (Goldfarb, 1981). These taxonomies were satisfyingly effective in explaining and systematizing various phenomena of text processing (Coombs, et al., 1987) and they played a crucial role in theorizing the campaign to promote the “content object” approach to designing text processing systems (DeRose et al., 1990), as well as the more philosophical claims about text ontology, such as the assertion that text is a structure of “content objects” (Renear et al., 1996).

The most important and influential distinction is between “descriptive” markup and “procedural” markup, first put forward by Charles Goldfarb (1981). On this well-known account descriptive markup describes or identifies a text component or textual feature of some kind, while procedural markup invokes (or commands) a formatting or rendering procedure to produce a rendering effect.

This distinction is arguably at the very heart of markup theory, closely related to the dominant recommendation that effective text processing requires modeling text as a structure of “content objects” and to the further ontological claim, mentioned above, that text *is* a structure of “content objects”. However, the distinction received little criticism or attention until the late 1990s when in a series of presentations and papers between 1998 and 2001 the present author argued that the distinction is flawed and that the remedies require considerable theoretical development of our notions of markup and text (Renear, 2000). In retrospect it is now apparent that the nature of this flaw confirms the importance of context and illocutionary force in understanding markup semantics.

### *2.2 The problem of markup that describes formatting*

In the 1980s markup users within the TEI humanities text encoding community noted an awkwardness with the descriptive/procedural distinction. Consider a transcriber using SGML/XML markup to indicate a renditional feature of a source document, that a title is centered or that a word is in bold. What sort of markup is this, descriptive or procedural?

In common with typical cases of descriptive markup, this sort of markup is describing or identifying rather than invoking or instructing. However there was an almost

universal hesitation about calling this markup “descriptive” – the sort of thing being described is a formatting effect, the sort of thing procedural markup was concerned with. Descriptive markup, after all, is the vehicle for implementing the “content object” view of text, according to which text is composed of things like sections, titles, and paragraphs – not things like typeface. It might be argued (and it was) that “rendered-in-bold” is descriptive markup nonetheless, that subject matter of the description is irrelevant. But this would have the unsettling effect of severing the connection between the descriptive/procedural distinction and the theories of text representation strategies with which descriptive markup is traditionally associated.

Another analysis of the problem is available. We may observe upon closer inspection that while the descriptive and procedural categories are disjoint, they are not exhaustive. Moreover, the reason they fail to partition the markup space is that they are conjunctively composed of two features: mood (whether markup describes something, or invokes/commands something) and domain (what sort of thing being described or commanded), and these are (or at least appear to be for the most part) logically independent. So we can now say that descriptive markup is markup that is in the “indicative” mood (it describes), and has a “logical” domain (it is about text components). Procedural markup on the other hand is in the “imperative” mood (it commands) and has a “renditional” domain (it is about graphical presentation features or formatting).

Markup that means “rendered-in-bold” is thus a third kind of markup, neither descriptive nor procedural. It is in the indicative mood (like descriptive markup, and unlike procedural markup), but with a renditional domain (like procedural markup, and unlike descriptive markup). So while it may be descriptive in the narrow sense of being in the indicative mood, it is not descriptive in the usual, and broadly explanatory, sense – the sense in which descriptive markup is a boolean composite of indicative mood and logical domain.

This analysis also suggests a fourth combinatorial possibility: markup that is imperative and logical – and we may wonder: are there any examples in nature?

### *2.3 Puzzles about markup that creates rather than describes*

Consider again SGML/XML transcription and encoding of primary source documents. The transcriber thinks she sees a title, and marks it as such in her transcription. She is asserting that something is a title. She might be wrong (it is an epigram, not a title) or she might be right (it is indeed a title).

Now consider a different context, the development of an original document by its author. The author desires to begin a section with a title, and adds the words of the title, and the encoding which identifies it as a title. Now it might not be a good title, and the author may be wrong about many things associated with this title, but could he be mistaken, in exactly the same sense the transcriber above might be, about whether or not it is a title? Could what he is at that moment authoring, and intends to be a title, actually be something else entirely (an epigram, abstract, or citation for instance) which he is mistaking for a title?

Most of us would say not. Upon reflection, we can observe that this sort of authorial markup with the logical domain, such as that used to create a title, typically has these two features: (i) it is not, really, a true or false assertion, not, in fact, an assertion at all; and (ii) it is in fact being used to actually *create* or bring into existence the textual component in question, not to report its independent and prior presence.

The first feature suggests that authorial markup of this sort, despite its similarity to true descriptive markup, is not in the indicative mood, and the second suggests a connection with actions and subsequent effects that resonates with our notion of imperatives like commands and instructions. But closer inspection suggests that despite the resonance the author is not commanding or instructing some words to be a title she is making them a title, she is *titling*. So this is not an example of markup in the imperative mood after all; not, that is, an representative of the fourth combinatoric possibility we identified above.

However we now observe that the two features, (i) and (ii) above, which are characteristic of authorial markup are precisely the features that the British philosopher John Austin noted picked out a special class of linguistic utterances, which he called “performatives” (Austin, 1962). These are utterances, such as “promising”, which when used in the first person present tense do not report a fact about which we might be either correct or mistaken (as they would if we were to say “He promised...” or “I promised...”), but instead actually *promise*. That is, an author is, through the linguistic event in question, actually bringing about the existence a promise, and not reporting the existence of one.

#### 2.4 Concluding Remarks

The elaboration of the grammatical notion of mood into theories of speech acts and “illocutionary force” has been widely studied in philosophy and linguistics since Austin. It is clear from the foregoing that these notions will bring much illumination to markup semantics, and we have hinted at some of these interesting applications elsewhere (Sperberg-McQueen et al., 2000; Renear, 2000).

The present point is this: even when we identify a particular bit of markup syntactically and then assign a simple semantics by associating it with the property it attributes and a mechanism for determining the referent of attribution, we still have not identified its full meaning unless we know its illocutionary force - and that can only be determined by context of use.

### 3. PUZZLES OF BIBLIOGRAPHIC ENTITY ASSIGNMENT

#### 3.1 Introduction

In what follows we identify a difficulty in assigning XML documents to an entity class in a particular model of the bibliographic domain. As with the preceding case we will analyze this difficulty as resulting from an ambiguity in markup meaning, where apparently identical markup elements vary in their illocutionary force with context of use.

### 3.2 The FRBR Group 1 Entities

The influential *Functional Requirements for Bibliographic Records* (FRBR), released by the International Federation of Library Associations and Institutions in 1998 is an entity-relationship model of the entities, attributes, and relationships of the domain library cataloging theorists have called the “bibliographic universe”. FRBR recognizes four groups of entities, identifying attributes that characterize these entities and binary relationships that hold between them. We are concerned here with the “Group 1” entities: *work*, *expression*, *manifestation*, *item* - which FRBR defines as follows:

*Work*: “... a distinct intellectual or artistic creation”

*Expression*: “... the intellectual or artistic realization of a work in the form of alphanumeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms” (roughly: text).

*Manifestation*: “the physical embodiment of an expression of a work”. (roughly: an edition)

*Item*: “a single exemplar of a manifestation” (a physical copy).

These notions can be further clarified by considering what properties are proper and essential to each entity. Works are understood to have as proper and essential attributes such things as titles, authors, and genre, but not, for instance, language: translations are different expressions of the same work. Expressions on the other hand do have language as an essential characteristic: French and German translations of the same work are different expressions. Manifestations, but not expressions, have as proper and essential properties such things as typeface and type size, and items have properties such as condition and location.

Named relations hold between adjacent entities and (oversimplifying just a bit) these are one-to-many: a single work may be *realized* by a number of expressions, each of which may be *embodied* in a number of manifestations, and each of those may *exemplified* by many items.

### 3.3 XML documents: conflicting considerations

XML documents are now one of the most common objects found in the contemporary “bibliographic universe”. Informally an XML document is a combination of text and XML markup. More precisely it is a string that matches a particular formal grammar and observes several other constraints (W3C, 1999). To which FRBR Group 1 entity class does an XML document belong? Obviously an XML document is not a FRBR item as it is not a concrete physical object. And almost as obviously it is not a FRBR work, an intellectual creation independent of any particular symbolic realization. That leaves two remaining entity classes: expression and manifestation. As something defined by a formal grammar an XML document would appear to be a straightforward, even exemplary, notational entity (like a text in a natural language), and therefore, a FRBR expression. Consistent with this assignment manifestation-specific features such as typeface and car-

rier material are certainly not properties of an XML document *per se*, even though XML documents may be rendered in a particular typeface and on a particular carrier.

But on the other hand Renear et al. (2003) argue that the markup of an XML document might be understood

“as functioning exactly like rendering events (font shifts, type size changes, vertical and horizontal whitespace, etc.) to effect, expedite, and disambiguate the recognition (whether by humans or computer) of the underlying textual objects. This would make the XML document seem more like a manifestation that embodies an expression; an expression that might in fact be embodied differently – for instance, with orthographically different, but semantically equivalent XML markup, or with traditional layout devices. That an XML document does not appear to have a number of attributes that are associated with manifestations, such as typeface, collation and so on only argues that it is not be a completely determinate manifestation, but rather a class of manifestations, or a part of a manifestation.” (Renear et al., 2003)

We now have conflicting considerations, with one set supporting identifying an XML document as a FRBR expression and the other set supporting identification as FRBR manifestation. Which is correct? Or is this otherwise quite plausible and widely used taxonomy flawed?

### 3.4 *The dual aspect theory*

In response to this puzzle Renear et al argue that the concept of an XML document does not “univocally identify something which is an appropriate candidate for identification with a FRBR entity”.

[...] “a particular XML document can only be assigned to one of the two candidate FRBR entity classes - manifestation or expression – with respect to certain features of its context or intended use. Apart from such a context, XML documents have a “dual aspect”, understood from one perspective an XML document is a manifestation ... understood from another it is itself an expression” (Renear et al., 2003).

Here is the context in which an XML document is a manifestation:

[...] “a work about, say, whales, consists of a body of abstract intentional acts: assertions about whales, questions about whales, warnings about whales, and so on, arranged in a rhetorical structure. When an author writes a book about whales, she creates an expression the meaning of which is this particular structure of intentional acts. She accomplishes this by creating an exemplar of a manifestation [...] that in a reader will effect the apprehension of the expression

that in turn has as its meaning the body of intentional objects which is the work about whales. Generally, such a manifestation will have, as well as the marks indicating linguistic characters, specific features that make the apprehension of the embodied expression reliable and efficient. In familiar printed books, these features are, most importantly, graphic devices such as changes in horizontal and vertical spacing, font shifts, color, and so on.

XML markup may also perform the same function as these graphic devices, making the apprehension and cognitive navigation of an expression efficient and unambiguous. Consider [...] an authorial manuscript about whales, prepared in an XML element set, such as TEI Lite. The author uses the markup to indicate the identities and boundaries of textual components, anticipating that subsequent processing by XML/TEI software will be able to follow these cues and process the (digital) manuscript appropriately. [...] In this case, XML markup for, say, a paragraph, seems to be functioning exactly like extra vertical leading or an initial em space. What is important to note in this scenario is the obvious fact that the author is creating a book about *whales*, and not about paragraphs, titles, lists, and such. That is, the author is not using markup to describe an expression, and in particular not using markup to describe the expression that realizes a work about whales. If she were she would be creating another work, on another subject – a work not about whales, but about an *expression*” (Renear et al., 2003).

On the other hand consider a context in which the same XML document is an expression:

[...] “a scholarly textual editor is preparing, using the TEI XML vocabulary, a new critical text of an important cultural work. [...] The text [of the work being edited] [...] may well be about whales, but our scholarly editor is probably not a whaling specialist, and, in any case, is not herself creating or modifying a work about whales. The editor is rather creating a work about an expression that realizes a work about whales. To do this, the editor creates an expression that realizes the work that is about the expression that realizes the work about whales. The editor’s own creative achievement, we note, is not zoological; it is philological.

Looking at the phenomenology of the situation more closely, we imagine that the editor or transcriber puzzles over the pages of a physical document and, drawing on her erudition, her knowledge of the relevant history, graphic vocabulary, the content and context of the specific textual item, and so on, comes to a conclusion: “<p>”. Here “<p>” itself is a bit of notation that in context is clearly being used to express an assertion about a expression. The XML document in this context is functioning as an expression, but an expression that realizes a work *about an expression*.

To appreciate how the subtlety of this ambiguity has impeded its identification note that the expression created by the scholarly editor is [...] a “second-order”



expression: an expression that realizes a work that is itself about an expression - and, specifically, about the very expression the XML document would embody if interpreted as a manifestation.” (Renear et al., 2003).

In one context, the first order authorial context where a book about whales is being authored, the markup is being used performatively to bring into existence a particular structure of linguistic objects, and is therefore a manifestation level feature of the work (about whales) in question. In the other context, the philological establishment of a text, the markup is being used as make assertions about a text of a work about whales, is an element of the expression which realizes a philological, not zoological, work. Orthographically the XML documents are in each case identical, and no distinctions can be identified by syntactic or (narrow) semantic analysis - only the context of use determines the FRBR entity class to which the document belongs.

#### 4. CONCLUSION

The preceding analysis raises as many problems as it purports to solve and there are undoubtedly errors and confusions. Nevertheless the principal claim seems to be confirmed: the full meaning of document markup is determined not just by syntax and (propositional) semantics, but also by *illocutionary force* – and it is the *context of use* that establishes illocutionary force. The further development of this speech act approach to markup analysis may well not only resolve other problems in markup studies, but contribute to the larger theory of textual communication in general.

#### 5. BIBLIOGRAPHY

- Austin J. L. (1962). *How to do things with words*. Cambridge, Harvard.
- Buzzetti D. (1999). Rappresentazione digitale e modello del testo. In *Atti del Convegno “Il ruolo del modello nella scienza e nel sapere”*. Roma, Accademia Nazionale dei Lincei. (Contributi del Centro Linceo Interdisciplinare “Beniamino Segre”, N. 100) 1.127-161. English translation to appear in *New Literary Criticism*.
- Buzzetti D. (2000). Ambiguità diacritica e Markup: Note sull’edizione critica digitale. In S. Albonico (ed.). *Atti del Seminario di studi “Soluzioni informatiche e telematiche per la filologia”*. Pavia, Università degli Studi di Pavia. (Pubblicazioni telematiche, n. 1).
- Caton P. (1999). Using <TEXT>. In *TEI markup, Abstracts of the Annual Joint Conference of the the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing (ACH/ALLC 1999)* University of Virginia.
- Caton P. (2000). Markup’s current imbalance, *Proceedings of Extreme Markup Languages 2000*. Graphic Communications Association.

- Coombs J. H., Renear A. H., DeRose S. J. (1987). Markup systems and the future of scholarly text processing. *Communications of the Association for Computing Machinery*, 30, pp. 933-947.
- DeRose S. J., Durand D., Mylonas E., Renear A. H. (1990). What is text, really? *Journal of computing in higher education*, 1 (2), pp. 3-26.
- Goldfarb C. F. (1981). A Generalized Approach to Document Markup. *Proceedings of the ACM SIGPLAN--SIGOA Symposium on Text Manipulation*. New York: ACM 1981.
- Huitfeldt C. (1995). Multi-dimensional texts in a one-dimensional medium. *Computers and the humanities* 28 (4-5), pp. 235-241.
- International Federation of Library Associations (IFLA). (1998). *Functional requirements for bibliographic records: final report*. UBCIM Publications-New Series. Vol. 19. München, K.G. Saur.
- Piez W. (2001). Beyond the descriptive vs. procedural distinction. *Markup languages: theory and practice*, 3 (2), pp. 141-172 (32).
- Renear A. H. (1995). Theory and metatheory in the development of text encoding. *The Monist 80:3 Interactive Issue*; <http://hhobel.phl.univie.ac.at/mii/>. 1995.
- Renear A. H., Mylonas E., Durand D. (1996). Refining our notion of what text really is. In N. Ide and S. Hockey (eds.), *Research in humanities Computing 4*. Oxford, Clarendon Press.
- Renear A. H. (2000). The descriptive/procedural distinction is flawed. *Markup languages: theory and practice*, 24, pp. 411-420.
- Renear A. H., Dubin D., Sperberg-McQueen C. M., Huitfeldt C. (2002). Towards a semantics for XML markup. In R. Furuta, J. I. Maletic, E. Munson (eds). *Proceedings of the 2002 ACM Symposium on Document Engineering*, pp. 119-126.
- Renear, A. Phillippe H. C., Lawton P., Dubin D. (2003). An XML document corresponds to which FRBR Group 1 entity? In B. T. Usdin and S. R. Newcomb (eds). *Proceedings of Extreme Markup Languages 2003*. Montreal, Canada.
- Sperberg-McQueen M. (1991). Text in the Electronic Age: Textual Study and Text Encoding, with Examples from Medieval Texts. *Literary and linguistic computing literary and linguistic computing*, 6 (1), pp. 34-46.
- Sperberg-McQueen M. C., Huitfeldt A., Renear H. (1999). Meaning and interpretation of markup. *Proceedings of Extreme Markup Languages 2000*. Graphic Communications Association.
- W3C. (1999). Extensible markup Language (XML) 1.0 (second edition) Recommendation. Published by the W3C, November 1999.

#### 4. NOTES

- <sup>1</sup> This section presents work reported on at Extreme Markup Languages 2000 in Montreal and published in Renear (2000). Earlier versions were presented in May 1998 at the HIT Center at the University of Bergen and in July 1998 at the Oxford University Humanities Computing Unit. Many debts are owed to Dino Buzzetti, Paul Caton, Claus Huitfeldt, and Michael Sperberg-McQueen.

- <sup>2</sup> This section presents work carried out by Allen Renear, Christopher Phillippe, Pat Lawton, and David Dubin (Renear et al., 2003); we thank Michael Sperberg-McQueen, Claus Huitfeldt, Kevin Hawkins, and other members of the GSLIS Electronic Publishing Research Group for insights of various kinds. Any errors or departures from previously reported findings are the author's responsibility alone.

# Diacronía y reconstrucción computacional

*Francisco A. Marcos-Marín*

Universidad Autónoma de Madrid

Università di Roma "La Sapienza"

(francisco.marcos.marin@uam.es)

*The time machine has been considered as impossible for centuries. However, the human brain works constantly backwards, trying to reconstruct primitive layers of the history of Mankind. Linguistics always tried to understand the connection between form and meaning (old Etymologia), or form and language families (historical linguistics and comparatistics). Language reconstruction appeared as an ultimate goal in the feverish minds of linguists of the 19<sup>th</sup> century. All sort of data were used to achieve that goal, the challenge was permanent. It is not surprising, therefore, that Classical scholars welcomed the arrival of the computer as the expected tool created to solve many of the problems enclosed in the corpora of Latin and Greek. New insights were developed, new programs written. This presentation will analyse the main issues concerning computers and language reconstruction: data, patterns, what has been done and what remains to be.*

*Keywords:* Computational linguistics, language reconstruction, historical linguistics

## 1. INFORMACIÓN DE CHOQUE Y BIBLIOGRAFÍA PRELIMINAR

La aplicación de métodos cuantitativos en la investigación humanística suele chocar con la incomprensión, la desconfianza e, incluso, la torpeza de quienes se mueven en las vagas nebulosas de los análisis subjetivos. No hace falta justificar este enfoque, cuando se están desentrañando las claves químicas (sintácticas) de la Genética o se opera en Física con unidades más deducibles que experimentables. Lo que hace falta es incrementar la formación matemática de los lingüistas y exponerlos a planteamientos que son, sencillamente, los que les corresponden. Con objeto de que se tenga en cuenta cómo el autor ha ido acercándose a estos estudios a lo largo de su carrera profesional, se recoge a continuación una selección de ejemplos de su obra de una serie de referencias que están disponibles en el internet, elementales, en comparación con lo que se puede encontrar en los vínculos específicos, pero por eso mismo ilustrativas e introductorias.

1. «Atlas lingüísticos plurilingües con ordenadores electrónicos.» *Bol. CCUM*, 23, 1973, pp. 12-15. (FMM en col. con Manuel Ariza, Ignacio del Campo, Isabel González, María Teresa Molina.)

2. «Un nuevo método lexicográfico: el ‘Diccionario del Español de México’». *Mundo Hispánico* 334, Enero de 1976, pp. 70-72.
3. «Apócope y leísmo en el castellano medieval. Aproximación cuantitativa.» *Estudios Ofrecidos a Emilio Alarcos Llorach*, Oviedo, Universidad, 1977, I, pp. 145-156.
4. «El orden de palabras en un texto latino altomedieval, o los peligros de la estadística», *RSEL*, 15, 1985, 177-179.
5. «UNITE: conjunto de programas para el tratamiento filológico de textos en verso», *Procesamiento del Lenguaje Natural*, (Sociedad Española para el Procesamiento del Lenguaje Natural) 4, 1986, 43-55.
6. «Metodología Informática para la Edición de Textos», *Incipit*, Buenos Aires, vi, 1986, 185-197.
7. «Amor, verdad, geometría: el enigma de Laura», *Insula*, 497, 1988, 5-6.
8. «El Libro de Alexandre: Notas a partir de la primera edición unificada por ordenador», *Actas del I Congreso Internacional de Historia de la Lengua Española*, Madrid: Arco Libros, 1988, 1025-1064.
9. «ADMYTE (Archivo Digital De Manuscritos y Textos Españoles); The Digital Archive of Spanish Manuscripts and Texts,» *Literary & Linguistic Computing*, 6/3, 1991, (News and Notes) 221-224.
10. «Computers and Text Editing: A Review of Tools, an Introduction to UNITE and Some Observations Concerning its Application to Old Spanish Texts», *Romance Philology*, XLV/1, 1991, 102-122, (Bibliography: 205-237).
11. «Determinación de parámetros tipológicos de los nombres del número», *Anuario de Letras*, XIX, Homenaje a Dámaso Alonso, 1991, 323-369.
12. «La biblioteca electrónica en el Archivo Digital de Manuscritos y Textos Españoles,» *Lexis*, XVII, 1993, 33-56.
13. «Admyte (Archivo Digital de Manuscritos y Textos Españoles)», *Le Médiéviste et l'Ordinateur*, 28, 1993, 9-11.
14. «Editar el *Libro de Buen Amor*: ¿técnicas tradicionales o informática?, *JHR*, 1, 1992-1993, 269-273.
15. «Análisis cuantitativo y percepción de lo literario», *Lenguajes Naturales y Lenguajes Formales*, X, *Actas del X Congreso de Lenguajes Naturales y Lenguajes Formales*, 1994, Barcelona: Promociones y Publicaciones Universitarias S.A. 45-58.
16. «Aproximación cuantitativa al estudio de las jarchas en muwaxahas árabes», *La corónica*, 24.1, 1995, 124-152.
17. «Prehistoria y Protohistoria de las Lenguas: Metodología», *Actas del III Congreso Internacional de Historia de la Lengua Española*, Salamanca, 22-27 de noviembre de 1993. Madrid: Arco Libros, 1996, 765-784.
18. «Establecimiento de la fecha del *Libro de Alexandre*», *ZrP*, 112, 1996, 424-437.
19. «Edición crítica electrónica,» en José Romera Castillo, Francisco Gutiérrez Carbajo y Mario García-Page (eds) *Literatura y Multimedia*, Actas del VI Seminario Internacional del Instituto de Semiótica Literaria, Teatral y Nuevas Tecnologías

de la UNED, Cuenca, UIMP, 1-4 de julio, 1996. Madrid: Visor Libros, 1997. 91-148.

20. "Etimologías engañosas y sus consecuencias: los numerales y el *Libro de Alexandre*", *La Corónica*, 26.1, Fall 1997, 87-103.

21. "Rojo bobo, rufo curvo: paradigma léxico, estructura semántica y cambio fonético desde un análisis informático de diccionarios", *Lexicografías iberorrománicas: problemas, propuestas y proyectos*, ed. M<sup>a</sup> Teresa Fuentes Morán y Reinhold Werner, Frankfurt: Vervuert & Madrid: Iberoamericana, 1998, 201-211.

22. "Presente y futuro de la Filología Electrónica en la recuperación de la Colección Foulché-Delbosc de la Biblioteca Nacional Argentina", *Boletín de la Academia Argentina de Letras*, LXIII, 1998 [1999], 15-52.

23. "GVNDISALVVS INFORMATICVS. Accesos y tratamientos electrónicos de la obra de Gonzalo de Berceo", *La corónica* 28.1, 1999, 83-92.

24. "Diccionario formal frente a diccionario conceptual: jerarquización de las formas para acceso electrónico a lemas", *Investigación y didáctica del léxico*, ed. Juan de Dios Luque Durán, Francisco José Manjón Pozas, Granada, Granada Linguística, Universidad, 1999 [2000], 1-11.

25. "Where is electronic philology going? The present and future of a discipline", *New Media and the Humanities: Research and Applications*. Proceedings of the first seminar *Computers, literature and philology*, Edinburgh, 7-9 September 1998, ed. Domenico Fiormonte & Jonathan Usher, University of Oxford: Humanities Computing Unit, 2001, 11-22.

26. «La determinación cuantitativa del léxico: reconstrucción desde la sincronía», en Elvira N. de Arnoux, Ángela Di Tullio, *Homenaje a Ofelia Kovacci*, Buenos Aires: Eudeba, 2001, 331-344.

27. «Documentos electrónicos y textos de las cámaras legislativas», *Revista General de Información y Documentación*, 12/1, 2002, 245-249.

28. «El *Cantar de Mio Cid* en edición hipertextual», *La Corónica*, 31, 1, Fall, 2002, 123-127.

29. La conservación del patrimonio bibliográfico iberoamericano. El archivo digital de manuscritos y textos españoles, Admyte, y la colección Foulché-Delbosc de la biblioteca nacional de la república argentina," págs. 155-164, en *Informatica umanistica. Dalla ricerca all'insegnamento*, a cura di Domenico Fiormonte, Roma, Bulzoni, 2003.

30. «La construcción de un instrumento informático para la edición crítica», *Letras de Deusto*, vol. 33, n. 100, Julio-Septiembre 2003, 81-108.

Las referencias anteriores tienen un valor muy pequeño, en el conjunto del problema; pero al menos garantizan una dedicación a los estudios cuantitativos desde fecha relativamente temprana y están pensadas, más que para la lectura de todas, para que cada estudioso pueda acudir a alguna de ellas que le interese particularmente en sus planes de trabajo o que esté relacionada, al menos por el título, con ellos. Hay, naturalmente, estudios impresos del autor que no se recogen, sencillamente porque no están digitalizados.

La mayor parte de esas referencias puede encontrarse en la bibliografía de los trabajos en formato electrónico y el resto está disponible en <http://www.campusred.net/uniroma1/articulo/index.html>.

## 2. ITINERARIO

La lectura de alguno de los trabajos anteriores y la curiosidad por seguir alguno de los vínculos que se incluyen al final deben haber provocado un interés por ver cómo se puede organizar cada estudioso el acceso a los estudios de reconstrucción modélica de lenguas. Es conveniente, por ello, esbozar un itinerario, para poder seguirlo con más comodidad que la que se encuentra en un recorrido campo a través, por muy incitante que éste sea.

## 3. PROBLEMAS BÁSICOS

La investigación sobre el cambio lingüístico se plantea, a partir de los neogramáticos, en el siglo XIX, como la necesidad de establecer un *sistema de reglas* que relacionen un conjunto de *unidades* y establezcan unos *parámetros* de la evolución a partir de unos *rasgos* que se van modificando. La insistencia está en el carácter regular, lo que da pie al concepto de *ley fonética*, fuera de la ley fonética queda lo que sería explicable por *analogía*. El avance posterior de la lingüística, con la aparición de la Fonología, introdujo el concepto de *fonema*, de tal manera que, más que hablar de una Fonética Histórica, el objetivo de la reconstrucción pasa a ser un componente de la Fonología Diacrónica. El procedimiento se inicia con el *método comparativo*, sobre una base léxica, es decir, mediante la construcción de *tablas* alineadas de vocablos que se corresponden y, partir de los resultados que se obtienen por comparación, se postula una ley que explique cómo se ha llegado regularmente a esa diferencia, de tal manera que, aplicando la ley hacia atrás, revirtiéndola, se pueda reconstruir la etapa anterior, cuando se desconoce. La aportación de la lingüística histórica, frente a la etimología tradicional (preocupada por el “sentido verdadero” de las palabras) es que importa menos el desarrollo del ejemplo individual, del caso concreto, para pasar a detenerse en lo que es común en la evolución, lo que afecta a un conjunto de elementos lingüísticos que comparten una serie de características.

La investigación para desarrollar fórmulas que establezcan matemáticamente la razón de cambio en las lenguas corresponde a la Glotocronología y a la Lexicoestadística. Si se determina la velocidad de cambio de las palabras a lo largo de la historia, se puede establecer el momento en el que se separaron dos lenguas, según sus índices de relación entre palabras. Pese a los innegables avances de estas disciplinas, la Glotocronología, especialmente, ha contado con numerosos contrarios y con notables artículos escritos en su contra, sin atender al hecho de que las limitaciones temporales de los conocimientos humanos no invalidan su interés, sólo posponen la obtención de resul-

tados fiables. La aportación de estos estudios se plasma, como objetivos concretos, en la realización de *listas y tablas* de correspondencias entre elementos de las lenguas.

Los postulados estructuralistas y funcionalistas enriquecen el concepto de regla, al delimitar perfectamente las unidades y los rasgos. Los desarrollos posteriores de la Fonología, el concepto de fonema como *haz de rasgos*, o la introducción de las teorías métricas, no hacen más que añadir precisión y rigor a la ciencia de la expresión. Véase el siguiente ejemplo, tomado de Marcel Schmuki, en su introducción a Etymo.

[-sonore, +tendu]	[-sonore, -tendu]	[+sonore, -tendu]	perte
/t/	→ /t/	→ /d/	→ -
/p/	→ /p/	→ /b/	( <i>évolution spéciale</i> )
/k/	→ /k/	→ /g/	→ -

En lo que se refiere al cambio lingüístico y, en el mundo románico, Jacob Malkiel desarrolló y aplicó dos nociones, precisándolas, la de *cambio fonético fuerte* frente a *cambio fonético débil* y la de *regularización paradigmática del cambio lingüístico*. El concepto de cambio fonético débil explicaba por qué el alcance de las reglas no era universal y la regularización paradigmática sustuía con ventaja a la analogía, para explicar por qué la consideración en el eje vertical de las unidades y los rasgos permitía una mejor interpretación de todos ellos.

#### 4. LA FORMALIZACIÓN

La fonética o fonología históricas desarrollaron una colección de herramientas sobre un planteamiento heurístico. Los problemas se ven fácilmente, pero su solución es mucho más compleja. Los lingüistas dedican enormes esfuerzos a soluciones perfectas, muy ingeniosas, pero no se alcanza un nivel de formalización que permita responder afirmativamente a la pregunta básica de los informáticos, que versa sobre la existencia o no de un algoritmo. Si se quiere desarrollar un algoritmo para la reconstrucción, si se quieren tener modelos informáticos que expliquen el cambio lingüístico, no se encuentran en los neogramáticos ni en los comparatistas, ni en los lingüistas históricos: hay que construirlos. El proceso se inició muy pronto, nada menos que en 1964, con el algoritmo de Martin Kay, y continúa hoy.

#### 5. EL EJEMPLO DE ETYMO

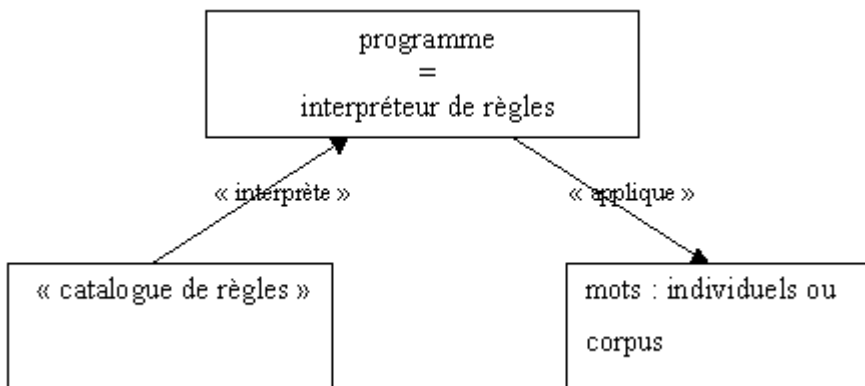
Se traduce, a continuación, mínimamente interpretado, el resumen de Etymo preparado por su autor. Se han conservado los gráficos en francés o en inglés, puesto que, en el marco de un seminario, parece innecesario traducirlos y rehacerlos. El punto de partida fue escribir un programa que permitiera calcular la evolución de las lenguas romances a lo largo del tiempo, puesto que estas lenguas tienen una serie de características ventajosas:



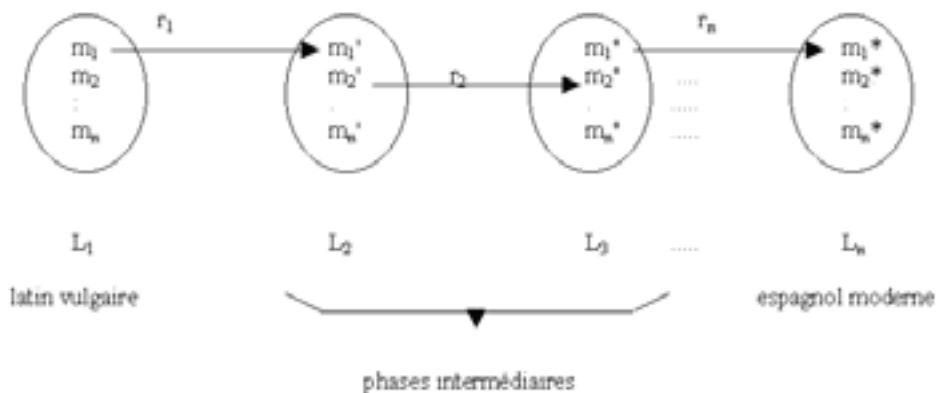
1. Tanto la lengua madre, el latín, como sus hijas (gallego-portugués, español, catalán, francés, provenzal, rético, sardo, italiano, dálmata, rumano) son bien conocidas.
2. También se conoce el proceso evolutivo, con documentación de las etapas intermedias de la evolución.

La mejor definición de Etymo es la que lo considera como una «máquina etimológica abstracta» que, gracias a cierto «lenguaje» formal, puede «programarse» para producir etimologías en una lengua natural concreta. Desde el punto de vista informático, se puede considerar Etymo como un «intérprete de fórmulas lingüísticas»: En la fase primera, de *parsing*, el programa lee las definiciones (que contiene lo que se llama un «catálogo de reglas») y que traduce a estructuras internas ejecutables. En combinación con el intérprete, esas estructuras forman un «algoritmo complejo» que permite calcular las etimologías de una lengua dada. Ese algoritmo se puede ejecutar (fase *interpretativa*, hablando propiamente, bien mediante su aplicación a unidades léxicas aisladas (palabras individuales), bien a conjuntos de palabras mayores (corpus). Esquemáticamente, todo cálculo efectuado por Etymo requiere, por tanto, dos informaciones:

1. un «catálogo de reglas» que define la evolución diacrónica de una lengua y
2. el material lingüístico, -palabras individuales o en un corpus- al que se aplica el cálculo.



Se puede decir de una manera muy sencilla que el proceso evolutivo del tipo de lenguas al que pertenecen las románicas se puede describir mediante un *conjunto de reglas ordenadas* que se aplica a un *conjunto de palabras (m)*. El cálculo comienza con la primera regla ( $r_1$ ), que transforma el conjunto ( $L_1 > L_2$ ). Este nuevo conjunto de palabras ( $L_2$ ) se convierte en la entrada o input de la regla siguiente ( $r_2$ ). Tras aplicar la última regla se obtiene el conjunto  $L_n$  que representa la lengua en la forma final de la derivación regulada:



Hasta aquí se presenta el problema de modo muy simple, ya que es evidente que puede haber palabras en  $L_1$  que ya no se conserven en  $L_n$  (las palabras perdidas a lo largo del proceso evolutivo). Por otro lado, también puede haber palabras en  $L_n$  que no existieran en  $L_1$  (los neologismos, por ejemplo). No obstante, este modelo puede ser útil para el cálculo de muchas palabras y por ese motivo fue elegido para el programa Etymo.

## 6. LÍMITES DEL PROYECTO

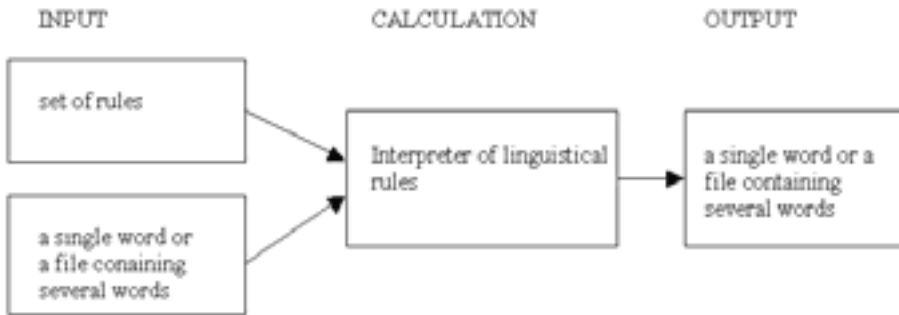
Puesto que los desarrolladores de Etymo sabían desde el principio que el problema de la evolución de las lenguas naturales es muy complejo, decidieron limitar el proyecto. Se calcularían sólo los fenómenos *fonéticos*, dejando a un lado los gramaticales, analógicos o de otro tipo en la evolución. De todos modos, no es que el programa desprecie esa información, de la que puede ocuparse en cierto modo, es sólo que queda excluida del cálculo.

## 7. EL CONCEPTO DEL PROGRAMA

El programa se separa conceptualmente, desde el principio, de las reglas lingüísticas, de manera deliberada. Es sencillamente un *intérprete de reglas lingüísticas*. Por lo tanto, hay un trabajo previo, puramente lingüístico, que tiene que determinar qué reglas se aplican, lo que Etymo hace es calcular en qué medida esas reglas son exactas y pueden emplearse en la reconstrucción. Este intérprete necesita dos tipos de información:

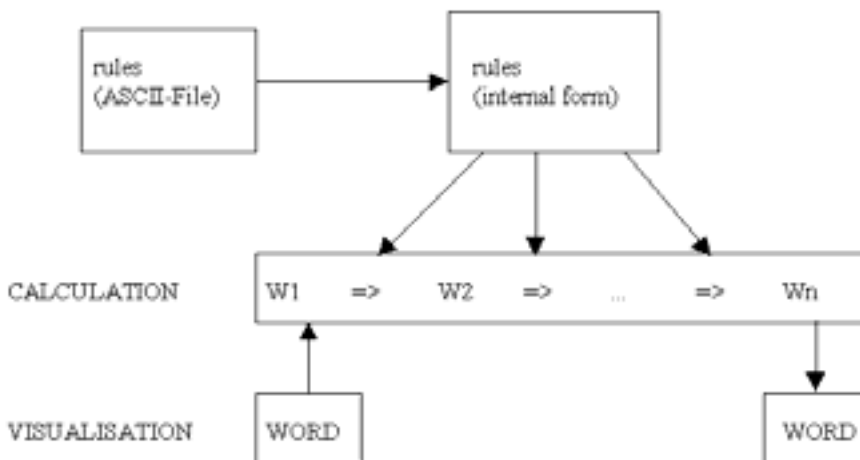
1. definiciones de los *rasgos* y *sonidos* que se emplearán
2. las reglas que describen la evolución de una lengua. La evolución de *cualquier lengua* se puede calcular mediante la modificación de las definiciones de los

sonidos y el conjunto de reglas. Es decir, mediante los elementos que deben aportar los lingüistas interesados en el proceso. Los módulos informáticos se limitarán a cambiar esas entradas de datos y, por lo demás, permanecerán idénticos. La única restricción que señalan es que tiene que ser posible describir los sonidos de cada lengua mediante *rasgos distintivos*.



## 8. LA EVOLUCIÓN COMO MODIFICACIÓN DE RASGOS DISTINTIVOS

Para el cálculo se convierte cada palabra en una lista de rasgos distintivos. Las reglas se traducen también a una estructura interna que facilita su interpretación al programa. Durante el cálculo, las reglas introducen modificaciones en los rasgos distintivos de la palabra. Para que los resultados del cálculo se puedan percibir con claridad, un último proceso reconvierte la palabra en una cadena de fonemas.



## 9. EL ESPAÑOL COMO LENGUA DE PRUEBA

Para probar el programa se definió un conjunto de reglas válido para la evolución del latín al español. La razón de elegir el español es la relativa sencillez de su evolución desde el latín, comparada a la de otras lenguas románicas, especialmente el francés; pero también porque se contaba con que el conjunto de reglas propuesto por C. Lleal (véase la bibliografía de Etymo) facilitaba notablemente el trabajo. (Nótese que el archivo está en código ASCII).

## 10. VÍNCULOS SOBRE INFORMÁTICA Y RECONSTRUCCIÓN DE LAS LENGUAS

Sigue una lista básica, incompleta, de software específico para la investigación diacrónica en lingüística, a partir de la recopilación de J.B. Lowe.

[www.linguistics.berkeley.edu/~jblowe/REWWW/RE\\_TOC.html](http://www.linguistics.berkeley.edu/~jblowe/REWWW/RE_TOC.html)

Lexicoestadístic

[www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC2](http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC2)

Combinatoria

<http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC3>

Programas, proyectos y bases de datos

[www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC4](http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC4)

1. algoritmo de Martin Kay (1964)

[www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC5](http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC5)

2. El neogramático electrónico: el experimento del protoalgonquino de Hewson (1973-)

[www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC6](http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC6)

3. Iberochange (Eastlack, 1977)

[www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC7](http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC7)

4. COMPASS (Frantz, 1970)

[www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC8](http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC8)

5. COGNATE (Guy, 1978-)

[www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC9](http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC9)

[programa] [www.linguistics.berkeley.edu/~jblowe/REWWW/](http://www.linguistics.berkeley.edu/~jblowe/REWWW/)

6. DOC: Chinese Dialect Dictionary on Computer., Wang et al 1970-)

[www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC10](http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC10)

7. Programas para CARP: Computer Aided (Historical Linguistic) Reconstruction in Phonology (Veatch, 1993)

<http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC11>

8. PHONO: a program for testing models of sound change (Hartman, 1981).

- www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC12  
[programa] <<http://www.linguistics.berkeley.edu/~jblowe/REWWW/>
9. WORDSURV (Wimbish, 1989-).  
www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC13  
[programa] <[www.linguistics.berkeley.edu/~jblowe/REWWW/](http://www.linguistics.berkeley.edu/~jblowe/REWWW/)>
10. MARIAMA (Nicolai, 1993?-)  
www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC14
11. CUSHLEX (Gragg, 1990?-)  
www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC15
12. RE: The Reconstruction Engine (Lowe and Mazaudon, 1989-)  
www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC16  
[portal] <<http://bantu.berkeley.edu/REWWW/REintro.html>>,  
[versión en francés] <<http://lacito.vjf.cnrs.fr/reconstr/prjre.htm>>
13. Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) (1987-)  
<<http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC17>>,  
[portal] <[trill.berkeley.edu/lingdept/research/stedt/](http://trill.berkeley.edu/lingdept/research/stedt/)>
14. The Comparative Bantu Online Dictionary (CBOLD) (1994-)  
<<http://www.linguistics.berkeley.edu/~jblowe/REWWW/PriorArt.html#RTFToC18>>.  
[portal] <<http://bantu.berkeley.edu/>>
15. Etymo - un interpréteur de formules dans le domaine de la linguistique diachronique. Avec une application à l'évolution du latin à l'espagnol (2000)  
<http://homepage.sunrise.ch/homepage/mschmuki/official/main.html>

# Esperimento per un'edizione diplomatico-interpretativa computerizzata dei *Ricordi* di Francesco Guicciardini

Simona Casciano

Università di Roma "Tor Vergata"  
(simona.cas@tiscalinet.it)

*During the first half of the XXth century the Ricordi by Francesco Guicciardini have been at the center of an important philological debate. Four different versions were written by the author from 1512 to 1530, one of which has been lost. According to the principal expert in the Ricordi philology (Michele Barbi) one of Guicciardini's grandsons took one of the manuscripts from the house archive in order to print it. This manuscript, that is the lost one and is therefore the basis for the early printed editions of Ricordi in the XVIth and XVIIth centuries. Conjectures of reconstructions have been proposed both by Michele Barbi and Raffaele Spongano on the basis of the collations of the three manuscripts with the printed editions. The paper proposes an xml based electronic edition which allows the simultaneous access to the four versions. Various functions will be available: collation, text search, concordances and indexes, allowing a multi-perspective study of the text.*

*Keywords:* Language reconstruction, xml based editions, philology

## 1. I *RICORDI* DI FRANCESCO GUICCIARDINI: BREVE INTRODUZIONE

I *Ricordi* di Francesco Guicciardini sono stati al centro di un importante dibattito critico-filologico che ha interessato il primo cinquantennio del Novecento, concludendosi (almeno ufficialmente) nel 1951 con la pubblicazione dell'edizione critica di Raffaele Spongano<sup>1</sup>.

Attraverso lo studio delle edizioni cinquecentesche Barbi si accorse di differenze sostanziali tra queste e i manoscritti presenti nell'Archivio Guicciardini. Inoltre il confronto con i testi originali evidenziò che le edizioni del Corbinelli (1576)<sup>2</sup>, di fra Sisto da Venezia (1582)<sup>3</sup> e del Sansovino (1583)<sup>4</sup>, erano molto differenti per ordine dei pensieri, per numero (158 del Corbinelli e 145 di fra Sisto) ed anche per argomenti trattati: bisognava stabilire quale redazione manoscritta fosse stata usata per queste edizioni.

Guicciardini non pubblicò i *Ricordi* ma dopo la sua morte il nipote fece uscire uno dei manoscritti dalla casa; Barbi cita a proposito un fatto conosciuto già nel 1812: "Messer Piero di mess. Nicolò Guicciardini dette copia de' suddetti Avvertimenti al

sig. don Flavio Orsini auditore della Camera Apostolica, che fu poi cardinale, e dopo la morte di messere Piero se ne sparsero copie, con poca soddisfazione de' Guicciardini, che gli avevano tenuti segretissimi”<sup>5</sup>. Secondo la ricostruzione critico-filologica di Barbi le redazioni dei *Ricordi* sono quattro:

1. Redazione Q (1512): il quaderno che la contiene è conservato intatto nell'Archivio Guicciardini<sup>6</sup> ed in esso è possibile distinguere due fasi redazionali risalenti allo stesso anno (Q1 e Q2).
2. Redazione A (1523-1525): è stata ricostruita congetturamente da Barbi e Spongano sulla base delle edizioni cinquecentesche e secentesche; secondo i due studiosi sarebbe questa la serie dei *Ricordi* pervenuta a Flavio Orsini e fonte della tradizione non autorizzata dell'opera. Non si conserva l'autografo.
3. Redazione B (1527)<sup>7</sup>.
4. Redazione C (1530)<sup>8</sup>.

Secondo le ricerche fatte da Barbi risultò che Flavio Orsini fu uditore della Camera nel 1561 e cardinale nel 1565, quindi sarebbe questo il periodo in cui i *Ricordi* uscirono dall'archivio di famiglia: ma in quale redazione? Quella Del '28? del '30? O in una ulteriore versione andata poi perduta?

Quest'ultimo dubbio e la diversa interpretazione che venne data ad una nota autografa posta in calce da Guicciardini nella redazione B, accesero il dibattito tra Barbi e Palmarocchi. Questa la nota guicciardiniana, che si trova nella prima pagina del quaderno B:

“Scritti innanzi al 1525 ma in altri quaderni che in questo, ma ridotti qui nel principio dell'anno 1528, nel grandissimo ozio che avevo, insieme con la più parte di quelli che sono indietro [prima aveva scritto innanzi, N.d.R.] in questo quaderno”<sup>9</sup>.

Per il Palmarocchi gli 'altri quaderni' sono i due che si conservano ancora nell'archivio Guicciardini, e che contengono ventinove ricordi scritti nel 1512 (la Redazione Q); mentre sembra più logico, secondo Spongano che, poiché il ricordo B 138 è datato '3 febbraio 1523'<sup>10</sup>, i quaderni a cui l'autore attingeva “comprendessero già tanti ricordi fino a quella data, e che avesse continuato a scriverli anche nell'anno seguente”<sup>11</sup>.

Inoltre Palmarocchi, dopo aver effettuato la *collatio* tra le redazioni e le principali stampe cinquecentesche, ritenne che le differenze esistenti fossero dovute all'arbitrio dei copisti (o dei tipografi) più che ai ripensamenti dell'autore<sup>12</sup>: così le copie dei *Ricordi* dei secoli XVI e XVII sarebbero derivazioni dell'autografo B e non di una stesura precedente rispetto a B.

Ma Barbi fece notare tuttavia che se il Palmarocchi avesse esteso il suo studio a tutti i *Ricordi* si sarebbe accorto di differenze molto grandi<sup>13</sup> fra quelle edizioni e la redazione B, non attribuibili certo solo ai copisti.

La questione sembra essersi conclusa a favore della tesi del Barbi e l'edizione critica di Spongano ne è la conferma.

Dunque concordando con la tesi Barbi-Spongano, l'esperimento per l'edizione ipertestuale dei *Ricordi* si potrà basare sul manoscritto del 1530, l'ultima redazione voluta dall'autore e conservata in ottime condizioni nell'Archivio di casa Guicciardini. Partendo dallo studio di questo autografo ed applicando il linguaggio di codifica informatica XML, è stata creata un'edizione diplomatico-interpretativa del manoscritto (EDIC)<sup>14</sup>.

Il progetto finale prevede tuttavia di riportare in vita la caratteristica mobilità dei *Ricordi*, applicando lo studio informatico anche alle altre redazioni conservate nell'Archivio, con lo scopo di creare un ipertesto critico di quest'opera.

## 2. RICORDI ED IPERTESTO

La vicenda redazionale dei Ricordi e la possibilità di poter confrontare gli autografi con le ipotesi critiche che li riguardano, rendono questo testo quasi unico tra i manoscritti della nostra letteratura.

Attraverso l'ipertesto il lettore potrebbe scoprire la vivacità variantistica dei manoscritti confrontandoli tra loro, entrando così in contatto con una tecnica scrittoria molto diversa da quella moderna. Potrebbe nascere un interessante studio linguistico confrontando le diverse redazioni dei Ricordi, vedere se e come cambia la lingua di Guicciardini (visto il lungo periodo di gestazione dell'opera) attraverso uno studio di tipo ortografico-calligrafico.

Un'edizione ipertestuale permetterebbe infatti il confronto simultaneo tra le varianti redazionali con i relativi manoscritti, e anche con studi fatti nel corso degli anni su di essi. Il lettore potrà creare indici di concordanze, fare spogli linguistici, ma, soprattutto, confrontarsi con il testo originale dell'autore, vederne direttamente il manoscritto, senza la mediazione del critico e della sua edizione. L'edizione critica diverrebbe peraltro un supporto da utilizzare qualora il lettore lo ritenesse necessario.

Dunque creando ipertesti di questo tipo si restituirebbe al testo la sua dinamicità ed il suo spessore variantistico, permettendo una lettura 'intelligente', basata sul confronto e sull'osservazione degli 'strati compositivi' dell'opera.

Ma la creazione di un ipertesto implica il trattamento del testo stesso con un linguaggio tecnico-informatico appropriato: tale processo viene detto codifica.

Ogni processo di codifica è un atto interpretativo, dunque vincolato dalla forma mentis di chi ne fa uso, e comporta ogni volta la costruzione di 'modelli di testo' (dove per modello si intende una delle tante possibili rappresentazioni della realtà).

Per la codifica dei *Ricordi* è stato utilizzato l'XML, acronimo per eXtensible Markup Language, rilasciato dal W3C (World Wide Web Consortium) e progettato per essere del tutto indipendente dai sistemi informativi e dai supporti digitali presenti e futuri; in questo modo le informazioni testuali codificate attraverso l'XML godono della massima portabilità nello spazio e nel tempo. Tale linguaggio non è soltanto uno strumento per il WEB ma è qualcosa di più perché permetterà la condivisione totale dei dati, intesi come pura informazione, a prescindere dal tipo di visualizzazione e di utilizzo che se ne farà in futuro<sup>15</sup>.



### 3. PROBLEMI DI TRASCRIZIONE E DI CODIFICA PER L'EDIZIONE DIPLOMATICO-INFORMATICA DEI RICORDI

La difficoltà della trascrizione e, di conseguenza, della codifica informatica dei *Ricordi* nasce dal fatto che la grafia di Guicciardini, come in genere la scrittura cinquecentesca, è ricca di abbreviazioni e simboli, che la rendono molto complessa.

Per questo lavoro è stata conservata la punteggiatura del manoscritto, gli a capo, gli spazi, le posizioni delle aggiunte e delle soppressioni, le forme puramente etimologiche o grafiche (ad esempio l'uso dell'*h* in parole come *havevano*, *pocho*, etc.) che nelle edizioni moderne sono state normalizzate.

È stato fatto uno studio linguistico relativo alla redazione C, che si potrà estendere anche alle altre redazioni conservate autografe, ossia Q1-Q2 e la redazione B. Per la redazione A (ricostruita congetturalmente) il lavoro risulterebbe meno utile perché non è stato conservato il manoscritto e, dunque, si dovrebbe lavorare soltanto sulla versione ricostruita filologicamente da Barbi e da Spingano.

Per meglio comprendere la difficoltà della trascrizione ed il tipo di approccio informatico adottato, riporto di seguito solo pochi esempi di forme abbreviate utilizzate da Guicciardini e la relativa codifica informatica:

a. Pronomi dimostrativi:

1. *questa/questi*: li abbrevia con la lettera *a* e la lettera *i* messe in apice *q<sup>a</sup>/q<sup>i</sup>*.

b. Altre forme:

2. *ancora*: si può trovare scritta come *ancra* (con *ra* scritti in apice), oppure *acra* (con la tilde su *ac* ed *ra* scritto in apice);

3. *che*: scrive *ch* tagliando l'ultima lettera con un tratto orizzontale;

4. *fiorinoli*: lo indica con la *f* corsiva;

5. *huomini*: la forma abbreviata si presenta come *huoi* con la tilde sopra la *u* e la *i*;

6. *Imbasciadori*: si trova la forma abbreviata *Imb* con la *r* e la *i* finali della parola scritti in apice;

7. la forma *in modo* la scrive sempre *im<sup>o</sup>*, con la tilde sulla *i*;

8. la nasale dentale *n* e la bilabiale nasale *m* sono spesso volte sostituite con la *tilde*, ad esempio: *in* diventa *i* con un tratto orizzontale sopra la *i*, *manco* è scritto *maco* con un tratto orizzontale sopra la *a*, ecc.

9. *Messere*: rappresentato con una *M* maiuscola tagliata dal basso verso l'alto, da sinistra verso destra;

10. *per*: lo sostituisce con la *p* tagliata trasversalmente da sinistra verso destra e dal basso verso l'alto, che dovrebbe sostituire le lettere *e* ed *r*;

11. *quando*: *qu* con un tratto orizzontale sulle due lettere *q* ed *u*;

12. *secondo/seconda*: *sdol/sda*, con *do* e *da* scritti in apice;

13. *seri*: usa la formula abbreviata *S* tagliata trasversalmente da sinistra verso destra;

14. *tutto, tutte, tutti*: a volte usa la formula abbreviata *tto, tte, tt'*.

Per molte di queste forme ho associato delle *identità* create appositamente inserendole nel *tag* specifico per le abbreviazioni <abbr>. Tenendo presente le descrizioni di cui sopra, si può confrontare la seguente tabella con le relative codifiche:

Segni del Manoscritto	Segni tipografici corrispondenti	Codifica <sup>16</sup>
	-	&cbar;
	/	&vbar;
	:	&ddot;
	ancora: anc(ora) a(n)c(ora)	anc&rsup;&casup; ac&bar;&rsup;&casup;
	ch(e)	ch&bar;
	(Eccetera)	&ecc;
	Ex(emplo)	ex&osup;
	(et)	&et;
	h(uomin)i	huo&bar;i
	Imb(asciadori)	Imb&rsup;&cisup;
	M(essere)	M&vbar;
	m(odo)	M&osup;
	p(er)	P&barbl;
	qu(ando)	qu&bar;
	q(uest)a	q&vbar;&casup;
	q(uest)o	q&vbar;&casup;
	q(uest)i	q&vbar;&cisup;
	-s	&ssup;
	-ss	&ssup;



s(econ)d(o)  
s(econ)d(a)

s&cdsup;&cvbar;&cosup;  
s&cdsup;&casup;

s(er)

s&cvbar;

(tu)tt(a)

tt&casup;

(tu)tt(i)

tt&isup;

(tu)tt(e)

tt&esup;

Oltre al *tag* per l'abbreviazione sono stati utilizzati quelli specifici per:

1. nomi di persona: <persName type="real">Medici</persName>
2. luoghi: <placeName type="real"> Firenze </placeName>
3. date: <date value="1457">1457</date>
4. numeri: <num type="cardinale" value="3">3</num>
5. note dell'autore: <note place="margine sinistro" resp="aut">nota dell'autore</note>
6. numeri di riga: <lb n"1,2,3,..."/>
7. numeri di pagina del manoscritto: <pb n="3r"/> (dove *r* sta per *recto*, mentre *v* indica il *verso*)

### 3.1 Le varianti testuali

Per rappresentare le varianti testuali della redazione C dei *Ricordi* ho adottato i seguenti criteri:

1. Nel caso in cui una parola è stata scritta e poi cancellata ho utilizzato il *tag* <del> con l'attributo *type*, per indicare il tipo di depennamento effettuato.  
Ad esempio nel *Ricordo* 10 si ha *el naturale* depennato, che tradotto con il linguaggio di codifica diventa:

<del type="sbarrato">el naturale</del>

2. Nel caso di una parola aggiunta all'interno del testo, ho utilizzato il *tag* <add> con l'attributo "place", che permette di specificare il luogo in cui è stata scritta (*supralinea*, *interlinea*, *margine destro*, ecc.).  
Ad esempio sempre nel *Ricordo* 10 in luogo della parola depennata *el naturale*, si ha:

<add place="supralinea">l'accidentale</add>

Esistono altri criteri di codifica per le varianti testuali molto più complessi ma per i *Ricordi* sono stati sufficienti quelli del punto 1 e 2. Questi, uniti agli altri *tag* ed *entità* hanno permesso una codifica molto vicina all'originale manoscritto.

#### 4. LA DTD DEI *RICORDI*<sup>17</sup>

Dopo aver scelto *tag* ed *entities* secondo le esigenze del testo, si è costruito il "contesto informatico": a tal proposito sono state create una DTD ed una *teiHeader* apposite.

La DTD<sup>18</sup> (Document Type Definition), sulla quale si basa la codifica del manoscritto dei Ricordi, è stata costruita seguendo sia lo schema di codifica della TEI, sia introducendo le entità specifiche del progetto MASTER<sup>19</sup>; inoltre sono state aggiunte alla DTD una serie di entità create per la codifica delle abbreviazioni e per i segni di punteggiatura propri del manoscritto.

La fusione tra le varie componenti è stata realizzata grazie a "Pizza Chef", un programma realizzato da Sebastian Ratz ed utilizzabile in rete, sul sito della TEI, all'indirizzo <http://www.tei-corg/pizza.html>.

Grazie al progetto MASTER si è potuto identificare il manoscritto in tutte le sue parti; di seguito riporto la descrizione dei tag e la loro posizione all'interno della *teiHeader*<sup>20</sup>:

- <catchwords>: indica le parole di rimando nel manoscritto;
- <dimensions>: contiene qualunque indicazione di dimensione;
- <heraldry>: utilizzabile per la descrizione di simboli araldici nel manoscritto;
- <locus>: indica un luogo all'interno del manoscritto;
- <material>: contiene la descrizione del supporto materiale del manoscritto;
- <msIdentifier>: per l'identificazione del manoscritto;
- <name>: contiene qualsiasi nome;
- <origDate>: utilizzato per la specifica della data di stesura del manoscritto;
- <origPlace>: utilizzato per la specifica del luogo di origine del manoscritto;
- <secFol>: contiene una parola tratta da un luogo specifico di un manoscritto per identificarlo in maniera univoca;
- <signatures>: contiene la descrizione della segnatura del manoscritto.

Lo schema MASTER si basa su un elemento <msDescription>, utilizzabile sia all'interno della *teiHeader*, sia all'interno del corpo del documento.

L'attributo *status* permette di specificare, attraverso valori predefiniti, se il manoscritto è unitario, miscelaneo e completo. I valori possibili sono *uni* (unitario), *compo* (miscelaneo), *frag* (mancante della maggior parte dei fogli), *def* (mancante di un piccolo numero di fogli) o *unknown* (se non si hanno informazioni a riguardo). L'attributo *type* contiene un testo libero che specifica il tipo di oggetto in questione (codice, fascicolo, etc.). Per i *Ricordi* si tratta di un manoscritto unitario.

Il *tag* <msDescription> è strutturato, al suo interno, nel seguente modo:

- <msIdentifier>: per l'identificazione del manoscritto (obbligatorio);
- <msHeading>: per la descrizione della struttura del manoscritto;
- <msContents>: per la descrizione del contenuto intellettuale del manoscritto;
- <physDesc>: per la descrizione dettagliata delle caratteristiche fisiche del manoscritto;
- <history>: per le informazioni riguardanti la storia (origine, provenienza, possessori) del manoscritto;
- <additional>: per le informazioni amministrative sulla attuale collocazione del manoscritto e sul suo eventuale materiale aggiuntivo;
- <msPart>: è un elemento annidato per la descrizione di un frammento di un manoscritto composito.

#### 4.1 L'identificazione del manoscritto

La segnatura del manoscritto dei *Ricordi* è stata codificata attraverso il *tag* <msIdentifier> (l'unico obbligatorio in <msDescription>). All'interno di <msIdentifier> possono comparire i *tag* per la descrizione della collocazione geografica del manoscritto (<country>, <region>, <settlement>, <institution>, <repository> e <collection>) e quelli per l'identificazione convenzionale del manoscritto (<idno> e <altName>). La collocazione geografica del manoscritto è definita con sempre maggiore specificità partendo dalla nazione in cui esso si trova (<country>), attraverso la regione (<region>), la città (<settlement>), l'istituzione (<institution>), l'edificio (<repository>) fino alla collezione (<collection>) di cui fa parte. L'identificazione all'interno della collezione avviene attraverso il numero di segnatura assegnato al manoscritto, espresso nel *tag* <idno>. Il nome alternativo con cui il manoscritto è comunemente noto nella comunità scientifica è codificato attraverso <altName>.

Qui di seguito è fornito un esempio di codifica dell'identificazione del manoscritto dei *Ricordi C*:

```
<msIdentifier>
  <country>Italia</country>
  <region>Toscana</region>
  <settlement>Firenze</settlement>
    <institution>Archivio Guicciardini</institution>
    <repository>Casa Guicciardini</repository>
    <collection>Carte Francesco Guicciardini</collection>
  <idno>V.a,1845,IX</idno>
  <altName>Ricordi</altName>
</msIdentifier>
```

#### 4.2 La struttura del manoscritto

Il *tag* <msHeading> è utile per fornire un breve sommario descrittivo del manoscritto. Al suo interno si possono fornire informazioni relative all'autore (<author>), al

titolo dell'opera (<title>) e alla responsabilità specifica di altre persone nella confezione del manoscritto (<respStmt>). Per autore, in questa sede, si intende il responsabile del prodotto intellettuale contenuto nel manoscritto:

```
<author attested="yes" accepted="n"><name>Francesco Guicciardini</name></author>
```

Il tag <respStmt> permette di indicare, ad esempio, il nome del copista:

```
<respStmt>
  <resp>Copiato da: <name> </name></resp>
</respStmt>
```

I tag <origPlace> ed <origDate> consentono di definire il luogo e la data di origine del manoscritto. In particolare attraverso gli attributi *notBefore* e *notAfter* di <origDate> si possono codificare le date *post quem* ed *ante quem* per la confezione del manoscritto. Attraverso i valori predefiniti dell'attributo "evidence" si può invece specificare se le prove della datazione sono interne (*internal*), esterne (*external*) al manoscritto, oppure frutto di congettura (*conjecture*).

La lingua in cui è scritto il testo è codificata tramite il tag <textLang>. Il tag <note>, utilizzabile anche altrove, è utile per specificare qualsiasi altra informazione sommaria che si ritenga utile. Nel caso dei *Ricordi* si ha:

```
<origPlace>Firenze</origPlace>
<origDate notBefore="1529" evidence="internal">
XVI
</origDate>
<textLang> Italiano volgare</textLang>
```

La descrizione del contenuto intellettuale del manoscritto è codificata attraverso il tag <msContents>, all'interno del quale si possono utilizzare il tag <msItem>, con l'attributo "defective", ed i suoi valori predefiniti "yes", "no" ed "unk" (che specificano se il testo in questione è completo oppure no).

Anche in <msContents> è possibile utilizzare <author> e <title> per codificare l'autore ed il titolo dell'opera riportata nella porzione di manoscritto esaminato.

Tramite il tag <bibl>, all'interno di <msContents>, si fornisce l'indicazione bibliografica dell'edizione del testo in questione, nel caso specifico:

```
<bibl>
  <author>Raffaele Spongano</author>
  <title>Ricordi</title>
  <publisher>Firenze, G.C. Sansoni</publisher>,
  <date value="1951">1951</date>
</bibl>
```

Infine, il tag <summary> contiene una breve descrizione del contenuto dell'opera in questione.

### 4.3 L'aspetto fisico del manoscritto

Il tag `<physDesc>` consente di descrivere l'aspetto fisico del manoscritto con la seguente struttura interna:

```
<physDesc>
  <form></form>
  <support></support>
  <extent></extent>
  <collation></collation>
  <layout></layout>
  <msWriting></msWriting>
  <decoration></decoration>
  <foliation></foliation>
  <condition></condition>
</physDesc>
```

Il tag `<form>` specifica il formato del manoscritto.

All'interno del tag `<support>`, relativo al supporto del testo, è stato utilizzato il tag `<material>` per specificare il materiale utilizzato, carta o pergamena:

```
<support><p><material>carta</material></p></support>
```

Le dimensioni del manoscritto sono codificate con il tag `<extent>`, all'interno del quale è annidato il tag `<dimension>`, che a sua volta contiene i tag `<height>` e `<width>`:

```
<extent>242 fogli
  <dimensions type="leaves" scope="all">
    <height units="mm">349</height>
    <width units="mm">249</width>
  </dimensions>
</extent>
```

Gli attributi *type* e *scope* indicano rispettivamente l'oggetto della misurazione, in questo caso i fogli, e l'applicabilità della misurazione. Quest'ultimo attributo ha dei valori predefiniti: *all* (tutto il manoscritto), *most* (la maggior parte) o *range* (una parte specifica).

All'interno di `<dimension>` si possono utilizzare i tag `<height>` e `<width>` (in cui l'attributo *units*, con valore predefinito *mm*, millimetri, fornisce l'unità di misura adottata).

Il tag `<collation>` si utilizza per fornire informazioni su fenomeni particolari presenti nel manoscritto.

Con `<layout>`, unito agli attributi *columns* e *writtenLines*, si può codificare la disposizione del testo sulla pagina. In questo caso la codifica è la seguente:

```
<layout columns="1" writtenLines="29-33">
```

`<p>` Il testo dei Ricordi, disposto su una colonna singola, si estende da pagina 2r a pagina 23r. Eccetto i ricordi C1 a pagina 2r, che ha un numero di 19 righe, e C2 a

pagina 2v, composto da 27 righe, il numero di righe per pagina nel manoscritto varia da 29 a 33. </p>

</layout>

Si possono descrivere anche il numero di mani intervenute nella stesura del testo (non è questo il caso) con il tag <msWriting>.

L'attributo "hands" specifica il numero delle mani e di conseguenza quanti tag <handDesc> si troveranno all'interno di <msWriting>.

Gli attributi utilizzati sono:

- *scribe*: utilizzato per definire il ruolo della persona responsabile della scrittura;
- *script*: la definizione del tipo di scrittura secondo le convenzioni paleografiche;
- *medium*: l'inchiostro utilizzato;
- "*cope*": l'ampiezza di utilizzo della mano in questione, con i valori predefiniti "sole" (unica), *major* (maggior parte del testo) *minor* (minore parte).

In questo caso specifico si ha:

```
<msWriting hands="1">  
<handDesc script="scrittura italica minuscola corsiva" medium="inchiostro" scope="sole"  
scribe="Francesco Guicciardini">  
<p>Il testo dei Ricordi è scritto in una scrittura italica minuscola ricca di abbreviazioni, con  
accentuato andamento corsivo.</p>  
</handDesc>  
</msWriting>
```

Come appare dall'esempio, è possibile fornire, all'interno di ciascun <handDesc>, una breve descrizione della scrittura utilizzata.

Gli elementi decorativi presenti nel manoscritto sono stati codificati all'interno del tag <decoration>. Per elemento decorativo si intende qualsiasi fatto grafico diverso dalla scrittura. Ogni singolo elemento decorativo è descritto tramite un tag <decoNote>, del quale sono stati utilizzati gli attributi *id*, *technique* e *type*.

L'attributo *technique* indica la tecnica grafica adottata, mentre *type* indica la funzione dell'elemento decorativo nel testo.

```
<decoration>  
  <decoNote technique="disegno a penna" type="struttura">  
  <p>Dopo il numero che indica il ricordo si trova un segno non comprensibile, seguito dal  
  capolettera.</p>  
  </decoNote>  
</decoration>
```

La numerazione dei fogli è stata codificata con il tag <foliation>, all'interno del quale si può fornire una breve descrizione del sistema in uso nel manoscritto.

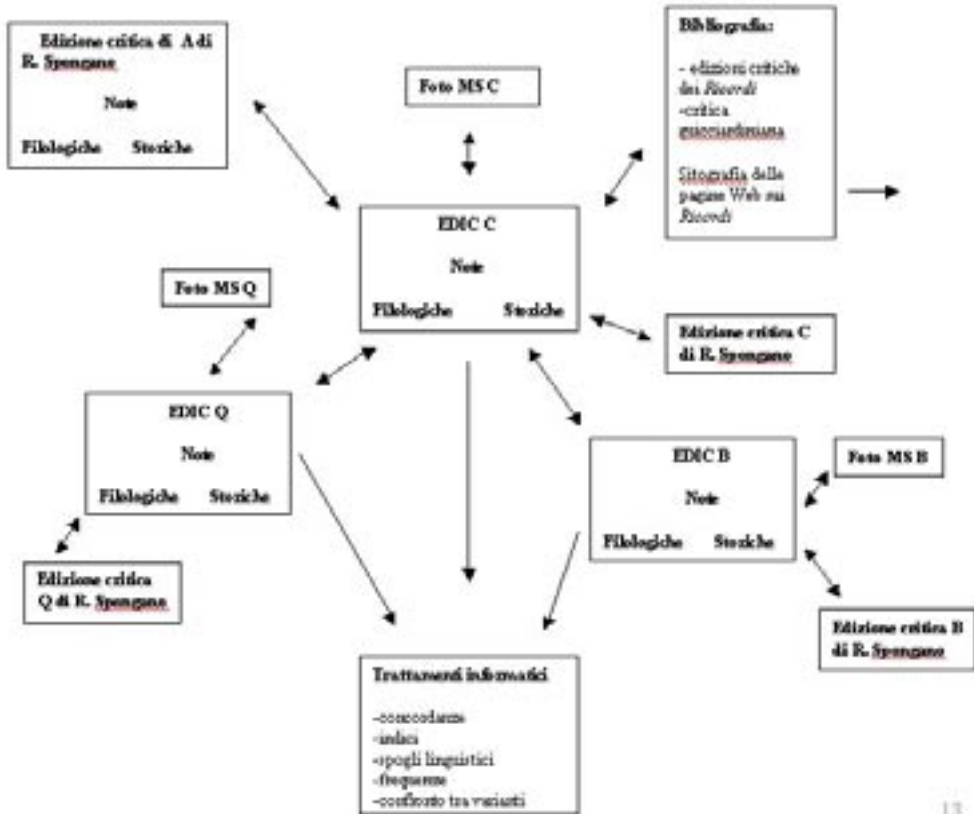


La descrizione fisica del manoscritto termina con l'indicazione delle condizioni di conservazione codificata nel *tag* <condition>.

## 5. IL MODELLO IPERTESTUALE PER UN'EDIZIONE INFORMATIZZATA DEI *RICORDI* DI FRANCESCO GUICCIARDINI

Per concludere, dopo aver parlato dei *Ricordi* e dei criteri di codifica adottati, riporto di seguito il progetto finale del testo, che interesserà non soltanto la redazione C ma tutte le redazioni presenti nell'Archivio Guicciardini.

Modello ipertestuale:



Secondo questo modello, il lettore avrà la possibilità di poter interagire con l'ipertesto attraverso le diverse redazioni *Ricordi*; di ogni redazione è presente l'EDIC all'interno della quale sarà possibile consultare le note filologiche e storiche relative a ciascun ricordo, l'edizione critica e le immagini del manoscritto. Sarà presente una bibliografia relativa alla critica guicciardiniana ed una sitografia che si potrà aggiornare direttamente dal WEB.

Inoltre (e questa la considero la parte più entusiasmante del progetto) si potranno effettuare dei trattamenti informatici utili per lo studio linguistico e variantistico dei *Ricordi*.

L'ipertesto deve permettere al lettore di scoprire la vivacità variantistica dei manoscritti, attraverso il confronto dei *Ricordi* codificati e delle loro rispettive immagini, sottolineando l'idea di dinamicità del testo (che è propria dell'ipertesto) da troppi secoli dimenticata e sostituita dalla rigidità connessa all'invenzione di Gutenberg.

Secondo questo modello, il lettore avrà la possibilità di poter interagire con l'ipertesto attraverso le diverse redazioni *Ricordi*; di ogni redazione è presente l'EDIC all'interno della quale sarà possibile consultare le note filologiche e storiche relative a ciascun ricordo, l'edizione critica e le immagini del manoscritto. Sarà presente una bibliografia relativa alla critica guicciardiniana ed una sitografia che si potrà aggiornare direttamente dal WEB.

Inoltre (e questa la considero la parte più entusiasmante del progetto) si potranno effettuare dei trattamenti informatici utili per lo studio linguistico e variantistico dei *Ricordi*.

L'ipertesto deve permettere al lettore di scoprire la vivacità variantistica dei manoscritti, attraverso il confronto dei *Ricordi* codificati e delle loro rispettive immagini, sottolineando l'idea di dinamicità del testo (che è propria dell'ipertesto) da troppi secoli dimenticata e sostituita dalla rigidità connessa all'invenzione di Gutenberg.

## 6. RIFERIMENTI BIBLIOGRAFICI

- AA.VV. (2002). *Computers and the Humanities*, Vol. 36. London, Kluwer Academic Publishers.
- Adamo G. (1989). *Trattamento, edizione e stampa di testi con il calcolatore*. Roma, Bulzoni.
- Antonelli R. (1992). Esercizi di lettura di Gianfranco Contini. In *Letteratura italiana. La ricerca letteraria*, vol. IV, pp. 339-404. Torino, Einaudi.
- Avalle d'Arco S. (1975). *Principi di critica testuale*. Padova, Antenore.
- Barbi M. (1938). Per una nuova compiuta edizione dei Ricordi politici e civili del Guicciardini. *Studi di filologia italiana*, vol. III. Ora in: Id. (1938). *La nuova filologia e l'edizione dei nostri scrittori da Dante a Manzoni*, pp. 125-160. Firenze, Sansoni.
- Barbi M. (1994). *La nuova filologia e l'edizione dei nostri scrittori da Dante a Manzoni*, Firenze, Le Lettere.
- Bentivogli B., Vecchi Galli P. (2002). *Filologia italiana*. Milano, Bruno Mondatori.
- Cadioli A. (1997). *Le soglie dell'ipertesto*. In P. Nerozzi Bellman (ed.). *Internet e le Muse*, pp.33-54. Milano, Mimesis.
- Cadioli A. (1998). *Il critico navigante. L'ipertesto e la critica letteraria*. Genova, Marinetti.
- Capelli A. (1988). *Cronologia, Cronografia, e Calendario perpetuo*. Milano, Ulrico Hoepli.
- Cazalé Bérard C., Mordenti R. (1997). La costituzione del testo e la «comunità degli interpreti»: libertà e responsabilità del critico/editore/ermeneuta in ambiente elettronico inter-attivo. In P. Nerozzi Bellman (ed.). *Internet e le Muse*, pp. 13-37. Milano, Mimesis.
- Cicchetti A., Mordenti R. (1985). *I libri di famiglia in Italia. Filologia e storiografia letteraria*. Roma, Edizioni di Storia e Letteratura.

- Ciotti F. (1997). Testo, rappresentazione e computer. Contributi per una teoria della codifica informatica dei testi. In P. Nerozzi Bellman (ed.). *Internet e le Muse*, pp. 221-250. Milano, Mimesis.
- Contini G. (1970). *Varianti e altra linguistica. Una raccolta di saggi (1938-1968)*. Torino, Einaudi.
- De Caprio V. (1984). Nei testi del '27: conoscere attraverso la crisi. *Annali d'italianistica*, vol. II, pp. 81-93.
- Fiormonte D., Usher J. (2001). *New Media and the Humanities: Research and Applications*. University of Oxford, Humanities Computing Unit.
- Fubini M. (1941). Le quattro redazioni dei 'Ricordi' del Guicciardini (Contributo della formazione del linguaggio e dello stile guicciardiniano). *Civiltà moderna* XIII, pp. 105-124, e pp. 247-271. Ora in Id. (1971). *Studi sulla letteratura del Rinascimento*, pp. 126-177. Firenze, La Nuova Italia.
- Giavieri M. T., Grésillon A. (eds) (1994). *I sentieri della creazione*. Reggio Emilia, Diabasis.
- Gigliozzi G. (1977). Filologia. In *Enciclopedia del Novecento*. Roma, Istituto della Enciclopedia Italiana.
- Gigliozzi G. (1985). *La critica del testo*. Bologna, Il Mulino.
- Gigliozzi G. (1992). *Breviario di ecdotica*. Torino, Einaudi.
- Gigliozzi G. (1996). *Il testo e il computer. Manuale di informatica per gli studi letterari*. Milano, Bruno Mondadori.
- Gigliozzi G. (1997). *Critica letteraria e nuove tecnologie*. In P. Nerozzi Bellman (ed.). *Internet e le Muse*, pp. 55-73. Milano, Mimesis.
- Gigliozzi G. (1997). *Il testo e il computer. Manuale di informatica per gli studi letterari*. Milano, Bruno Mondadori.
- Gigliozzi G. (1997). La Galassia di Von Neumann: il testo tra piombo e byte. I nuovi orizzonti della filologia. Ecdotica, critica testuale, editoria scientifica e mezzi informatici elettronici. In T. Orlandi (ed.). *Il problema della formalizzazione*. Roma, Accademia Nazionale dei Lincei.
- Guicciardini M. F., Lottini M. G. F., Sansovino M. F. (1583). *Proposizioni, ouero Considerationi in materia di cose di Stato, sotto titolo di Avvertimenti etc.* Venezia, Salicato.
- Guicciardini F. (1933). *Scritti politici e Ricordi*. A cura di R. Palmarocchi. Bari, Laterza.
- Guicciardini F. (1949). *Ricordi*. With an English Translation by Ninion Hill Thomson. New York, S.F. Vanni.
- Guicciardini F. (1857). *Opere inedite*, vol. I., A cura di P. e L. Guicciardini. Firenze, Barbera.
- Guicciardini F. (1951). *Ricordi*. A cura di R. Spongano. Firenze, Sansoni.
- Guicciardini F. (1977). *Ricordi*. A cura di M. Fubini. Milano, Biblioteca Universale Rizzoli.
- Guicciardini P. (1946). *Contributo alla bibliografia di Francesco Guicciardini*. Firenze, Olschki.
- Landow G. P. (1998). *L'ipertesto. Tecnologie digitali e critica letteraria*. A cura di P. Ferri. Milano, Bruno Mondadori.

- Malagoli L. (1939). *Francesco Guicciardini*. Firenze, La Nuova Italia.
- Mordenti R. (2001). *Informatica e critica dei testi*. Roma, Bulzoni.
- Nerozzi Bellman P. (ed.) (1997). *Internet e le Muse*, Milano, Mimesis.
- Orlandi T. (1990). *Informatica umanistica*. Roma, La Nuova Italia Scientifica.
- Palumbo M. (1984). *Gli orizzonti della verità, saggio su Guicciardini*. Napoli, Liguori.
- Pasquali G. (1988). *Storia della tradizione e critica del testo*. Firenze, Le Lettere.
- Ridolfi R. (1931). *Archivio della famiglia Guicciardini*. Firenze, Leo S. Oschky.
- Ridolfi R. (1978). *Studi guicciardiniani* (vol. 45). Firenze, Biblioteca dell'Archivium Romanicum.
- Ridolfi R. (1982). *Vita di Francesco Guicciardini*, Milano, Rusconi.
- Roncaglia G. (1997). *Alcune riflessioni su edizioni critiche, edizioni elettroniche, edizioni in rete*. In P. Nerozzi Bellman (ed.). *Internet e le Muse*, pp. 251-276. Milano, Mimesis.
- Sasso G. (1984). *Per Francesco Guicciardini. Quattro studi*. Roma, Istituto storico italiano per il Medio Evo.
- Scarano E. (1980). *Le redazioni dei 'Ricordi'*. Ora in: Id., *La ragione e le cose. Tre studi su Guicciardini*, pp. 89-178. Pisa, ETS Università (ma già pubblicato in *Giornale Storico di Letteratura italiana*, vol. CXLVII, anno 1970, p. 290).
- Segre C. (1985). *Avviamento all'analisi del testo letterario*. Torino, Einaudi.
- Spirito U. (1970). *Machiavelli e Guicciardini*, Firenze, Sansoni.
- Spongano R. (1948). *Per l'edizione critica dei 'Ricordi' del Guicciardini*. Firenze, Sansoni.
- Stussi A. (1994). *Introduzione agli studi di filologia italiana*. Bologna, Il Mulino.
- Timpanaro S. (1981). *La genesi del metodo Lachmann*. Padova, Antenore.
- Treves P. (1931). *Il realismo politico del Guicciardini*, Firenze, La Nuova Italia.
- Treves P. (1932). *Francesco Guicciardini*. Roma, A. F. Formiggini.
- Zarri G. P. (1968). *Linguistica algoritmica e meccanizzazione della "collatio codicum"*. *Lingua e stile*, vol. III, n. 3, pp. 21-40.

## 7. NOTE

- <sup>1</sup> Benché, dopo un accurato confronto tra i manoscritti e l'edizione Spongano, ho potuto riscontrare un errore del critico relativo all'attribuzione di un intero periodo cancellato al *Ricordo* 166, anziché al *Ricordo* 165.
- <sup>2</sup> *Più Consigli et Avvertimenti di M. FR. Guicciardini Gentilhuomo Fior. In materia di re publica et di privata, nuovamente mandati in luce*, Parigi, Morello, 1576.
- <sup>3</sup> *Considerazioni civili sopra l'Historie di M. Francesco Guicciardini e d'altri storici trattate per modo di discorso da M. Remigio Fiorentino etc. con CXLV Advertimenti di M. F. G. nuovamente posti in luce*, Venezia, Zenaro, 1582.
- <sup>4</sup> Guicciardini M. F., Lottini M. G. F., Sansovino M. F. (1583).
- <sup>5</sup> Barbi M. (1938), p. 133.
- <sup>6</sup> L'Archivio si trova nella casa dei Guicciardini a Firenze.
- <sup>7</sup> Autografo conservato nell'Archivio di casa Guicciardini.
- <sup>8</sup> Conservato nell'Archivio di casa Guicciardini.
- <sup>9</sup> Guicciardini F. (1951), p. 17.
- <sup>10</sup> Per segnalare la data i fiorentini usavano lo stile della 'Incarnazione' (detto anche 'stile fiorentino'), che prendeva come principio d'anno il giorno 25 marzo, festa dell'Annunciazione di Maria Vergine.

In questo modo viene posticipata la cronografia odierna di due mesi e 24 giorni: quindi la data '3 febbraio 1523' dovrebbe corrispondere al 3/02/1524. Lo 'stile fiorentino' fu messo in vigore nel X sec. e rimase fino al 1749 incluso; in seguito venne adottato il metodo odierno. Le notizie relative alla cronografia fiorentina sono state tratte da: Capelli A. (1988), pp. 9-13.

<sup>11</sup> Guicciardini F. (1951), cit., p. 17.

<sup>12</sup> Barbi M. cit., p. 156.

<sup>13</sup> *ivi*, pp. 157-159.

<sup>14</sup> EDIC è l'acronimo per Edizione Diplomatico Interpretativa Computerizzata.

<sup>15</sup> L'applicazione dell'XML sui testi letterari è regolata dalla TEI (*Text Encoding Initiative*), un progetto che ha visto coinvolte le maggiori organizzazioni internazionali dedicate all'Informatica Umanistica.

<sup>16</sup> L'inizio della codifica coincide con il carattere  $\text{Ⓔ}$  e la sua fine con il carattere.

<sup>17</sup> Ringrazio il Dott. Paolo Rembadi dell'Università "La Sapienza" di Roma per avermi aiutato nella formulazione di questo capitolo.

<sup>18</sup> La definizione di un tipo di documento in SGML viene formalizzata sotto forma di una *Document Type Definition* (DTD). Essa è costituita da un elenco di dichiarazioni (*markup declaration*) espresse nella sintassi SGML, che descrivono la struttura del documento.

<sup>19</sup> Il progetto MASTER (Manuscript Access through Standars for Electronic Records) vide la luce all'interno del progetto European Union Framework IV tra il gennaio del 1999 ed il giugno del 2001. Esso è stato concepito per la catalogazione e la descrizione fisica dei manoscritti medievali e moderni dell'area dell'Europa Occidentale.

Per una descrizione più dettagliata si veda il sito internet del progetto MASTER, consultabile alla pagina <http://www.cta.dmu.ac.uk/projects/master>.

<sup>20</sup> La *teiHeader* si trova in ogni testo TEI ed offre informazioni analoghe a quelle fornite dalla copertina di un testo stampato.

# Tempo e genesi del testo. Un esperimento in rete<sup>1</sup>

*Domenico Fiormonte*

Università Roma III  
(fiormont@uniroma3.it)

*Cinzia Pusceddu*

University of Edinburgh  
(mpuscedd@staffmail.ed.ac.uk)

Né il pensiero può ricostruire l'antica situazione per confrontarla con la nuova, giacché non ha più campo libero: la conoscenza che abbiamo acquisita, il ricordo dei primi, insperati minuti, le parole che abbiamo udite, sono là a ostruire l'ingresso della nostra coscienza, controllando gli sbocchi della memoria assai più di quelli della nostra fantasia, retroagendo sul nostro passato – che non siamo più padroni di vedere senza tener conto di loro – più che sulla forma, rimasta libera, del nostro avvenire.

*Marcel Proust*<sup>2</sup>

Mi piace dire che all'equilibrio la materia è cieca, lontana dall'equilibrio comincia a "vedere".

*Ilya Prigogine*<sup>3</sup>

*This paper discusses the birth and development of the new textual paradigm embodied today in digital forms of writing. The historical roots of this paradigm-shift can be traced back to the theoretical and literary movements of the 1930s, especially in Italy and France. It follows an exploration of the links and relationships between this open and dynamic conception of the literary work and that of the psychology of composition. The article concludes with the illustration of an online experiment for the representation of the textual genesis of several poems of the Italian writer Valerio Magrelli.*

*Keywords:* textual genesis, writing process, document digitalization

## 1. LA FILOGIA DEL NOVECENTO E IL NUOVO PARADIGMA DEL TESTO

### *1.1 Crisi e rinnovamento della critica testuale in Italia*

Gli inizi del Novecento sono anni di crisi per la filologia positivista, scienza della restituzione dell'originale perduto: in Francia Bédier sferra un primo attacco al metodo lachmanniano, seguito a breve distanza da Quentin. La critica bédieriana alla

soggettività insita nel procedimento, compromette irrimediabilmente il primato della scuola tedesca, non solo sul piano ecdotico della restituzione del testo ma, e soprattutto, su quello ermeneutico: la scelta di pubblicare il *bon manuscrit* implica infatti una riduzione dell'esercizio critico dell'editore, spostando in realtà l'attenzione sull'autore, e sul manoscritto in quanto documento storico.

La crisi del metodo lachmanniano si ripercuote anche in Italia. La cosiddetta scuola storica capeggiata da Rajna, che aveva visto l'applicazione rigorosa di tale metodo e prodotto un numero notevole di edizioni critiche, inizia il suo declino. Un inizio di rinnovamento si registra a partire dagli anni '30, e si svolge lungo gli anni a cavallo del secondo conflitto mondiale. Antonelli (1985: 207) definisce la filologia di questi anni come 'materiale', ben indicando con tale termine la rinnovata attenzione alla storicità del manoscritto e alla sua fisicità in quanto documento accertato. A gettare i semi di questo nuovo indirizzo sono Pasquali (1934) e Barbi (1938). Anche se riconfermando il valore del metodo lachmanniano (che in Italia non verrà mai completamente abbandonato, ma semmai rivisitato), entrambi ne avvertono i limiti. Barbi riconosce la necessità di considerare i documenti nella loro individualità storica; Pasquali, filologo classico, si mostra invece più critico e invita a studiare i rapporti tra i codici senza isolarli dal contesto storico-culturale in cui sono stati prodotti.

In questo clima di attenzione materiale al manoscritto non stupisce il rinnovato interesse anche per i manoscritti in redazioni successive, o con varianti d'autore.

E proprio il Pasquali dedica nel 1934 (e, non casualmente, proprio nello stesso testo in cui propone un approccio globale ai codici) un intero capitolo a tale fenomeno. In esso il filologo documenta un notevole numero di casi di varianti d'autore in ambito classico, spostandosi poi alla registrazione del fenomeno in presenza di manoscritti autografi come nel caso di Petrarca e Boccaccio.

Se Pasquali ha il merito di porre in evidenza, tra i primi, il problema delle varianti d'autore, Santorre Debenedetti (1937) ha quello di essere il primo a pubblicarle in edizione critica: i suoi *Frammenti dell'Orlando Furioso* del '37 mostrano in apparato le aggiunte apportate dall'Ariosto all'ultima versione del poema, documentate su due autografi. Ma è nello stesso anno che un venticinquenne e brillante allievo di Debenedetti cambia radicalmente la prospettiva secondo cui considerare il fenomeno delle varianti d'autore, inaugurando un filone di ricerca e un metodo completamente nuovo nel panorama dell'ecdotica italiana.

Parliamo naturalmente di Gianfranco Contini. È lui a compiere il passo decisivo: invece che limitarsi a considerarle come un dato empirico è il primo a chiedersi: "Che significato hanno, per il critico, i manoscritti corretti degli autori?" (Contini, 1970: 233-234)<sup>4</sup>, e dunque ad interpretarle e attribuire ad esse una funzione fondamentale nel processo di analisi testuale. Con lui nasce la critica delle varianti o variantistica (o critica degli scartafacci, come la definirà sprezzantemente Croce).

Il primo nucleo della riflessione continiana è proprio in un commento del '37 alla pubblicazione dei *Frammenti* di Debenedetti. Contini osserva che tale edizione rappresenta "in modo subito evidente, *nella sua esatta cronologia, tutto il lavoro elaborativo e correttivo dell'Ariosto*" (1970: 232, corsivo nostro). Egli individua dunque un mo-

vimento temporale nell'atto creativo, dato dalla successione dei cambiamenti al testo; e poco più avanti, in un passo ormai celeberrimo, porta a compimento questa sua intuizione affermando che l'opera letteraria è da considerarsi dinamica, "una perenne approssimazione al valore". Se l'opera si inserisce nel flusso del tempo, in una dimensione diacronica, allora compito del critico deve essere di restituire questa dimensione temporale che attraversa il testo. Questa è l'operazione rivoluzionaria di Contini: il testo oggetto di analisi critica cessa di essere statico, un dato, "come un oggetto o un risultato" e acquisisce gli attributi di mobilità e fluidità.

La seconda colonna portante della critica delle varianti poggia sul concetto di testo-sistema, che viene definito qualche anno più tardi, nel '41. Per Contini ciascuna variante non deve essere considerata singolarmente, separata dalle altre e isolata in sede d'analisi: tutte le variazioni e i rifacimenti costituiscono il testo, che viene dunque a configurarsi come un sistema di elementi interagenti tra loro. Ciascuna variante è dunque uno spostamento di forma e senso che si riverbera su *tutto* il testo, una rappresentazione parziale e dinamica del suo farsi nel tempo. Il che implica che ciascun rifacimento e modifica d'autore cambia aspetto e significato *complessivi* e non *locali* del testo. In tal senso Contini accoglie e porta al punto più alto la lezione di Bédier e Pasquali: non solo sposta ulteriormente l'attenzione sull'autore, ma inserisce il testo nel flusso della storia. Ciascuna singola redazione d'autore è *il testo*, documento storicamente accertato e originale, non uno scarto o una versione deteriore rispetto alla perfezione dell'ultima stesura.

Ne consegue che anche il risultato finale dell'analisi del critico di testi in redazioni successive, l'edizione critica, è "nel tempo": l'apparato sarà perciò diacronico, e renderà ragione dell'intera evoluzione del testo attraverso la registrazione di tutte le modifiche effettuate.

Per Contini l'attività di analisi su testi moderni non sarà in contrapposizione o disgiunta dalla pratica ecdotica su testi a trasmissione manoscritta privi di originale. Ed egli sarà un innovatore anche in ambito più strettamente critico-testuale, operando una revisione del metodo lachmanniano<sup>5</sup>: ciò implica l'esistenza, anche nei futuri sviluppi metodologici e pratici della variantistica, d'una continuità tra filologia tradizionale e filologia d'autore.

In questi che sono i punti cardine della variantistica si possono senz'altro cogliere influssi e suggestioni che i movimenti e fermenti culturali di quel tempo esercitarono su Contini, anche se egli li riutilizza in modo assolutamente originale. Il testo come entità dinamica e fluida e l'atto creativo come processo in fieri sono infatti il manifesto di una nuova generazione di poeti moderni quali Mallarmé e Valéry, attivi proprio in quegli anni, cfr. Corti 1997<sup>6</sup>: 115), mentre il concetto di sistema è il fulcro della riflessione filosofica della scuola di Tartu. Se nel primo caso possiamo legittimamente parlare di influsso e intravedere l'ascendenza di tali poeti sulle riflessioni continiane, poiché lo stesso Contini la riconosce piuttosto esplicitamente<sup>6</sup>, nel secondo caso ci pare più opportuno invece parlare di suggestioni ed echi<sup>7</sup>.

La scuola della critica delle varianti si rivelerà lungamente produttiva e fondamentale nel panorama filologico italiano (e non solo) futuro: la nuova generazione di filo-



logi strutturalisti degli anni Settanta troverà su tale impianto metodologico una base e un terreno fertile su cui impiantare le teorie linguistiche e strutturaliste. E l'elemento comune sarà proprio il concetto di testo come sistema, come affermerà lo stesso Contini. I rapporti tra critica delle varianti e strutturalismo verranno individuati da Benvenuto Terracini (1968) e D'Arco Silvio Avalle (1970) negli anni Settanta e rimarcati più recentemente da Antonelli (1985) e Caprettini (1985). Saranno in particolare Segre, Avalle e la Corti, coagulatisi attorno alla rivista *Strumenti Critici* nata negli anni Settanta, a sviluppare e arricchire col proprio personale apporto i rivoluzionari concetti di Contini. Se il merito di Contini è quello di concentrare l'attenzione sul segmento Autore-Testo del circolo ermeneutico, Segre (1979) con la sua concezione di diasistema, individua il movimento testuale sul versante opposto, cioè nel segmento Testo-Lettore/Critico. La critica delle varianti individua il dinamismo nel processo compositivo e produttivo del testo; la teoria dei diasistemi di Segre nell'atto di lettura/critica e ricezione del testo. Non a caso questo filologo, come Contini, medierà tra le posizioni bédieriane e lachmanniane. Avalle invece, più incline alle sperimentazioni formalistiche, svilupperà ulteriormente la riflessione sull'opposizione testo-dato/testo-processo e l'inafferrabilità dell'originale. La Corti, coordinatrice di un centro di archiviazione di manoscritti di autori contemporanei, raccoglierà l'insegnamento della filologia d'autore nel modo più applicativo.

### 1.2 *Il processo di scrittura: la critique génétique in Francia*

A distanza di più di tre decenni dalla critica continiana, nella Francia degli anni Settanta accesa da fervori, stimoli e rivoluzioni culturali, si sviluppa un nuova metodologia letteraria volta allo studio del processo di scrittura: la *genèse du texte*. Due sembrano essere gli eventi cruciali e significativi per cui ascriverle questa data di nascita:

1. L'Istituzione nel 1969 all'interno del CNRS di un gruppo di ricerca franco-tedesco per l'edizione del *corpus* di manoscritti di Heine, acquisiti dalla Bibliothèque Nationale; tale gruppo costituirà il primo nucleo dell'attuale ITEM, *Institut des Textes et Manuscrits Modernes*, ente ufficiale e cuore della ricerca critico-genetica francese.
2. La pubblicazione del volume *Le texte et l'avant-texte* di Bellemin-Noël, (1971), cui si deve l'introduzione del termine e del concetto di avantestato quale insieme di tutto il materiale testuale, preparatorio e redazionale, di mano d'autore, che precede la versione ultima.

Secondo la ormai cospicua letteratura<sup>8</sup>, l'obiettivo della filologia genetica è ricostruire le tappe successive della genesi di un'opera letteraria. L'origine dell'atto creativo si appunta fin dai primissimi e informi abbozzi, documentati sotto forma di disegni, note, appunti di lettura, schemi, progetti di lavoro, prove di scrittura, ecc., passa poi attraverso gli strati di versioni successive da rintracciare nelle cancellature, riscritture, aggiunte presenti nello spazio grafico dell'autografo, per giungere fino all'ultima versione composta dall'autore. L'avantestato o dossier genetico rappresenta dunque mate-

rialmente il processo di lavoro dell'autore e il divenire nel tempo del testo. E poiché un tale tipo di corpus documentario si possiede soprattutto per gli autori contemporanei, sono questi che la *genèse* indaga, con una restrizione ad una fascia temporale che ha sollevato molti e stimolanti interrogativi.

I punti cardine della metodologia genetica – la concezione dell'opera letteraria come evoluzione nel tempo, l'obiettivo di ricostruire il processo di scrittura, il ritorno alla storia del manoscritto – sono evidentemente gli stessi della critica delle varianti e sono ormai riconosciuti all'interno della scuola francese: Falconer (1993) individua un unico movimento genetico che si snoda tra gli anni Venti e Settanta, mentre nel volume di *Romanic Review* del 1995 dedicato alla *genèse* i richiami a Contini quale precursore della filologia genetica provengono da diversi studiosi, in particolare Compagnon (1995)<sup>9</sup>. Certo, non mancano le differenze: per esempio gli opposti rapporti che variantistica e critica genetica intrattengono con lo strutturalismo e con la filologia tradizionale, di continuità per la prima, di rottura per la seconda (cfr. Grésillon, 1999).

## 2. L'INCONTRO FRA LA “CRITICA MATERIALE” E L'INFORMATICA

### 2.1 *Testo invariante e testo mutante*

Come abbiamo visto uno dei tratti comuni fra critica genetica francese e variantistica italiana è il riconoscimento della multidimensionalità del documento scritto, ovvero dei suoi aspetti contestuali (psicologici, sociali, ecc.) e fisici: grafia, *outils*, tipo e consistenza della carta, cancellature, immagini e disegni. Questa multidimensionalità è sottolineata anche dalla scuola anglo-americana della *textual bibliography* (fondata sulla tradizione materiale del testo [Stoppelli 1987: 7-31]) che sembra orientata da tempo verso il concetto di *mobile text*<sup>10</sup>.

Si direbbe che l'insieme di queste posizioni preannunci la nascita di quella nuova sensibilità di tipo “post-testuale” (Ricciardi, 1998: 130-132) che è oggi la caratteristica delle scritture di rete. La scoperta della fluidità ha come effetto la rimessa in discussione dell'autore e del testo unico (e delle pratiche filologiche che ne derivano) e lo spostamento del baricentro dal *prodotto* al *processo* di scrittura. Sarebbe sbagliato sottovalutare quest'evento, le cui cause sono molteplici, ma del quale certamente il computer ha accelerato l'evoluzione (Cerquiglini, 1989: 48). Questo non vuol dire che l'informatica non abbia fornito alla concezione del testo come “struttura informativa” strumenti potenti, come il *data base*. D'altra parte però il computer, nel momento in cui sembrava fornire l'alleanza più solida al mondo del testo-prodotto, ne ha svelato (per esempio attraverso il problema della codifica) i condizionamenti e la fragile storicità delle strutture. Diverso il discorso sulla dialogicità e le processualità della comunicazione in rete, un utilizzo dell'informatica che alla “stabilità” testuale non offre nessuna sponda e anzi che favorisce la nascita di forme espressive ibride (Chat, MUD, MMS e altri tipi di “electronic script acts”)<sup>11</sup>.

In conclusione, ciò che possiamo notare è che la crisi della filologia in quanto strumento della ricostruzione della “verità” del testo (crisi che da Joseph Bédier giunge sino a Jerome McGann [2001]) interseca il versante attuale della produzione testuale legata alla processualità, interattività e collaboratività dalle nuove forme della comunicazione digitale. Dove è difficile – e forse inutile – rintracciare la prevalenza di una definita e individuale volontà autoriale.

In questo preciso momento il modello di interpretazione della realtà testuale proposto dai fautori dei linguaggi di marcatura basati sull’architettura SGML/XML e del collegato paradigma del text-retrieval (testo = informazione) sembra essere quello vincente. In questo modello, frutto dell’incontro (soprattutto in Italia) fra informatica e critica strutturalista, non sembra esserci molto spazio per altri elementi e dimensioni della comunicazione. Tre in particolare: il ruolo dei codici visuali (intesi sia come elemento esterno sia come elemento interno del testo), l’elemento processuale e quello dialogico-contestuale<sup>12</sup>. È da notare come questo approccio trovi alleati e ispiratori proprio all’interno di quella comunità di studiosi che più ha contribuito alla nascita di una nuova sensibilità testuale. Prendiamo ad esempio la definizione che compare all’interno di un ‘manuale’ ormai classico:

“Se consideriamo i segni grafici (lettere, interpunzione, ecc.) come significanti di suoni, pause, ecc., e riflettiamo sul fatto che questi segni possono essere trascritti più volte e in vari modi (per esempio con grafia e caratteri diversi), *restandone immutato il valore*, possiamo concludere che *il testo è l’invariante*, la successione di valori, rispetto alle variabili dei caratteri, della scrittura, ecc. [...] Il testo è dunque *una successione fissa di segni grafici*. Questi significati grafici sono poi portatori di significati semantici [...]; ma occorre insistere in partenza su questa costituzione originaria. Occorre insistervi, perché le ricchissime, praticamente infinite implicazioni di un testo, quelle che richiamano lettori ai testi anche per secoli e millenni, sono tutte racchiuse nella letteralità dei significati grafici. Di qui l’importanza della filologia, che s’impegna nella conservazione il più possibile esatta di questi significati. Il fatto che la sopravvivenza dei testi implichi inevitabili guasti nella loro trasmissione deve sollecitare ancor più lo sforzo di tutelarne la genuinità.” (Segre, 1985: 29-30; corsivi nostri).

Forse oggi appare scontato che una parola come ‘invariante’ riferita a un testo rimandi a un mondo perduto – per analogia e contrasto vengono in mente le “scritture mutanti” della rete, che si situano agli antipodi della “successione fissa di segni”. Ma anche rimanendo al testo stampato, considerare i segni grafici delle varianti che rimandano a un “valore immutato”, vuol dire considerare la forma e l’aspetto del testo come elementi accessori che non influiscono sul significato e la ricezione<sup>13</sup>. E se, come viene ammesso, il testo può variare, allora com’è possibile che il suo valore possa rimanere immutato nel tempo? In effetti questa intangibilità può essere attinta solo da un modello ideale del testo.

In conclusione, mi pare che in questo passaggio (tenendo ovviamente in conto che si tratta di una formulazione “di servizio”) Segre faccia un passo indietro rispetto alla sua definizione di testo come “concetto limite” (Segre, 1981; cfr. nota 20), quasi a delineare un conflitto, o comunque una dialettica, fra le incursioni del teorico-produttore e le preoccupazioni del filologo-ricostruttore. D'altronde leggendo la storia del testo come “trasmissione del guasto” si è costretti ad ammettere l'esistenza di un testo ideale che, come nella caverna platonica, ci sia accessibile solo attraverso la sua ombra. E allora il compito del filologo non potrà essere altro che quello di (ri)costruire il “modello” a partire dall'ombra.

## 2.2 *La digitalizzazione del documento: dal prodotto al processo*

Quella appena riassunta sembra essere la concezione di testo alla base delle riflessioni e degli sforzi applicativi della maggior parte teorici della codifica.<sup>14</sup> Per Lou Burnard (Burnard, 2001: 29-36) lo scopo del *markup* è quello di esplicitare essenzialmente tre classi di caratteristiche: 1) caratteristiche composizionali (*compositional features*), cioè legate all'aspetto del testo; 2) caratteristiche contestuali (*contextual features*); 3) caratteristiche interpretative (*interpretative features*). Burnard, ricalcando l'idea del “testo invariante”, definisce “accidents of the medium” (Fiormonte, 2003: 185-186) tutto ciò che accade “all'esterno” del testo, ovvero tutte le caratteristiche materiali che esso accumula nel corso della propria vita (sia genetica sia storica). E tuttavia egli afferma che non esiste un approccio unico (*unified approach*), ma esistono tante codifiche per quanti sono i testi e le domande che a essi intendiamo porre. Ciò che viene codificato infatti non è il testo, ma una certa *lettura* di quel testo. Dunque la codifica mediante linguaggi di *markup* non è in grado (perché è “troppo complicato”) di porre domande diverse in uno stesso *tempo*, per esempio rappresentare insieme, come vedremo più avanti, la dimensione temporale e quella materiale del testo. Il concetto di modello euristico proposto dall'informatica umanistica degli anni '80-'90 (Gigliozzi, 1999) perciò funziona solo se supponiamo che da qualche parte – iperuranio o caverna che sia – esista un testo “veritiero” fissato una volta per sempre e del quale “tutelare la genuinità”.

Ma lo strumento analogico (l'occhio o la fotografia) rimane ancora indispensabile per leggere gli aspetti multidimensionali della ‘vita’ del testo. Questi aspetti sono centrali nella filologia moderna e contemporanea, dove l'attenzione si sposta dal prodotto al processo. È probabilmente per questi motivi che la scuola genetica francese ha esplorato in questi anni più le possibilità di visualizzazione che quelle di retrieval del testo, arrivando solo recentemente alla implementazione di archivi marcati. E si spiega perché le tesi “antirealiste”, cioè quelle che portano all'estremo il paradigma codifica/interpretazione, vengono sostenute da chi studia autori contemporanei.<sup>15</sup> Uno strumento che tende a ricostruire e mappare le relazioni gerarchiche è meno adatto allo scopo di un'edizione genetica o, per esprimerci nei termini della *writing science* (Levy, Ransdell, 1996), a quello della rappresentazione del processo compositivo. Se concepiamo e interpretiamo l'opera come processo e non come prodotto e, soprattutto, la inquadriamo in un contesto di interazione con l'utente/fruitori (com'è per alcune

tipologie di scritture online), possiamo dire per la scrittura ciò che viene detto per altri media: “quanto accade nella pratica non può essere dedotto semplicemente da quanto avviene nei testi e nelle strutture” (Newcomb, 1999: 3).

D'altronde la stessa scrittura non è una mera trascrizione del parlato, ma ci offre un modello *concettuale* della dimensione verbale: “writing is in principle metalinguistics” (Olson, 1997: 19). Allo stesso modo in cui la scrittura fornisce un modello standard per la lingua (Duranti, 2000: 118), il testo codificato offre un modello altrettanto (e forse ancora più) “concettuale” del testo originario ottenuto attraverso “metalinguaggi” – i *markup languages*. Ma, scrive Olson, “knowledge of those aspects of linguistic structure for which our script provides a model and about which it permits us to think, has imparted an important bias to our thought and the development of our document culture” (Olson, 1997: 19). Così come la scrittura fornisce il supporto per lo sviluppo di determinati stili cognitivi, i linguaggi dichiarativi forniscono un preciso modello di che cosa sia la realtà documentale (e di come vada rappresentata).

A me pare che né decostruzionisti né anti-decostruzionisti (e neo-strutturalisti) abbiano interpretato correttamente il senso e la direzione del *bias* indicato da Olson e altri, elaborando, come sarebbe necessario, un'adeguata cornice teorica per la nuova relazione che nella dimensione digitale si instaura fra *processi* e *prodotti*. Al cospetto di una scrittura che si fonde con altre forme di comunicazione, adottando sempre più “criteri operativi misti” (come la mescolanza di semasiografico e alfabetico; cfr. Valeri, 2001: 206-211), o a testi pensati per essere consultati come banche dati (Berners-Lee et al., 2001), che cosa avrà senso “codificare” in futuro? Quale passaggio di supporto potrà garantire la fedeltà della fonte, e come?

Ma limitandoci al presente, sono varie le conseguenze metodologiche che questa riconfigurazione dell'oggetto-testo porta con sé: innanzitutto le idee di *conservazione* e *restituzione* del testo.

McGann nel 1985 aveva denunciato come gli orientamenti allora attuali della critica testuale (fra cui proprio “the ideology of final intentions” [McGann, 1985: 37]) frenassero la nascita di un modo diverso di trasmettere, e dunque di leggere, i testi. Tanselle criticò in modo sprezzante quelle posizioni e pur ammettendo che ogni metodo fosse lecito in scienza, nei fatti (re)indicò un'unica strada, quella del *rationale*: “McGann believes that ‘to see ‘author’s intention’ as the basis for a ‘rationale of copy-text’ is to confuse the issues involved; [...]’; one should rather say that confusion is promoted by maintaining that an undefined mixture of two distinct approaches constitutes a useful rationale” (Tanselle, 1987: 132). Non sorprende che il nucleo della critica di Tanselle ritorni qualche anno dopo in un articolo sui rapporti fra *textual criticism* e *critique génétique*, nonostante un apprezzabile avvicinamento agli ambienti liquidati altrove come “sociologici”. Forse a questo punto dovremmo parlare non di critica, ma di timore per i pericolosi germi decostruzionisti, annidatisi – chissà – anche nella *genèse*.

Quella di Tanselle non è soltanto una filologia, ma una visione della letteratura.<sup>16</sup> Un'idea di opera d'arte come successione di stati e quantità *separabili* e *interpretabili*, che dà enorme fiducia all'autore e alla comunità interpretante. Di qui la cautela nei

confronti dell'edizione ipertestuale (dove per Tanselle il cambiamento metodologico è "of degree, not of kind" [Tanselle, 1995: 591]). Questa cautela discende da un sospetto e da un timore profondi nei confronti di un'ermeneutica che dubiterebbe – a cominciare *da e insieme* ai suoi autori – della sua irreversibilità e fissità. A partire dalla nuova concezione dell'opera nell'arte, nell'estetica e nella filosofia del primo Novecento non è stato più possibile per una certa critica vedere letteratura e filologia come entità *teoreticamente* separabili. *Edizione e produzione* non sono sempre due momenti scindibili della storia del testo – cioè di un fenomeno che si fa *nel tempo* e di cui, se è legittimo ritagliare lo spicchio sincronico dell'edizione critica, è ugualmente legittimo rifiutarne l'esigente intangibilità meta-storica.

## 2.2 Verso una critica testuale dinamica

In *Orientations to Texts* Peter Shillingsburg (2001) ha proposto una classificazione delle edizioni in base ai diversi "interessi" del critico testuale. Tali interessi si esplicitano in cinque diversi "orientamenti al testo", che corrispondono a altrettanti modi di concepire la costituzione di un testo: il sociologico, il documentario, il bibliografico, l'agenziale (*Agential*) e l'estetico. A parte l'ultimo, si tratta in tutti i casi di orientamenti che mettono al centro la storicità del documento. Ciascuno di questi modelli, secondo Shillingsburg, esprime interessi legittimi e non deve imporsi sull'altro. Questa impostazione aperta indubbiamente risente delle discussioni sulle nuove tendenze della critica testuale<sup>17</sup>, ma soprattutto sembra nascere dal bisogno di mettere ordine nel 'caos' generato dall'utilizzo delle nuove tecnologie, dove questi vari orientamenti tendono a confondersi.

Volendo semplificare, è possibile vedere lo schema di Shillingsburg come la traduzione di quattro principali "esigenze" storiche dettate dalla contingenza del medium:

- l'esigenza della conservazione dell'opera;
- l'esigenza della trasmissione dell'opera;
- l'esigenza della conservazione dei manufatti dell'artista;
- l'esigenza storico-strumentale dell'edizione critica o dell'edizione *tout court*.

Più di dieci anni fa, confrontandosi con queste esigenze, alcuni filologi avevano cominciato a esprimere insoddisfazione per gli strumenti e le metodologie tradizionali. Come accennavamo all'inizio del paragrafo queste perplessità spinsero anche studiosi della tradizione a stampa alla 'forzata' riscoperta del processo testuale: "I hope I have made my point that the Quarto and Folio *Lears* are artificial, if not arbitrary, abstractions from the debris of evidence left by the history of the unstable text" (Brockbank, 1991: 96). Insoddisfatto per le soluzioni tipografiche adottate per l'edizione delle varianti del *Lear*, Brockbank propone di sfruttare l'appena nata tecnologia CD-ROM per un'edizione *Variorum* di Shakespeare (Brockbank, 1991: 102).

La filologia assistita dal calcolatore o computazionale (Perilli, 1995; Bozzi, 1997) a partire dagli anni Ottanta (Marcos Marín, 1985; Gigliozzi, 1987) aveva cominciato a dare le prime risposte. Raul Mordenti usa, quasi in contemporanea, la stessa espres-

sione di Brockbank (“testo mobile”) commentando la sua edizione informatizzata<sup>18</sup> del *Dialogo della mutazione di Firenze* di Bartolomeo Cerretani (cfr. Mordenti, 1992: 266-268). L’edizione elettronica è ormai nell’aria, e le strade francesi e italiane si rincrociano. Jean-Louis Lebrave, filologo della scuola della *genèse*, progetta un modello sperimentale di edizione ipertestuale per rendere navigabile l’eterogenea mole manoscritta dell’*Hérodias* di Flaubert. Ben al di là della constatazione delle potenzialità dei nuovi strumenti si spinge la bibliografia materiale e sociologica nordamericana, dove il giudizio sulla rigidità della stampa è ancora più netto (cfr. Price, Smith, 1997).

Oggi l’incontro fra informatica e critica materiale<sup>19</sup>, attraverso le sue applicazioni e le continue verifiche teoriche che ne derivano, ci obbliga ad aggiungere a quelle originali quattro esigenze una quinta voce: la necessità della rappresentazione della genesi testuale e del processo di scrittura.

### 3. PSICOLOGIA DELLA COMPOSIZIONE E VARIANTI

Il progetto *Digital Variants* nasce dalla contaminazione teorica di due filoni di ricerca: antropologia della scrittura e filologia d’autore. Due sono i nomi verso i quali ci sentiamo in debito: Giorgio Raimondo Cardona e Gianfranco Contini. Contini come abbiamo visto nei paragrafi precedenti fu tra i primi filologi in Italia a spostare il punto di vista del critico dal prodotto al processo, a inserire cioè il “tempo” dentro l’orizzonte dell’ermeneuta.

Tutta la filologia europea del Novecento potrebbe essere definita come la storia della tensione dialettica fra Testo, Autore e Lettore, fra entità reali e storiche e un oggetto astratto<sup>20</sup>. Il cammino percorso da ciascuna scuola critica converge su un punto: la svolta avviene con gli autori moderni. Sono autori come Flaubert, Proust, Montale, Dickinson o Joyce a guidare la riflessione teorica su un nuovo terreno, che è quello della concezione dinamica del testo. Ma lo spostamento descritto da Contini è accompagnato anche dall’intuizione del contributo ‘epistemico’ e pedagogico della variante (Contini, 1974: 233-234).

E Cardona scriveva nel 1988: “L’attività letteraria ci offre un’opportunità di vedere il pensiero nel corso del suo funzionamento [...]. Anche il cosiddetto errore – in realtà una divergenza fra il circuito del pensiero e quello del linguaggio – ci interessa” (Cardona, 1990: 356-357). Cardona, tra i fondatori europei dell’antropologia della scrittura, era convinto, conoscendo gli studi di variantistica italiani e francesi, che in alcuni materiali come manoscritti, autografi, “scartafacci” e appunti degli scrittori, fosse possibile seguire scie e indizi del movimento del pensiero e attraverso questi risalire a precisi fenomeni della lingua; concepiva dunque la scrittura come “attività”, come oggetto dinamico e non semplice “trascrizione” del parlato. Credeva che la scrittura potesse fornire “nuovi concetti e categorie” per ragionare sulla lingua (Olson, 1997: 5). Ma se su quest’ultimo fronte si andavano producendo ricerche (Miller, 1993), egli fu tra i primi a porre il problema della ‘variante’ in rapporto al processo di scrittura e di questo in rapporto al linguaggio. Come abbiamo visto si conosceva l’importanza

epistemologica del processo ricostruttivo, ma il cammino a valle, quello compositivo, rimaneva in gran parte inesplorato.

A onta dei ripetuti riferimenti al processo compositivo, nelle ricerche di storici della lingua, filologi e critici del testo brilla l'assenza di riferimenti alla psicologia della composizione e alle prime importanti ricerche del cognitivismo (cfr. Gregg, Steinberg, 1980). Il progetto della *genèse* è timido e insieme sospettoso nei confronti di una scienza generale della produzione scritta (la Grésillon parla di “zones d'interférence”, ma dedica poche righe alle ricerche dei cognitivisti [Grésillon, 1994: 220]). Rispetto alla filologia la *composition* fa il cammino inverso: studia la scrittura dal punto di vista di chi scrive, mentre la filologia, per secoli, aveva studiato la scrittura dal punto di vista di chi legge – l'editore e la sua edizione critica. Ma che ne era dello *struggle* dello scrittore?

*Genèse* e variantistica fanno un passo in più verso l'autore, ma non compiono il balzo decisivo: al lettore rimane al massimo la sensazione di essere un “guardone” del testo. Il passaggio successivo – e il cambio di prospettiva – è compiuto dalla psicologia e dalle *cognitive sciences*, questa nuova “scienza di scienze” coagulatesi oltreoceano, che viene a colmare e a rafforzare gli spazi di riflessione comuni alle scienze editoriali e a quelle autoriali (perché naturalmente filologia e psicologia si ignorano proprio sul fronte comune del processo di scrittura).

Questa ignoranza reciproca, frutto più delle consuetudini che dei fatti, apre il campo alle prime incursioni.

In *The psychology of written composition* Carl Bereiter e Marlene Scardamalia (1987) costruiscono il primo solido edificio ribadendo quello che diventerà l'ovvio assioma di ogni programma didattico: la scrittura è una competenza complessa che si acquisisce attraverso molteplici fasi processuali<sup>21</sup>.

È l'attività di *knowledge transforming*, ma soprattutto la concezione del testo come “tappa” di un processo che ci riconducono alle riflessioni dei filologi moderni (Branca, Starobinski, 1977: 82). La critica testuale moderna, se si prescinde da alcune resistenze a riconoscere l'informatica come luogo privilegiato per l'espressione, la modellizzazione e lo studio dei segni, sarebbe ‘scienza cognitiva’ per eccellenza e la psicologia della composizione ne rappresenterebbe il naturale compimento (e *complemento*). Potremmo quasi dire che la seconda compie la profezia contenuta nella prima: entrambe assumono un punto di vista diacronico – la psicologia come scienza sperimentale, la critica del testo come scienza storica – procedendo in direzioni opposte lungo lo stesso cammino.

Il progetto Digital Variants<sup>22</sup> nasce dall'incontro di queste due “scienze cognitive”, con l'obiettivo di recuperare (e sfruttare) il dinamismo della scrittura, ovvero il testo fotografato e interpretato come durata. Riflettendo sui limiti e la forza di critica delle varianti-*genèse du texte* e psicologia della composizione, a Edimburgo abbiamo provato a fondere le due esperienze.

### 3.1 La Magrelli Genetic Machine

Sul sito *Digital Variants* sono disponibili una serie di testi, autografi e documenti di autori italiani e spagnoli consultabili attraverso diversi strumenti di visualizzazione.



L'obiettivo principale del sito infatti, pur nel rispetto delle basilari norme ecdotiche, non è quello di *conservare* i materiali, ma di renderli direttamente *fruibili* da parte degli utenti. La nostra impressione infatti è che a fronte delle grandi opportunità di ricerca e analisi automatica del testo offerte da molte biblioteche digitali, sia stata dedicata scarsa o nulla attenzione al problema della lettura. In termini informatici ciò vuol dire che lo sforzo nella direzione dell'*information retrieval* non è bilanciato da un' inadeguata progettazione delle interfacce-utente.

La rinuncia all'adozione di standard è stata dettata anche (e forse soprattutto) dalla forte eterogeneità dei documenti. Ciascun/a autore/autrice ha donato all'archivio materiali assai diversi fra loro. Si va da un racconto in otto passaggi di scrittura di Francesca Sanvitale (genesì di un testo, con un'unica stesura pubblicata) alla raccolta di racconti *La gente* (Torino, Einaudi, 1993) di Vincenzo Cerami, di cui l'archivio possiede tutte le stesure intermedie (testi che hanno subito modificazioni e sono stati pubblicati in tempi e contesti diversi).

La *Magrelli Genetic Machine*, l'ultimo degli esperimenti in ordine di tempo, nasce dall'esigenza di esplorare nuove soluzioni per la resa del processo di scrittura. A questo scopo abbiamo utilizzato il programma *Flash*, normalmente impiegato in siti commerciali per effetti di animazione. Due parole su questa scelta tecnica – che non vuol dire “neutra”, perché la scelta degli strumenti implica e presuppone sempre precise scelte teoriche. L'adozione di *Flash* appare giustamente criticabile da un punto di vista scientifico (ed etico) perché non basata su risorse aperte (software *open source*). Tuttavia anche se è in corso la codifica XML-TEI delle poesie e dei relativi avantesti, il nostro obiettivo principale rimaneva quello di costruire un'interfaccia efficace e facilmente fruibile online. I fogli di stile (XSL e XSLT), che permettono di trasformare in HTML testi codificati in XML, richiedono una programmazione a parte spesso molto complessa che non garantisce la rappresentabilità di tutti i fenomeni testuali. Un problema tipico di XML infatti è quello della rappresentazione di “overlapping textual phenomena”, come per esempio varianti di struttura temporale (e grafica) particolarmente complessa, come nel caso dei testi di Magrelli<sup>23</sup>. *Flash*, nonostante i suoi limiti, sembra prestarsi in modo naturale al *display* e alla visualizzazione della *mouvance* testuale – oltre a consentire un notevole risparmio di costi e tempi. Il fatto poi che l'azienda che produce questo software (Macromedia) si stia aprendo al mondo XML apre nuove prospettive anche per l'utilizzo di documenti codificati con tale linguaggio.

I materiali utilizzati nel progetto della *Genetic Machine* sono i *brouillons d'écriture*, gli appunti e le diverse stesure della raccolta di poesie *Ora Serrata Retinae* (Milano, Feltrinelli, 1980) messi a disposizione da Valerio Magrelli.

Il dossier completo consiste in un quaderno con le prime stesure originali autografe e in varie versioni stampate o trascritte a macchina corrette dall'autore prima di giungere a quella definitiva. Ciascuna poesia ha dunque una storia genetica diversa. Nel progetto generale ogni poesia della raccolta di Magrelli disporrà di una *Genetic Machine*, vale a dire di un insieme di strumenti, appositamente creati a seconda della sua storia editoriale, che permettano di illustrarne ed esplorarne la dinamicità.

Finora sono stati pubblicati online con questo sistema tre avantesti di *Ora serrata retinae* (*Il corpo è chiuso*, *Essere Matita* e *Molto sottrae*)<sup>24</sup>.

Nella finestra a sinistra (Fig. 1) c'è sempre l'immagine di un originale, manoscritto oppure a stampa. L'originale scelto può essere letto in comparazione incrociata, sia con la trascrizione diplomatica del manoscritto nella finestra in basso sia con le trascrizioni delle varie versioni a stampa (fino a quella definitiva) nella finestra a destra. Una legenda in basso a destra indica i segni diacritici che sono stati usati nella trascrizione diplomatica. Sono inoltre disponibili altri effetti. *Floating variants* mostra ciascuna delle due versioni a stampa in pannelli scorrevoli che possono essere spostati a piacimento; *Fade transcription* consente di leggere sul manoscritto le cancellature effettuate dall'autore al semplice scorrimento del mouse; *Zoom* infine permette di spostarsi con una lente di ingrandimento sull'autografo. È inoltre in fase di implementazione la possibilità per l'utente/ricercatore di inserire dei commenti attraverso un modulo *ad hoc*.



Fig. 1. Esplorazione con Flash dell'avantesto di *Il corpo è chiuso* di Valerio Magrelli. Passando il mouse sulle cancellature dell'autografo viene visualizzato, con un effetto di fading, la trascrizione del testo sottostante (vedi quarta riga "gettato in se").

Non c'è spazio qui per una puntuale analisi critica di *Il corpo è chiuso*<sup>25</sup>. Tuttavia si può notare come questo tipo di rappresentazione del movimento testuale possa aprire la strada *di per sé* a specifiche e diverse interpretazioni del testo. Si osservi l'autografo in comparazione con le versioni successive. Al di là delle attese differenze strutturali (es. la composizione del verso) e linguistiche (cambiamenti lessicali, spostamenti, soppressioni, sostituzioni, ecc.), esso presenta una serie di elementi di contorno e

commento, sia visuali che testuali (vedi il fondo della pagina) che contestualizzano e illuminano i significati del testo. Queste immagini e appunti, permettono di leggere e interpretare il componimento seguendo, per contrasto o per affinità, il discorso parallelo fra mezzo grafico e mezzo linguistico (occhio-ginocchio, ciglio-portale, ecc.).<sup>26</sup> Nella versione definitiva (1980) vengono soppressi i primi due versi (“Splendido l’occhio / Questo è il suo segreto”), conservati fino all’edizione su rivista del 1979. Eppure dall’esame del manoscritto sembra proprio che da questi due versi scaturisca la traccia fonica (“occhio”), iconica (il ginocchio-occhio disegnato) e tematica (“il suo segreto”) della poesia. Solo attraverso una comparazione delle varie versioni dunque è possibile accorgersi di come l’autore proceda, per successive scremature, su un cammino che va dall’esplicito all’implicito; l’autografo da questo punto di vista si offre come una “mappa cognitiva” nella quale sono intessuti e depositati, in forma di nodi e flussi di pensiero, tutti i motivi e i temi sviluppati – o scartati – nelle stesure successive.

Ma a nostro giudizio non è solo la genesi della poesia la protagonista di questo strumento di rappresentazione. È tutto il baricentro del testo a spostarsi, creando un nuovo equilibrio in cui tutti gli elementi – tutte le parti – sono ugualmente importanti. L’intra-testualità rivelata ed esibita dalla *Machine* diventa un oggetto *a sé*, poiché l’insieme dei legami fra testi e avantesti così montati costruisce e forma una nuova esperienza del testo/dei testi. Percepire e fruire il testo definitivo senza guardare o leggere le galassie sorelle che lo attorniano diventa un’operazione innaturale – e probabilmente vana.

Riprendendo la famosa analogia con le scienze fisiche di Bachtin (2000: 313, 316) si potrebbe dire che fino ad oggi abbiamo analizzato il testo letterario secondo le leggi dell’universo pre-heisenbergiano, ovvero dentro un sistema stabile nel quale l’osservatore non modifica l’oggetto osservato: tutti i fattori in gioco – autore, opera, lettore, supporti – conservano la loro indipendenza e distanza. Il computer, attraverso strumenti semplici o complessi (ma tutti resi possibili dalla malleabilità del bit), permette all’osservatore di abbandonare il mito di questa neutralità ed *entrare* nel tempo del testo. Che cosa altro è un’edizione su carta – diplomatica, critica o genetica che sia – se non il tentativo di immortalare e *giustificare* una specifica e storicamente determinata concezione del tempo del testo? Ma quando la scrittura, attraverso la digitalizzazione, viene restituita alla dimensione processuale (o ad una simulazione di essa), e rientra nel flusso tempo, l’unico appiglio per la stabilità del testo rimane la cronologia.

#### 4. RIFERIMENTI BIBLIOGRAFICI

- Anis J., Lebrave J.-L. (1991). Text et Ordinateur. Les Mutations du Lire-Écrire. In *Actes du colloque interdisciplinaire tenu à l’Université de Paris X*. Paris, Éditions de l’Espace Européen.
- Antonelli R. (1985). Interpretazione e critica del testo. In *Letteratura Italiana*, vol. IV. *L’interpretazione*, pp. 141-243. Torino, Einaudi.
- Avalle D’A. S. (1970). *L’analisi letteraria in Italia. Formalismo, strutturalismo, semiologia*. Milano-Napoli, Ricciardi.

- Barbi M. (1938). *La nuova filologia e l'edizione dei nostri scrittori da Dante al Manzoni*. Firenze, Sansoni.
- Bellemin-Noël J. (1971). *Le texte et l'avant-texte*. Paris, Larousse.
- Bleuca J. M., Claveria G., Sánchez C., Torruella J. (eds) (1999). *Filologia e informatica. Nuevas tecnologías en los estudios filológicos*. Barcelona, Editorial Milenio-Universidad Autónoma de Barcelona.
- Bereiter C., Scardamalia M. (1987 [1995]). *The psychology of written composition*. Mahwah (NJ), Lawrence Erlbaum. (trad. it. *La psicologia della composizione scritta*. Firenze, La Nuova Italia).
- Berners-Lee T., Hendler J., Lassila O. (2001). The Semantic Web. In *Scientific American* at <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>.
- Bozzi A. (ed.) (1997). *Better Access to Manuscripts and Browsing of Images*. Bologna, Clueb.
- Branca V., Starobinski J. (1977). *La filologia e la critica letteraria*. Milano, Rizzoli.
- Bridwell-Bowles L. S., Johnson P., Brehe S. (1987). Composing and Computer: Case Studies of Experienced Writers. In Matsuhashi A. (ed.). *Writing in Real Time: Modelling Production Processes*, pp. 81-107. Norwood (NJ), Ablex.
- Brockbank P. (1991). Towards a mobile text. In I. Small, M. Walsh (eds). *The theory and practice of text-editing. Essays in honour of James T. Boulton*, pp. 90-106. Cambridge-New York, Cambridge University Press.
- Budor D., Perrus C. (eds) (2000). *Arzanà. Le Texte: genèse, variantes, édition*. Paris, Presses de Sorbonne Nouvelle.
- Burnard L. (1995). Text Encoding for Information Interchange. An Introduction to the Text Encoding Initiative (document TEI J31). Paper presented at the *Second Language Engineering Conference*. Oxford, Oxford University Computing Services. URL: <http://www.uic.edu/orgs/tei/info/teij31>.
- Burnard L. (2001). On the hermeneutic implications of text encoding. In D. Fiorimonte, J. Usher (eds). *New Media and the Humanities: Research and Applications*, pp. 29-36. Oxford, Humanities Computing Unit of University of Oxford.
- Busby K. (ed.) (1993). *Towards a Synthesis? Essays on the New Philology*. Amsterdam: Rodopi.
- Buzzetti D. (2000). Ambiguità diacritica e markup. Note sull'edizione critica digitale. In *Atti del seminario del Dipartimento di Scienza della Letteratura. Soluzioni informatiche e telematiche per la filologia*. Facoltà di Lettere e Filosofia, Università di Pavia.  
[http://lettere.unipv.it/dipslamm/pubtel/Atti2000/dino\\_buzzetti.htm](http://lettere.unipv.it/dipslamm/pubtel/Atti2000/dino_buzzetti.htm).
- Caprettini G. P. (1985). Le strutture e i segni. Dal formalismo alla semiotica letteraria. In *Letteratura Italiana, vol. IV. L'interpretazione*, pp. 495-548. Torino, Einaudi.
- Cardona G. R. (1990). *I linguaggi del sapere*. Roma-Bari, Laterza.
- Ceri P., Borgna P. (eds) (1998). *La tecnologia per il XXI secolo. Prospettive e rischi di esclusione*. Torino, Einaudi.
- Cerquiglioni B. (1989). *Éloge de la variante. Histoire critique de la philologie*. Paris, Seuil.

- Chartier R. (1999). Testi, forme, interpretazioni. In McKenzie F. *Bibliografia e sociologia dei testi*, pp. 98-107. Milano, Edizioni Silvestre Bonnard.
- Cherchi P. (2001). Filologie del 2000. *Rassegna Europea di Letteratura*, n. 17, pp. 135-153.
- Ciotti F. (2001). Text Encoding as a theoretical language for text analysis. In D. Fiormonte, J. Usher (eds). *New Media and the Humanities: Research and Applications*, pp. 39-47. Oxford, Humanities Computing Unit of University of Oxford.
- Compagnon, A. (1995). Introduction. In *Romanic Review*, Vol 86, 3, pp. 393-401.
- Contini G. (1970). *Varianti e altra linguistica. Una raccolta di saggi (1938-1968)*. Torino, Einaudi.
- Contini G. (1974). *Esercizi di lettura*. Torino, Einaudi.
- Corno, D. (1995). Teorie della scrittura, tra psicologia e semiotica. In Bereiter C., Scardamalia M. *Psicologia della composizione scritta*, pp. 9-43. Firenze, La Nuova Italia.
- Corti M. (1976-1997<sup>6</sup>). *Principi della comunicazione letteraria*. Milano, Bompiani.
- Corti M. (1997). *Per una enciclopedia della comunicazione letteraria*. Milano, Bompiani.
- de Biasi P. M. (2000). *La Génétique des textes*. Paris, Nathan.
- De Maria L. (ed.) (1988). *All'ombra delle fanciulle in fiore* (trad. G. Raboni), p. 133. Milano, Mondadori.
- Debenedetti S. (ed) (1937). *I frammenti autografi dell'Orlando Furioso*. Torino, Chiantore.
- Duranti A. (1997). *Linguistic Anthropology*. Cambridge, Cambridge University Press.
- Falconer G. (1993). Genetic Criticism. *Comparative Literature*, vol. 45, n. 1, pp. 1-21.
- Ferrer D. (1995). Hypertextual Representations of Literary Working Papers. *Literary and Linguistic Computing*, vol. 10, n. 2, pp. 143-145.
- Fiormonte, D., Usher J. (eds) (2001). *New Media and the Humanities: Research and Applications. Proceedings of the first "Computers, literature and philology" seminar*. Oxford, University of Oxford Humanities Computing Unit.
- Fiormonte D. (2003). *Scrittura e filologia nell'era digitale*. Torino, Bollati Boringhieri.
- Giaveri M. T. (1993). La critique génétique en Italie: Contini, Croce et 'étude des paperasses. *Genesis*, vol. 3, pp. 9-29.
- Gigliozzi G. (ed.) (1987). *Studi di codifica e trattamento automatico di testi*. Roma, Bulzoni.
- Gleißgen M. D., Lebsanft F. (eds) (1997). Alte und Neue Philologie. In *Editio (Beihfte)*. Tübingen, Niemeyer.
- Gregg L. W., Steinberg E. R. (eds) (1980). *Cognitive Processes in Writing*. Hillsdale (NJ), Lawrence Erlbaum.
- Grésillon A. (1994). *Éléments de critique génétique. Lire les manuscrits modernes*. Paris, PUF.
- Grésillon A. (1999). Philologie et critique génétique: ressemblances et différences. In *Atti del convegno Internazionale dell'Accademia Nazionale dei Lincei. "I nuovi oriz-*

- zonti della filologia. Ecdotica, critica testuale, editoria scientifica e mezzi informatici elettronici”, pp. 53-58. Roma, Accademia Nazionale dei Lincei.
- Hay L. (ed.) (1989). *La naissance du texte*. Paris, José Corti.
- Lebrave J. L. (1991). L’hypertexte et l’avant-texte. In: J. Anis, J. L. Lebrave (eds) *Texte et: les mutations du lire-écrire*, pp. 101-117. Nanterre, Centre de Recherches Linguistiques de Paris.
- Levy C. M., Ransdel, S. (eds) (1996). *The science of writing. Theories, methods, individual differences, and applications*. Mahwah (NJ), Lawrence Erlbaum.
- Lisa T. (2004). *Scritture del riconoscimento. Su Ora serrata retinae di Valerio Magrelli*. Roma, Bulzoni.
- Martiradonna V. (2004). *La codifica elettronica dei testi. Un caso di studio*. Tesi di laurea, Facoltà di Scienze Umanistiche, Università di Roma “La Sapienza”.
- McGann, J. J. (1985). *A critique of modern textual criticism*. Charlottesville and London, The University Press of Virginia.
- McGann J. J. (2001 [2002]). *Radiant Textuality. Literature after the World Wide Web*. New York, Palgrave (trad. it. *La letteratura dopo il World Wide Web*. Bologna, Bononia University Press).
- McKenzie D. F. (1986 [1999]). *Bibliography and the sociology of texts*. London, The British Library (trad. it. *Bibliografia e sociologia dei testi*. Milano, Edizioni Sylvestre Bonnard).
- Magrelli V. (1980). *Ora Serrata Retinae*. Milano, Feltrinelli
- Magrelli V. (2002). *Vedersi vedersi. Modelli e circuiti visivi nell’opera di Paul Valéry*. Torino, Einaudi.
- Marcos Marín, F. (1985). Computer-Assisted Philology: Towards a Unified Edition of *Osp. Libro de Alexandre*. In: *Proceedings of the European Language Services Conference on Natural-Language Applications*. Copenhagen, IBM Denmark.
- Miller J. (1993). Spoken and Written Language: Language Acquisition and Literacy. In Scholes R. J. (ed.). *Literacy and Language Analysis*, pp. 99-139. Hillsdale (NJ), Lawrence Erlbaum.
- Mordenti R. (1992). Informatica e filologia. In AA.VV. *Calcolatori e Scienze Umane*, pp. 236-272. Milano, Etas Libri.
- Morrás M. (1999). Informática y crítica textual: realidades y deseos. In Blecua J. M., Clavería G., Sánchez C., Torruella J. (eds.). *Filología e informática. Nuevas tecnologías en los estudios filológicos*, pp. 189-210. Barcelona, Editorial Milenio-Universidad Autónoma de Barcelona.
- Nichols S. G. (ed.) (1990). The New Philology. *Speculum. A Journal of Medieval Studies*, vol. 65, n. 1.
- Olson D. R. (1997). On the relations between speech and writing. In: Pontecorvo, C. (ed.). *Writing development. An interdisciplinary view*, pp. 3-20. Amsterdam-Philadelphia, John Benjamins.
- Orlandi T. (1995). Alla base dell’analisi dei testi: il problema della codifica. In M. Ricciardi, (ed.). *Scrivere comunicare apprendere con le nuove tecnologie*, pp. 69-86. Torino, Bollati Boringhieri.

- Pasquali G. (1934). *Storia della tradizione e della critica del testo*. Firenze, Le Monnier.
- Perilli L. (1995). *Filologia Computazionale*. Contributi del Centro Linceo Internazionale "Beniamino Segre". Roma, Accademia nazionale dei Lincei.
- Perrus C. (2000). Gianfranco Contini et l'approche de l'oeuvre in fieri. In D. Budor, C. Perrus (eds). *Le Texte: genèse, variantes, édition*. pp. 13-29. Paris, Presses de la Sorbonne Nouvelle.
- Petrucci, A. (1985). La scrittura del testo. In *Letteratura Italiana*, vol. IV. *L'interpretazione*, pp. 285-308. Torino, Einaudi.
- Price K. M., Smith M. N. (1997). Whitman, Dickinson, and Teaching American Literature with New Technologies. [http://warthog.cc.wm.edu/Whitman/FIPSE/1997\\_FIPSE\\_Funding\\_proposal.html](http://warthog.cc.wm.edu/Whitman/FIPSE/1997_FIPSE_Funding_proposal.html).
- Renear A. (1997). Out of Praxis: Three (Meta)Theories of Textuality. In Sutherland K. (ed.). *Electronic text. Investigations in method and theory*, pp. 107-126. Oxford, Clarendon Press.
- Renear A., McGann J., Hockey S. (1999). What is a text? A debate on the philosophical and epistemological nature of the text in the light of humanities computing research. In *ACH-ALLC Conference*, University of Virginia. [http://www.humanities.ualberta.ca/Susan\\_Hockey/ACHALLC99.htm](http://www.humanities.ualberta.ca/Susan_Hockey/ACHALLC99.htm).
- Renear A. (2001). Literal transcription – can the text ontologist help? In D. Fiormonte, J. Usher (eds). *New Media and the Humanities: Research and Applications*, pp. 23-30. Oxford, Humanities Computing Unit of University of Oxford.
- Ricciardi M. (1998). Le comunità virtuali e la fine della società testuale. In P. Cesi P. Borgna (eds). *La cultura per il XXI secolo*, pp. 130-155. Torino, Einaudi.
- Robinson P., (1994). *The Transcriptions of Primary Textual Sources Using SGML*. Oxford, Office for Humanities Communication Publications (6).
- Ryan M. L. (ed.) (1999). *Cyberspace Textuality. Computer Technology and Literary Theory*. Bloomington and Indianapolis, Indiana University Press.
- Schmid M. (1998). *Process of literary creation. Flaubert and Proust*. Oxford, Legenda.
- Schmidt D. (2004). A Graphical Editor for Manuscripts.  
URL: <http://www.wittgen-cam.ac.uk/cgi-bin/text/mseditor.html>.
- Schreibman S. (2002). Computer-mediated Texts and Textuality: Theory and Practice. *Computers and the Humanities*, vol. 36, n. 1, pp. 283-293.
- Segre C. (1985). *Avviamento all'analisi del testo letterario*. Torino, Einaudi.
- Segre C. (1979). *Semiotica filologica. Testo e modelli culturali*. Torino, Einaudi.
- Segre C. (1981). *Testo*. In *Enciclopedia Einaudi*, vol. 14, pp. 269-291. Torino, Einaudi.
- Shillingsburg P. (1997). *Resisting texts*. Ann Arbor, University of Michigan Press.
- Shillingsburg P. (2001). *Orientations to Texts Edition*, n. 15, pp. 1-15.
- Sperberg-McQueen C. M., Burnard L. (eds) (1997). *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*. Chicago, Text Encoding Initiative.
- Stoppelli P. (ed.) (1987). *Filologia dei testi a stampa*. Bologna, Il Mulino.
- Tanselle G. T. (1987). *Textual criticism since Greg. A chronicle 1950-1985*. Charlottesville and London, The University Press of Virginia.

- Tanselle G. T. (1995). Critical editions, hypertexts, and genetic criticism. In *Romanic Review*, Vol.86, n. 3, pp. 581-593.
- Terracini B. (1968). Stilistica al bivio? Storicismo *versus* strutturalismo. *Strumenti Critici*, Vol. V, pp. 1-37.
- Tizzi E. (1996). *Fermare il tempo*, p. 2. Milano, Raffaello Cortina.
- Valeri V. (2001). *La scrittura. Storia e modelli*. Roma, Carocci
- Vanhoutte E. (2003). Display or Argument: Markup Visualisation for Electronic Scholarly Editions. In *Standards und Methoden der Volltextdigitalisierung*, pp. 71-96. (Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse. Einzelveröffentlichung. Nr. 9; Akademie der Wissenschaften und der Literatur Mainz).
- Vygotskij L. S. (1974-1990). *Storia dello sviluppo delle funzioni psicologiche superiori e altri scritti*. Firenze, Giunti.
- Vygotskij L. S. (1978 [1987-1997]). *Mind in society. The development of higher psychological processes*. Cambridge (Mass.) and London, Harvard University Press (trad. it. *Il processo cognitivo*. Torino, Bollati Boringhieri).
- Vygotskij L. S. (1998<sup>4</sup>). *Pensiero e linguaggio*. Roma-Bari, Laterza.
- Wardrip-Fruin N, Montfort N. (eds) (2003). *The New Media Reader*. Cambridge (Mass.), MIT Press.
- Zervos C. (1936). *Picasso 1930-1935*. Paris, Editions "Cahiers d'art".

## 5. NOTE

- <sup>1</sup> Sebbene il lavoro sia frutto di una discussione e di una progettazione comuni, per quanto riguarda la redazione si devono a Cinzia Pusceddu il paragrafo 1 e a Domenico Fiormonte i paragrafi 2 e 3. Alcune sezioni del paragrafo 2 riprendono e aggiornano parte del cap. 5 di Fiormonte (2003).
- <sup>2</sup> L. De Maria (acd) (1988), p. 133.
- <sup>3</sup> Prefazione a E. Tizzi (1996), p. 2.
- <sup>4</sup> Questo passo famoso viene giudicato una sorta di manifesto teorico (cfr. Corti, 1997<sup>6</sup>: 115 e Antonelli, 1985: 222).
- <sup>5</sup> Per questo filologo infatti in presenza di un tale tipo di testi l'atteggiamento del critico sarà neo o translachmanniano (cfr. Antonelli, 1985), vale a dire che il metodo lachmanniano conserva la sua validità per l'edizione di testi antichi, anche se mediato dalle riflessioni bédieriane: l'edizione critica sarà infatti da considerarsi come "un'ipotesi di lavoro" piuttosto che l'originale ricostruito.
- <sup>6</sup> "La scuola poetica uscita da Mallarmé, e che ha in Valéry il proprio teorico, considerando la poesia nel suo fare, l'interpreta come un lavoro perennemente mobile e non finibile, di cui il poema storico rappresenta una sezione possibile, a rigore gratuita, non necessariamente l'ultima. È un punto di vista di produttore, non d'utente. Sennonché, se il critico intende l'opera d'arte come un 'oggetto', ciò rappresenta soltanto l'oggettività del suo operare, il 'dato' è l'ipotesi di lavoro morale della sua abnegazione; e una considerazione dell'*atto* poetico lo porterà a spostare dinamicamente le sue formule, a reperire direzioni piuttosto che contorni fissi, dell'energia poetica. Una direttiva, e non un confine, descrivono le correzioni degli autori." (Contini, 1970: 5).
- <sup>7</sup> Cfr. Perrus (2000) e Caprettini (1985).
- <sup>8</sup> Cfr. Hay (1989), Grésillon (1994), de Biasi (2000), pubblicazioni e rivista ufficiale dell'ITEM, *Genesis* (<http://www.item.ens.fr/contenus/publications/PUBaccueil.htm>).



- <sup>9</sup> Ma in termini altrettanto espliciti si era già espressa Giaveri (1993) in un contributo che ripercorre, in modo assai più sistematico e preciso di quanto si è potuto fare qui, i legami storici e teorici fra critica delle varianti italiana e critica genetica francese.
- <sup>10</sup> “[...] because the concept of a stable work and a stable text is fundamentally flawed.” (Shillingsburg 1997: 167). Per una succinta ma utile panoramica delle connessioni fra critica genetica, *textual bibliography* anglo-americana e le altre scuole filologiche europee si veda Morrás (1999).
- <sup>11</sup> Prendo in prestito l’espressione dal saggio (in corso di stampa) di Peter Shillingsburg: *Gutenberg to Google: Electronic Representations of Print Literature*. Un’ancora valida introduzione al testo letterario in ambito digitale è Ryan (1999), mentre molto utile è la recente antologia di N. Wardrip-Fruin e N. Montfort (2003).
- <sup>12</sup> Il richiamo è a M. Bachtin (cfr. Fiorimonte, 2003: 244). In uno dei suoi ultimi scritti (“Il problema del testo nella linguistica, nella filologia e nelle altre scienze umane”, Bachtin 2000: 291-319) il critico russo parla della “interrelazione dinamica” e della “lotta” fra i due momenti che determinano “il testo come enunciazione: il suo progetto (intenzione) e l’attuazione di questo progetto” (Bachtin, 2000: 292). Il rapporto con il senso di qualsiasi comunicazione per Bachtin è sempre dialogico: “La stessa comprensione è già dialogica” (311). Una concezione del testo, che si sarebbe tentati di definire “processuale” – anche se Bachtin non adopera mai questa espressione – e che sembra offrire un modello di analisi delle scritture digitali interattive: “L’evento della vita del testo, cioè la sua autentica essenza, si svolge sempre *sul confine tra due coscienze, tra due soggetti*. [...] È un incontro di due testi: di quello pronto e di quello che si crea in reazione al primo, quindi un incontro di due soggetti, di due autori. Il testo non è una cosa...” (295).
- <sup>13</sup> Se si esclude il breve paragrafo su “l’uso iconico della sostanza” (fisicità delle lettere, ecc., cfr. Segre, 1985: 55-57) manca un riferimento alla forma del testo e ai supporti materiali, un argomento che non trova riscontri nell’area della critica strutturalista. Proprio un anno dopo la pubblicazione dell’*Avviamento a Londra* esce *Bibliography and the sociology of texts* di Donald McKenzie (1986): “Un testo [...] è sempre iscritto in una qualche materialità: quella dell’oggetto scritto che lo trasmette, quella della voce che lo legge o lo recita, quella della rappresentazione che lo fa intendere. Ciascuna di queste forme è organizzata secondo strutture proprie che svolgono un ruolo essenziale nel processo di produzione del senso. Per restare nell’ambito dello scritto stampato, il formato del libro, la disposizione dell’impaginazione, i modi di segmentazione del testo, le convenzioni tipografiche sono tutti elementi investiti di una ‘funzione espressiva’ che contribuiscono alla costruzione del significato. [Per McKenzie] ‘Forms effect meaning’...” (Chartier, 1999: 99).
- <sup>14</sup> Per una discussione sui problemi teorici della codifica informatica del testo si rimanda ai contributi di Buzzetti (2000), Ciotti (2001), Renear et al. (1999), Orlandi (1995), Renear (1997 e 2001). Su TEI e SGML/XML si vedano gli ormai classici Robinson (1994), Sperberg-McQueen, Burnard (1997) e Burnard (1995).
- <sup>15</sup> Mi riferisco qui a Alois Pichler e Claus Huitfeldt, curatori e codificatori del Nachlass di Wittgenstein (cfr. Renear, 1997: 122-123). Daniel Ferrer esprime in modo chiaro le difficoltà dell’editore genetico di fronte alle forme di rappresentazione lineari: “[...] the first page of a novel is naturally linked to the second page (regardless of the network of semantic and formal connections it may weave with other parts of the text). But what about the first page of the draft of a novel? It is naturally linked to the second page of this draft – but just as naturally, albeit in another manner, with the second version of the first page. So, a narrative order, or more generally a textual order, is opposed to a genetic order.” (Ferrer, 1995: 143).
- <sup>16</sup> “Historical investigation into the growth of literary works must start with the physical objects that attempt to convey the texts of those works, but it must move on to reconstructions that aim to bring the preserved texts into closer agreement with what was intended by someone at some past time. When we talk about literature (not just as editors, but as readers), we are inevitably referring to critically reconstructed texts. Historical reconstructions are never certain, nor are the texts of literary works in any of their stages. But those uncertainties, those critical judgments, are what we have to live with as students of literature.” (Tanselle, 1995: 592-593).
- <sup>17</sup> Il riferimento qui è alle discussioni pro e contro la *New Philology* (che tanto dovrebbe all’*Éloge de la variante* di Cerquiglini). Si vedano Nichols (1990); Busby (1993); Gleßgen; Lebsanft (1997) e Cherchi (2001).

- <sup>18</sup> Cioè edizione condotta con l'ausilio del computer, e non edizione su supporto elettronico. Il gruppo di ricerca di Mordenti vi arriverà più avanti con l'edizione ipertestuale degli Zibaldoni di Boccaccio (<http://rmcisadu.let.uniroma1.it/boccaccio/ipertesto/modello.htm>).
- <sup>19</sup> È per brevità che riassumiamo così l'insieme delle scuole teoriche che valorizzano i "renditional features" (Schreibman, 2002: 285): bibliografia materiale-sociologica anglo-americana, *critique génétique* francese, critica delle varianti italiana.
- <sup>20</sup> "Si comprende che la parola *textus* sia stata elaborata in un mondo cristiano – per questo aspetto, anzi, giudaico-cristiano – che riteneva le tavole della legge 'scritte con il dito di Dio' (Esodo, 31, 18), il quale con questo rende sacro lo stesso atto dello scrivere. [...] ma è utile ribadire sin dall'inizio che *la natura del testo è condizionata dai modi della sua produzione e riproduzione, che insomma il testo non è una realtà fisica ma un concetto limite.*" (Segre 1981: 269; corsivo nostro).
- <sup>21</sup> Scrive Dario Corno nell'introduzione all'edizione italiana: "[...] dunque la scrittura si trova a lavorare, al suo livello semiotico più elevato di processo che presiede non solo al modo di "dire le cose" ma allo stesso "modo di pensare" [...]. La scrittura, a questo livello (*knowledge transforming*), esercita un potente influsso cognitivo: fa notare le cose, crea contenuti e induce a prestare attenzione ad aspetti e proprietà che prima erano sfuggite o non si erano proprio viste." (Corno, 1995: 36). Si tratta di riflessioni di cui la psicologia della composizione è debitrice nei confronti di uno dei padri della psicologia moderna, Lev Semenovič Vygostkij (cfr. Fiormonte, 2003: 220-225, 240-243).
- <sup>22</sup> Sito ufficiale: <http://www.selc.ed.ac.uk/italian/digitalvariants>. *Mirror* italiano: <http://www.digitalvariants.org>.
- <sup>23</sup> Sul problema della rappresentazione delle strutture temporali attraverso i linguaggi di markup (e in particolare con XML-TEI) cfr. Schimdt (2004) e Vanhoutte (2001).
- <sup>24</sup> La *Genetic Machine* è raggiungibile all'indirizzo: <http://www.digitalvariants.org/home.htm> [http://www.selc.ed.ac.uk/italian/digitalvariants/autori/magrelli/mag\\_index.htm](http://www.selc.ed.ac.uk/italian/digitalvariants/autori/magrelli/mag_index.htm)  
La complessità di una codifica XML-TEI dell'avantesto di *Ora Serrata Retinae*, e in particolare di alcuni autografi, è stata messa in luce da un recente tesi (Martiradonna, 2004).
- <sup>25</sup> Per un'analisi critica di *Ora serrata retinae* vedi Lisa (2004). Una versione preliminare del saggio introduttivo dedicato all'avantesto è disponibile sul sito (vedi nota precedente).
- <sup>26</sup> Un po' come avviene per l'autore studiato e amato da Magrelli, Paul Valéry: "[le manuscrit] montre l'une (parmi près d'une quinzaine) des tentatives de Paul Valéry pour commencer son poème *Été*. Mots et graphisme se conjuguent (ou se concurrent) pour figurer l'imaginaire – paysage de mots, paysage de traits – ou pour signifier la pensée: dynamomètre et vecteurs empruntent à la physique et aux mathématiques le symbolisme de la tension, du couple de forces.» (Hay, 1989: 11).



# Lo stato dell'analisi testuale informatizzata: riflessioni a partire da un esperimento

*Daniele Silvi*

Università degli studi di Roma "Tor Vergata"  
(silvi@email.it)

*The paper summarizes the state of the art of present technologies for textual analysis describing the main tools with reference to their specific functionalities. The paper also presents the use of one of the main de facto standards in the field (the TACT program) for the study of textual variants and for applications in the field of spoken language.*

*Keywords:* textual analysis, concordances, textual variants

## 1. LO STATO DELL'ANALISI TESTUALE INFORMATIZZATA

L'analisi testuale si definisce come l'incontro di tre discipline: la critica letteraria tradizionale (nell'aspetto specifico della testualità), l'informatica e la statistica. Gli strumenti di queste discipline si devono quindi incontrare e generare una sinergia che possa sviluppare, di fatto, strumenti nuovi.

Analizzare un testo, a qualunque livello e con qualunque mezzo, significa porgli delle domande. Queste domande possono cambiare a seconda del ricercatore, dei tempi in cui vive e delle concrezioni di significato che si sono stratificate sull'opera.

Si tratta dunque di una operazione antica che gode oggi i vantaggi offerti dalle moderne tecnologie e che ha, di conseguenza, allargato i proprio orizzonti. La discussione aperta è se questa nuova disciplina può ampliare i "limiti dell'interpretazione" comunemente accettati. Tuttavia l'avvento del computer nelle scienze umane non da tutti è stato accolto con favore: spesso questo nuovo, misterioso oggetto è stato salutato con diffidenza e anche con timore:

[...] "perché pare quasi una sorta di contaminazione l'accostamento del mondo della qualità (letteraria dei testi) con quello della quantità (individuata, elaborata, per mezzo del computer)"<sup>2</sup>.

Tuttavia dagli albori di padre Busa ad oggi, gli strumenti dell'analisi testuale si sono affinati e importanti traguardi sono stati raggiunti: un decisivo passo avanti si sta facendo, ad esempio, nel campo delle reti semantiche, che sembrano finalmente

risolvere – almeno parzialmente – l’annoso problema dell’analisi semantica e contestuale, finora tralasciata dagli attuali *software*, che non riconoscono la parola in quanto portatrice di uno o più significati, quanto piuttosto la “forma grafica”<sup>3</sup>. In Italia il progetto più importante è quello portato avanti dalla *Expert System*, attiva dal 1989. Il cuore delle tecnologie sviluppate da questa società è il “sensigrafo”, una rete semantica strutturata su base concettuale e non in ordine alfabetico (come un tradizionale dizionario). Parole e nomi sono raggruppati ed intercorrelati in nodi semantici (tramite rapporti di peronimia ed iponimia), nodi della rete che si arricchiscono delle caratteristiche e del significato (letterale o traslato) dei termini a cui sono collegati. Attualmente tale rete conta all’incirca 300.000 nodi. A questa base vengono poi applicati diversi programmi che permettono l’analisi delle frasi, l’interpretazione del linguaggio naturale, la catalogazione per argomenti, ecc<sup>4</sup>.

Ancora oggi è vero che i *software* per l’analisi testuale sono stati sviluppati maggiormente per i sistemi Ibm/Dos/Windows, piuttosto che per quelli Macintosh/Unix/Linux, tuttavia il notevole impulso di questi ultimi, soprattutto nel campo della ricerca, fornisce valide motivazioni alle produzioni future proprio in questa direzione.

L’uso dei programmi di analisi testuale è stato finora proposto più come esperimento, piuttosto che come lavoro sistematico, per mostrare le potenzialità dei programmi e le possibilità di applicazione. Questi “esperimenti” hanno però sollevato diversi problemi teorici; un esempio è la presentazione dei dati e la ricombinazione dei medesimi: infatti il poter “ricostruire” un testo tramite gli *output* di un programma di analisi testuale può portare anche a fraintendimenti. Inoltre i risultati possono venire presentati con dei *software* di visualizzazione, che nascono in ambienti puramente ingegneristici o matematici. Il problema della presentazione e interpretazione dei risultati è stato discusso da John Bradley e Geoffrey Rockwell alla conferenza di ALLC/ACH svoltasi a Parigi nel 1994<sup>5</sup>.

Tra i numerosi problemi teorici forse il primo da affrontare è quello relativo alla distinzione tra autore semantico e autore editoriale: la seconda voce racchiude molti aspetti (casa editrice, tipografia, redazione, collana). Ci si chiede quindi a cosa viene applicata l’informatica? Quale è il “testo” reale oggetto di questa disciplina? Quali aspetti vanno attribuiti all’editore e quali all’autore? Possiamo dire, con buona approssimazione, che capitoli, paragrafi, strofe, stanze, ecc possono essere di buon grado attribuiti all’autore semantico; salti di riga e salti di pagina, caratteri usati, ecc sono di paternità editoriale. Questo problema teorico ha immediatamente delle conseguenze pratiche (ed ecco il motivo del nostro interesse) poiché è lecito chiedersi se è sempre vero che vanno conservati solo gli aspetti del “testo che si legge”, quando invece conservare salti di righe e di pagina può aiutare a fare la storia di un’edizione e a capire se ci troviamo di fronte ad una ristampa o ad una riedizione. Queste scelte vanno fatte prima di iniziare l’analisi testuale informatizzata e compongono le delicate fasi della costruzione del “modello”.

Infatti, prima di procedere all’analisi di un testo, è necessario avere chiari quali sono gli scopi della nostra operazione. In altre parole dobbiamo sapere cosa andare a ricercare, qual è l’ipotesi di partenza, quali sono gli strumenti di cui necessitiamo;

questo per non correre il rischio di perdersi in una quantità sovrabbondante di dati numerici che finirebbero col perdere di significato e confondere le idee. Si tratta dunque di costruire un “modello”, schematizzare cioè quale parte del testo si intende mettere in rilievo per poi interrogarla; dunque, sostanzialmente, compiere una ricerca mirata ad aspetti specifici, quali possono essere l’analisi delle varianti, oppure l’isolamento di determinate aree semantiche, o ancora la frequenza e distribuzione di sostantivi di particolare interesse. I risultati ottenuti devono essere poi valutati dal critico, alla luce dei suoi studi e della sua pratica, senza aspettarsi quindi che il computer possa generare giudizi definitivi e inoppugnabili, poiché affidare tutte le proprie sicurezze ad uno strumento, per quanto potente, potrebbe distoglierci dallo scopo, che è comunque quello di “fare letteratura” e non di diventare degli ingegneri: il connubio tra letteratura ed informatica è tutt’altro, è una “terra di mezzo” in cui due mondi si incontrano e procreano una disciplina nuova, che si occupa di letteratura ma anche di informatica: l’informatica umanistica, il cui principale oggetto di studio è il testo elettronico, ovvero la sua produzione, elaborazione ed analisi.

Naturalmente esistono diversi livelli di edizione elettronica di un testo (dipendenti anche dal tipo di ricerca che si intende effettuare); possiamo principalmente distinguere in:

1. Edizione meccanica (immagine digitalizzata da ingrandire, stampare, ruotare, ritagliare)
2. Edizione diplomatica (riproduzione dell’esemplare)
3. Edizione critica (restituire il testo alla sua originalità)

Nel passaggio alle seconde due forme si rischia la proliferazione di prodotti a basso profilo, poiché non esiste un controllo sui testi che vengono distribuiti sulla rete Internet.

Questo pericolo ultimamente si è maggiormente delineato nella sua pericolosità poiché molti siti amatoriali sono nati sotto la stella della pubblicazione di testi (letterari) in rete. Naturalmente il bisogno di consultazione di un testo elettronico ha favorito la fruizione di qualunque versione sia reperibile in rete, senza che l’utente si preoccupi della “qualità” di questa. Il problema nasce per la generale disinformazione che dilaga tra gli utenti non esperti, che non sanno come poter riconoscere un prodotto “buono” da uno approssimativo o inaffidabile: tutte le massaie sanno come riconoscere un prodotto alimentare sano da uno di scarsa qualità, e lo stesso discorso può essere fatto per i prodotti della casa, ecc., ma quanti studenti o insegnanti di scuola media (superiore e inferiore) sono in grado di riconoscere un “buon” testo elettronico da uno infimo? Da queste domande nasce la riflessione sulla necessità di una alfabetizzazione sistematica o di un maggior controllo sulla qualità dei prodotti che finiscono in rete.

Gli interventi sull’analisi testuale si sono moltiplicati, soprattutto sul web, come quelli (numerosi) pubblicati da Francesca Tomasi su *GriseldaOnLine*<sup>6</sup>, o i diversi articoli pubblicati sulle riviste ufficiali come *Literary and Linguistic Computing*<sup>7</sup> o *Computers and the Humanities*.<sup>8</sup> Riguardo quest’ultima vale la pena di citare un articolo risalente al 1998 che affronta un tema decisamente importante per l’analisi dei testi: l’attribuzione autorale

di un'opera.<sup>9</sup> Anche questo campo meriterebbe un saggio apposito, visti i problemi che veicola e che, come emerge dall'articolo di Rudman, sono sempre gli stessi: approccio dilettantistico alla disciplina, scarsità di materiali validi da analizzare, mancanza di tecniche standard di analisi, ecc. Ovviamente con l'aiuto di un computer e di una solida banca di testi elettronici si possono fare confronti stilometrici impensabili al lavoro umano e questo significa un grosso passo avanti nell'attribuzione dei testi di incerto autore. Si tratta di un campo ancora non molto praticato ma che sicuramente offre vaste possibilità applicative.

Per concludere questa breve rassegna citiamo un recente lavoro pubblicato dalla S.I.S.M.E.L.<sup>10</sup> dal titolo *L'analisi testuale con TACT: un'esperienza nel campo dell'inglese antico*, di Loredana Teresi.<sup>11</sup> In questo caso il programma di analisi testuale (Tact, di cui si parlerà più diffusamente nei prossimi paragrafi) viene utilizzato per ricostruire lo *stemma codicum* di alcuni manoscritti che si suppongono avere un progenitore comune.

Nel prossimo paragrafo descriveremo le diverse forme del *text retrieval*, vale a dire i principali indici misurabili che possiamo ottenere da un testo. Grazie a questi indici il ricercatore può servirsi dei metodi statistici per operare valutazioni stilometriche sulle opere.

## 2. LE FORME DEL TEXT RETRIEVAL

Stabilito che il *computer* è una macchina, e che l'operazione che più gli è propria è quella di fare dei calcoli, dobbiamo ora capire cosa, in un testo, può essere misurato e matematicamente (numericamente) espresso. Il punto di partenza è il "mattoncino di base" di ogni testo: la parola<sup>12</sup>.

A partire dal conteggio di queste si possono generare:

1. **Indice:** elenco delle parole contenute in un testo, ordinate secondo un criterio definito, avulse dal proprio contesto;
2. **Concordanza:** elenco delle parole contenute in un testo, ordinate secondo un criterio definito, inserite in un contesto che permette di coglierne il senso, accompagnate da indicazioni che ne permettono la localizzazione all'interno della struttura (precedentemente definita) del testo;
3. **Lista di frequenze:** elenco delle parole contenute in un testo, accompagnate dal numero di occorrenze con cui ricorrono nel testo stesso, ordinate secondo il criterio alfabetico (diretto o inverso) o quello di frequenza;
4. **Indice lemmatizzato:** elenco dei lemmi contenuti nel testo, ottenuti riconducendo le forme flesse a quelle di base. L'ordine può essere alfabetico o per categorie semantiche (a seconda di come viene realizzata la lemmatizzazione).

## 3. I SOFTWARE DISPONIBILI

Prima di procedere all'analisi di un testo è necessario avere chiari quali sono gli scopi della nostra operazione. In altre parole dobbiamo sapere cosa andare a ricercare, qual è l'ipotesi di partenza, quali sono gli strumenti di cui necessitiamo; questo per non correre

il rischio di perdersi in una quantità sovrabbondante di dati numerici che finirebbero col perdere di significato e confondere le idee. Si tratta dunque di costruire un modello, schematizzare cioè quale parte del testo si intende mettere in rilievo per poi interrogarla; dunque, sostanzialmente, compiere una ricerca mirata ad aspetti specifici, quali possono essere l'analisi delle varianti, oppure l'isolamento di determinate aree semantiche, o ancora la frequenza e distribuzione di sostantivi di particolare interesse. I risultati ottenuti devono essere poi valutati dal critico, alla luce dei suoi studi e della sua pratica, senza aspettarsi quindi che il computer possa generare giudizi definitivi e inoppugnabili, poiché affidare tutte le proprie sicurezze ad uno strumento, per quanto potente, potrebbe distoglierci dallo scopo, che è comunque quello di "fare letteratura" e non di diventare degli ingegneri: il connubio tra letteratura ed informatica è tutt'altro, è una "terra di mezzo" in cui due mondi si incontrano e procreano una disciplina nuova, che si occupa di letteratura ma anche di informatica: l'informatica umanistica, il cui principale oggetto di studio è il testo elettronico, ovvero la sua produzione, elaborazione ed analisi.

Oltre ai programmi che analizzano il testo ne dobbiamo distinguere altri progettati per la "collazione automatica": si tratta di metodologie informatiche volte al confronto di più redazioni dello stesso testo. Vari sono stati gli esperimenti in questa direzione negli ultimi dieci anni, ma ancora non si è approdati ad un vero e proprio sistema generalizzato e standardizzato. Nell'elenco che segue i *software* sono descritti in base alle loro principali caratteristiche e alle forme di text retrieval utilizzate:

**Collate 2**, è un programma di supporto alla preparazione di un'edizione critica basata su un gran numero di manoscritti, elaborato da P. Robinson nel 1994. L'applicazione è in grado di eseguire la collazione simultanea di un numero massimo di 100 manoscritti ed è in grado di gestire facilmente tutte le note accessorie che riguardano il testo, come per esempio i commenti dell'editore o l'indicazione di una glossa, ecc. L'editore naturalmente può intervenire sui risultati della collazione automatizzata, che non sempre potrà risultare soddisfacente, dal momento che il sogno di un "filologo elettronico" sembra destinato a rimanere tale ancora per molto tempo. È inoltre possibile in qualsiasi momento mutare il testo scelto come base per la collazione; la registrazione delle varianti può segnalare tutti i casi in cui un manoscritto si accorda con il testo base o i casi in cui se ne discosta. I risultati della collazione possono essere visualizzati e stampati in diversi formati: il formato standard presenta un tradizionale apparato critico, ma è anche possibile visualizzare le righe di ogni manoscritto su linee diverse, separando le singole parole in modo da vedere a colpo d'occhio le diverse varianti; infine nel *formatted style* all'editore viene lasciata ampia libertà di determinare quale sarà l'aspetto della collazione, determinando quali dati e in quali ordine dovranno apparire nell'apparato critico; alcuni modelli di formati consentono di pubblicare edizioni conformi alle raccomandazioni della Text Encoding Initiative (TEI), di interagire con il programma EDMAC, o di pubblicare un'edizione elettronica ipertestuale. Il programma Collate 2 è disponibile per il momento solo per la piattaforma Macintosh; il progetto per una completa versione per Windows, che avrebbe dovuto



esser pronta per il 1997, non sembra per il momento aver portato ad alcun risultato concreto, almeno in base alle informazioni disponibili in linea.

**Tustep** (Tuebingen System of Text Processing Programs), iniziato ad essere sviluppato nel 1966, il Tustep è in grado di effettuare la collazione automatica di più testi, di eseguirne le correzioni in base a precise modalità istruite alla macchina, scomporre il testo, creare elenchi bibliografici, preparare indici, ordinare in vari criteri parole o frasi o porzioni di testo, convertire i formati dei *files* di testo per renderli accessibili ad altri programmi. Rimane da notare che il programma è in lingua tedesca e ancora non ne esistono traduzioni, fatta eccezione per una traduzione in lingua spagnola del solo manuale. Esistono le versioni per Windows e per Unix (solo la vers. 2002).

**EDMAC**, è un'applicazione creata da J. Lavagnino e D. Wujastyk che raccoglie sequenze complesse di comandi (comunemente chiamate macro) per il sistema plain TeX per consentire la preparazione di un'edizione critica dall'aspetto assai sofisticato: grazie ad EDMAC è possibile per esempio avere una numerazione del testo linea per linea, lemmi con riferimento al numero di linea, sei livelli di note a piè di pagina, ognuno dei quali con una sua formattazione particolare.

**Critical Edition Typesetter**, è un programma di supporto alla creazione di edizioni critiche per MS-DOS, OS/2, Windows 95 e Windows NT o versioni superiori. Il programma si compone di 6 moduli:

1. Un *word processor* per la digitazione del testo da pubblicare
2. Un preprocessore che converte il testo digitato in un file TeX
3. Un programma che consente di comporre il testo secondo il sistema TeX
4. Il pacchetto EDMAC, che permette di utilizzare le funzionalità speciali di TeX per la creazione di un apparato critico
5. Un convertitore di file in formato PostScript
6. Un interprete dei files PostScript che consente in primo luogo di stampare il testo anche su stampanti che non sono in grado di riconoscere questo formato e inoltre anche di visualizzare le pagine sullo schermo.

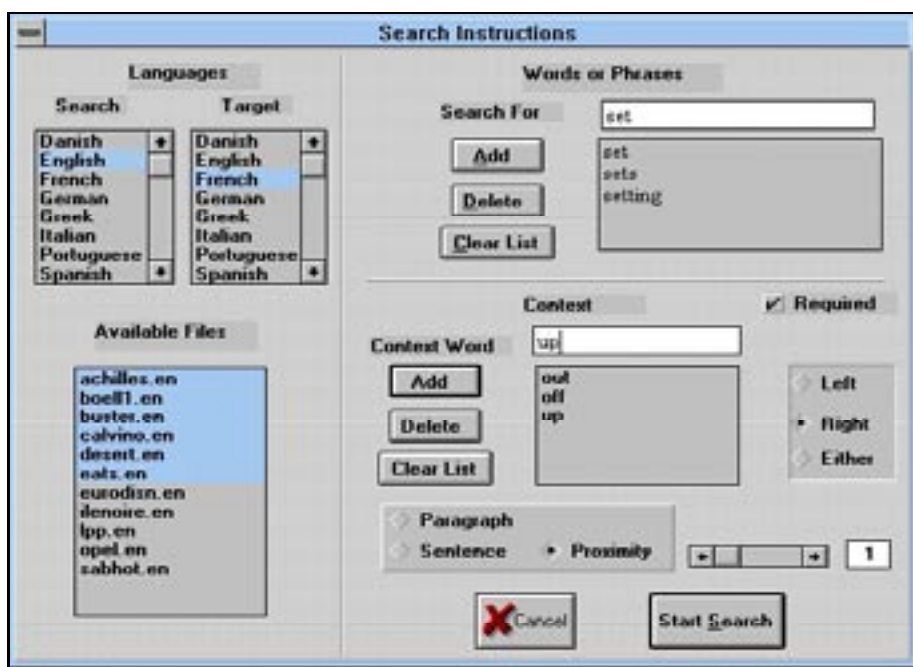
**Classical Text Editor**, (CTE), è un programma di word processing scritto da S. Hagel esplicitamente pensato per la redazione di edizioni critiche e commenti di testi classici, anche in formato elettronico. L'applicazione funziona per la piattaforma Windows, ma può girare anche sui computer Macintosh in cui sia montato il software di emulazione Virtual PC. Secondo quanto riportato nella pagina Classical Text Editor Features, il programma consente tra l'altro di:

1. Creare un apparato critico su dodici livelli e riferimenti alla linea di testo rilevante dall'apparato critico e dagli indici.
2. Creare una copia dell'edizione camera-ready per la stampa o in una versione postscript; è possibile esportare la propria fatica in formato HTML o in formato TEI, per la pubblicazione elettronica su CD-ROM o in Rete.

3. Lavorare in modo flessibile con le sigle dei manoscritti, per esempio definendo le sigle per ciascun testimone o gruppo di essi.
4. Lavorare con testi in scrittura sinistrorsa, come l'ebraico o l'arabo.
5. Creare automaticamente indici con riferimenti al testo e alle note.

**Verbum.** Il programma “Verbum 2.0 – Analisi Testuale”, adatto per PC con Windows (3.1, 95 e 98), è un “data base testuale” o “text processor”, per l’analisi di testi di ogni genere. Esso funziona sotto DOS, interagisce con Windows e con Word, e consente, a partire da un testo dato, una facile realizzazione di indici, concordanze, ricerche contestuali, e alcune analisi statistiche, di leggibilità, di significatività delle parole e delle frasi; crea inoltre delle tabelle parole-frequenza esportabili nei più comuni programmi di statistica avanzata tipo SPAD oppure SPSS. Il programma lavora secondo pre-codifiche e suddivisioni del testo impostate con un comune word processor tipo Word per Windows. Può essere d’ausilio allo studio critico della letteratura, ma l’impiego tipico è nell’elaborazione di questionari a risposte libere, di colloqui clinici, di sceneggiature cinematografiche e teatrali, nella “prova su strada” dei testi di facile lettura destinati alle scuole elementari, nell’auto-verifica del proprio stile di scrittura col *word processor*, e così via. Insieme al programma è fornito un manuale in formato Word con le istruzioni dettagliate per l’uso, ed alcuni file di esempio.

**Multiconcord,** permette di confrontare automaticamente testi originali e la loro traduzione in dieci lingue europee. Sostanzialmente opera il raffronto di diverse traduzioni dello stesso testo, allineate, e raffrontabili in sinossi. Il ricercatore può inserire una o più



parole da ricercare ed il programma ne visualizzerà le occorrenze nelle diverse lingue e nei diversi contesti<sup>13</sup>. Il sistema è progettato per girare sotto ambiente Windows.

*Tact*, Questo programma deve il suo nome al fatto di essere non un programma, bensì un sistema di sedici programmi riuniti in un unico pacchetto ed accessibili dallo stesso menu.<sup>14</sup> Progettato per operare sotto il sistema operativo Ms-Dos,<sup>15</sup> questo software è in grado di effettuare ricerche e analisi sulla base di testi letterari opportunamente codificati; in genere i ricercatori lo usano per elencare le occorrenze di una parola, combinazioni di parole o gruppi. Quello che il programma è in grado di generare sono liste, tabelle, grafici, concordanze, passando dalla semplice elencazione della frequenza di parole, lettere o frasi fino ad arrivare alle statistiche del rapporto *type-token*, creare anagrammi, dizionari lemmatizzati, collocazione delle parole in base ai loro rapporti con il resto del testo, ecc.

Tact è un sistema multilingue, il che significa che per supportare lingue straniere si serve della tabella ASCII estesa ed è in grado, con l'uso di opportuni *editor* di *font*, di coprire tutti i linguaggi europei moderni (francese, tedesco, greco, etc.); questo perché nella tabella ASCII standard si trovano le cifre numeriche, le lettere maiuscole e minuscole (maiuscole e minuscole hanno codici ASCII differenti) la punteggiatura, i simboli aritmetici e altri simboli (\$, &, %, @, #, ecc.). Essendo stata concepita in America, tale tabella non comprende le lettere accentate (sconosciute all'ortografia inglese). I primi 32 byte della tabella standard sono inoltre riservati per segnali di controllo e funzioni varie.

Tuttavia Tact utilizza una particolare estensione del codice ASCII, e cioè il formato testo Ms-Dos. Il set di caratteri utilizzato dall'Ms-Dos corrisponde, per i primi 128 caratteri, alla tabella ASCII standard, i secondi 128 caratteri sono invece individuati da tabelle che cambiano a seconda dei paesi ed etichettate con dei codici (code page). Il Dos usava dei codici a 8 bit che estendevano l'ASCII standard in modo differenziato a seconda dell'area geografico-linguistica di riferimento; le varie tavole venivano indicate con un numero (850 quella per l'italiano, 437 per l'inglese...). Il codice della nazione è molto spesso il prefisso telefonico della nazione, come risulta da questa (parziale) tabella riassuntiva:

Code page	Definizione
437	Latin MS-DOS (U.S.A.)
737	Greco Windows 95
850	Latin 1 MS-DOS (Europa Ovest)
852	Latin 2 MS-DOS (Europa Est)
861	Islandese MS-DOS
866	Cirillico MS-DOS (Russia)
932	Giapponese Shift-JIS MS-DOS
936	Cinese semplificato MS-DOS
949	Coreano Hangul MS-DOS
950	Cinese tradizionale MS-DOS

Tabella 1

Per preparare un testo all'analisi di Tact bisogna dunque, come prima cosa, salvarlo nel formato Ms-Dos. Per compiere questa operazione tramite il diffuso Microsoft Word<sup>16</sup> si devono seguire alcuni semplici passi:

1. Fare clic sul menu "strumenti"
2. Selezionare "opzioni"
3. Scegliere la cartella "standard"
4. Barrare la casella "conferma conversione all'apertura"
5. Salvare il testo come "testo normale"
6. Scegliere la codifica "ms-dos"

In alcuni casi bisogna fare particolare attenzione, poiché spesso Word si arroga il diritto di fare delle sostituzioni automatiche che Tact non riconosce. Questi sono i casi più frequenti:

1. L'apostrofo inclinato ' o '. Word (tramite il correttore automatico) assegna automaticamente l'inclinazione all'apostrofo usato a seconda che apra o chiuda una porzione di testo. Il formato Ms-Dos non contempla queste forme ma ne conosce una sola, comune a tutti gli usi, l'apostrofo diritto o verticale '. Pertanto (con Office XP) al momento del salvataggio nel formato ms-dos,<sup>17</sup> il box ci avvertirà che questi segni sono incompatibili con il formato scelto. Per sostituirli si deve utilizzare lo strumento "trova/sostituisci" e copiare nella riga del testo da cercare i due tipi di apostrofo sbagliati (prima uno e poi l'altro) ed inserire nella riga del testo da sostituire il codice ascii ^39, che corrisponde all'apostrofo diritto.

2. Il tratto unificatore lungo: —. Questo segno non è contemplato nel set di caratteri utilizzato dal Dos e serve ad unire una parola composta da due elementi. Va sostituito con il segno unificatore corto, normalmente usato e presente nelle tastiere: -. Per la sostituzione è sufficiente copiare e incollare il tratto lungo nella riga “trova” e inserire il normale trattino nella riga “sostituisci”.
3. I tre puntini di sospensione: ... Sempre in virtù del correttore automatico, Word sostituisce la sequenza di tre puntini con una simile, dove i puntini sono però più distanziati tra loro. Questa nuova entità carattere non è contemplata



dal dos, che vede solamente i puntini in quanto tali e non li interpreta come “di sospensione”. Per operare la sostituzione, ancora una volta, copiare ed incollare i tre puntini di sospensione “creati” da Word ed inserire nella riga del “sostituisci” tre puntini normali, digitandoli direttamente.

Gli esperimenti presentati nel par. 4 sono stati portati avanti con l’uso di Tact, pertanto ci sembra giusto dedicare alla descrizione di questo programma uno spazio maggiore. Va subito detto che gli autori (John Bradley e Lidio Presutti) hanno iniziato la progettazione del software nel 1984 presso l’Università di Toronto. Nel 1995 il programma fu adattato alla rete per poter fruire di database testuali che si trovassero *on-line*: il risultato di questa operazione fu la creazione di TACTweb. In questo modo chiunque avrebbe potuto accedere alle maschere di ricerca e alle opzioni del programma senza possedere il programma stesso. Il *demonstration server* di Tact è accessibile all’indirizzo <<http://tactweb.humanities.mcmaster.ca/>>.

Anche gli esperimenti di ricerca sull'analisi statistica del testo si sono moltiplicati, giungendo ad esempi eccellenti, dai primi lavori, come quello di John Bradley<sup>18</sup> and Geoffrey Rockwell che nel 1992 hanno presentato i risultati di una ricerca sul testo di D. Hume *Dialogues Concerning Natural Religion*, un'opera di circa 120 pagine<sup>19</sup>.

## GATTO

Dal 1985, col nome Centro Studi Opera del Vocabolario Italiano, l'attuale Istituto Opera del Vocabolario Italiano (OVI) del CNR ha ereditato dall'Accademia della Crusca il compito di realizzare il Vocabolario Storico della Lingua Italiana.

Il progetto iniziale era quello di un vocabolario che andasse dalle origini ai giorni nostri. In seguito apparve evidente che, data la limitatezza delle risorse disponibili, era opportuno procedere concentrandosi inizialmente sulla sezione cronologica antica, ovvero il Tesoro della Lingua Italiana delle Origini, limitando per il momento le risorse destinate alle fasi successive. L'oggetto della ricerca per la quale il programma è stato implementato è la lingua dei primi secoli, valutabile in 18-20 milioni di occorrenze. Si è dunque trattato, in prima istanza, di tradurre in formato elettronico questa vasta documentazione; i testi sono stati inoltre corredati delle necessarie codifiche ed in seguito si è effettuata una lemmatizzazione, in gran parte manuale. Questo perché l'estrema varietà linguistica presente nei testi presi in considerazione rende del tutto inapplicabile il ricorso a lemmatizzatori automatici, risultando di gran lunga più efficiente una lemmatizzazione manuale, sia pure assistita dal programma.

L'esigenza di disporre di uno strumento in grado di gestire in maniera integrata le diverse fasi di questo lavoro ha portato alla nascita di GATTO, il cui nome è per l'appunto un acronimo che sta per "Gestione degli Archivi Testuali del Tesoro delle Origini". GATTO opera sotto tutte le versioni di Windows a 32 bit, a partire da Windows 95 ed è articolato sotto i tre seguenti, principali, aspetti:

1. Gestione base dati
2. Lemmatizzazione
3. Ricerche ovvero interrogazione della base di dati

I requisiti di memoria dipendono essenzialmente dal sistema operativo in uso. Nel caso di corpora testuali di dimensioni rilevanti, i requisiti di memoria considerati minimi per le diverse versioni di Windows potrebbero rivelarsi non del tutto adeguati.

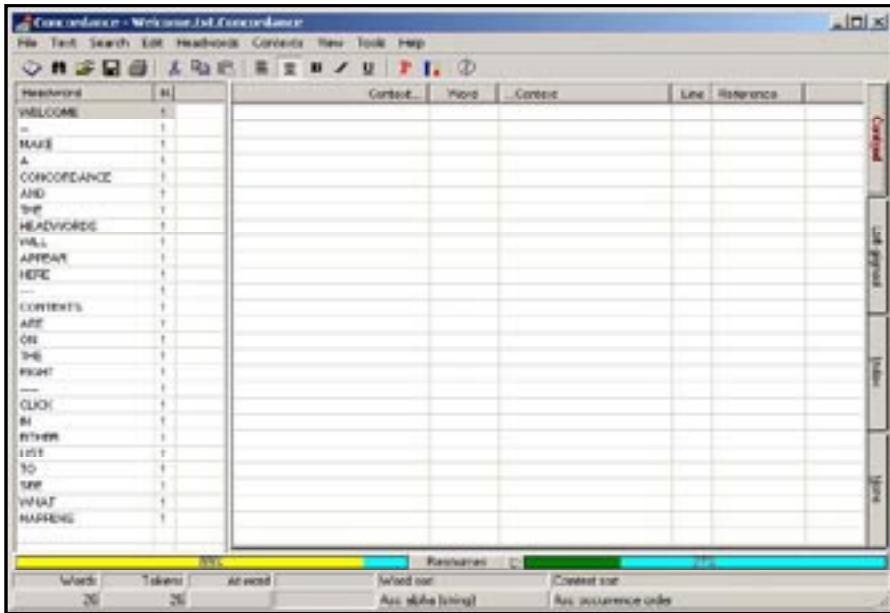
Il programma crea ed utilizza *directory* aventi nomi scelti dall'utente ed estensione "GAT". È consigliabile evitare la presenza, nei dischi del PC e in quelli delle macchine ad esso collegate, di altre *directory* aventi un nome includente l'estensione "GAT".

Per quanto riguarda le condizioni di utilizzo GATTO è un programma registrato ed è protetto dalle vigenti leggi italiane ed internazionali sulla tutela del software. È vietata ogni duplicazione o riproduzione non espressamente autorizzata. Più precisamente, può essere liberamente usato e riprodotto per scopi di studio e ricerca senza fini di lucro, con l'obbligo della citazione in tutte le pubblicazioni

che ne derivino e in tutte le applicazioni nelle quali venga utilizzato. Qualunque altro uso deve essere oggetto di un accordo scritto preventivo con l'Istituto Opera del Vocabolario Italiano.

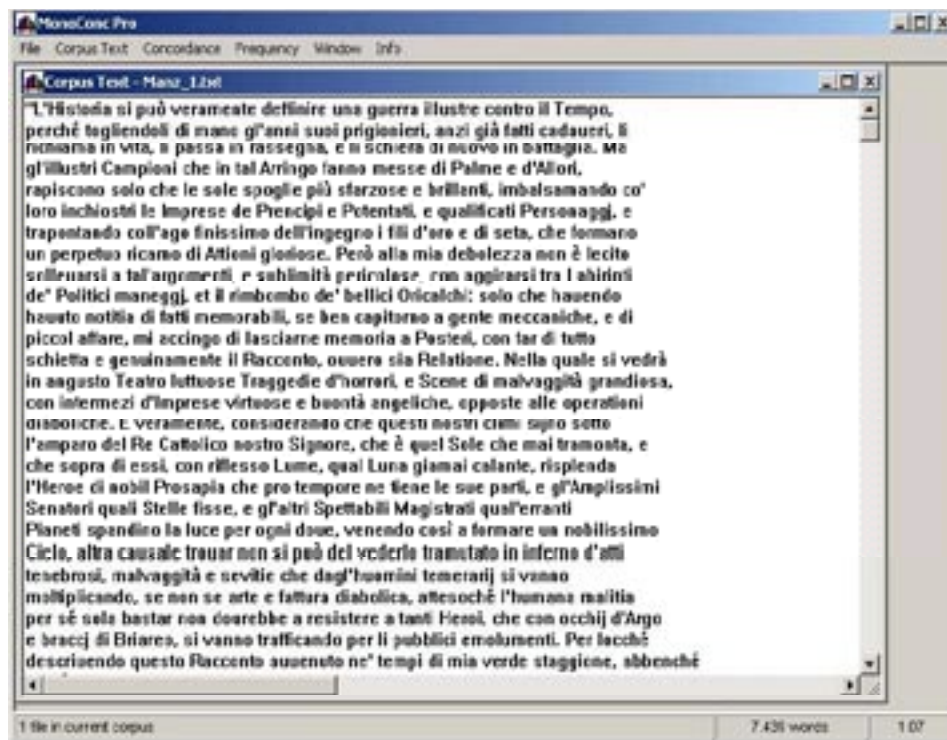
Riguardo al formato dei testi analizzabili questi devono essere di tipo "ANSI", nel testo va inserita la struttura tipografica, per cui bisogna introdurre al termine di ogni riga, dove voluto, un fine paragrafo, ottenuto normalmente col tasto <Invio> ed una serie di codifiche che ovviamente non devono essere modificate dal programma di videoscrittura utilizzato: ciò rende inadeguato Microsoft Word il quale, a partire dalla versione 97, introduce automaticamente una serie di modifiche al testo incompatibili con quanto detto; l'unica versione che funziona correttamente è la 6.0, nel formato 'Solo testo', recuperabile con l'uso di "WordPad".

**CONCORDANCE**, questo programma è stato sviluppato da R.J.C. Watt, università di Dundee (Scozia); gira sotto sistemi operativi Windows, versioni 95, 98, ME, 2000, XP ed è in grado di generare liste di frequenza, concordanze, Kwic, Indici, pagine html, lemmatizzazioni, gruppi.



Il software è a pagamento e se ne può scaricare una demo<sup>20</sup>, con limite di 30 giorni di uso; la registrazione costa \$ 89. Il programma è interamente tutelato dai diritti d'autore e la versione demo è distribuita solamente a scopo valutativo.

Tra le varie potenzialità di questo software è bene sottolineare la possibilità di convertire il set di caratteri utilizzato da DOS a Unix e viceversa, di leggere file di testo in formato ASCII, meglio se salvati con "interruzione di linea" (Word), e di aggiungere automaticamente alcuni caratteri all'alfabeto di base e controllare la lunghezza delle righe (max 255 cr.)<sup>21</sup>.



La preparazione del testo prevede dunque salvataggi del file in formato ANSI, ASCII o “Windows predefinita”, preferibilmente con le interruzioni di linea; la codifica MS-DOS non va bene! Inoltre il file deve risiedere nella *directory* del programma.

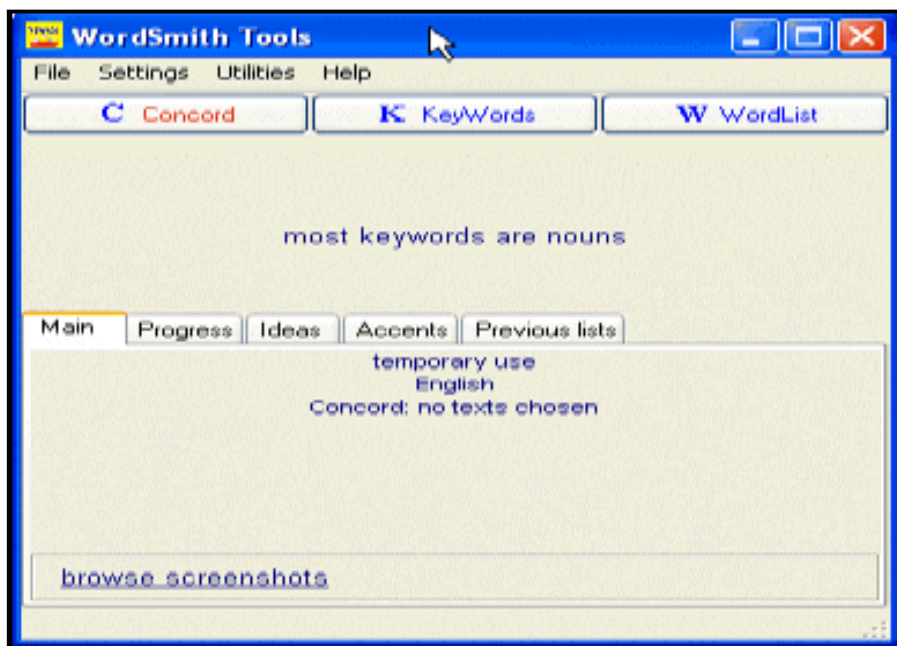
Tra i vantaggi che vanno accreditati a questo programma possiamo elencare i seguenti:

1. Si possono generare liste di nomi su cui operare la concordanza
2. Tutti i risultati possono essere stampati o salvati
3. L'aspetto grafico è molto flessibile
4. I risultati possono essere automaticamente convertiti in una pagina web
5. Il manuale viene distribuito gratuitamente in formato html e .pdf

Tra gli svantaggi elenchiamo invece:

1. Le occorrenze indicano come riferimento per la citazione il numero di linea del file di input utilizzato, anche nel caso di corpora, e non l'indicazione della suddivisione testuale alla quale quel contesto appartiene
2. È possibile utilizzare Concordance con i file XML, grazie al comando ‘ignore’ che esclude i caratteri compresi tra due marcatori, ma non è possibile operare ricerche che tengano conto dei tag come in WordSmith e MonoconC
3. Non si possono utilizzare gli operatori per la ricerca delle parole.





Questo programma è molto simile a WordSmith, anche se le maschere di lavoro sono un po' "spartane". È principalmente finalizzato alle concordanze, ma si possono ottenere anche liste di frequenza e statistiche. Il programma è tutelato dai diritti d'autore, la licenza d'uso costa \$ 85 e la versione "demo" può compiere tutte le operazioni ma limita i risultati ai primi 20.

Il programma è particolarmente versatile nella visualizzazione delle occorrenze, utilizza il formato KWIC oltre a quello per frase ed elimina le occorrenze che non interessano. Inoltre può visualizzare un contesto molto ampio oltre ad avere la possibilità di segnalare tag particolari all'interno del contesto medesimo. Uno speciale strumento permette di definire una ricerca avanzata tramite gli operatori, i separatori, ecc. È possibile ordinare i risultati delle concordanze e stabilirne l'aspetto tipografico con una (complicata) varietà di scelte, anche se in questo siamo facilitati dal fatto che è possibile mantenere attive più finestre con i risultati delle ricerche.

È possibile manipolare la lista delle occorrenze: in particolare si può 1) cancellare un'occorrenza (o una selezione di più occorrenze) dal menu *Display: Delete Item(s)*; 2) salvare o stampare una lista di occorrenze, con la possibilità di "nascondere" la parola cercata (opzione *conceal hits*); questa opzione è particolarmente pratica per costruire esercizi per i quali gli studenti devono indovinare funzione e significato della parola nascosta (come attività propedeutica a inferire il significato dal contesto).

**WORDSMITHTOOLS**, insieme di tool per l'analisi linguistica particolarmente versatile e complesso, sviluppato presso l'università di Oxford. Tra le sue peculiarità presenta alcune interessanti funzioni, come la possibilità di dividere il testo analiz-

zato in sezioni; statistiche complesse; una modalità di ricerca che utilizza operatori avanzati. È un insieme integrato di programmi per esaminare come le parole si comportano in un testo o insieme di testi ed è composto da vari strumenti; le statistiche possono essere ottenute sui singoli file oppure sull'intero *corpus* analizzato. Attualmente sono distribuite 2 versioni: la 3 e la 4.

Particolarmente interessante è la possibilità di lavorare con testi codificati in XML. I tag possono essere ignorati, ma anche utilizzati per ricavare statistiche dettagliate (ad esempio le concordanze del primo capitolo di un romanzo, oppure delle battute di un singolo personaggio, o ancora l'elenco dei nomi propri di un saggio).

La versione attualmente in uso e in distribuzione è la 3: si tratta di una versione a pagamento.

Estremamente interessante per chi studia lessicologia, ma anche ai fini dell'analisi testuale, è la possibilità di effettuare una lemmatizzazione automatica. Quest'ultima operazione richiede la stesura di un elenco delle desinenze della lingua italiana, in quanto il programma è impostato per la lingua inglese; particolarmente difficile è il caso dei lemmi verbali.

Altri software più vecchi, di ambiente DOS, non vengono riportati in questo elenco in quanto ormai rappresentano poco più che una curiosità storica. Possiamo citarne alcuni come *Intext 4.1* (il cui ultimo aggiornamento risale al 1998) e *Conc-A* (risalente al 1996, per Macintosh).

#### 4. ALCUNE APPLICAZIONI DI T.A.C.T.

In questo paragrafo vogliamo presentare due esperimenti di ricerca sviluppati con l'ausilio del programma Tact, descritto nel precedente paragrafo. Il primo riguarda l'analisi delle traduzioni che Giacomo Leopardi fa della *Batracomiomachia* pseudo-omerica. Il motivo della nostra attenzione verso questa particolare opera è che Leopardi traduce tre volte la *Batracomiomachia* (1815 – 1821/'22 – 1826), creando così un'interessante serie di varianti.

La codifica è abbastanza semplice: sono state codificate con il sistema Co.Co.A. le stanze e i canti; il segno % alla fine di ogni verso serve a contare i versi, mentre l'uso delle doppie parentesi acute "<< >>" indica righe di commento da ignorare nell'elaborazione. Quello che segue è il *file* .MKS che fornisce al programma le specifiche di codifica e di interpretazione:

```
[Alpha]
a b c d e f g h i j k l m n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
[DiacRet]
- \Acute\ \Grave\ \Circumflex\ \Cedilla\ \Umlaut\ \Tilde\ ""
[Title]
La guerra dei topi e delle rane [1815]
```

```

[RefTemplate]
canto/stanza $CS, v. $verso
[ReferenceBracket]
< > SupText NoWordSep
[IgnoreBracket]
<< >> SupText NoWordSep
[WordCounter]
WORD
0
[LineCounter]
LINE
0
[Reference]
Autore
Autore NonNumeric 9999
[Reference]
Titolo
Titolo NonNumeric 9999
[Reference]
Note
Note NonNumeric 9999
[Reference]
CS
Batra15 NonNumeric 9999
[Counter]
verso
1 SupText %

[Counter]
conta
1 NoSupText @CS
verso

```

TABELLA 2. *FILE\*.MKS*

Una volta ottenuto il database testuale possiamo operare le ricerche di cui *T.A.C.T.* è capace; negli esempi che si trovano in appendice abbiamo ottenuto le informazioni ritenute più utili al fine di notare le eventuali differenze tra le tre edizioni della *Batracomiomachia* leopardiana e trarne utili conclusioni.

Sempre a titolo di esempio aggiungiamo una seconda codifica che è stata effettuata sulle tre versioni della *Batracomiomachia*; detta codifica nasce dall'esigenza di analizzare in maniera più specifica le varianti, pertanto i tre testi (come si nota dall'esempio che segue) sono stati "fusi" in uno solo ed ogni terna di versi è stata etichettata con la sigla dell'anno di edizione compresa tra asterischi, inoltre sono stati utilizzati dei marcatori speciali per effettuare ricerche sulle varianti. Nella tabella che segue viene data spiegazione di tali marcatori speciali:

Sono state inserite le seguenti prolettere nel testo elettronico come codici di identificazione di tutte le diverse versioni:

1815:                   —

1821: #  
 1826: ^  
 1815/21: °  
 1821/26: +  
 1815/26 =

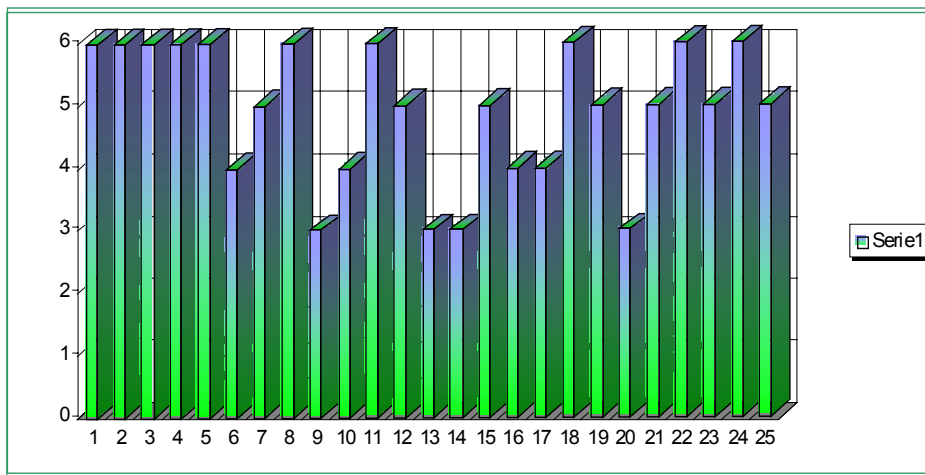
TABELLA 3. Caratteri speciali utilizzati nella codifica “Varianti”

Inoltre si è reso necessario modificare di conseguenza il *file* .MKS aggiungendo questi caratteri in fondo all’alfabeto ed eliminando l’accento circonflesso dai diacritici (in quanto non è presente come carattere alfanumerico nel testo in esame).

Quello che segue è il *file* risultante:

```
[Alpha]
a b c d e f g h i j k l m n o p q r s t u v w x y z _ # ^ ø + =
0 1 2 3 4 5 6 7 8 9
[DiacRet]
- \Acute\ \Grave\ \Cedilla\ \Umlaut\ \Tilde\ “”
[RefTemplate]
$Versione, canto/stanza $CS, v. $verso
[ReferenceBracket]
< > SupText NoWordSep
[IgnoreBracket]
<< >> SupText NoWordSep
[WordCounter]
WORD
0
[LineCounter]
LINE
0
[Reference]
Autore
Autore NonNumeric 9999
[Reference]
Titolo
Titolo NonNumeric 9999
[Reference]
Note
Note NonNumeric 9999
[Reference]
CS
15-21-26 NonNumeric 9999
[Counter]
verso
1 SupText %

[Counter]
conta
1 NoSupText @CS
verso
[Label]
```



Versione

\* \* NoSupText NoWordSep NonNumeric 9999

TABELLA 4. File .MKS utilizzato per la codifica “Varianti”

Una volta superata questa fase di “istruzione” del programma possiamo accedere alle interrogazioni, che svolgiamo mediante query, una funzione specifica del programma che permette di ricercare singole parole, porzioni di testo, segni diacritici e via discorrendo. Concentriamoci ora sulla codifica “Varianti” ed utilizziamo il file “3Vers.txt” che genera un omonimo database. Per mezzo del Query Dialog Box otteniamo quindi una distribution list per ciascun codice, impostando come unità di segmentazione la stanza e come sequenza di output la successione nel testo. Dall’analisi dei risultati emergono alcuni punti fondamentali, e cioè che col passare degli anni e l’arricchirsi delle letture il lessico leopardiano si arricchisce e si personalizza; i termini dell’epica classica vengono sostituiti da un vocabolario “realistico”, “concreto” e “satirico”; i personaggi assurgono a simbolismo politico e rispecchiano la realtà contemporanea. L’andamento del lessico e le conseguenti trasformazioni possono essere meglio apprezzate consultando i seguenti grafici, ottenuti mediante la trasposizione dei risultati numerici sul programma Excel.

L’asse orizzontale mostra le stanze del primo canto dell’opera, l’asse verticale mostra il numero di varianti. Come si può notare tutte le stanze del primo canto (preso qui a titolo di esempio) vengono ritoccate nel passare dalla prima all’ultima redazione.

Inoltre possiamo confrontare alcuni dati numerici che riguardano i tre testi relativamente ai rapporti token/type:

	Redazione 1815	Redazione '21-'22	Redazione 1826
Number of Types	1322	1334	1370
Number of Tokens	3121	3184	3217

Type/Token ratio	0,424	0,419	0,426
Token/Type ratio	2,361	2,387	2,348
Hapax Legomena	906	955	987
Hapax Dislegomena	197	172	196

Inoltre, possiamo generare un quadro molto più complesso e completo di statistiche: il programma TACTstat fornisce un complesso quadro numerico di statistiche del testo, generando tabelle, istogrammi e misurazioni statistiche sui token, i type, la lunghezza delle parole, la prima e l'ultima lettera delle parole e su tutte le lettere che compongono le parole. Per *default* queste informazioni statistiche vengono memorizzate in un file che ha lo stesso nome del database e l'estensione .STA.

Una volta lanciato il programma si deve inserire il nome completo del database testuale da utilizzare, selezionare l'output dei risultati (file o stampante) e specificare (se si vuole) il nome del file di output, che altrimenti prenderà automaticamente lo stesso nome del database testuale. L'output di TACTstat si compone di cinque parti: (1) tabella e statistiche della frequenza delle parole, (2) un istogramma della lunghezza delle parole, misurate in lettere, (3) istogramma alfabetico e a frequenza discendente e statistiche per le prime lettere di tutte le parole, (4) istogramma alfabetico e a frequenza discendente e statistiche per l'ultima lettera di tutte le parole, (5) istogramma alfabetico e a frequenza discendente e statistiche per tutte le lettere che compongono le parole.

Queste statistiche si sarebbero potute costruire anche manualmente, ma con enorme fatica, tempi lunghissimi ed elevate probabilità di errore. Prima di passare all'analisi di dette tabelle è opportuno chiarire però il significato e le differenze delle parole *token* e *type*. In sostanza quando si vuole valutare la ricchezza del linguaggio in un testo si deve valutare il rapporto tra le parole e la classe a cui queste parole appartengono: il rapporto parola/classe è appunto il rapporto *token/type*.

Il *token* quindi è l'occorrenza (insomma quante volte la singola parola si ripete nel testo) ed il *type* è una via di mezzo tra l'occorrenza e il lemma, sostanzialmente la forma grafica. Il numero totale dei *type* ci fornisce quindi il numero totale di parole (forme grafiche) diverse usate nel testo.

Vediamo ora la tabella delle statistiche per quanto riguarda la versione della *Batracomiomachia* del 1826 e spieghiamo tutte le voci presenti:

Frequency Rank	Observed Frequency of Rank	Words in Frequency	Types Total	Tokens Total	Percentage of Types	Percentage of Tokens	Percentage of word in Frequency
1	987	987	987	987	72.04	30.68	30.68
2	196	392	1183	1379	86.35	42.87	12.19
3	71	213	1254	1592	91.53	49.49	6.62
4	27	108	1281	1700	93.50	52.84	3.36
5	16	80	1297	1780	94.67	55.33	2.49

6	17	102	1314	1882	95.91	58.50	3.17
7	9	63	1323	1945	96.57	60.46	1.96
8	8	64	1331	2009	97.15	62.45	1.99
10	3	30	1334	2039	97.37	63.38	0.93
ecc.	ecc.	ecc.					
90	1	90	1369	3077	99.93	95.65	2.80
140	1	140	1370	3217	100.00	100.00	4.35

Number of Types = 1370  
 Number of Tokens = 3217  
 Type/Token ratio = 0.426  
 Token/Type ratio = 2.348  
 Hapax Legomena = 987  
 Hapax Dislegomena = 196  
 Hapax Legomena/Dislegomena ratio = 5.0357  
 Hapax Legomena/Number of Types = 0.7204  
 Hapax Legomena/Number of Tokens = 0.3068  
 Hapax Legomena cubed/Types squared = 512.2834  
 Variance ( S.D. squared ) = 45.9992  
 Standard Deviation (S.D.) = 6.7823  
 Coefficient of skewness = 11.7400  
 Coefficient of kurtosis = 177.6344  
 Herdan's characteristic = 0.0780  
 Yule's characteristic = 681.4803  
 Carroll TTR (Types / Sqrt of 2 X Tokens) = 17.0797  
 Most Frequent word "e" occurred 140 times  
 repeat rate (Tokens / frequency most frequent word) = 22.9786

Statistica delle frequenze nella versione 1826

- *Frequency Rank*, fornisce in ordine crescente le frequenze dei *token* del testo. Nella tabella non c'è, ad esempio, il numero 9, ciò significa che non ci sono *token* che hanno frequenza 19. In pratica non ci sono parole che compaiano 9 volte.
- *Observed Frequency of Rank*, fornisce, per ogni riga, la somma totale dei *types* che contengono i *tokens* con la stessa frequenza, riportata dalla prima colonna. Si tratta della somma totale delle parole che compaiono un *tot* di volte. Nell'esempio in esame ci sono, in altri termini, 987 parole che compaiono una sola volta, 197 che compaiono 2 volte, e così via.
- *Words in Frequency*, fornisce, per ogni riga, il numero complessivo dei *tokens* contenuti nei *types* riportati dalla seconda colonna. Per ottenere tale numero è sufficiente moltiplicare i dati numerici delle prime due colonne.
- *Types Total*, fornisce, in ordine crescente, il totale dei *types* presenti nel testo. Ogni riga si ottiene sommando progressivamente i dati della seconda colonna.

- *Tokens Total*, fornisce, in ordine crescente, il numero totale dei *tokens* presenti nel testo. Ogni riga si ottiene sommando progressivamente i dati della terza colonna.
- *Percentage of Types*, fornisce, in ordine crescente, la percentuale dei *types* presenti nel testo.
- *Percentage of Tokens*, come la precedente ma per quanto riguarda i *tokens*.
- *Percentage of word in Frequency*, fornisce la percentuale dei *tokens* rispetto alla totalità dei medesimi presenti nel testo.

Per spiegare a cosa si riferiscono tutti questi numeri possiamo, ad esempio, leggere la prima riga di quest'ultima tabella nella seguente maniera: 987 parole differenti (col. 2) ricorrono una volta (col. 1) per un totale di 987 volte (col. 2 moltiplicato col. 1=col. 3). Queste ripetizioni coprono il 30,68 % di tutte le parole presenti nel testo (col. 8). Inoltre questo primo grado di frequenza (il grado di frequenza "1") contiene in tutto 987 parole (col. 4) che rappresentano il 72,04 % del vocabolario totale (col. 6) di 1.370 parole (ultima riga della colonna 4 oppure prima voce della didascalia, "number of types"). Sempre in questo primo grado di frequenza possiamo quindi contare un totale (tra ripetizioni e hapax) di 987 parole (ovviamente il numero è sempre lo stesso perché ci troviamo sul primo grado di frequenza. L'esempio sarà più chiaro con il secondo grado) che corrisponde al 30,68 % del totale<sup>22</sup> (col. 7) espresso nell'ultima riga della colonna 5, cioè di 3.217 parole totali (comprese di ripetizioni).

Facciamo, per maggiore chiarezza, la stessa lettura sulla seconda riga, o secondo grado di frequenza: la frequenza 2; si tratta insomma delle parole che ricorrono due volte nel testo.<sup>23</sup> Dunque abbiamo che 192 parole differenti (col. 2) ricorrono due volte (col. 1) per un totale di 392 occorrenze (col. 2 moltiplicato col. 1=col. 3). Queste ripetizioni coprono il 12,19 % di tutte le parole presenti nel testo (col. 8). Inoltre questo secondo grado di frequenza contiene in tutto 1.183 parole (col. 4, che rappresenta la somma dei valori della colonna 2, fin qui incontrati) che rappresentano l'86,35 % del vocabolario totale (col. 6) di 1.370 parole (ultima riga della colonna 4 oppure prima voce della didascalia, "number of types"); in sostanza stiamo dicendo che, tra parole che ricorrono una sola volta e quelle che ricorrono due volte abbiamo già coperto un buon 86% del testo. Già da questo dato si può dedurre che si tratta di un testo dal vocabolario ricco e variato e molto poco ripetitivo. Sempre in questo grado di frequenza possiamo quindi contare un totale (tra ripetizioni e hapax) di 1.379 parole (ora il numero rappresenta la somma dei valori del primo grado con quelli del secondo) che corrisponde al 42,87 % del totale (col. 7) espresso nell'ultima riga della colonna 5, cioè di 3.217 parole totali (comprese di ripetizioni). Se generiamo questi risultati anche per le precedenti due redazioni del testo in esame e le mettiamo a confronto, possiamo evidenziare alcuni aspetti, qui riportati a titolo di esempio:

1. Nei prospetti che sono in calce ad ogni tabella figura la voce *Hapax Le gomena* che indica quante parole compaiono una volta sola nel testo: nella versione 1815 sono 906, nella versione 1821 sono 955 e nella versione 1826 sono



987. Tale numero è andato sempre crescendo, e in maniera evidente se confrontiamo la prima con l'ultima versione, segno chiaro che il testo nel corso degli anni si è arricchito di nuovi termini ed è stato sfrondata di ripetizioni. L'andamento di questi risultati numerici denuncia quindi un arricchimento lessicale.

2. Sempre nelle voci a fine tabelle si nota la *Type/Token ratio* che esplicita, per l'appunto, il rapporto tra il numero dei *types* ed il numero dei *tokens*. Questo dato numerico ci fornisce un'idea ancora migliore della ricchezza del testo:

“Una delle indicazioni alle quali aggrapparci per interpretare la ricchezza del linguaggio di un determinato testo è il rapporto tra le parole e le classi nelle quali possono essere catalogate (in inglese token e type). [...] Una conseguenza abbastanza immediata (ma da prendere con le molle, come tutte le misurazioni) è che un testo in cui il rapporto token/type sia alto deve essere un testo in cui vi sono pochi type rispetto ai token e quindi potrà essere considerato un testo ripetitivo, con un vocabolario non molto ricco e tendenzialmente poco faticoso da leggere<sup>24</sup>.”

Anche in questo caso il dato è significativo, poiché nel passaggio dalla versione 1821 alla versione 1826 il rapporto in questione si è abbassato di 0,039, passando da 2,387 della versione 1821 a 2,348 della versione 1826. Per di più dobbiamo tenere conto del fatto che questo decremento è avvenuto in un testo abbastanza breve, il che ne aumenta l'importanza:

[...] “notare come una cosa sia l'aggiunta di un token in un testo di 10 parole, tutt'altra faccenda l'inserimento di un termine in un lavoro di 1000 parole. L'incremento dei termini diverrebbe significativo nel primo caso e di scarso interesse nel secondo<sup>25</sup>”.

Il secondo esperimento cui accenniamo è un tentativo di applicazione del programma Tact nel campo della psicologia. Questo progetto si è sviluppato in sinergia con il prof. Marco Zanasi, dell'Università di Roma “Tor Vergata” e si prefigge di analizzare linguisticamente i sogni dei pazienti per ricavarne una comunanza di espressioni linguistiche che possano indicare una determinata patologia. Il sogno del paziente viene trascritto fedelmente, con la sola aggiunta delle pause; per la codifica delle pause abbiamo elaborato la “scala dei silenzi”, che prevede tre codifiche a seconda della durata della pausa in secondi: 0-3 = #; 3-6 = ## ; oltre 6 = ###. Nello sviluppare questo progetto ci siamo posti le seguenti domande:

1. I pazienti malati utilizzano un particolare schema linguistico?
2. Si può assegnare uno schema linguistico ad ogni determinata patologia?
3. Si può ricavare un test diagnostico standard da utilizzare per il futuro?

Se queste domande avranno una risposta, anche grazie all'uso di Tact, potremmo dire che l'analisi testuale avrà avuto anche una utilità terapeutica.

## 5. CONCLUSIONI

Computer e critico letterario (o filologo) possono lavorare in simbiosi con risultati sorprendenti, poiché questo nuovo strumento – il computer – possiede, rispetto agli strumenti tradizionali, delle potenzialità enormi e non ancora del tutto esplorate. Tuttavia è bene tenere presente che non dobbiamo delegare al computer tutta la responsabilità dell'analisi di un testo, non si tratta di un cilindro magico dal quale fuoriescono statistiche e tabelle: esso va istruito e guidato<sup>26</sup> dal critico affinché possa poi svolgere per lui il lavoro di ricerca e di analisi che si è proposto.

Vale la pena di raccontare un aneddoto, una storia che riferisce lo scrittore Stephen King nella sua autobiografia; lo strumento in questione in questo caso non è il computer – di cui peraltro King fa largo uso per comporre i suoi romanzi – ma una scrivania, altro oggetto di indubbia necessità per un umanista:

“Per anni avevo sognato di possedere una di quelle tavole di quercia in legno massiccio che dominano la stanza: basta con gli scrittoi per bambini nello sgabuzzino di una roulotte, basta con le ginocchia intorpidite in un vano troppo angusto in una casa in affitto. Nel 1981 mi procurai quella che volevo e la piazzai al centro di uno studio spazioso illuminato da un lucernario. Per sei anni mi sono seduto a quella scrivania o ubriaco o rimbambito, come uno skipper che dirige la sua barca verso nessun posto. Un paio di anni dopo, a sangue ripulito, mi sbarazzai di quella mostruosità e al suo posto allestii una suite, scegliendo i pezzi e un bel tappeto turco con l'aiuto di mia moglie. [...] Ho un'altra scrivania, costruita a mano, molto bella, grande la metà del dinosauro precedente. L'ho sistemata nell'angolo ovest dello studio, sotto il soffitto spiovente. [...] Ci sono seduto sotto ora, cinquantenne, con gli occhi malandati, una gamba sifolina e senza postumi di sbornie. Faccio quello che so come si fa e lo faccio come meglio lo so fare, [...] ora vi racconterò tutto quello che posso sul mio mestiere. Come promesso, non ci metterò molto. Comincia così: sistemate la vostra scrivania nell'angolo e tutte le volte che vi sedete lì a scrivere, ricordate a voi stessi perché non è al centro della stanza. La vita non è un supporto per l'arte. È il contrario<sup>27</sup>”.

## 6. RIFERIMENTI BIBLIOGRAFICI

- Ciotti F., Roncaglia G. (2003). *Il mondo digitale*. Bari, Laterza.  
D'Angelo E., Stella F. (eds) (2003). *Poetry of the early medieval Europe: manuscripts, language and music of the rhythmical latin texts*. Firenze, Sismel.  
Eco U. (1990). *I limiti dell'interpretazione*. Milano, Bompiani.  
Fiormonte D. (2003). *Scrittura e filologia nell'era digitale*. Torino, Bollati Boringhieri.  
Gigliozzi G. (1997). *Il testo e il computer*. Milano, Mondadori.  
King S. (2000). *On writing*. Milano, S&K.

- Lana M. (1994). *L'uso del computer nell'analisi dei testi*. Milano, FrancoAngeli.
- Lancashire I. (1996). *Using TACT with electronic texts*. New York, Modern Language Association of America.
- McGann J. (2002). *La letteratura dopo il World Wide Web*. Bologna, Bonomia University Press.
- Rudman J. (1998). The State of Authorship Attribution Studies: Some Problems and Solutions. *Computers and the Humanities*, Vol. 31, pp. 351-365.
- Spina S. (2001). *Fare i conti con le parole*. Perugia, Guerra.
- Teresi L. (2003). *L'analisi testuale con TACT: un'esperienza nel campo dell'inglese antico*. S.I.S.M.E.L.

## 7. NOTE

- <sup>1</sup> Eco U. (1990).
- <sup>2</sup> Lana M. (1994), p. 23.
- <sup>3</sup> Sul concetto di *type* torneremo più avanti. Cfr. *infra* p.
- <sup>4</sup> <http://www.expertsystem.it>
- <sup>5</sup> Il testo dell'intervento è consultabile presso l'indirizzo:  
<http://www.humanities.mcmaster.ca/~grockwel/ictpaper/paris.htm>
- <sup>6</sup> <http://www.griseldaonline.it>
- <sup>7</sup> <http://www3.oup.co.uk/litlin/>
- <sup>8</sup> <http://www.ach.org/>
- <sup>9</sup> Rudman J. (1998).
- <sup>10</sup> Società Internazionale per Studio del Medioevo Latino.
- <sup>11</sup> Teresi L. (2003).
- <sup>12</sup> Ai fini dei conteggi che un software di analisi testuale effettua una parola è definita come una sequenza alfanumerica isolata da due spazi (uno iniziale e uno finale).
- <sup>13</sup> [http://artsweb.bham.ac.uk/pking/multiconc/l\\_text.htm](http://artsweb.bham.ac.uk/pking/multiconc/l_text.htm)
- <sup>14</sup> Questo software è prodotto dal *Centre for Computing in the Humanities* dell'Università di Toronto, ed è distribuito gratuitamente.
- <sup>15</sup> Attualmente lo si può far girare con tranquillità sotto Windows98, per i successivi sistemi operativi c'è qualche difficoltà per via dell'assenza di una modalità MS-DOS reale. Per i sistemi Windows2000 e WindowsXP si deve utilizzare il comando *forcedos*.
- <sup>16</sup> Facciamo qui riferimento alle versioni "2000" e "XP" del pacchetto "Office".
- <sup>17</sup> Con Office 2000 non ci sarà nessun avvertimento, quindi bisogna prestare maggiore attenzione.
- <sup>18</sup> Uno dei creatori del programma.
- <sup>19</sup> I risultati di questa ricerca possono essere consultati all'indirizzo:  
<http://www.humanities.mcmaster.ca/~grockwel/ictpaper/learneds.htm>
- <sup>20</sup> <http://www.rjcw.freerve.co.uk/>
- <sup>21</sup> Il problema della lunghezza delle righe si verifica anche con il programma Tact. Tuttavia quest'ultimo, a dispetto di quanto dichiarato nel manuale, non sembra in grado di effettuare una corretta decurtazione automatica delle righe stesse.
- <sup>22</sup> Vale anche in questo caso la considerazione fatta poco prima per il dato numerico non percentualizzato: il numero rimane lo stesso perché ci troviamo nel primo grado delle frequenze.
- <sup>23</sup> Che gli ideatori del programma, con un curioso ossimoro, chiamano "hapax dislegomena".
- <sup>24</sup> Gigliozzi G. (1977), pp. 193-194.
- <sup>25</sup> *Ibidem*, p. 194.
- <sup>26</sup> Per esempio al momento della codifica del testo.
- <sup>27</sup> King S. (2000), pp. 94-95.

# The C-ORAL-ROM multilingual corpus of spontaneous speech: validation of the prosodic annotation by expert transcribers

*Massimo Moneglia*

Università degli Studi di Firenze  
(moneglia@unifi.it)

*Antonietta Scarano*

Università degli Studi di Firenze  
(scarano@lablita.dit.unifi.it)

*Marius Spinu*

Università degli Studi di Firenze  
(ms@exetec.it)

*C-ORAL-ROM is a multilingual corpus of spontaneous speech for the main romance languages and puts together four collections of Italian, French, Portuguese and Spanish spoken texts (123h35'40" for about 1,200,000 words). The corpus is the main outcome of an IST project (IST-2000-26228) and features innovative solutions for storing and exploiting multilingual speech resources in digital multimedia format. The C-ORAL-ROM format is based on the hypothesis that speech events (utterances) in spontaneous speech are marked by perceptively relevant prosodic breaks. This property is assumed to be consistent at a multilingual level in the frame of the romance languages and its detection enables an easy and effective synchronization of each transcribed utterance with its acoustic counterpart. The paper presents the result of validation experiments showing that the perception of prosodic breaks by expert annotators, especially terminal breaks, is consistent at inter-linguistic levels (Italian, Spanish and Portuguese) and that this kind of prosodic tagging reaches a high level of inter-annotator agreement.*

*Keywords:* Multilingual Speech Resources, Validation, Synchronization

## 1. INTRODUCTION

The C-ORAL-ROM project (IST 2000-26228) is a multilingual corpus of spontaneous speech for the main romance languages. Four collections of Italian, French, Portuguese and Spanish corpora have been delivered by national providers<sup>1</sup> The corpus (123h35'40") has been recorded in free situations and will be distributed by ELDA

(European Language Resource Distribution Agency) in a multimedia format that can be considered as a standard for multilingual speech resources in digital format.

The multimedia format of C-ORAL-ROM is based on a relation between the prosodic annotation and the text to speech synchronization unit that is assumed to be meaningful at a multilingual level. The multilingual corpus (1,200,00 words) is tagged with respect to prosodic breaks and it is assumed that utterance boundaries are marked by perceptively relevant prosodic breaks. More specifically a one to one correspondence is assumed between the limit of each utterance and the presence of a terminal prosodic break; i.e. terminal prosodic breaks necessarily mark the end of each utterance (see classic works like Karcevsky, 1931; Crystal, 1975 and Cresti, 2000 for justifications regarding the standard approach in C-ORAL-ROM).

Each word sequence ending with a terminal prosodic break is therefore considered an utterance and is synchronized with its acoustic counterpart, giving rise to a data base of meaningful linguistic entities. The utterance based segmentation of speech and its synchronization to the acoustic signal allows to appreciate the syntax of spontaneous speech and its proper semantic interpretation at a multilingual level. For example the brief verb-less utterances and isolated discourse markers which are typical of everyday dialogues, can be represented in their proper linguistic function and domain.

For example the following sequence of words belonging to an Italian dialogic turn of the C-ORAL-ROM corpus does not show a clear structure on the basis of the textual information.

\*EST: no ascolta qui sopra [no listen up here]

On the contrary this structure becomes clear once terminal (//; ?) and non-terminal (/) the prosodic breaks are integrated .in the transcript:

\*EST: no // ascolta / qui sopra?

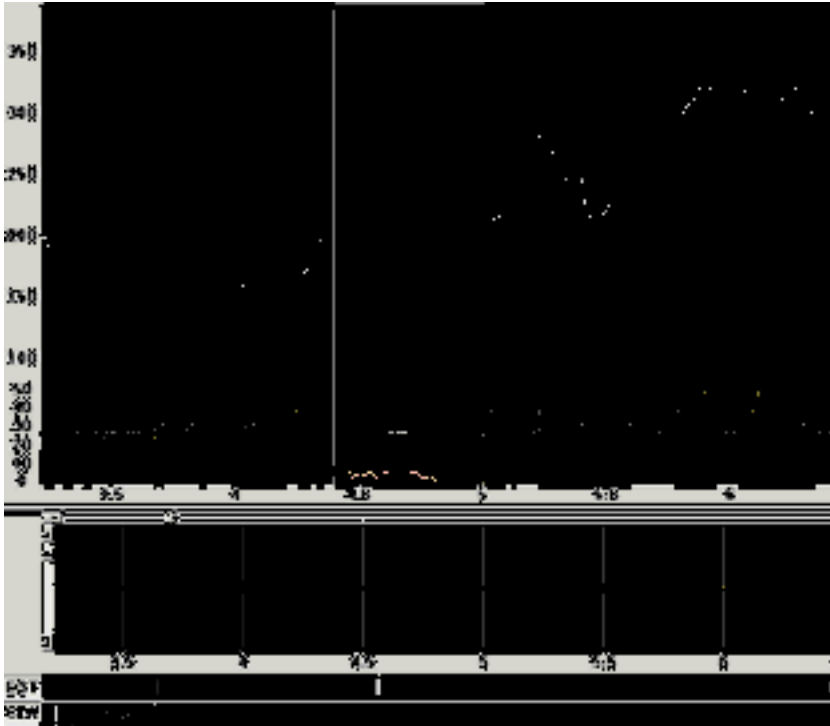
%ill: reassurance(1) question introduce by a discourse marker (2)

The alignment per utterance follows this prosodic tagging. As the following figure shows, the alignment specifies one synchronization unit for each sequence ending with a terminal break allowing a meaningful storage of the speech resource in an utterance data base.

This synchronization principle is applied in C-ORAL-ROM through 120 hours of the multilingual corpus. Therefore the marking of terminal breaks has a strong role within the frame of a multilingual speech resource because it determines the data base of all utterances that is otherwise underdetermined.

The prosodic tagging is based on perceptual judgments by competent speakers (transcribers). (See Oostdijk et al., 2002 for a similar approach with a 250,000 words sub-corpus of the Spoken Dutch Corpus). In C-ORAL-ROM it is assumed that competent speakers can perceive prosodic breaks in the speech continuum of spontaneous speech and that linguistic perception allow one to discriminate between *terminal* and *non-terminal* prosodic breaks (see 't Hart et al., 1990 and Cresti, 2000). A specific prosodic tag is inserted in the transcription each time a terminal or a non terminal prosodic break (or possibly a fragmentation break) is perceived by the

transcribers in the speech flow<sup>2</sup>. On the basis of the information provided by terminal breaks each utterance is aligned to its acoustic counterpart, so generating the data basis of all utterances in the resource. In C-ORAL-ROM each tagged transcription is revised by at least by one or two expert transcribers, and a third operator checks terminal breaks during alignment.



## 2. GOAL

The validation test aims to verify whether, under normal circumstances, judgments regarding the presence of terminal and non-terminal breaks in the speech flow are assigned coherently by the expert transcribers who tagged the four romance corpora.<sup>3</sup>

The terminal breaks are assumed to be highly prominent to perception; that is they should also be considered *strong breaks*. The assignment of the quality “terminal” to a prosodic breaks is probably not only a matter of direct perception, but rather a linguistically determined one (linguistic perception) and comes from the detection of specific prosodic cues that enable the a judgment of the termination of speech acts. The validation aims to ensure the maximum agreement among transcribers with respect to terminal prosodic tags.

Non-terminal breaks may be in principle strong or weak<sup>4</sup>. Non-terminal breaks, in many circumstances, are assumed to be much less determined on the basis of perceptual

factors, although in other circumstances the perception of non-terminal breaks may be also strong. For this reason the opposition “terminal vs non-terminal” does not clash with “weak vs. strong” adopted in similar works (see See Buchman et al., 2002). The circumstances of weak definition of non-terminal breaks and the distinction between terminal prosodic breaks and strong non-terminal breaks are not the object of this validation test.

The validation test aims to detect the level of agreement reached by expert transcribers in C-ORAL-ROM with no prevision as to the possible adequacy level.

### 3. TASK

Judgments should be verified to be coherent in typical contexts of speech segmentation that occur in C-ORAL-ROM, which represent common contexts of spontaneous spoken language corpora. The test corpus must therefore represent typical cases of speech variation in the C-ORAL-ROM corpus. This corpus has been selected making reference to both corpus structure and internal prosodic properties of texts; i.e. the socio-linguistic variation and the frequency of the prosodic breaks as well as the relative proportion between terminal and non-terminal breaks.

Two samples of each romance corpus have been selected:

- roughly 750 words of an informal dialogue;
- roughly 750 words of a formal monologue (in the follow up section of the web site in WP3);
- the acoustic quality of both samples is bound to Excellent (A) or Good (B).

In other words the two samples represent, in each corpus, two opposite situations with respect to socio-linguistic properties (informal dialogue vs. formal monologue). Moreover each selection should presents, in principle, the same degree of difficulties from the point of view of the prosodic tagging. To this end, in all dialogues the number of words for each dialogic turn (MLTw) is comparable. This choice ensures that a comparable level of difficulty is encountered by each group of transcribers in passing from an objective partition of the speech flow (division in turns) to the perceptually based segmentation in tone groups. In the formal monologue, where there is not such an objective partition, the level of difficulty is controlled through the restriction to the same gender (conference or teaching session)

The choice has been oriented also to ensure enough representation of the C-ORAL-ROM sampling with respect to the main standard measurements of spoken language variability<sup>5</sup>. The two text have been selected among those which ranges over average values on the two opposite sides of variability. The measures considered are the following:

Number of prosodic breaks with respect to the number of words, which specify different cases of the frequency of occurrence of a prosodic break in the speech flow.

Mid-length of Utterance, (MLU) which specifies differences between texts in the relative proportion between terminal and non-terminal breaks.

	Italian		Portuguese		Spanish	
	Informal	Formal	Informal	Formal	Informal	Formal
	Ifamd16	Inatte01	Pfamd12	Pnatte03	Efamdl22	Enatte04
MLU	7.18	28.12	15.16	19.49	6.90	11.59
MLtone	2.49	2.05	3.29	4.50	3.55	2.94
MLTw	11.84		17.32		11.24	

## 4. PROCEDURE

In both texts proposed to transcribers for the validation procedure the prosodic tagging was not marked.

The texts have been tagged by four expert transcribers in each group in accordance with the practice adopted in C-ORAL-ROM.

The text is tagged marking terminal and non-terminal prosodic breaks (and fragmentation breaks in the Spanish and Italian corpora) on the basis of perceptual judgments.

Judgments are given on the basis of the acoustic signal.

The degree of accuracy required in the validation test is the same used for transcription in C-ORAL-ROM

The audio is played by the transcribers themselves, selecting temporal segments through a familiar sound editor (a default duration of around 10 seconds per segment, possibly modified by the transcribers according to his/her perceptual needs).

Each selected segment is played no more than three times.

Each transcriber performs the task independently of the other.

## 5. CRITERIA OF ANALYSIS

### 5.1 Measurements

The average length of prosodic unit (MLTone) specifies the frequency of prosodic breaks, while the average length of utterance (MLU) specifies the frequency of terminal breaks within the previous set.

The compatibility of the previous measures in the four comparable collections with respect to general constraints on both linguistic notions guarantees that the same criteria of prosodic tagging has been applied by transcribers to the four selections for both terminal and non-terminal breaks.

In the evaluation of both indices we refer to two very general constraint:

- the number of words in each tone group must be compatible with the number of syllables that can be realized within an expiration event;
- in formal texts the MLU range over higher values than in informal texts.



## 5.2 Coherence in the application of prosodic tagging criteria

The level of agreement within each collection with respect to perceptual judgments that allow to mark a position in a text with a prosodic break ensures the coherence of application of criteria by transcribers. The following are the criteria of analysis for the evaluation of results.

- Each *virtual position* (that is a possible point for inserting tags in text) tagged by at least one transcribers is a *position*.
- Each position tagged by at least two transcribers is a *convergent position*.
- Each convergent position marked by at least two transcribers with a terminal break is defined as a *terminal position*<sup>6</sup>.

The following is calculated for each selection and for each formal and informal corpus in each selection<sup>7</sup>:

- percentage of positions marked by only one transcriber (*index of random application*);
- percentage of terminal positions marked by all transcribers (*index of high prominence terminal*);
- percentage of not-terminal positions marked by all transcribers (*index of high prominence not terminal*);
- for each terminal position the percentage of terminal, not terminal, or empty choices by the four transcribers is marked (*relative prominence of terminal breaks*);
- for all convergent positions that are non-terminal the set possible choices realized by transcribers is put on a scale and discussed (*relative prominence of non-terminal breaks*).

## 6. DISCUSSION OF THE RESULTS

### 6.1 Measurements

As shown in the tables below the length of the prosodic unit (MLTone) registered in the four collections (from 2.5 to 4 words) turns out to be compatible in each collection with the limit of breath. The relative difference found in the collection nicely fit the different syllabic structures of the four language<sup>8</sup>.

Moreover such length is registered extremely constant in all transcribers within each collection, thus ensuring that the number of tone breaks is perceived with a similar frequency. This is true for both formal and informal texts as the low values of standard deviation show (0.10-I; 0.31-S; 0.065-P in the informal text and 0,15 -I; 0,14-S; 0.28 in the formal text).

The range of variation of MLU through the collections in the informal and in the formal texts is compatible with different range of values registered in formal and informal texts in previous studies (Cresti, 2000), that is lower values of MLU in the informal domain.

The similarity of values of MLU and MLTone obtained by four transcribers in each collection and the relative difference between MLU in the formal and informal text is one of the indices of coherence in the application of criteria within each corpus.

Informal		ITALIAN					
ifamd116	Words	Utterances	Dialogic turns	Tone units	MLTone	MLU	MLTw
transcriber A	779	73	34	306	2,55	10,67	22,91
transcriber B	778	76	39	338	2,30	10,24	19,95
transcriber C	779	66	38	304	2,56	11,80	20,50
transcriber D	778	84	40	308	2,53	9,26	19,45

Formal		ITALIAN					
inatte01	Words	Utterances	Dialogic turns	Tone units	MLTone	MLU	MLTw
transcriber A	813	42	1	339	2,40	19,36	8,13
transcriber B	813	53	1	369	2,20	15,34	8,13
transcriber C	813	36	1	338	2,41	22,58	8,13
transcriber D	812	32	1	306	2,65	25,38	8,12

Informal		SPANISH					
efamd122	Words	Utterances	Dialogic turns	Tone units	MLTone	MLU	MLTw
transcriber A	754	99	67	183	4,12	7,62	11,25
transcriber B	753	89	69	191	3,94	8,46	10,91
transcriber C	753	102	69	207	3,64	7,38	10,91
transcriber D	753	91	70	228	3,30	8,27	10,76

Formal		SPANISH					
enatte04	Words	Utterances	Dialogic turns	Tone units	MLTone	MLU	MLTw
transcriber A	715	64	2	205	3,49	11,17	357,50
transcriber B	715	60	2	210	3,40	11,92	357,50
transcriber C	715	67	2	216	3,31	10,67	357,50
transcriber D	715	67	2	231	3,10	10,67	358

Informal		PORTUGUESE					
pfamd112	Words	Utterances	Dialogic turns	Tone units	MLTone	MLU	MLTw
transcriber A	823	61	30	238	3,46	13,49	27,43
transcriber B	820	50	31	228	3,60	16,40	26,45
transcriber C	821	61	26	240	3,42	13,46	31,58
transcriber D	820	60	30	234	3,50	13,67	27,33

Formal		PORTUGUESE					
pnnatte03	Words	Utterances	Dialogic turns	Tone units	MLTone	MLU	MLTw
transcriber A	770	42	1	216	3,56	18,33	770
transcriber B	769	35	1	218	3,53	21,97	769
transcriber C	769	52	1	258	2,98	14,79	769
transcriber D	769	45	1	257	2,99	17,09	769

Data shows, however, that the range of values in question turns out to be more constant among each group of transcribers in the case of Mltone, rather than in

MLU. That is the detection of prosodic breaks (terminal or not terminal) in the speech flow is extremely constant in all transcribers, so confirming the significance to perception of prosodic breaks. Some variation may occur when the supplementary judgment on terminal nature of prosodic breaks is in question. For some group the Standard deviation in MLU is therefore more significant, specifically in the formal text (0,51-S; 2,59-P; 3,7-I), while only a few differences can be appreciated in the informal text (0,44-S; 0,91-I; 1,24-P).

More substantial data can therefore be derived by a close study of the agreement between transcribers with respect to the positions marked.

### 6.2 Index of random application

In the informal texts, on the total positions (354-I, 235-S, 260-P) only a low percentage (7.9%-I; 9.8%-S; 6.9%-P) is not convergent. In the formal texts, on the total positions (411-I; 257-S; 276-P) the percentage of not convergent positions may rise a little (11.2%-I; 7.6%-P) or decrease (8.6%-S), but it remains low. In other words, only a few judgments come from a random application of criteria (for example only 28 judgments out of a total of 1400 in the Italian texts).

Moreover no position in the not convergent set (informal 7.9%-I; 6.9%-P, 9.8%-S) (Formal 11.2%-I; 8.6%-S; 7.6%-P) in both formal and informal texts is marked *terminal* (0%); that is random applications are rare and are never terminal.

This first result ensures:

- the high perceptual relevance of terminal breaks, that are always detected by more than one transcriber;
- the coherence in the application of perceptual criteria is strong even with non-terminal breaks.

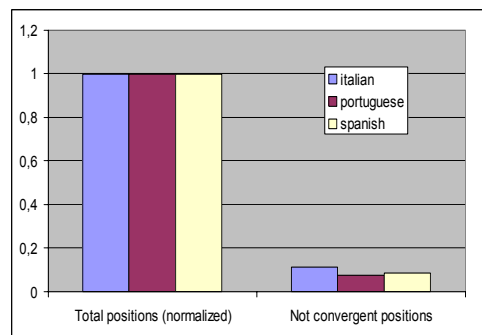
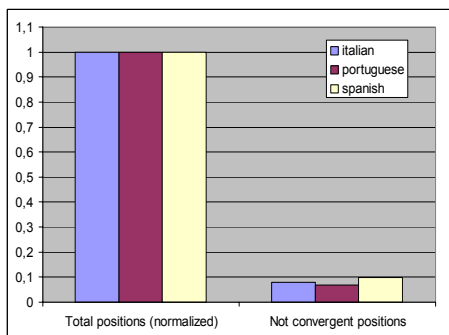


Fig. 1 Informal, incidence of random judgments Fig. 2 Formal, incidence of random judgments

In principle the existence of weak non-terminal breaks in spontaneous speech could lead to a certain degree of under-determinacy in the application of prosodic

tagging criteria, thus causing random applications on spontaneous speech corpora. The result shows that the statistic relevance of weak positions is necessarily low. From this point of view, the relevance of the criteria can be assumed despite the noted degree of under determinacy.

### 6.3 *Indexes of high and relative prominence*

In the informal texts the terminal position (110--I; 115-S; 59-P) represent respectively 31%-I 48,9%-S and 22.7%-P of the total positions). The agreement among the four transcribers (high prominence) is complete in 63.6 %-I ; 77.4%-S, 62.7%-P of terminal positions of the informal dialogue. Very generally, in 2/3 of non-random positions a complete agreement is reached, while in the remaining 1/3, obviously, at least 2 transcribers agree. The complete agreement as regards the presence of a prosodic break may be stronger as in the case of non-terminal breaks of the Italian and Portuguese group, that is 71%-I 74%-P of positions not defined as terminal, or weaker (52%-S). The status of the non-terminal prosodic breaks is more complex to identify (see below). The analysis of the set in which the complete agreement on terminal breaks is not reached (1/3) shows details that are important for the validation of the utterance data base of the C-ORAL-ROM corpus. *In the subset of terminal positions (110-I and 115-S 59-P positions) respectively for 440, 460 and 236 possible choices*, the possibility to have non-terminal or empty positions is restricted by definition to a range of from 0 to 220-230-118). In 86.5%-I – 88.7%-S – 89.8%-P of terminal positions at least three transcribers agree on the terminal nature of the position; that is, their relative prominence is very high. The remaining part is distributed as follows: only on 7-I; 1-P and 3-S occasions does the transcriber not mark the terminal position at least with a prosodic break and on no occasion is the position left un-marked by two transcribers together. In other words the possibility that a terminal position will turn out not marked in C-ORAL-ROM is irrelevant and, moreover, when a terminal position occurs, no agreement between two competent transcribers is possible as regards about the absence of prosodic breaks in that position. This will occur in both formal and informal texts (see below). Therefore we can argue from the previous data that terminal prosodic breaks are definitively perceived as strong breaks by transcribers. The remaining part is distributed as follows: only on 7-I; 1-P and 3-S occasions does the transcriber not mark the terminal position at least with a prosodic break and on no occasion is the position left un-marked by two transcribers together. In other words the possibility that a terminal position will turn out not marked in C-ORAL-ROM is irrelevant and, moreover, when a terminal position occurs, no agreement between two competent transcribers is possible as regards about the absence of prosodic breaks in that position. This will occur in both formal and informal texts (see below). Therefore we can argue from the previous data that terminal prosodic breaks are definitively perceived as strong breaks by transcribers.

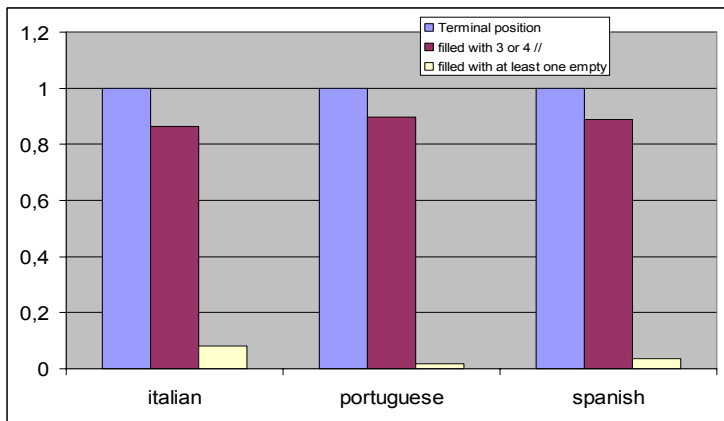


Fig. 3 Terminal breaks in Informal texts

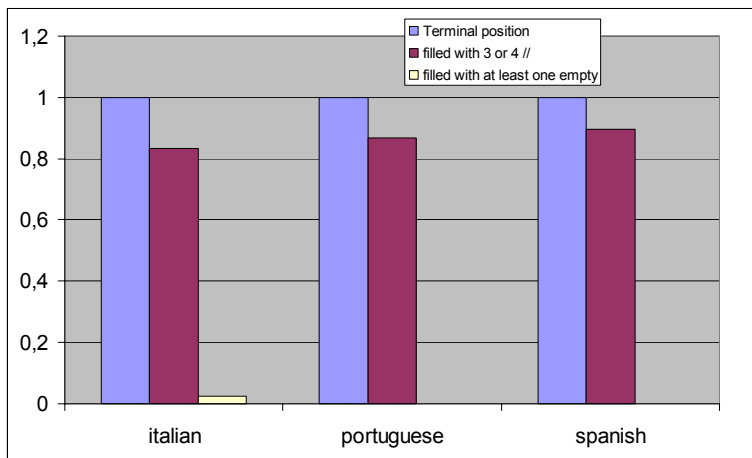


Fig. 4. Terminal breaks in Formal texts

Formal texts are in general characterized by a lower percentage of terminal positions. In the case in question 42-I (roughly 10% of the total positions) 45-P (16.30% of total positions) and 86-S (33.5% of the total positions).

In the formal texts the agreement by all transcribers is reached in 59.5%-I, 68.9-P and 76.7%-S of terminal positions. Terminal positions are marked in more than 80% of cases by at least three transcribers (83.3%-I , 86.7-P and 89.5%-S). Significantly, even more clearly than in informal texts, it never happens that transcribers omit to mark the prosodic break in a terminal position, thus confirming the strong nature of terminal breaks.

#### 6.4 Prominence of non-terminal breaks

With respect to non-terminal breaks the agreement is complete in 58%-I, 63%-P and 82%-S of position not defined as terminal. Despite a little discre-

pancy with respect to the total agreement index in terminal breaks, the relative prominence even of non-terminal breaks is high in both formal and informal texts, i.e. at least three transcribers agree on the presence of a non-terminal break in 87%-I; 88%-P; 96%-S in the formal text and 77%-S; 90%-I and 90.1%-P of the informal ones.

This data are important for two reasons: a) the high frequency of non-terminal breaks in the formal texts; b) the fact that the weak or strong nature of non-terminal breaks has been left undefined in the present framework.

With regard to the first point the fact that the average of relative prominence of non-terminal breaks is equal to or higher than the relative prominence of terminal breaks ensure an high level of validation of prosodic criteria also with non-terminal breaks.

However, the validation task shows interesting results also with respect to the second qualitative question. From a close analysis of the data it turns out that the relative frequency of omissions of prosodic tagging in all positions not marked as terminal in which three or two position are marked is significant in both informal (21%-I 36%-S 21%-P) and formal (34%-I; 15%-S; 33%-P) texts. More specifically in the majority of positions in which three transcribers agree on the presence of a non-terminal break (relative prominence), and in almost all positions in which two transcribers agree on the non-terminal nature of a the prosodic break, the remaining transcribers agree on the absence of breaks. Few judgments omit to mark a break when a terminal position occurs. Therefore we must point out that when a position is terminal the main alternative to the transcriber is that the position may be not terminal, when a position cannot be marked as terminal the alternative is not the reverse, but rather that there is no prosodic break.

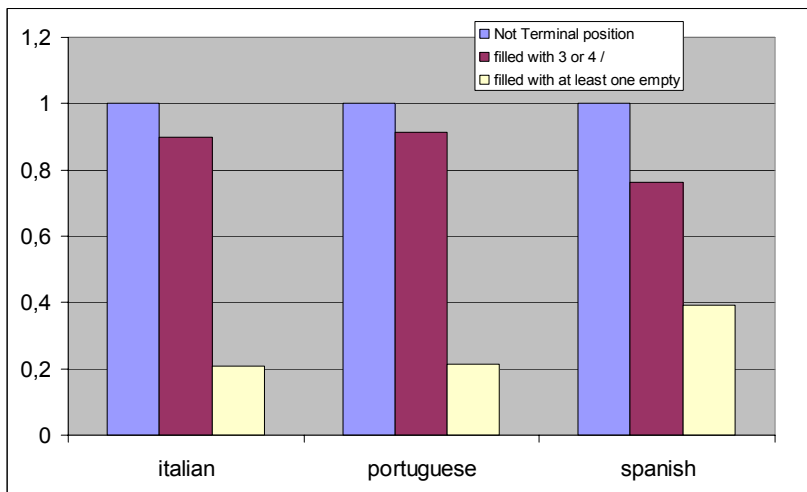


Fig. 5 Non-terminal breaks in Informal texts

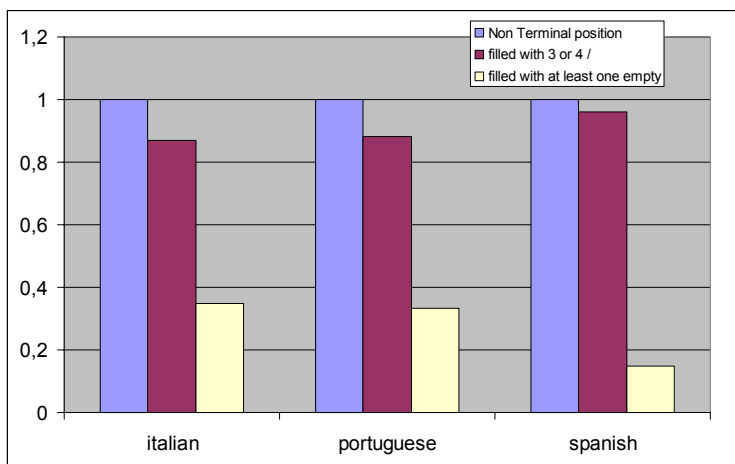


Fig. 6 Non-terminal breaks in Formal texts

In other words, in both formal and informal texts, only in the case of non-terminal breaks, can we verify the presence of a certain amount of positions in which the break is felt weak.

## 7. CONCLUSIONS

The validation test highlights a set of main conclusions with respect to the coherence of the application of prosodic tagging in C-ORAL-ROM in the main case of terminal contours (in both formal and informal texts):

- terminal positions are almost never left without a prosodic tag;
- at least three of the four transcribers agree in a very high percentage, i.e. from 83% to 96% of terminal positions;
- the frequency of non-terminal choices in terminal positions is extremely low.

Given that each C-ORAL-ROM texts has been revised by at least three transcribers the probability of leaving a terminal position unmarked is irrelevant and the probability of missing a not terminal-mark is strongly reduced. Therefore the validation test shows that the prosodic tagging in C-ORAL-ROM is reliable and coherent as regards the marking of terminal prosodic breaks.

From another point of view the low percentage of random application, the high coherence in marking also of non-terminal breaks, together with the very similar frequency of tone breaks registered in all collections also ensure the reliable application of this secondary tag.

## 8. REFERENCES

- Buchmann J., Caspers J., van Heuven V., Hoekstra H. Martens J.-P., Swerts M. (2002). Annotation of prominent words, prosodic boundaries and segmental lengthening by no-expert transcribers in the spoken Dutch corpus. In M. C. Rodriguez, C. Suarez Araujo (eds). *Proceedings of III<sup>o</sup> International Conference on Language Resources and Evaluation*, pp. 779-785. Paris, ELRA.
- C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Languages. At <http://lablita.dit.unifi.it/coralrom> and <http://www.htlcentral.org/projects/coralrom>
- Cresti E. (2000). *Corpus di italiano parlato*, vol. I-II, CD-Rom. Firenze, Accademia della Crusca.
- Cresti E., Moneglia M., Bacelar F., Choucri K, Mapelli V., Sandoval A. M., Veronis J. (2002). The C-ORAL-ROM Project. New methods for spoken language archives in a multilingual romance corpus. In M. C. Rodriguez, C. Suarez Araujo (eds). *Proceedings of III<sup>o</sup> International Conference on Language Resources and Evaluation* pp. 2-10. Paris, ELRA.
- Crystal D. (1975). *The English tone of voice*. London, Edward Arnold.
- 't Hart H., Collier R., 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. *Travaux du Cercle linguistique de Prague*, IV.
- Oostdijk N., Goedertier W., Van Einde F., Boves L., Martens J.-P., Moortgat. Baayen H. (2002). Experiences from the Spoken Dutch Corpus Project. In M. C. Rodriguez, C. Suarez Araujo (eds). *Proceedings of III<sup>o</sup> International Conference on Language Resources and Evaluation*, pp. 340-347. Paris, ELRA.

## 9. NOTES

- <sup>1</sup> See Cresti et al. (2002) and the Project official web sites: <http://lablita.dit.unifi.it/coralrom> and <http://www.htlcentral.org/projects/coralrom> for detailed information.
- <sup>2</sup> Rectively “/” for terminal and “/” for not terminal. Fragmentation breaks may mark retracting “[/” or interruptions “+”.
- <sup>3</sup> In the cases of Italian and Spanish corpora the validation may regard also fragmentation breaks. In the more general system retracting breaks are mapped on not-terminal breaks, while interruption are mapped on terminal breaks.
- <sup>4</sup> The distinction between weak and strong prosodic breaks is the only perceptually based prosodic annotation that has received a sufficient validation by both expert and not expert transcribers. See Buchman et al. (2002), and in particular the results showed in the oral presentation.
- <sup>5</sup> We refer tentatively to the measures calculated on the base of the revision of the first corpus delivery (June 2002). To such references only an heuristic value must be assigned (to avoid circularity in the argument).
- <sup>6</sup> It is assumed that as far as perception is concerned, the concordance of two expert transcribers in assigning a terminal break to a position gives enough evidence that prosodic cues allowing such judgments do exist.



- <sup>7</sup> The results for the four corpora have been processed in Florence and are stored in the follow-up section of the project's official web site.
- <sup>8</sup> The text choice from the Portuguese Informal collection, with an higher MLU, represent the more typical genre in that collection, that is interviews.
- <sup>9</sup> Moreover the in all convergent positions not defined as terminal where an empty choice occurs, almost never transcribers choose a terminal sign (5 cases on 1178 positions).

# Plurilingüismo, traducción y diccionarios

*María Carlota Nicolás Martínez*

Università degli Studi di Firenze  
(carlota.nicolas@unifi.it)

*With this brief article I would like to present some findings about the relationship between the Spanish and Italian languages. In particular I wish to analyze some areas where these can be seen: the double translations of important writings of the beginning and end of the XX century; translations of contemporary novels; the presence of Spanish words in Italian texts on the Internet and the way in which Spanish words of strong cultural meaning are considered in bilingual dictionaries. From the information gleaned some questions have been raised and these have underlined some present day multilingual tendencies.*

*Keywords:* multilingualism, dictionary, translation

## 1. DEFINICIÓN DE LOS TÉRMINOS Y ESTRUCTURA DEL TRABAJO

Es difícil decidir lo que en nuestros días se entiende por plurilingüismo, se puede tratar de un simple fenómeno de “coexistencia de varias lengua en un país o territorio”<sup>1</sup>, pero en muchas ocasiones la lengua se considera el elemento más representativo de una cultura y la “coexistencia” de estas culturas supera el “país”, sobrepasando el límite de la nación por lo que se podría hablar de plurilingüismo europeo. Nosotros vamos a ser aún más osados y de la definición vamos a quitar los límites territoriales que hoy no influyen como en el pasado en el intercambio cultural, fruto lógico de la convivencia.

Los otros dos términos principales de este artículo, *diccionario* y *traducción* no es necesario definirlos porque serán utilizados con su significado habitual.

El tema de este Seminario “*La gestión unitaria de la herencia cultural multilingüe europea y su difusión en la red*” coincide con uno de los temas más en boga de los últimos tiempos sobre todo en el campo social, debatido en las instituciones políticas y, claro, también en el mundo universitario. Para respetar el tema del Seminario hemos querido hacer un análisis del uso del léxico en un campo muy delimitado, esperamos que sirva para observar un punto de contacto entre culturas o lenguas. Hemos procedido, en un microcosmos lingüístico, al análisis de dobles traducciones de textos modernos, de traducciones de novelas contemporáneas, y a la búsqueda de algunas voces en la red y en algunos diccionarios bilingües.

## 2. PROCESO DE ANÁLISIS

### 2.1 *Análisis de dobles traducciones de textos modernos*

En esta parte del trabajo nos proponemos analizar dos traducciones<sup>2</sup> en italiano de textos literarios modernos españoles, por lo que hemos extraído de estas traducciones las palabras que el traductor ha dejado en la lengua original y que ha indicado en general en cursivo. Veremos a qué campos semánticos pertenecen estas palabras y hasta que punto son intraducibles por corresponder a aspectos específicos de la cultura española.

Mediante la extracción de estos datos querríamos poder corroborar o rechazar la hipótesis que: *el comportamiento del traductor con respecto a términos propios de la cultura ha cambiado de principios a finales del siglo XX: en la actualidad se ha dejado de considerar necesario hacer su traducción e incluso se ha dejado de dar una descripción detallada del término, simplemente se ha hecho una transferencia, dejando el término sin traducir.*

Los autores elegidos son de finales del XVIII y principios del XIX. Desafortunadamente no son tantas las traducciones y menos aún las dobles traducciones de estos clásicos españoles; hemos llegado a la selección de cinco obras que nos han parecido interesantes por su variedad.

*El político* de Azorín

*Gustavo el incongruente* de Ramón Gómez de la Serna

*Pepita Jiménez* de Juan Valera

*Tristana* de Benito Pérez Galdós

*Sonata de Primavera* de Ramón María del Valle-Inclán

La búsqueda ha sido hecha consultando los archivos de las bibliotecas de Florencia; no se ha incorporado un número de textos mayor, pues como veremos podemos sacar algunas deducciones a partir de este número; sería sin duda interesante hacer una catalogación y estudio exhaustivo de las traducciones de obras de este periodo, resultaría muy útil para la práctica y la teoría de la traducción.

Hemos cotejado las dos traducciones y el original y hemos recogido las transferencias encontradas en la siguiente tabla. Los asteriscos indican que la palabra va acompañada de una nota del traductor.

Subrayemos algunas de las características de estas obras y traducciones para ayudarnos a comprender mejor la tabla anterior.

*El político* es un ensayo publicado en 1908 y como vemos la primera traducción es de 1910 y la segunda del año 2001; en ninguna de las dos hay transferencias, es un ensayo de carácter muy general<sup>5</sup> sobre el comportamiento adecuado del político. La segunda traducción tiene una bibliografía y un índice comentado de nombres, es una traducción comunicativa (Newmark 1999: 72), solo guiada por la necesidad de transmitir el contenido del texto y que, como veremos, no siempre sabe transmitir el contenido de lo tratado<sup>6</sup>; se modernizan, en ella, algunos términos con connotaciones culturales del pasado<sup>7</sup> y podría tratarse de una obra original contemporánea. Es significativo que el editor en una banda que rodea el libro escriba: “Un manuale di ieri per politici di oggi”.

Gustavo, el incongruente		Pepita Jiménez		Tristana		Sonata de primavera	
1928	1991	1936	1991	1970	1991	1992	1995
duros*	duros	fandango	“fandango o di rondegna**”	hidalgo duro*	“gentiluomo”	“vestito alla chamberga**”	“vestito alla carlona”
pesetas	pesetas <sup>3</sup>				“cinque pesetas”		
flan							
plaza de toros	plaza de toros	“scudi” <sup>4</sup>	duros	pesetas	pesetas		
termilana	termilana	corrida	cacique	carabineros	carabineros		
cursilerias*				banderillas			
socorro*	socorro			doña			
navaja	navaja			jerez			
xilofon				malagueña			
pijama				verónica			
visillos				moriscos*	malagueña*		
				señò	verónicas*		
				monadas*	“mori		
				“turibolo”	scacciati”		
				“calzoni alla valenziana”	botafumeiro*		
					zaragüelles*		

Vemos, en cambio, que el traductor de la primera traducción italiana de la obra se concede libertades literarias sin ser siempre fiel al texto original<sup>8</sup>. La belleza léxica que Azorín derrocha queda disminuida también en esta obra<sup>9</sup>.

La obra de Gómez de la Serna *Gustavo, el incongruente* es una obra satírica; la acción está bien enmarcada en un periodo histórico y en ella hay muchos movimientos escénicos lo que favorece el uso de términos culturales. En ambas traducciones encontramos numerosas palabras extranjeras, especialmente francesas, transferencias del original. La edición de 1928 es muy cuidada, hecha por un traductor importante de la época a juzgar por el número de obras importantes que tradujo; tiene cinco notas, cuatro de ellas dedicadas a explicaciones sobre el significado o la forma en el original. El caso de la transferencia *visillos* es interesante pues se trata de un doblete “nei loro *visillos*, cioè nelle cortine” 44a, frente a la traducción de 1991: “nelle tendine” 23b. Tiene también interés la palabra *cursileria* con una nota que explica su procedencia y su significado, liquidada con “goffi oggetti” 25b. Ninguna transferencia tiene acento (*xilofon*, *cursileria*). *Socorro* se pone en español con una nota que explica el juego de palabras que se hace en el texto entre el nombre propio y el común.

En la traducción de 1991 no hay ninguna nota ni una introducción a la obra y vemos que algunas transferencias están en cursiva y otras no<sup>10</sup>. Se trata de una colección comercial de clásicos de humor, quizás esto nos explique el bajo número de transferencias frente a la traducción anterior.

Otra obra que nos ha ofrecido una interesante comparación ha sido *Pepita Jiménez*<sup>11</sup>. Las traducciones estudiadas son de 1936 y de 1991, ambas son ediciones

cuidadas y la más reciente tiene unas 40 notas sobre datos topográficos, biográficos y literarios. En el original encontramos algún fragmento en el que se presentan términos culturales gastronómicos, caso prototípico de uso de transferencias, que sin embargo no se han utilizado:

“Hubo hojuelas, pestiños, gajorros, rosquillas, mostachones, bizcotelas, ... se regaló con almíbares, chocolate, miel de azahar, y miel de prima, y varios rosolis y mistelas aromáticas y refinadísimas”. 283 (*mistelas* tiene una nota que explica que se refiere a tipos de aguardiente).

“Paste, fritelle, panocchie di granoturco arrostite, dolci d’ogni genere, biscotti e vino .... gustarono sciropi, confetture, cioccolato, miele di fior d’arancio, nonché rosoli ed altre miscele aromatiche e raffinate”. 270a.

“Vi furono paste, fritelle col miele, dolci, ciambelle, mostaccioli, biscottini leggeri, ... offerti cioccolata, frutta candita, miele di zagara, melassa, vari rosolii e bibite aromatiche e raffinatissime” 220b.

Estas traducciones no respetan siempre el significado: en ocasiones son equivalencias culturales, en otros casos funcionales (Newmark, 1999: 119-120). En alguna ocasión parece que se busca el equivalente fónico como en: “mistelas” (un tipo de aguardiente) traducido por *miscela*.

Las traducciones de *Tristana* son de 1970 y de 1991, esta última es una edición bilingüe. La primera edición de esta novela es de 1892.

La edición del 70 tiene 27 notas, tres lingüísticas y las demás culturales.

La edición del 91 tiene 53 notas la mayoría culturales o sobre la procedencia de citas; muy pocas son lingüísticas, como por ejemplo, la justificación del uso de *nomadi* en lugar de *monadas* para conseguir un juego de palabras.

Las traducciones de la *Sonata de Primavera* son de 1992 y de 1995; la primera edición en lengua original es de 1944. Es probable que el carácter poético de la obra influya en el escaso número de transferencias<sup>12</sup>. La edición de 1995 no tiene ni una sola nota sobre la traducción, sin embargo tiene muchas notas de carácter filológico. La edición de 1992 tiene una sola nota es del traductor y explica el uso de la transferencia *chamberga*.

Los datos recogidos hasta ahora nos llevan a las siguientes reflexiones:

Parece lógico pensar que estas obras, al ser de grandes autores, suelen tener traducciones de carácter filológico, como hemos visto la mayoría tiene notas; algunos traductores pertenecen al mundo académico. De las ocho traducciones cinco tienen notas del traductor a las transferencias hechas.

El carácter poético y ensayístico de dos de estas obras puede haber determinado el poco uso de transferencias.

La traducción más antigua de la obra satírica presenta transferencias del español muy interesantes, pues palabras como *visillos* y *pijama* son casos limítrofes por lo que se refiere a los campos semánticos habituales en los que encontramos las transferencias, que no son ni la ropa interior ni el mobiliario.

Quedan dos obras, *Tristana* y *Pepita Jiménez*, que pueden aportarnos alguna luz ya que son las obras con más transferencias. La traducción más reciente de *Pepita Jiménez* tienen un número mayor, mientras sucede lo contrario en *Tristana*. Hay que tener en cuenta que las traducciones de *Tristana* se pueden considerar contemporáneas al ser de 1970 y de 1991; estas dos traducciones son las que tienen entre todas las vistas el número mayor de transferencias.

Podemos concluir que no tenemos datos para afirmar que en las traducciones más recientes hay más transferencias. Podemos decir que la mayor presencia del traductor, es decir los casos en los que hay bastantes notas, son los casos que coinciden con un buen número de transferencias. Este binomio, traducción profesional transferencia, resulta interesante.

Lo que no podemos afirmar, para corroborar la hipótesis planteada, es que el número de transferencias es siempre mayor en las traducciones de finales del XX, pues vemos una obra de 1928 con un número bastante grande de ellas, por lo tanto, creemos que no depende del periodo sino de cada caso concreto.

## 2.2 Análisis de traducciones de novelas contemporáneas

Seguimos nuestro camino de extracción de información y análisis de otros datos. Presentamos en las siguientes tablas las transferencias encontradas en cursiva en otros textos traducidos. Hemos elegido al azar diez novelas de contemporáneos españoles e hispanoamericanos. Las obras originales son de 1979 a 1995 y las traducciones de 1988 a 1996.

Estos resultados nos permiten llegar a una interpretación, primero veamos los datos: dos de las ocho traducciones carecen de transferencias; entre las que tienen un número medio las hay de 8, 11, 15 y 19; en el extremo hay dos con 48 y 40; próximas a cero hay una con 1 y otra con 2. No podemos decir que es el estilo del traductor lo que determina el número, pues G. Felice traduce dos obras una con 48 y otra con 15 transferencias. La traductora I. Carmignani hace dos traducciones con 8 transferencias una y 11 la otra. Evidentemente el número de transferencias no depende tampoco de la procedencia del autor, pero sin duda puede depender del tipo de obra. Entre las muchas dudas que se plantean podemos hacer una afirmación: el número de traducciones con gran número de transferencias es, en general, mayor en traducciones de autores contemporáneos que en las de autores modernos.

Los campos de referencia de estas palabras son efectivamente culturales, referidos la mayoría a gastronomía, espacios vitales y personas; bastante menor es el número de palabras del mundo animal, vegetal, folclórico, las referidas al vestuario, a las monedas y medidas. Es interesante observar los significados de las otras transferencias.

	<i>El misterio de la cripta embrujada</i> 1979 (Mendoza, 1990) español pp. 136 (0)	<i>Arráncame la vida</i> 1986 (Mastretta, 1988) mejicana pp. 223  (40)	<i>El invierno en Lisboa</i> 1987 (Muñoz Molina, 1995) español pp. 211 (0)	<i>Las edades de Lulú</i> 1989 (Grandes, 1990) española pp. 240  (11)	<i>El viejo que leía novelas de amor</i> 1989 (Sepúlveda, 1993) chileno pp. 133  (8)
gatronomía		tamales* mole en chile pasilla chiles en nogada chalupas chileatole tinga chilaquiles pozole epazote pipián tinga tortillas muérganos anis cecina		cocido leche merengada tocino de cielo pastaza aceitazo harinaza picatostes empanada	
lugares		zócalo* barrio rancho plaza de toros sierra			
personas o grupos por procedencia, oficio, apelativos		chipileños novillero peones mariachi costeño* hombre damas y caballeros (con doblote) charro gringa capacito		picadores banderilleros	jíbaros
folclore		corrido paso doble			
monedas, medidas		centenarios*		duros	sucres
ropa		huipil*			
animales, plantas		pirú (árbol)			(animales) ti- grillo saínos bagre guaca- mayo tzanza
otros		terital (tela) mal del pinto* (ma- lattia) anticristera (lucha) piñatas (fiestas)			anets (poemas)

	<i>Amirbar</i> 1990 (Mutis, 1994) colombiano pp. 137  (2)	<i>Corazón tan blanco</i> 1992 (Marías, 1996) español pp. 287  (1)	<i>Cuando ya no im- porte</i> 1993 (Onetti, 1994) uruguayo pp. 197  (19)	<i>La hora sin sombra</i> 1995 (Soriano, 1996) peruano pp. 220  (15)	<i>Las semanas del jardín. Un círculo de lectores</i> (Goytisolo, 1997) pp. 165  (26*+22)
gastronomía			locro* empanadas asado mates*	asados mate dulce de leche	alboronía* bastela* anís alcuzcuz
lugares	camino real		embarcadero ranchos	estancia rancho	riad zebala aduar* alhama* alquibla* jaima* mechuar* zagüía*
personas o grupos por procedencia, oficio, apela- tivos			gringos gringa doña che patroncito patrón peón caudillos	guerrillero montoneros estanciero peronchos guerrilla che	emjzní señoritos almahal camarada afrancesados dueñas meninas moras (fuerzas) alfaquí* amazigh* Glaiú* hedillista* áscari* ausente* jerifiano* mojamé* morabo* moro* mudéjar* nesrani* requetés* tabor* tercio* gnaua hedaua
folclore		chotis*			
monedas, medidas			peso		riyal arrobas
vestuario			ponchos		chilaba mantillas
animales, plantas	guayacán (árbol)			talas (planta) suribí (animal) chimangos (pájaro) benteveos (pájaro)	
otros			(colores de caballos) rosillo tubiano pangaré		casida (poema) pronunciamento alzamiento azulejos (dar ) el paseo alifato* baraca* (don divino) budali (lengua?) azalá* (rezo)



La presencia de estas transferencias confirma la hipótesis enunciada al principio de este artículo. Podemos pensar que en la actualidad se tiene una idea de proximidad cultural que lleva al traductor a dar una transferencia creyendo que no provoca la pérdida de un contenido importante para el lector, o bien porque son palabras conocidas por el lector medio o bien porque por el contexto se pueden deducir los significados; aunque se pierda la referencia perfecta al objeto mencionado se puede suponer de qué se trata, como en el caso de plantas o animales, de monedas o medidas o ropas.

Veremos como incluso cuando existe un equivalente funcional que “consiste en utilizar una palabra culturalmente neutra y añadir, a veces, un nuevo término específico” (Newmark, 1999: 120) o un equivalente cultural “es la traducción de un término cultural de la LO por un término cultural de la LT [...] son equivalentes culturales aproximados” (Newmark, 1999: 119), el traductor decide no traducir estas palabras; las razones que lo mueven pueden ser muchas y las más lógicas son estas: para dar mayor colorido, por exotismo, para facilitar la lectura, para resolver el problema o, simplemente, por considerarlo más correcto.

### 2.3 Control de voces en diccionarios bilingües

Pasamos a controlar cómo el diccionario bilingüe trata estas palabras. Hemos decidido buscar en varios diccionarios las transferencias encontradas en dos de las traducciones con un porcentaje medio y de los dos traductores que traducen dos de las obras elegidas casualmente, lo que nos hace suponer que son profesionales serios. Los diccionarios usados son de tres dimensiones muy diferentes: El Ambruzzi 1973 (a) suponemos que tiene en cada parte del diccionario entre 40.000 y 50.000 entradas o subentradas, el Calvo, Giordano 1993 (b) tendrá alrededor de 35.000 y el Lavacchi, Nicolás Martínez 2000 (c) declara unos 18.000.

Hemos llamado explicación a lo que comprende más de un sintagma nominal, como “cavaliere che ferisce il toro con la lancia nella corrida” (c)<sup>13</sup>; en el Calvo, Giordano se presentan tales explicaciones entre comillas tras un signo matemático de igualdad: = “Brodo di carne, salsicce speciali, prosciutto crudo, ceci e verdure” (b). Llamamos *equivalente extenso* a un sintagma nominal de tipo *Cocido*: “vivanda a base di carne, ceci, verdure, ecc” (a) o *Estanciero*: “padrone della *estancia*” (azienda agricola e pastorile) (a). Vemos por ejemplo la diferencia entre un *equivalente* normal y uno *extenso*: *Duro*: “scudo da argento di 5 pesetas / moneta di altro metallo da 5 pesetas” (a), “moneta di cinque pesete” (b). Un ejemplo de *equivalente con glosa* es este: “sfoglia (di pasta di pane ripiena)” (c).

De este cuadro se puede deducir que en la mayoría de los casos estas transferencias no han sido necesarias: solo las cuatro plantas y las tres palabras referidas a alimentos no han encontrado un equivalente en el diccionario. Es evidente que los derivados de palabras sencillas como son *aceite*, *harina* y *pasta* se pueden traducir, pues existe la connotación que se quiere dar mediante el sufijo; *peronchos* también es una palabra derivada y que se puede traducir si se está seguro de la connotación que tiene. *Picadores* y *banderilleros* tienen un equivalente en singular en el Calvo, Giordano, *picador*,

*banderillero*, pero poner estas palabras en plural no es fácil al oído italiano, son pues justificables como transferencias obligadas.

	Ambruzzi	Calvo–Giordano	Lavacchi–Nicolás
asados	equivalente	equivalente	equivalente
mate	explicación	equivalente	
dulce de leche	equivalente		
estancia	equivalente	equivalente	
rancho	equivalente	equivalente	equivalente con glosa
guerrillero	equivalente	equivalente	equivalente
montoneros	equivalente		
estanciero	equivalente extenso	equivalente	
peronchos			
guerrilla	equivalente	equivalente	equivalente
che	explicación	equivalente	
talas (planta)	explicación		
chimangos (pájaro)	explicación		
benteveos (pájaro)	explicación		
suribí (animal)			
cocido	equivalente extenso	explicación	
leche merengada	explicación		
tocino de cielo	explicación	explicación	
pastaza			
aceitazo			
harinaza			
picatostes	equivalente extenso	equivalente extenso	explicación
empanada	explicación	equivalente	equivalente con glosa
picadores	explicación	equivalente	explicación
banderilleros	explicación	equivalente	
duros	equivalente extenso	equivalente	equivalente

Quedan por lo tanto 7 palabras cuya traducción resulta farragosa y cuya transferencia se justifica por este motivo. Son 26 las palabras examinadas 4 las derivadas, de las 22 restantes solo hay 7/9 que no se pueden traducir mediante un equivalente. Esto nos lleva a deducir con cierta seguridad que el uso de transferencias es una cuestión de elección libre del traductor.

Otro posible control lo hacemos en la base de datos en la que está contenido el Lavacchi–Nicolás. En esta base de datos se han agrupado en un campo las entradas que no han encontrado un equivalente y han sido descritas buscando una explicación: 46 palabras de

gastronomía, 30 de costumbres, fiestas y grupos sociales, 14 de tauromaquia, 11 de literatura, 12 de arquitectura, 9 de personas, 4 de oficios, 4 de mobiliario y 2 de vestuario.

Vemos que coincide con los ámbitos de las transferencias, salvo por lo que se refiere a los animales o las plantas que no aparecen en absoluto, sin duda porque este diccionario no recoge palabras procedentes de hispanoamérica. Hay monedas y medidas que tienen una traducción clara como *pesetas*, mientras que las medidas antiguas no se hallan en este diccionario que recoge el lenguaje común actual.

De estos datos se puede deducir que también el lexicógrafo siente la necesidad de respetar las palabras propias de una lengua sin hacer esfuerzos por buscar falsos equivalentes.

Los datos de los tres análisis hechos hasta ahora nos indican que en la actualidad la presencia de transferencias es mayor y más constante, e inmotivada al menos en parte. Lo que puede incluso autorizar dos opuestas interpretaciones: la primera es que las características propias de una cultura reflejadas en estas palabras se tratan con enorme respeto; la segunda que se dejan sin traducir porque para esta convivencia entre culturas es suficiente que se alcance una idea vaga de lo que quieren decir.

Quizás sean dos modos extremos de interpretar los hechos, pero pueden ser las dos caras de lo que hoy se entiende por intercultura.

#### 2.4 Frecuencias en la red

Seguiremos nuestro camino observando en otro medio, en la red (que hoy es un buen metro de tendencias de las lenguas) el número de palabras culturales españolas que aparecen en textos italianos. Hemos buscado en Google.it en las páginas *en italiano* y en Google.es *en español* las mismas palabras. Aunque no estamos seguros de que en realidad sean solo textos en italiano, podemos considerar esta tabla un dato indicativo, pero no representativo de la presencia de estas palabras españolas en la lengua italiana.

Hay que saber que *rancho*, *banderillero*, *picador*, *estancia*, *montonero* son préstamos aceptados ya en la lengua italiana.

Otras palabras en páginas *en italiano* que nos han provocado curiosidad y sorpresa son las siguientes:

*churro* 178; *churros* 568; *puchero* 3.050; *tapas* 14.200; *gazpacho* 5.820;  
*merendero* 3.730; *merenderos* 462; *bares* 3.700; *paseo* 22.500;  
*cacique* 2.280; *caciques* 274; *morisco* 3.580; *moriscos* 678; *chulo* 11.500; *chulos*  
147;—.

Los datos nos indican la presencia de estas palabras, incluso de las que tienen una posibilidad clara de traducción. Notamos la presencia de palabras que tienen una posibilidad clara de traducción. Observamos, también, que se pueden encontrar en la red términos que no están recogidos en el diccionario de la RAE, por ejemplo de *peronchos* y *suribí*.

	<i>en italiano</i>	<i>en español</i>
Asados	219; singular 3.670	101.000; singular 198.000
dulce de leche	686	60.400
Estancia	32.200	706.000
Rancho	11.900	323.000
Guerrillero	557; singular 308	218.000; singular 185.000
Montoneros	434	25.200
Peronchos	215 (en toda la red)	214
Guerrilla	48.700	276.000
chimangos (pájaro)	7; singular 296	735; singular 11.900
benteveos (pájaro)	1; singular 147	17; singular 1570
suribí (animal)	1	17
Cocido	5.370; plural 57	155.000; plural 83.800
leche merengada	5	897
tocino de cielo	11	2.270
Pastaia	470	43.500
Aceitazo	57 (en toda la red)	55
Harinaza	1 (en toda la red)	1
Picatoses	10; singular 16	816; singular 3760
empanada	400	63.700
Picadores	339; singular 2820	19.200; singular 21.400
Banderilleros	207; singular 94	7.890; singular 7970
duros	3.040	776.000

Los datos contenidos en este monstruo de información son difíciles de interpretar; creemos que la red al ser un contenedor ilimitado y poco concreto solo nos permite extraer las frecuencias relativas y observarlas como hechos.

Vemos que el uso de palabras de origen español en las páginas italianas en comparación con su uso en páginas españolas es de un 1% de *chimangos*, *picatoses*, *dulce de leche*, *empanada* aly 4% de *duros*; y en cambio de para palabras ya reconocidas como préstamos encontramos que es el 18% en de *guerrilla*, el 5% de *estancia* y el 2% de *montoneros*.

### 3. CONCLUSIONES

Hemos querido poner en relación traducciones y plurilingüismo, puesto que “La traduzione nasce da una distanza culturale, per lieve che sia:” (Terracini, 1996: 43) y debe respetar dos formas de cultura: “possiamo giungere così al vero e proprio tradurre, cioè da lingua a lingua, che dunque non sarà riprodurre formalmente il linguaggio altrui, ma trasportarlo da una forma culturale ad un'altra, giacché ogni lingua, considerata storicamente, ci appare come il prodotto

elaborato dalla tradizione di una particolare forma di cultura” (Terracini, 1996: 44).

Sabemos que en el trabajo del traductor se refleja perfectamente el fenómeno del plurilingüismo. “Per il traduttore invece le due lingue e le due culture coesistono, nettamente distinte l’una dall’altra, e con grado di assoluta parità” (Terracini, 1996: 45). La coexistencia de estas dos lenguas en el traductor tiene una neta distancia, pero quizás en él se concentre la máxima atención para mantener bien separadas las dos realidades culturales: “nel caso della traduzione le due forme di cultura permangono distinte e contrapposte l’una all’altra.” (Terracini 1996: 45) Incluso Terracini cree que llega a crearse un conflicto de personalidad. “Tutti sappiamo quanto sia difficile fare l’esatta separazione tra ciò che è nostro e ciò che è d’altri nel discorso normale, in conseguenza del naturale, fatale sovrapporsi di una personalità ad altra. Orbene, nel caso del tradurre che per definizione, di proposito, tende ad annullare la propria personalità a tutto vantaggio del suo autore, questa difficoltà non sussiste, o per lo meno è attenuata dal fatto che il traduttore sta con gli occhi aperti per evitarla.” (Terracini 1996: 42) del que el traductor tiene que ser consciente de eso.

Una vez afirmado esto volvemos a los datos recogidos anteriormente., Hemos visto que en las traducciones de obras literarias contemporáneas se utilizan un buen número de transferencias no todas necesarias, por varios motivos: por tratarse de préstamos ya en la lengua italiana (*picador/picadores*), por ser palabras ya italianas (*picarol/picari* o *guerrigliero*), por tener equivalentes claros como *asados=arrosti*. Por tanto existe la posibilidad de no usar muchas de las transferencias encontradas. Pero si consideramos lo dicho anteriormente sobre el traductor y su conciencia de mantener bien diferenciadas las culturas con las que trabaja puede que el uso que hace de las transferencias sea justo un modo de mantener mejor las distancias entre ambas culturas. , La realidad es que el uso de la cursiva con las transferencias puede ser un modo de protección del patrimonio cultural cuando quien lo usa es un traductor que quiere conscientemente mantener estas distancias culturales. Es, porque es justo la cursiva la que hace aislar estas palabras como en una convivencia marcada por una cierta distancia. Sin duda de esta manera es más clara y fuerte la coexistencia de las dos culturas, se pone más de manifiesto que si se crean traducciones usando equivalencias funcionales o culturales, llegando a compromisos satisfactorios. El comportamiento del traductor frente a libros no contemporáneos es probablemente muy distinto, como ya hemos visto en el caso de obras modernas; la distancia temporal es una perspectiva de lejanía cultural, la misma que cada persona tiene con las obras antiguas de su propia cultura; el traductor es consciente de que ya no hay convivencia cultural que proteger, puede pues hacer uso de más equivalentes y de menos transferencias (usadas solo en casos más justificados). Esto coincide con el análisis de las varias traducciones, hecho al principio de este artículo, tras el cual no se ha podido confirmar la hipótesis de que en la actualidad, en todo tipo de traducción literaria, el uso de transferencias es mayor que en el pasado.

Las palabras españolas vistas en las páginas italianas en la red no las podemos considerar transferencias, por lo menos no creemos que estén allí escritas con la misma

conciencia de diferenciación y convivencia de culturas que creemos que mueve al traductor literario, es decir, quizás no estén allí por respeto de las diferencias culturales., pProbablemente la coexistencia de términos culturales, sin conciencia de diferenciación, es mayor en la red, puesto que no sabemos si los que escriben en la red son conscientes de esta diferenciación: no sabemos si lo son, y vemos por ejemplo el uso enorme de palabras que existen en italiano (, *guerriglia/ guerrilla*). Quizás una última pregunta que nos podemos hacer es si todas estas palabras usadas como transferencias en la red se entienden o solo se medio entienden, con una cierta vaguedad que quizás caracteriza los tiempos, ricos de medios y carentes de contenidos.

Por lo tanto es probable que el mundo de las traducciones literarias y el de la red sean dos fuerzas distintas, una con conciencia de diferenciación de las dos culturas, la otra sin ella, pero creemos que ambas contribuyen a un mayor convivencia de términos de las varias culturas en la propia, . eEsta fuerte convivencia parece una clara tendencia de nuestros tiempos.

#### 4. REFERENCIAS BLIOGRÁFICAS

##### Traducciones

- Goytisolo J. (2004). *Le settimane del giardino. Un circolo di lettori*. Torino, Einaudi (trad. Glauco Felice).
- Grandes A. (1990). *Le età di Lulù*. Parma, Guanda (trad. Ilide Carmignani).
- Marías J. (1996). *Un cuore così bianco*. Roma, Donzelli (trad. Bianca Lazzaro).
- Mastretta A. (1988). *Strappami la vita*. Milano, Feltrinelli (trad. Sandro Ossola).
- Mendoza E. (1990). *El mistero de la cripta stregata*. Milano, Feltrinelli (trad. Gianni Guadalupi).
- Muñoz Molina A. (1995). *L'inverno a Lisbona*. Milano, Feltrinelli (trad. Elena Liverani).
- Mutis A. (1994). *Amirbar*. Torino, Einaudi (trad. Fulvia Bardelli).
- Onetti J. C. (1993). *Quando ormai nulla più importa*. Torino, Einaudi (trad. Raúl Crisafio e Pier Luigi Crovetto).
- Sepúlveda L. (1993). *Il vecchio che leggeva romanzi d'amore*. Parma, Guanda (trad. Ilide Carmignani).
- Soriano O. (1996). *L'ora senza ombra*. Torino, Einaudi (trad. Glauco Felice).

##### Dobles traducciones

- Azorín (2000). El Político  
<[http://omega.ilce.edu.mx:3000/sites/fondo2000/vol2/03/htm/sec\\_3.html](http://omega.ilce.edu.mx:3000/sites/fondo2000/vol2/03/htm/sec_3.html) >
- Azorín (2001). *Il politico*. Firenze, Le Lettere (trad. Lia Ogno).
- Gómez de la Serna R. (1928). *Gustavo, l'incongruente*. Milano, edizioni "Corbaccio" (trad. Gilberto Beccaci).
- Gómez de la Serna, R. (1991). *L'incongruente*. Roma, Lucarini Editore (trad. Lucio D'Arcangelo).

- Martínez Ruiz J. (1910). *Il politico*. Firenze, Ferrante Gonelli (trad. G. Beccari)
- Pérez Galdós B. (1970). *Tristana*. Milano, Adelphi (Versión de di Italo Alighiero Chiusano).
- Pérez Galdós B. (1991). *Tristana*. Venezia, Marsiglio (trad. Augusto Guarino) edición bilingüe.
- Valera J. (1936). *Peppina Jiménez*. Milano, Mondadori (trad. Mario Puccini).
- Valera J. (1994). *Pepita Jiménez*. Milano, Fabbri editori (trad. Gabrielli Morelli) (1ª ed. 1991).
- Valera J. (2001). *Pepita Jiménez*. Madrid, Grupo Santillana de Ediciones.
- Valle-Inclán R. (1944). *Sonata de Primavera*. Madrid, Espasa-Calpe.
- Valle-Inclán R. (1992). *Sonata di Primavera*. Firenze, Passagli (ed. Oreste Macrì).
- Valle-Inclán R. (1995). *Sonata di Primavera*. Venezia, Marsiglio (ed. Giovanni Battista De Cesare) edición bilingüe.

### Bibliografía citada

- Ambruzzi L. (1973). *Nuovo dizionario Spagnolo-Italiano, Italiano-Spagnolo*. Torino, Paravia.
- Calvo Rigual C., Giordano A. (1996). *Diccionario italiano*. Barcelona, Editorial Herder.
- Lavacchi L., Nicolás Martínez M. C. (2000). *Dizionario Spagnolo-Italiano, Italiano-Spagnolo*. Firenze, Le Lettere.
- Newmark P. (1999). *Manual de traducción*. Madrid, Cátedra.
- Pierini P. (1999). La ritraduzione in prospettiva teorica e pratica, en P. Pierini *L'atto del tradurre*. Roma, Bulzoni pp. 51-72.
- Pierini P. (ed.) (1999). *L'atto del tradurre*. Roma, Bulzoni.
- Real Academia Española, (Vigésima segunda ed.). *Diccionario de la Real Academia Española de la Lengua*. Versión 1.0 CD.ROM
- Terracini B. (1996). *Conflitti di lingue e di cultura*. Torino, Einaudi.

## 5. NOTAS

- <sup>1</sup> Diccionario de la Real Academia Española de la Lengua. Versión 1.0 CD.ROM, (Vigésima segunda edición).
- <sup>2</sup> Consideramos de gran interés el estudio de las varias traducciones de una obra: “[la ritraduzione] offre interessanti spunti di ricerca; ad esempio, il censimento della ritraduzione dà un dato TP può delineare la storia della sua diffusione, del suo influsso e dalla sua interpretazione” (Pierini, 1999: 53). En este artículo concreto nos limitaremos a hacer un estudio no global de la “ritraduzione” de una obra.
- <sup>3</sup> Estas transferencias no están en cursiva en la traducción.
- <sup>4</sup> Cuando los términos están entre comillas es para compararlos con la otra traducción, presentada en la misma línea; en este caso duos se encuentra en la otra edición por “scudi”.
- <sup>5</sup> Sin duda, a pesar de ser un ensayo, hay términos que se podían haber dejado en español: “Todo este mundo de los pequeños alhauines o tejedores, de los peltreros, de los percoceros o modestos plateeros, de los herreros, de los carpinteros, tiene su incanto”: tutto questo mondo di piccoli tessitori, di peltrai, di modesti argentieri, di fabbri 154a; Tutto questo mondo di piccoli cardatori o tessitori, di peltrai, di bulinatori o modesti argentieri, di fabbri 130b.

El traductor del 2001 traduce alhaquín como si fuera un cardador; la solución de 1910 es no traducir todas las palabras de la lista ya que en dos casos se trata de sinónimos.

- <sup>6</sup> La primera edición da un equivalente mientras el falso amigo ha engañado la elección del segundo traductor: “galopín”, “galopines”: malvagio 99a, igli scelti 90a; galoppino 94b, i galoppini 92b.
- <sup>7</sup> Un caso de traducción de equivalencia cultural anacrónica: “pero no lo pongamos tampoco debajo del celemín”: ma non lo mettiamo neppure sotto i piedi 96a; ma non lo mettiamo nemmeno sotto lo zerbino 92b. Un caso de pérdida de colorido del texto original: “sepa sus picardihuelas y malicias”: sappia le loro maliziette e furberie 114 a; sappia le loro malizie e le loro sottefurgi 105b.
- <sup>8</sup> Incluso los títulos son tratados con mucha libertad “Fingir conformidad”: Ostentar rassegnazione 88a, Fingere rassegnazione 88b.
- <sup>9</sup> Vemos “atildamiento melindroso”: “eleganza esagerata” 126a; “ornamento affettato” 113b.
- <sup>10</sup> Notamos también algún despiste: “zerza russa” (curiosa la creación de un doble error a partir probablemente de “berza roja”); en la edición de 1928 “zerza rossa”.
- <sup>11</sup> Agradezco a la alumna Marta Simoncelli, que en este momento está haciendo su tesina sobre esta la doble traducción de este autor, su ayuda prestada.
- <sup>12</sup> Hemos anotado alguna palabra de las poquísimas con connotación cultural como por ejemplo “casa terrena” traducida como casa di campagna; casupola; “birreta” cappello a tre punte, berreto; “silla de posta” corriere, carrozza, en la edición de 1992 y 1995 respectivamente.
- <sup>13</sup> En el formato electrónico, aún no publicada, de Lavacchi-Nicolás Martínez aparece en una ventana distinta; está en la lista de tauromaquia que se puede extraer de la base de datos que contiene las voces del diccionario.





# L'editoria digitale della Firenze University Press

*Patrizia Cotoneschi*

Firenze University Press  
Università degli Studi di Firenze  
(cotoneschi@unifi.it)

*Antonella Farsetti*

Firenze University Press  
Università degli Studi di Firenze  
(antonella.farsetti@unifi.it)

*Since the early stage of its life Firenze University Press (FUP) has as its main objective digital publishing and the perspective of new modalities of text fruition. Thanks to a fulltext research systems FUP allows searches of forms and lemmas and the extraction of frequency lists within its catalogue. After reserach and localization the customer can buy a text, or part of it, for downloading or as print on demand. In the FUP Catalogue each work receives a description following the main international standard available for e-publishing, to allow maximum interoperability and long term accessibility to texts. At present an xml based structuring of the catalogue is being developed. More specifically the definition of a DTD for journals will allow a better multiformat visualization of content and its long term diffusion.*

*Keywords:* e-publishing, metadata, xml based archives

## 1. IL CONTESTO DELL'EDITORIA ACCADEMICA E DELL'INFORMAZIONE SCIENTIFICA

La Firenze University Press, casa editrice dell'Università degli Studi di Firenze, condivide i motivi della sua esistenza con tutte le altre University Press nate in fasi alterne nelle varie università del mondo, vale a dire, la diffusione della conoscenza attraverso *publishing*<sup>1</sup>.

Al momento attuale le University Press nel mondo sono circa 1000 con diverse caratteristiche, ma tutte unite dal comune intento di offrire risposte alla richiesta di diffusione dei risultati della ricerca che il mondo dell'editoria tradizionale non soddisfa. Esse si sono sviluppate soprattutto nelle università degli Stati Uniti d'America, in Spagna, nell'America latina e nei paesi dell'Europa orientale, lasciando il resto delle università ad appoggiarsi pressoché totalmente sull'editoria locale.

L'Italia fino ad alcuni anni fa, era senz'altro uno di quei paesi nei quali le università avevano delegato (e continuano a delegare in molti casi anche oggi) questo compito di veicolazione del sapere alle diverse realtà editoriali non istituzionali, sparse sul territorio nazionale.

L'editoria commerciale ha attualmente in mano la quasi totalità delle pubblicazioni scientifiche mondiali, a seguito di un fenomeno creatosi subito dopo la seconda guerra mondiale.

Infatti nel dopoguerra, a causa del notevole incremento della richiesta di pubblicazioni per l'allargamento del circuito informativo, si è avuto il graduale passaggio delle pubblicazioni di associazioni scientifiche e università, nelle mani degli editori commerciali. Questo fenomeno in Italia, ed anche all'estero, ha portato come conseguenza per le università alti costi di pubblicazione e una scarsa circolazione dell'informazione scientifica che nasce all'interno delle università stesse.

L'informazione scientifica, frutto delle attività di ricerca, viene pubblicata (spesso a pagamento) presso editori commerciali, i quali a loro volta la rivendono alle biblioteche delle stesse università che la hanno prodotta.

I costi di questa spirale sono ulteriormente maggiorati se si tiene conto del fatto che le università devono sostenere per le loro pubblicazioni una serie di costi aggiuntivi come ben evidenzia Sevenjie:<sup>2</sup>

- il costo della risorsa umana che produce e scrive la pubblicazione;
- il costo di eventuali revisioni e correzioni richieste dall'editore;
- il costo di acquisto della pubblicazione (spesso sotto forma di estratti a pagamento);
- il costo del ri-utilizzo della pubblicazione riacquistata spesso in pacchetti di abbonamenti;
- l'archiviazione della pubblicazione (spesso nelle biblioteche).

Quindi le università pagano ben più di una volta, il costo del lavoro prodotto dai propri studiosi e ricercatori. Questo meccanismo andato consolidandosi negli anni, ha creato una spirale dalla quale sembra pressoché impossibile uscire.

Tuttavia negli anni novanta si assiste nel mondo anglosassone, e un quinquennio successivo anche in Italia, alla presa di coscienza del fatto che l'aver delegato ad editori esterni all'università la gestione delle pubblicazioni scientifiche accademiche, stava portando effetti devastanti alla circolazione del sapere. Le università non erano più in grado di ricomprare l'informazione che i ricercatori avevano ceduto agli editori.

Il fenomeno dell'aumento esponenziale dei prezzi delle riviste costituisce allora, la bandiera issata sulle barricate innalzate dai bibliotecari nel mondo, che tra i primi analizzano la spirale dei costi dell'editoria scientifica, i cui contenuti sono prodotti, nella maggior parte dei casi, proprio all'interno delle istituzioni di cultura e ricerca, quali le università<sup>3</sup>.

È in questo contesto che le istituzioni universitarie sembrano divenire sempre più consapevoli delle potenzialità di mercato del frutto del proprio sapere.

Un ulteriore elemento che va ad incrementare questa presa di coscienza delle istituzioni accademiche, è sicuramente legato al vantaggio, che le università hanno, di essere esse stesse fonti produttive delle loro opere di ingegno: libri e periodici.

Inoltre, un ulteriore aspetto svantaggioso che caratterizza la spirale dei costi, è legato al fatto che le università non ricevono compensi economici, per la cessione dei diritti delle loro opere d'ingegno, ma solo la possibilità che esse vengano diffuse e distribuite all'interno del circuito accademico culturale di riferimento.

L'editoria universitaria ha bisogno, quindi, di uscire dalla spirale dei costi creatasi con molti degli editori commerciali, ma ha anche bisogno di garanzie, per far sì che gli obiettivi che guidano la scelta dell'editore, siano quelli del controllo di qualità delle pubblicazioni e della loro diffusione, e non siano solo obiettivi economico lucrativi<sup>4</sup>.

## 2. LA NASCITA DELLA FIRENZE UNIVERSITY PRESS

Quando il progetto di editoria elettronica della Firenze University Press (FUP) fu approvato nel febbraio 2000, la scommessa sull'uso dell'ambiente digitale fatta da FUP, voleva trovare soluzione a tre problemi ritenuti nodali, legati alla produzione editoriale dell'Ateneo fiorentino: il recupero del *copyright*, la riduzione dei costi e l'incremento della visibilità delle pubblicazioni dell'Università di Firenze.

L'esistenza di questi problemi da risolvere ha trovato anche riscontro, due anni più tardi, nello studio condotto grazie all'*Osservatorio sull'editoria elettronica italiana*<sup>5</sup> nato per iniziativa di editori commerciali come Casalini digital.

Da questo studio riguardante sia l'editoria accademica sia quella commerciale, emergono aspetti della realtà editoriale universitaria sui quali è necessario soffermarsi. Innanzi tutto l'attività accademica italiana appare frammentata, affidata spesso all'attività dei singoli dipartimenti senza un servizio editoriale centralizzato, i contenuti scientifici sono spesso diffusi tramite singoli siti locali e poco valorizzati avendo una visibilità limitata; le pubblicazioni maggiormente presenti nell'ambiente accademico risultano essere le riviste ed il materiale didattico. Da questo stesso osservatorio è emersa anche la nascita di alcune *University Press*, con l'ambizioso obiettivo di riappropriarsi dei contenuti intellettuali, frutto del lavoro delle università stesse.

Grazie alla collaborazione di un gruppo di docenti dell'Ateneo fiorentino che ebbe chiaro fin dal suo nascere l'importanza di una University Press elettronica, fortemente voluta dal Rettore Paolo Blasi, fu possibile pubblicare un piccolo numero di opere e un paio di riviste. Fu così che nacque la Firenze University Press con i seguenti obiettivi:

- assistere gli autori universitari nelle pubblicazioni elettroniche offrendo un servizio editoriale;
- facilitare l'accesso e la diffusione delle pubblicazioni elettroniche dell'Università con l'utilizzo di avanzate tecnologie digitali;
- valorizzare la produzione editoriale dell'Ateneo garantendo la certificazione di autenticità e i diritti di proprietà intellettuale sia dell'edizione elettronica che cartacea;

- dare visibilità soprattutto alla ricerca e alle iniziative emergenti (giovani ricercatori e dottorandi e settori di nicchia) dove l'interesse del mercato commerciale è scarso.

### 3. TAPPE FONDAMENTALI E CARATTERISTICHE DELLA FUP

Firenze University Press nasce quindi nel febbraio 2000, quando il Consiglio di Amministrazione dell'Università di Firenze approva la proposta di progetto di editoria elettronica nell'ambito del Sistema Bibliotecario di Ateneo. Nell'aprile 2000 viene costituito il Comitato editoriale composto da docenti e tecnici del settore al fine di individuare gli obiettivi politico-strategici di questa nuova realtà. Nel dicembre 2001 si arriva alla stesura della Relazione finale, dalla quale si delineano i risultati del progetto, che volge ormai al termine, per lasciare spazio ad una vera e propria struttura editoriale creata grazie all'approvazione del Regolamento del Servizio Editoriale del Sistema Bibliotecario dell'Ateneo fiorentino, da parte del Senato Accademico nel dicembre 2002 e del Consiglio di Amministrazione nel marzo 2003, per giungere infine alla nomina del Consiglio Scientifico della Firenze University Press nel giugno 2003.

Firenze University Press è oggi un servizio editoriale indipendente gestito dall'Università di Firenze con proprio personale. A differenza di altre iniziative, nate successivamente in alcune Università italiane, quali le Edizioni Plus di Pisa e la Bononia University Press, essa non ha un editore alle spalle che partecipa al capitale e al lavoro.

FUP è guidata da un Consiglio Scientifico composto da un rappresentante per ciascuna Area Scientifica scelto tra i docenti dell'Ateneo, da un Presidente nominato dal Rettore fra i membri del Consiglio Scientifico e da un Direttore nominato dal Direttore Amministrativo fra il personale del Servizio Bibliotecario.

Il consiglio scientifico è responsabile della qualità delle pubblicazioni della FUP e gestisce il fondamentale momento del "referaggio" dei lavori che successivamente vengono inseriti in produzione seguendo il piano di sviluppo editoriale, valuta le proposte di pubblicazione pervenute alla FUP, ne delibera la realizzazione sulla base del piano editoriale e individua gli esperti per il referaggio di quelle opere che ne sono prive.

### 4. LE TIPOLOGIE EDITORIALI DELLA FUP

Le tipologie di pubblicazioni della Firenze University Press si possono ascrivere a monografie, collane, atti di convegno, rapporti tecnici e riviste.

La FUP ha delineato in particolare la propria produzione editoriale monografica, con una serie di collane che esprimono nella loro interezza l'articolazione dell'informazione scientifica dell'Ateneo fiorentino, dando rilievo ai diversi settori disciplinari.

La chiara definizione delle finalità e delle caratteristiche di ogni collana editoriale, offre agli autori e ai dipartimenti la possibilità di valutare in maniera trasparente le strategie editoriali che sostengono il piano editoriale varato annualmente.

La chiara definizione dell'articolazione della produzione editoriale della FUP intende anche aprire una comunicazione efficace con gli autori, al fine di facilitare la collaborazione con gli autori stessi, caratterizzando in maniera precisa ogni prodotto; dichiarandone il *layout*, le condizioni di pubblicazione, il sistema di "referaggio" e le modalità di recupero dei costi, ed offrendo quindi un supporto agli autori nella produzione di documenti già strutturati in modo da accelerare i tempi di pubblicazione e diminuirne i costi di produzione.

Tra le collane già varate per il piano editoriale 2004, ricordiamo la collana delle *Tesi di dottorato*, la collana *Quaderni di studi e ricerche* dove si intendono convogliare i risultati della ricerca condotta all'interno del nostro Ateneo, la collana *Quaderni per la didattica* nella quale inserire le pubblicazioni che sono il risultato dell'elaborazione di materiale usato per la didattica, durante corsi e/o seminari tenuti presso l'Università degli Studi di Firenze. Creando la collana *Quaderni per la didattica* si è pensato ad una evoluzione editoriale di ciò che spesso viene indicato come dispensa. Le pubblicazioni edite come *Quaderni*, maneggevoli anche nella forma, sono uno strumento volto ad agevolare i docenti e gli studenti nelle reciproche fasi di insegnamento e di apprendimento e possono esser pubblicate sia in versione cartacea che elettronica.

Le collane di dipartimento occupano un posto particolare nel piano editoriale di FUP, infatti esse possono considerarsi la tipica espressione dell'interazione fra la FUP e i dipartimenti, intende infatti dar voce alle pubblicazioni dei dipartimenti garantendo il recepimento delle diverse realtà nel rispetto del *marchio* di dipartimento.

La collaborazione si realizza attraverso la condivisione del lavoro, il responsabile della collana garantisce che l'autore fornisca un impaginato in *word* di buona qualità nel rispetto dei *templates* consegnati, bibliografia e impianto note consistenti. Il frontespizio riporta nell'intestazione il nome della collana e del dipartimento di appartenenza. Il "referaggio" è delegato ai comitati scientifici delle collane all'interno del dipartimento stesso.

Il marchio dell'Università di Firenze è sicuramente un richiamo di qualità per la vendita diretta dell'edizione sia cartacea sia elettronica.

Le riviste caratterizzeranno il piano di sviluppo 2004 con un progetto mirato che intende raccogliere adesioni da parte delle riviste dell'Ateneo fiorentino offrendo alle redazioni servizi aggiuntivi quali un sito web che ospiti la versione elettronica con funzionalità di ricerca nel *full-text* ma anche per autore, titolo e parole chiave ed inoltre possibilità di gestire *mailing list* e *newsletter* dedicate alla singola rivista.

## 5. LE PUBBLICAZIONI DELLA FUP

La Firenze University Press pubblica le sue opere in edizione digitale con stampa in *print on demand*.

Ogni opera ha una edizione elettronica e cartacea a richiesta. Questa duplice edizione delle pubblicazioni FUP ha rassicurato molto gli autori, specialmente durante i primi momenti di vita di FUP. Il formato cartaceo è sicuramente la panacea verso l'instabilità dell'ambiente elettronico.

Tuttavia la Firenze University Press nasce come servizio di editoria elettronica e ha fatto del deposito legale volontario presso la Biblioteca Nazionale Centrale di Firenze (BNCF) il proprio vessillo di promozione per l'uso del formato elettronico.

L'edizione digitale viene certificata e garantita attraverso il deposito legale alla BNCF attivato a seguito di un accordo tra la Biblioteca e l'Università di Firenze<sup>6</sup>. L'accordo prevede che la Biblioteca Nazionale Centrale di Firenze garantisca la certificazione dell'autenticità delle pubblicazioni elettroniche della Firenze University Press e la loro fissità, ossia correttezza del contenuto scientifico e permanenza nel tempo.

Tutto questo avviene tramite un meccanismo di harvesting che la BNCF attua scaricando direttamente dal Web le risorse pubblicate da FUP, archiviandole a lungo termine ed attribuendo loro un identificatore univoco calcolato secondo un algoritmo (md5) che garantisce l'inalterabilità della risorsa elettronica depositata.

Le potenzialità del formato elettronico consistono tuttavia nella maneggevolezza, nella facilità di uso da parte dell'utente finale, offerta dalla maggiore navigabilità e ricercabilità del testo.

In edizione elettronica sono state già pubblicate opere con struttura ipertestuale che offrono al lettore elementi di valore aggiunto rispetto all'edizione cartacea, con funzionalità di ricerca avanzate nel testo.

Nelle pubblicazioni in versione digitale si possono ricercare le occorrenze e frequenze dei lemmi grazie a un sistema di ricerca e di gestione dei documenti full-text.

L'utente dopo questa ricerca può decidere se acquisire l'opera completa o solo una parte di essa e scaricarla sul proprio *personal computer* per l'archiviazione nella propria biblioteca digitale o per la stampa.

All'interno del catalogo, ogni opera viene descritta secondo standard internazionali per la gestione delle pubblicazioni elettroniche che ne garantiscono interoperabilità e accessibilità a lungo termine.

Alla data attuale sono già state pubblicate 100 opere (di cui 5 collane di dipartimento e 7 riviste) tutte disponibili dal catalogo della FUP <<http://epress.unifi.it/index.html>>.

La maggior parte delle pubblicazioni riguardano soprattutto il settore umanistico e quello delle scienze sociali, ma anche un buon numero di atti di convegni e rapporti di ricerca relativi al settore tecnologico-scientifico.

## 6. I FORMATI ELETTRONICI

Le pubblicazioni di FUP sono edite sempre in formato elettronico al quale si abbina a richiesta dell'autore, il formato cartaceo.

L'esigenza di sposare le potenzialità di navigazione e ricerca tipiche del formato elettronico, con il bisogno di avere un'edizione cartacea, ha fatto sì che si prediligesse l'uso di file pdf per la messa in rete della pubblicazione e per l'invio al tipografo.

Benché file proprietario e quindi deprecabile<sup>7</sup> nell'uso, il file pdf offre delle potenzialità interessanti per il controllo dell'accesso ai documenti. Usando infatti Acrobat 5 i file pdf hanno visto potenziate le possibilità di navigazione nel file stesso. Inoltre

l'uso di Acrobat e-book, consente all'utente finale di scaricare i file sul proprio computer, salvarlo nel proprio archivio digitale, ma senza asportarlo e modificarlo, con somma gioia di alcuni autori.

Attualmente i testi sono pubblicati per la quali totalità in formato pdf, questo formato gestito all'interno di un *server* Adobe consente di poter dare ai nostri autori garanzia di non modificabilità della loro opera insieme ad una gestione agevole dei diritti digitali (autorizzazione alla lettura del documento solo a video senza possibilità di scarico né stampa; autorizzazione solo per scarico del documento o di parte di esso; scarico con o senza stampa; sola stampa dell'edizione elettronica).

### 6.1 XML: un miraggio?

Al fine di abbandonare l'uso quasi esclusivo di formati proprietari come il pdf e con l'obiettivo di massimizzare la visibilità delle pubblicazioni FUP migliorandone anche l'archiviazione a lungo termine, è iniziato uno studio di fattibilità sull'utilizzo del formato XML per le pubblicazioni FUP, in particolar modo per le riviste.

Sono state scelte le riviste come risorse informative sulle quali testare per prime l'uso di XML al fine di rispondere alle esigenze di maggiore visibilità presso le banche dati internazionali. Da un'indagine condotta presso alcune di queste, è scaturita l'esplicita richiesta di fornire alla banche dati e agli aggregatori internazionali, file in formato SGML oppure XML. Dato che questo per la FUP alla data attuale non è possibile, il Consiglio Scientifico ha deliberato un incremento dell'investimento sul settore riviste della FUP per migrare il ciclo produttivo verso l'uso di XML.

Questo sarà un processo laborioso e abbastanza lungo, ma sarà sicuramente un ottimo investimento dato che XML si sta sempre più affermando come standard per l'archiviazione a lungo termine ed offre ampie possibilità di gestione dei documenti essendo un formato aperto e piattaforma indipendente.

L'uso di XML permetterà grazie ad opportuni fogli di stile, una maggiore standardizzazione dei *layout* dei prodotti editoriali FUP e agevolerà la gestione stessa dei contenuti. Inoltre associato ad opportuni metadati (es. Dublin Core<sup>8</sup>) potrà facilitare l'archiviazione ed estrazione degli stessi, nonché una maggiore reperibilità dei documenti sul *Web*.

Attualmente è in corso la definizione di una comune DTD per le riviste che potranno avere un *layout* finale sia in pdf sia in HTML, per andare a rispondere alle esigenze di alcune delle redazioni che si sono rivolte a FUP e che ancora prediligono l'uso di HTML per i loro prodotti, spesso distribuiti gratuitamente in rete. Al fine di individuare una DTD comune si sono analizzate tutte le riviste attualmente edita da FUP, evidenziando le loro componenti strutturali a livello sia di testata di periodico che di singolo articolo parallelamente all'analisi di alcune DTD standard già esistenti, quali TEILITE, DocBook e XML Special.

La struttura generale dell'informazione contenuta nelle riviste potrebbe essere così riassunta:



- definizione e codifica metainformativa sia a livello di singolo articolo con metadati Dublin Core, sia a livello di testata di rivista;
- definizione e codifica gerarchico-strutturale sia a livello articolo sia a livello fascicolo (più file articoli formano un unico file fascicolo);
- elaborazione di fogli di stile per il formato HTML e per il pdf.

L'uso di XML permetterà inoltre una maggiore reperibilità in rete delle risorse ad accesso libero offrendo un accesso di qualità anche senza l'intermediazione delle banche dati, qualora siano associati alle pubblicazioni gli opportuni metadati standard.

## 7. DIFFUSIONE NON SOLO ELETTRONICA

Dalle esperienze di questi primi due anni di sperimentazione, peraltro supportata anche dalla letteratura di riferimento, è risultato che il contesto di diffusione del formato digitale non ha avuto la velocità di sviluppo ipotizzata sia per quanto riguarda il commercio elettronico che gli strumenti di lettura a video.

Inoltre da più indagini condotte<sup>9</sup> sugli utenti accademici nella loro veste di autori e lettori, è emerso che la versione elettronica non ha ancora un diffuso riconoscimento di autorevolezza culturale.

Pertanto al fine di garantire ai nostri autori i meccanismi di diffusione riconosciuti dalla comunità accademica, la FUP ha deciso di associare alla diffusione elettronica anche iniziative di distribuzione tradizionale sul territorio, in particolare per una selezione di titoli riguardanti monografie scientifiche e didattiche. Grazie alle nuove tecnologie di stampa digitale possiamo infatti pubblicare la versione elettronica e stampare un bassa tiratura della stessa, direttamente proporzionata alle richieste degli utenti finali.

## 8. L'ARCHIVIO DI 'E-PRINTS'

La Firenze University Press gestisce, in collaborazione con lo CSIAF, l'archivio istituzionale dei documenti elettronici dell'Università degli Studi di Firenze che contiene i lavori per la didattica e la ricerca prodotti da docenti e ricercatori dell'Ateneo. Gli autori, previa registrazione, depositano direttamente i propri documenti nell'archivio. L'accesso alle informazioni ed ai documenti è libero. L'archivio offre garanzie di stabilità e chiara affiliazione dell'informazione all'Università di Firenze e grazie al protocollo utilizzato (Open Archive Initiative Protocol for Metadata Harvesting) permette lo scambio e l'interoperabilità delle informazioni con i maggiori centri di ricerca a livello mondiale; quello di *Los Alamos* per i fisici e i matematici, *Repec* per gli economisti.

## 9. DIVERSIFICAZIONE DELL'ACCESSO

La creazione di una rete di trasmissione e di accesso agevolato ai documenti, è un messaggio di apertura verso modi alternativi di diffusione da associare a quella commerciale legata anche alla sostenibilità economica della struttura.

L'accesso libero (*open access*) ai documenti, quale quello dell'archivio *e-prints*, può incrementare in maniera sostanziale la fruizione e la diffusione dei risultati della ricerca e della didattica prodotti all'interno dell'Ateneo. L'adozione o sviluppo di servizi di *linking* permetterà il colloquio fra ambienti diversi e soprattutto risorse diverse agevolando la loro fruizione da parte dell'utente.

## 10. CONCLUSIONI

La sfida dell'editoria digitale accademica presenta molteplici contraddizioni ed aspetti ancora da analizzare. Tuttavia l'interesse del mondo accademico per le potenzialità dell'ambiente digitale è ormai evidente. Le potenzialità dei formati elettronici devono ancora essere mediate con la realtà dell'editoria cartacea che offre ancora stabilità e sicurezza. Tuttavia l'editoria elettronica grazie alle garanzie offerte dal deposito legale sarà sempre più presente in ambito accademico.

Inoltre l'adozione di formati aperti, non proprietari, quali anche XML, è sempre più importante in un contesto digitale che si evolve velocemente.

I formati aperti usati nella rappresentazione dell'informazione, uniti alle evoluzioni tecnologiche consentono oggi di realizzare un punto di accesso integrato fra ambienti di diversa natura, es. commerciale e gratuito, includenti diversa tipologia documentaria, es. documenti *preprint*, *working paper*, pubblicazioni monografiche, periodici. Tutto ciò sta delineando per l'utente la possibilità di accedere ad un sistema informativo nel quale sono abolite le barriere e dove potrà disegnare la propria biblioteca digitale senza difficoltà di dialogo e lettura poiché sistemi e documenti dialogheranno tra di loro.

## 11. RIFERIMENTI BIBLIOGRAFICI

- Arms W. Y. (2002). *Quality Control in Scholarly Publishing on the Web*. Journal of Electronic Publishing, 8, 1.  
<http://www.press.umich.edu/jep/08-01/arms.html>
- Bergstrom T. C. (2001). Free Labor for Costly Journals. *Journal of Economic Perspectives*, 15, 4, pp. 183-198.  
<http://www.econ.ucsb.edu/~tedb/Journals/jeprevised.pdf>
- Brian K., What are Xlink and Xpointer?, *Ariadne*, 16, <http://www.ariadne.ac.uk/issue16/what-is/intro.html>
- Brian K. What is ... xml?, *Ariadne*, 15.  
<http://www.ariadne.ac.uk/issue15/what-is/intro.html>
- Case M. M. (2001) Scholarly Communication. A system in crisis. <http://www.lib.ohio-state.edu/Staff/scholcom/case925>

- Cotoneschi P. (2001). *Electronic Publishing and Scholarly Communication: the potential of an e-print Archive for the Scientific Community of Florence University*, Newcastle, University of Northumbria.
- Crow R. (2002). *The Case for Institutional Repositories: A SPARC Position Paper*, SPARC. <<http://www.arl.org/sparc/IR/ir.html>
- Fisher J. H. (1999) Comparing Electronic Journals to Print Journals: Are there Savings? In: *Scholarly Communication and Technology Conference* Organized by The Andrew W. Mellon Foundation, Emory University, USA, <http://www.arl.org/scomm/scat/fisher.html>
- Fytton R. (1997). Print journals: fit for the future?, *Ariadne*, 7, 1997, <<http://www.ariadne.ac.uk/issue7/fytton/intro.html>
- Hai K., Palmer D., Dunlop C., A. IMS generator for the masses. *Ariadne*, 36, 2003. <http://www.ariadne.ac.uk/issue36/martini/intro.html>
- Harnad S. (2001). For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now. Paper presented at the Workshop on the Open Archive Initiative and peer review journals in Europe. Geneva. <http://www.cogsci.soton.ac.uk/~harnad/TP/resolution>
- Harter S. P. (1997). Scholarly communication and the digital library: Problems and issues, *JoDi*, 1 (1) <http://jodi.ecs.soton.ac.uk/Articles/v01/i01/Harter/>
- Lagoze C., Van de Sompel H. (2001). The Open Archives Initiative: Building a low-barrier interoperability network. Joint Conference on Digital Libraries. <http://www.openarchives.org/documents/oai>
- Lawrence S. (2001). Online or invisible?. *Nature*, 411, 6837 <<http://www.neci.nec.com/~lawrence/papers/online-nature01/>>
- Pelikan J. (1992) *The idea of the University*, a Re-examination. New Haven, CT: Yale University Press. Prickett, S. and Erskine-Hill, P. (eds) (2002)
- Peters S.. EPRESS scaling up electronic journal production. *Ariadne*, 23, <http://www.ariadne.ac.uk/issue23/epress/intro.html>
- Pinfield S. Managing electronic library services: current issues. In UK higher education institutions, *Ariadne*, 29, 2001. <<http://www.ariadne.ac.uk/issue29/pinfield/intro.html>>
- Revelli C. (2000). *Pubblicazioni periodiche, Biblioteche oggi*, 5, p. 44 <http://www.bibliotecheoggi.it/2000/20000504401.pdf>
- Savenije B. (in press) Who pays the ferryman? About new models for scientific communication. *Proceedings of the international symposium 'science & engineering libraries for the 21st century'*. Leuven, Leuven University Press.
- SPARC, Washington. <http://www.arl.org/sparc>
- Tenopir C. King T. (2001). Lessons for the Future of Journals. *Nature Web Debates: e-access*. <http://www.nature.com/nature/debates/e-access/Articles/tenopir.html>

- Tammaro A. M. (1998). Modelli economici per i periodici elettronici. *Biblioteche oggi*, 5, p. 58.
- Walker T. J. Walker. (2002). Author (or sponsor) pays. Presentation at the 18th International Learned Journals Seminar, ALPSP. London. <http://cssrvr.entnem.ufl.edu/~walker/epub/ALPSP.htm>
- Willinsky J. (2002). Copyright Contradictions of Scholarly Publishing. *First Monday*, 7, 11.  
[http://www.firstmonday.dk/issues/issue7\\_11/willinsky/index.html](http://www.firstmonday.dk/issues/issue7_11/willinsky/index.html)
- Willinsky J. (2003). Policymakers' Use of Online Academic Research. *Education Policy Analysis Archives*, 11, 2.  
<http://epaa.asu.edu/epaa/v11n2/>
- Willinsky J. (2003). Scholarly associations and the economic viability of open access publishing, *JoDi*, 4, 2.  
<http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Willinsky/>

## 12. NOTE

- <sup>1</sup> Pelikan J. (1992).
- <sup>2</sup> Sevenjje, Bas, *Who pays the ferryman? About new models for scientific communication*, in press.
- <sup>3</sup> Tenopir C., King T. (2001).
- <sup>4</sup> Willinsky J. (2003); Savenije, Bas, (in press).
- <sup>5</sup> <<http://digital.casalini.it/osservatorio.asp>>.
- <sup>6</sup> Il testo dell'accordo è visibile alla pagina < <http://epress.unifi.it/info.htm>>.
- <sup>7</sup> <<http://www.w3.org/>>.
- <sup>8</sup> <http://dublincore.org>.
- <sup>9</sup> Cotoneschi P. (2001).

---

# Text understanding from discourse model and internal processes

Rodolfo Delmonte

Università di Venezia "Ca Foscari"  
(delmont@unive.it)

*Text Understanding (hence TU) is viewed here as a linguistically-based process which activates Knowledge of the World only when needed, by triggering inferential processes and tapping encyclopaedic information. For a system based on linguistic principles, TU is a process requiring a pipeline of interleaved independent modules which deliver to the Semantic Processing Modules the preferred structural representation according to parsing strategies. These strategies can in turn access semantic information limited though to lexical and dictionary lookup. Semantic interpretation requires an intermediate module for Discourse processing, which takes as input the DAG structure produced by the parser, and the Discourse Model as asserted by the system, with the task of choosing the most relevant entity and make it the object of referential processes. A Centring-like algorithm is used to generate a stack of possible Topics of Discourse and then choose the one which fits best the current sentence. Semantic processes vary according to the referential nature of nominal expressions. We use a children story and other texts taken from newspapers to show how the TU system works.*

*Keywords:* Semantic Interpretation, Text Understanding, Discourse Models

## 1. INTRODUCTION

GETARUNS, the system for text understanding developed at the University of Venice, is equipped with three main modules: a lower module for parsing where sentence strategies are implemented; a middle module for semantic interpretation and discourse model construction which is cast into Situation Semantics; and a higher module where reasoning and generation takes place (Delmonte, 2000). The system work in Italian and English.

The system is based on LFG theoretical framework (Bresnan, 2001) and has a highly interconnected modular structure. It is a top-down depth-first DCG-based parser written in Prolog which uses a strong deterministic policy by means of a look-ahead mechanism with a WFST to help recovery when failure is unavoidable due to strong attachment ambiguity.

It is divided up into a pipeline of sequential but independent modules which realize the subdivision of a parsing scheme as proposed in LFG theory where a c-structure is built before the f-structure can be projected by unification into a DAG. In this

sense we try to apply in a given sequence phrase-structure rules as they are ordered in the grammar: whenever a syntactic constituent is successfully built, it is checked for semantic consistency, both internally for head-spec agreement, and externally, in case of a non-substantial head like a preposition dominating the lower NP constituent. Other important local semantic consistency checks are performed with modifiers like attributive and predicative adjuncts. In case the governing predicate expects obligatory arguments to be lexically realized they will be searched and checked for uniqueness and coherence as LFG grammaticality principles require (Delmonte, 2002). In other words, syntactic and semantic information is accessed and used as soon as possible: in particular, both categorial and subcategorization information attached to predicates in the lexicon is extracted as soon as the main predicate is processed, be it adjective, noun or verb, and is used to subsequently restrict the number of possible structures to be built. Adjuncts are computed by semantic cross compatibility tests on the basis of selectional restrictions of main predicates and adjuncts heads.

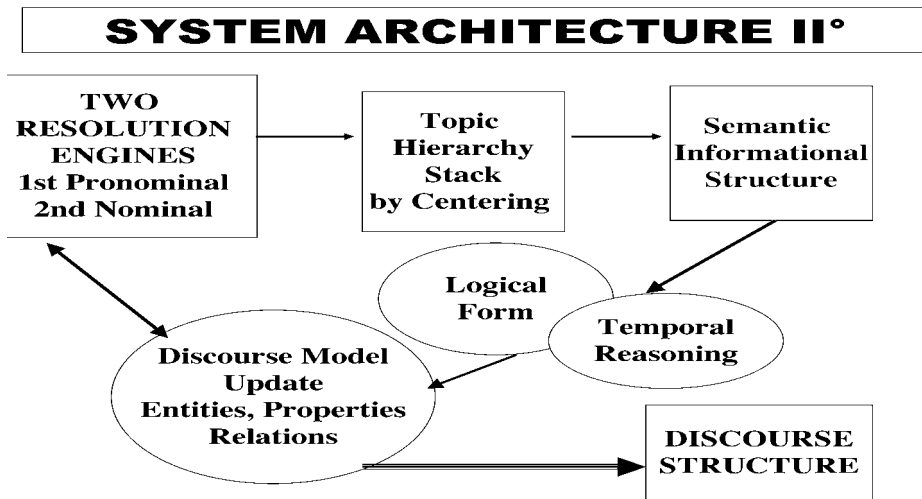
As far as parsing is concerned, we purport the view that the implementation of sound parsing algorithm must go hand in hand with sound grammar construction. Extragrammaticalities can be better coped with within a solid linguistic framework rather than without it. Our parser is a rule-based deterministic parser in the sense that it uses a lookahead and a Well-Formed Substring Table to reduce backtracking. It also implements Finite State Automata in the task of tag disambiguation, and produces multiwords whenever lexical information allows it. In our parser we use a number of parsing strategies and graceful recovery procedures which follow a strictly parameterized approach to their definition and implementation. Recovery procedures are also used to cope with elliptical structures and uncommon orthographic and punctuation patterns. A shallow or partial parser, in the sense of S. Abney (1996), is also implemented and always activated before the complete parse takes place, in order to produce the default baseline output to be used by further computation in case of total failure. In that case partial semantic mapping will take place where no Logical Form is being built and only referring expressions are asserted in the Discourse Model – but see below.

## 2. THE UPPER MODULE

GETARUNS, has a highly sophisticated linguistically based semantic module which is used to build up the Discourse Model. Semantic processing is strongly modularized and distributed amongst a number of different submodules which take care of Spatio-Temporal Reasoning, Discourse Level Anaphora Resolution, and other subsidiary processes like Topic Hierarchy which will impinge on Relevance Scoring when creating semantic individuals. These are then asserted in the Discourse Model (hence the DM), which is then used to solve nominal coreference together with WordNet. The system uses two resolution submodules which work in sequence: they constitute independent modules and allow no backtracking. The first one is fired whenever a

free sentence external pronoun is spotted; the second one takes the results of the first submodule and checks for nominal anaphora. They have access to all data structures contemporarily and pass the resolved pair, anaphor-antecedent to the following modules. Semantic Mapping is performed in two steps: at first a Logical Form is produced which is a structural mapping from DAGs onto of unscoped well-formed formulas. These are then turned into situational semantics informational units, infons which may become facts or sits.

Fig. 1 GETARUNS' Discourse Level Modules



### 3. REFERENCE RESOLUTION

The problem of assigning a suitable antecedent to pronominals and nominals or bridges, can be solved by a number of interleaved procedures which first search in the same utterance, then in the previous utterance's topic hierarchy (or local history list) – and this first process has been dubbed grammatical binding; in case this search fails the full history list will be searched, to look for another explicitly mentioned occurrence of the nominal head to corefer with, or worse of the pronominal head to agree with and in case this fails some inferential mechanism must be activated – and we may dub this semantic binding.

There is a number of possible processes that our system can successfully deal with a number of different procedures which we have classified as follows:

- Indirect Reference
- Implication
- Predicate-Argument Equality Match
- Selectional Restrictions Membership Match
- Meaning Postulate Inferences
- Recency Constraint



- Individual to Set Reference
- Domain Point-of-View Reference
- World Knowledge Inferential Reference

We shall discuss all these phenomena in this chapter in the following sections. However, before going into the description of what the system does, it is important that we point out what mechanisms have not been introduced which are usually adopted in a knowledge representation and artificial intelligence environment.

In particular, we are neither using frames nor plans to reason about the world, all our knowledge is encoded in lexical forms, where inherent semantic features are associated to noun predicates, and selectional restrictions to verb and adjective argument slots if present. In addition, the semantic module takes advantage of semantic roles and sometimes of grammatical functions to carry out its tasks; inherent features are also important, for instance to deal with relational nouns or with collective nouns. So we do not rely on scripts or on expectations generated by the previous context. In fact, we are not interested in building a dialogue system, nor do we actually want to ascertain whether the text conforms to a specific plan structure. As we said, we are only interested in narratives, stories, newspaper articles, etc. which should all make their internal plot, text structure clear to the reader.

The knowledge base that we access, when needed, has been heuristically built around a number of phenomena of inferential processes that call for external knowledge of the world. Conditions on this access are very strict:

- A. no match has been found by grammatical and semantic binding procedures;
- B. the nominal expression must be a definite singular NP.

In particular it is used for the following specific tasks:

- to infer the meaning a nominal head, not mentioned in the previous discourse, as for instance is required by a relational complex noun like “father-in-law”;
- to infer the referential import of attributes associated with a nominal head, and not mentioned in the previous discourse;
- to associate an adequate coreferential entity for metaphorical nominal heads;
- to reason about actions and causality in order to establish the appropriate discourse relation in a given clause.

In practice, we only built knowledge representations for the items that required them, but in theory we could have built similar representations for all linguistic items of the stories we analysed. However, the way in which we built our KB is highly specialized and no unwanted traversal can be triggered.

We could summarize our introduction in the following two schemes:

#### **How Inferential Processes are activated**

1. Grammatical properties of linguistic objects
  - i. Functional Features of NPs

- ii. Selectional Restrictions of NPs
  - iii. Grammatical Functions of NPs
  - iv. Semantic Roles of all Phrases
  - v. Semantic classes of propositions
  - vi. Aspectual and Temporal Representations of propositions
  - vii. Quantifier scope operator-variable pairs
2. Topic Hierarchies for referential expressions
  3. Domains of Point\_of\_View and Subject\_of\_Consciousness
  4. Discourse Model of facts locs and sits
  5. spatiotemporal locations for scenario effects
  6. External World Knowledge Base

### Goals and Aims of Inferential Processes

1. Execute Anaphora Resolution
2. Ascribe properties and roles to entities in the model
3. Determine non-extensional entities from a computation of opaque and intensional propositional domains
4. Determine superset, subset and inclusion relations
5. Determine spatiotemporal locations of entities and relations
6. Build up discourse structure, assign discourse relations and build reference intervals for temporal relations

## 4. SEMANTIC BINDING

The perspective we are working however requires us to cope with these problems from the context. Anaphors is the general term for a range of expressions that are context-dependent in that they either specify entities in an evolving model of the discourse that the listener is constructing, or they depend on other entities in the discourse model. They are called Discourse Anaphors (DA) in Webber's paper on Tense (1988).

The dependency of a DA on a discourse entity may result from the ontology of the specified entity, as well as from discourse structure and its focusing effect.

It has been argued extensively in the literature that definite NPs are exclusively used to corefer or cospecify entities already in the DM, whereas indefinite NPs can be used to introduce new entities in the world. However, we shall stress the need to ensure that both definite and indefinite NPs can be used to corefer or cospecify entities and relations in the DM. Besides, we should also note that both kinds of NPs can be used to introduce new entities in the world(see Viera and Poesio).

We now comment her example 5. where a definite NP is introduced as a new entity associated to an existing discourse entity:

- 12 a. A bus came round the corner.
- 12 b. I signalled to the driver to stop.

The definite NP “the driver” is here understood as an entity associated with the “bus” mentioned in the previous sentence. In our system this is achieved by an inference on the main location of the text. The driver is related to the bus, since there is the setting of a scenario by mentioning the location in which the driver could be inferred as being a part of, or better a role linked to that frame. Better examples of this problem are constituted by the texts provided in a paper by Garrod and Sanford (1988) – now *At the Restaurant Text* – where the scenario effect is more marked:

- 13 a. John went into a restaurant.
- 13 b. There was a table in the corner.
- 13 c. The waiter took the order.
- 13 d. The atmosphere was warm and pleasant.

The scenario is set at the beginning of the text, either by a title “At the restaurant”, or by explicitly mentioning the restaurant as an Oblique argument of a going relation where the main Topic is also introduced. When we get to the definite NP “the waiter”, an inclusion relation inference has already been fired by the location “the corner” which is understood as being a part of the main location, the restaurant. Thus, “the waiter” can now be computed as being in an inclusion relation with “the restaurant”, being a role pertaining to that scenario or frame. Another important side effect of the analysis is that the Main Topic, John, is not discarded in favour of the new entity “the waiter”, but persists in the Discourse Module. Thus, in case the text continues with what the authors call “psychological atmosphere statements”, we may impute the subjective judgment on the Main Topic, John, rather than on a possible Expected Topic, the waiter.

More complex examples can be derived from our texts, in which more than one character is present. In particular, the story of the three little pigs has the most intricate plot we found, since we are given at the same time six possible entities to be used for further processing. This is the beginning of an abridged version:

Segment A.

“Once upon a time there were three little pigs who lived happily in the countryside. But in the same place lived a wicked wolf who fed precisely on plump and tender pigs. The little pigs therefore decided to build a small house each, to protect themselves from the wolf. The oldest one, Jimmy who was wise, worked hard and built his house with solid bricks and cement. The other two, Timmy and Tommy, who were lazy settled the matter hastily and built their houses with straw and pieces of wood. The lazy pigs spent their days playing and singing a song that said, “Who is afraid of the big bad wolf?”.

And these are the entities already present at this point of the computation:

- i. a set of three little pigs
- ii. a set of three little houses

- iii. an individual, member of the set of little pigs, whose name is Jimmy
- iv. a subset made of two little pigs, always members of the set of three little pigs, whose names are Timmy and Tommy
- v. a single little house, member of the set of three little houses, made of bricks and cement, owned by Jimmy
- vi. a subset of two little houses, included in the set of little houses, made of straw and little pieces of wood, owned by the subset of two little pigs, named Timmy and Tommy

As the text proceeds, coreference is activated by evoking one of the entities either by a property, the fact of being brothers or being pigs, or by explicitly indicating cardinality.

#### Segment B.

“And one day, lo and behold, the wolf appeared suddenly behind their backs. “Help! Help!”, shouted the pigs and started running as fast as they could to escape the terrible wolf. He was already licking his lips thinking of such an inviting and tasty meal. The little pigs eventually managed to reach their small house and shut themselves in, barring the door”.

Consider the deictic singular NP “such an inviting and tasty meal”, which requires knowledge of the world to compute the metaphor. We should also note that the sentence is a case of idiomatic expression computed by the parser at grammatical level. Reference to houses in the following text, is expressed in the form of a singular little house associated to the two lazy pigs, rather than a plural noun. We should still be talking about a set of two houses, and this is suggested by the possessor’s cardinality. However the following text, reverts this decision, in that it makes us understand that the two little pigs are now both in the same house.

#### Segment C.

“In the meantime the wolf was thinking a way of getting into the house. He began to observe the house very carefully and noticed it was not very solid. He huffed and puffed a couple of times and the house fell down completely. Frightened out of their wits, the two little pigs ran at breakneck speed towards their brother’s house”.

Problems at this point arise in the reasoning mechanism since the only singular little house available is the one built by the wise brother, Jimmy, which could be picked up for coreference. This is clearly wrong, because this is not the house that gets destroyed since it is still there in the last sentence of Segment D. In order to prevent a failure in the reasoning process, we simply allow reference to a little house to be inferable as belonging to the set of two little pigs under discussion, on the assumption

that it is the closest one and is available for coreference in the previous portion of text. The inferential mechanism is always driven by a recency checker which estimates which is the closest topic being asserted: on that basis an inference is fired, and in case it is successful that topic is taken as being coreferent.

Another interesting case is exemplified by the second version of the story of the three little pigs, where we see that a definite NP is used to introduce one of the main characters in the story. Differently from what happens in the first story, where the big bad wolf is presented in the world at the very start together with the little pigs, in the second story the wolf only appears after the little pigs have built their houses. The story refers to the wolf as an intensional entity whose presence in the woods nearby is taken for granted:

- 14 a. Jimmy wanted a robust house because he knew that the bad wolf lived in the woods nearby.

This reference to the wolf is understood by the system as a sit related to an individual, and an appropriate identifier is assigned to it. In the following text, when the wolf reappears on the scene, it is a definite NP again and a fact is asserted. Coreference is enabled thus by virtue of the fact that an intensional entity was available in the model:

- 14 b. But as soon as Timmy opened the door, the bad wolf appeared out of the woods

Another interesting case is represented by Webber (1988, her example 4.) where a plural pronoun is made to corefer to a class related to a singular NP in the previous discourse by inferential processes,

- 16 a. The dachshund down the block bit me yesterday.  
16 b. They are really vicious beasts.

The pronoun is computed as the controller of the property “beasts” which is in the open complement of the verb “be”; this in turn should be taken as a generic property associated to the definite NP “the dachshund” which the sentence evokes. Clearly this case requires the presence of an ISA Relation in the ontology of the entities, constituting the extra-linguistic knowledge required to understand the relations intervening in the text at hand.

#### 4.1 *Indefinite NPs*

Consider now indefinite NPs: they may cospecify some previously asserted relation or they may introduce some generic property which is already inferable from the model:

- 17 a. John gave Mary a rose.  
17 b. She took it and put it in her hair.  
17 c. She knew that she had been given a present, something precious.

While in a. the indefinite NP is actually a new entity in the DM, in c. “a present” cospecifies the previous NP “a rose” and must be computed as a class entity in which “a rose” is included.

18 a. Mary picked up the phone and called Jason.

18 b. Her husband would have considered such a move as base.

The NP “a move” is an event noun coreferring to a discourse segment as Webber (1990) suggests discussing the role of deictics in discourse; in our system this is achieved at discourse structure where discourse segments are produced and can be recovered from the model.

19 a. They appointed John managing director.

19 b. This was a position he had been longing for for ages.

The NP “a position” is an activity noun cospecifying the role property “managing director” asserted in the previous discourse segment and associated to John.

20 a. Richard would marry Sara.

20 b. He felt strongly about all this.

20 c. She was the right person for a man like Richard.

The NP “a man” is a generic noun which corefers with Richard, while the deictic is again coreferring with the previously built discourse segment.

21 a. Once upon a time there were three little pigs.

21 b. They decided to build a little house each.

The NP “a little house” is asserted as a sit and not a fact and a cardinality is assigned to this set due to its being in the scope of the distributive quantifier “each”. In turn, the floating quantifier is computed as an open adjunct controlled by the subject “They” which corefers with the antecedent NP “three little pigs” in the previous segment of discourse. In another version of the same story, the little house become a straw hut and then a little house again, as shown in the following excerpt:

21 c. As they reached a nice wood, they decided to build each a comfortable little house.

21 d. Timmy didn't like working at all, so he thought to build quickly a straw hut.

21 e. Soon the little house was ready.

The NP “a straw hut” is understood to be one of the little houses already introduced in the previous discourse segments. However, knowledge of the world is called for in order to make the appropriate inference. Also notice that the following sentence uses a definite NP to corefer with the previous indefinite NP, and here again the same problem arises: the hut becomes a little house. For sure, we want the DM to be consistent and to understand that there is only one little house under discussion, which however is a kind of house, a hut.

Finally the most interesting case: an indefinite NP which is understood as a generic property of an entity introduced in the model by the same relation.

22 a. John gave Mary a rose.

22 b. She took it and put it in her hair.

22 c. She knew she had been given a present, something precious.

The NP “a present” cospecify the NP “a rose” which has been previously introduced in the model by a giving relation. The dependency is contextually determined by the presence of an entity Mary which is assigned the same semantic role in both sentences: she is always the Patient argument of the predicate GIVE. The system can also recover the Agent argument which has been omitted and is represented in f-structure as a lexically-bound existential quantifier “exist”. At the end of the computations, we know that “a rose” is also “a present” which has been given to Mary by John. In order for the inferential mechanism to draw the relevant inference, tense is used: the past perfect can be regarded as a presuppositional tense, i.e. a tense that indicates that some fact or event took place in the previous portion of text. This can be assumed also by the presence of a presuppositional verb “know” that governs the sentential complement clause in which the giving relation is used.

The second story refers to Jimmy’s house as an intensional entity which is however not yet fully in existence:

- Jimmy wanted a sturdy house because he knew that the bad wolf lived in the woods nearby

The system introduces in this case a set of infons which are computed as sits and are then ascribed to the perspective set up by Jimmy.

#### *4.2 Possession Relations*

The presence of a possessor is a property which adds some specificity to the head noun. In particular, it could enable the system to identify a certain object in the world, by the fact of its belonging to a certain possessor. In the text we analysed, the possessor is used frequently to tell different object included in a set apart, on the basis of the identity of the possessor. In particular, if we consider Segment D. again, where the text introduces a plural reference to house with the possessive pronoun “their”, there are two possible inferences:

- a. either the text refers collectively to a set of houses with two members;
- b. or they are referred to disjointly and there is no set, contrary to the linguistic form used in the text, “their houses”.

In fact, if we keep ourselves to the distributive reading induced by the presence of “each” in the previous portion of text, we are led to the conclusion that there should be three separate entities of the class “house” in the world.

However this is not actually what the text does, since I take the subsequent reference to a single house for Jimmy and a plural set for Timmy and Tommy to imply that we should consider the initial set as made up of two subsets. Since the two lazy pigs are always treated as a plural entity which is a set, also their possessions – their little houses – are so treated in the text. Thus, we know that each little pig has built its own little house and that there are three individual entities in the world, but for coherence and cohesion principles coreference is activated only in relation to a subset of two such entities. This is the reason of the transformation of a set with two members into a single entity, which can be used to corefer to the plural NP “their houses” in Segment B. with the NP “their small house” in Segment C. without any further specification. In turn, this latter NP is coreferred in Segment D. simply as “the house” being the more prominent house locally available or last mentioned due to recency effects.

In the same segment we note then that in order to distinguish the reference to this now singular “house” from the other little house present in the model, the latter is evoked with a possessor’s specification, “their brother’s house”.

To recover the identifier of that entity we let the system infer a property belonging to the class “brother” and search for a fully specified individual – someone who has been given a name – and possesses a house which has been introduced as a single entity. In case every little house were introduced both as set and as single individual there would be no straightforward way to draw the necessary inference. Or at least the inferential process required to recover that individual little house would have been by far more complicated.

In particular, one would have been obliged to compute first the individual entities who possess the property of being “brother”; then, by subtraction, calculate from the possessors’ identity – or their names – the single brother the genitive is now indicating. At this point the identity of the little house could be established.

We might regard the procedure by which we reached the conclusion to generalise the description of a set to the description of a single entity as a linguistically driven sloppy reading. There are empirical reasons that drove our decision: if we look at the longest version of the story of the three little pigs, we discover that it justifies our stance. This is the relevant portion of text:

Segment E.

“This is the story of three little pigs who went around the world seeking their fortune. Their names were Timmy, flute player, Tommy, violinist, Jimmy, great worker. As they reached a nice wood, they decided to build each a comfortable little house. Timmy didn’t like working at all, so he thought to build quickly a straw hut. Soon the little house was ready and Timmy decided then to go and see what his little brothers were doing. At first he met Tommy the violinist. Also he, himself, did not have much wish to toil, so he was building a simple little house with sticks of wood”.



Segment E.

“Very soon, also the house of wood was ready. Like that of straw, it was not very resistant. But the two little pigs lazy had managed to finish their work in a short time and now they could enjoy themselves freely. While Timmy was playing the flute, Tommy accompanied him with his violin and together they were having a lot of fun.”

Segment F.

“Then, tired to make merry, they decided to go and see what their little brother was doing. They started walking and soon they reached Jimmy. The clever little pig was building a little house. But since Jimmy was farsighted and did not fear working hard, he built it with bricks and cement.”

As we can see, the story starts by introducing the set of little pigs; then it assigns them names, and establishes thus their individuality. Also the little houses are introduced as sets, again by the use of a distributive quantifier. As the story continues, we see that each little pig builds its own little house separately. The identity of each little house is now preserved only by its specific property: in particular in Segment F. we see that in order to cospecify each of the two little houses previously introduced in text as single entities, the linguistic form used is “the house of wood”, even though this was the house now under discussion. The following sentence, uses a pronoun “it” to corefer to the most recently mentioned little house and a deictic with a property, “that of straw” to corefer to the other singular little house already existing in the model. The specification of some distinctive property is now required simply because in the model there are two entities of the same class “house”, both singular in number, i.e. existing as individuals, which are being further specified as “not very resistant”. This property, in force of the linguistic form used in the text, is thus distributed over the two single entities.

As the story proceeds, in Segment G. we see that coreference with the little brother (Jimmy) is achieved as in the previous version of the story, by the use of a possessive pronoun. In this case, we might deem that the identity of the individual brother is recovered by subtraction: first, the identity of the possessor is recovered, i.e. an inference is required from the property “brother” to that of their names; then, a singular brother is searched for, whose name does not match with the ones already assigned to the possessor. However, we might also consider the possibility that the use of “brother” here is simply due to the need to assert this property and to extend it to the set of little pigs. In fact, the following sentence makes it clear to the reader that the topic is now set to Jimmy.

An important thing to notice, at this point, is the fact that the introduction of another single entity belonging to the class “house” requires the use of an indefinite NP: another possibility could have been the use of a possessive, “his little house”. In both cases, the system understands that it should add a new individual of the class “house” to the model, since Jimmy does not yet own a little house – the first

mention to a set of three houses is computed as an intension, or a sit, and not as a fact. However, we take the use of an indefinite NP a much simpler way to achieve the same goal. In this case, the factivity of the proposition in which the NP is used, requires the indefinite NP to be regarded as a new entity, which however is included in the set of the three little houses mentioned at the beginning of the story.

### 2.3 Proper Names

Differently from what happens with definite or indefinite NPs, notably the fact that both types can be used either extensionally or intensionally to denote some entity, in the case of proper names we know that we are always dealing with rigid designator of the same individual in all possible worlds, as Kripke defined them. A proper name fixes the reference to an individual in that it designates unambiguously that individual for any further reference in the text.

However, the mechanism by which a proper name is used in a text is something that deserves further scrutiny. In our texts, there are at least two ways to associate proper names with individual entities. The first and more canonical method is the one that is represented by a copulative construction, as in,

Segment G.

“This is the story of three little pigs who went around the world seeking their fortune. Their names were Timmy, flute player, Tommy, violinist, Jimmy, great worker”.

Another method, is the one represented by a predicative adjunct, as in,

Segment H.

“The oldest one, Jimmy, who was wise, worked hard and built his house with solid bricks and cement. The other two, Timmy and Tommy, who were lazy, settled the matter hastily and built their houses with straw and pieces of wood”.

These might be considered as more or less direct methods for name association with a property already existing in the world. A more subtle and indirect way of obtaining the same result is shown in the following example,

Segment I.

“The three friends went all outside. As they were walking in the garden, John said to himself “Sara will marry that man”, without any resentment. Richard would marry her”.

In this latter example, the association is implicitly achieved by the semantic import of the structural organisation of the utterance. A pronoun is used in the subordinate clause to corefer to the property “friends” in the previous text; then, the subordinator

indicates coincidence of temporal relation between the main and the subordinate clause. The main clause, in turn makes the pronoun explicit and introduces proper names as prominent characters. The inference we are naturally drawing at this point is that the Subject pronoun “they” and the Subject of the main clause point at the same individual. However since the pronoun corefer with a set with cardinality 3, we are allowed to make further inferences: we assume that also “Sara” is the name of one of the entity denoted by the set of friends. The same applies to the appearance of “Richard” in the following sentence.

How are these assumption and inferences caused: I take the association rule for proper names to be formulated as follows:

#### **Rule for Proper Names Association**

1. Whenever a Proper Name appears in the world, check in the model if there is already some entity associated with that name;
2. In case the search fails, check whether there is an explicit (direct/indirect) association link with some entity in the current clause;
3. Else, check whether there is an implicit indirect association link with some entity in the local context.

In other words, we always require Proper Names to be associated with some previously asserted property in the local context. However, narratives show that a proper name could be simply introduced as such, and be a new individual in the world, as for instance in,

23. Mary picked up the phone and called Jason.

or as in,

24. John went into a restaurant.

In these last two example, proper names are used to introduce some new entity in the world and the property assumed is simply a generic class specification derived by default properties rules from selectional restrictions associated to that NP as argument of a given predicate.

Names are used freely in the following text to recover coreference to a given individual. In particular, they may be used in place of a pronoun, when the text would make its use ambiguous. In the first version of the story of the little pigs, names are introduced in the story, however no other mention is needed in the following text to recover the corresponding individuals. This is simply due to the fact that they are well distinguished as being either a set of three little pigs, a set of two little pigs, and a single individual: thus, the use of names becomes redundant.

However, the second version of the same story introduces the three little pigs at first as a set with cardinality three; but as the story progresses, each of the three little pigs is introduced separately by its name. This is possible because priorly there

has been an explicit association of names to each member of the set of three. At a certain point of the story, it would seem that talking of a single little pig induces ambiguity, however this is not so, as shown by Segment K, which we report here below,

Segment J.

“Then, tired to make merry, they decided to go and see what their little brother was doing. They started walking and soon they reached Jimmy. The clever little pig was building a little house. But since Jimmy was farsighted and did not fear working hard, he built it with bricks and cement.”

Rather than using the name for the first occurrence of a reference to Jimmy, we see that the text uses “their little brother”, which is clearly less individuating as the three little pigs are all brothers: but the use of the possessive makes coreference clear. In the following sentence, we see however that a name is used: is this required or is it redundant? We take the use of the proper name to be not cognitively but textually required, since the use of a common name like “the little pig” would sound unnatural. Also, note that the use of a pronoun is impossible, since “their little brother” has not been established as a current Topic.

## 5. INFERENCES AND TRIGGERS

In a more schematic way we list here below the most interesting cases of inferential processes performed by the system:

### Case 1 – Individual to Set Inference

1. The three friends went all outdoors.
2. As they were walking in the garden, John said to himself, “Sara will marry that man”, without any resentment.

**TRIGGER:** Free pronoun “they” coreferent with “friend” a set with cardinality 3 in a fronted subordinate clause temporally coincident with the main.

**INFERENCES TO BE DRAWN:** Proper nouns introduced in the main clause should cospecify the members of the set of friends.

### Case 2 – Domain of Point-of-View

1. Richard would marry Sara.
2. He felt strongly about that.

**TRIGGER:** Free pronoun “he” to be coreferred to the subject\_of\_consciousness rather than to the locally available masculine proper noun “Richard”. The deictic pronoun “that” with clausal reference.

**INFERENCES TO BE DRAWN:** The first utterance is to be computed as belonging to John's domain of `point_of_view`: this would allow the system to continue the same Main Topic which will be made available to the following utterance, thus correctly coreferring "he" to John rather than to Richard.

Discourse deixis is triggered by the presence of an external pronoun with indefinite antecedent: discourse structure will be invoked and a discourse segment will be taken as coreference to "that".

### **Case 3 – Domain of Point-of-View**

1. She was the right person for a man like Richard.
2. For himself he was absurd.

**TRIGGER:** external pronouns "she", "he".

**INFERENCES TO BE DRAWN:** Both pronouns require a search in the Discourse Model from where to recover the antecedent of "she"; as to "he", the domain of `point_of_view` is used and the `subject_of_consciousness` taken as the antecedent to which the pronoun corefers.

### **Case 4 – Predicate-argument equality match**

1. John gave Mary a rose.
2. She took it and put it in her hair.
3. She knew that she had been given a present, something precious.

**TRIGGER:** External pronoun "she" and the dummy logical predicate "exist" added by the grammar to account for the missing agent of the passive sentence. The search in the Discourse Model is also activated by the indefinite NP "a present" in presence of a governing presuppositional predicate "know", in addition to tense Pluperfect in which the passive is expressed.

**INFERENCES TO BE DRAWN:** To recover the complete propositional content of the passive sentence a search in the Discourse Model is performed to look for a similar relation, "give" in a previous portion of text in which the same participant, "Mary" was involved with the same semantic role that of "Patient". In addition the indefinite NP "a present" is computed as a property cospecifying "a rose" and not as a new property in the DM.

### **Case 5 – Set to Individuals Inference**

1. When Steve faced them saying: are you enjoying yourselves?.

**TRIGGER:** The external pronoun "them".

**INFERENCES TO BE DRAWN:** In the Discourse Model there must be two fully identified individuals – i.e. which have been assigned proper names – apart from the one currently being nominated which is however obviative with the reference of the pronoun. They shall have to be built into a set to be used as antecedent of the pronoun "them". It uses the History List made up of most recent topics.

### Case 6 – Domain of Point-of-View

1. It was horrible! It was shocking!
2. Not for herself.

**TRIGGER:** An external pronoun “it” with indefinite reference subject of copulative constructions with evaluative predicates; a logophoric anaphora “herself” which has no local antecedent in the adjacent portion of text.

**INFERENCES TO BE DRAWN:** The indefinite pronoun “it” should be treated as a case of discourse deixis to be bound to an appropriate discourse segment independently organized by discourse structure. As to the anaphoric “herself” the subject\_of\_consciousness will be as usual assumed to be the antecedent coreferent with it, in this case, Mary.

### Case 7 – Recency Constraint

2. She felt only hostility and his determination to ruin that wonderful moment.

**TRIGGER:** The external free pronoun “his”.

**INFERENCES TO BE DRAWN:** No indication whatsoever in the previous portion of text tells us who should be the antecedent for the pronoun, apart from the meaning of the text which however requires in itself quite a lot of understanding. The antecedent is picked out on a recency criterion: the most recent possible antecedent is Steve and there is no other intervening Topics which interfere with this decision. We use the History List again and look for an individual entity which has been fully specified.

### Case 8 – Selectional restrictions membership match

1. Mary picked up the phone and called Jason.
2. Her husband, she thought, would have considered such a move as untruthful and utterly base.

**TRIGGER:** The deictic NP “such a move” realized linguistically as an Object semantically dependent on the predication.

**INFERENCES TO BE DRAWN:** The grammatical nature of the NP is such that it does not allow it to be computed as an independently referring entity, so the previous discourse is searched for. Actually, the semantic class to which the predicate “move” belongs is an Event, this will cause the semantic interpretation to search for an event in the previous DM, and the relation “call” will be taken to be the cospecifying one.

### Case 9 – Selectional restrictions membership match and meaning postulate

1. Più tardi fu un ufficiale di successo.
2. Poi nel 1945 Avveduti si dimise dall’esercito.
3. I militari lo avevano deluso.
4. Avveduti, deposta la divisa, si iscrisse all’università.

**TRIGGER:** Singular definite NP “l’esercito” and plural definite NP “i militari”.

**INFERENCES TO BE DRAWN:** The reference to “esercito”/army is recovered in the global context by looking up into the DM for an instance of a domain dependent individual/set/role of the kind “military”: in this case “ufficiale” is the role which has been asserted as being an instance of an individual belonging to a “military” class. Selectional restrictions associated to “esercito” include “coll, hum\_set” which allow the system to assert the fact of the superset nature of “esercito” if related to “ufficiale”. At the same time, the plural definite NP “militari” in the third utterance is also understood as coreferring with “esercito” even though with a different grammatical number. The same applies to “divisa”/uniform in the fourth utterance which is also understood as a metonymic expression, thus included semantically in the meaning of “esercito”/army.

In addition, an important inference is drawn in the second utterance due to the presence of the achievement verb “dimettere”/resign. In force of the meaning of the main predicate - a meaning postulate - the role “ufficiale” is a property no longer applicable to the individual it was associated with. Thus, there will be a new fact in the DM with polarity set to “0” asserting that the property of possessing the role of “ufficiale” is no longer applicable to the same individual from a given spatiotemporal location.

For the complete semantic representation of the DM, see below in the section devoted to the Stories.

### Case 10 – Meaning postulate

1. Intanto a Verona aveva conosciuto Paola, figlia di Antonio Alberti, potente senatore democristiano, e la aveva sposata.
2. Gli piaceva parlare del suocero come di una facile occasione mancata che chiunque avrebbe sfruttato ma che lui, Avveduti, preferiva lasciare perdere.

**TRIGGER:** The singular definite NP “suocero”/father-in-law.

**INFERENCES TO BE DRAWN:** The head noun is taken together with the pronominal “gli” and two suitable antecedents are searched in the local context, but an inference is required in order to determine the reference of the definite NP. A meaning postulate is used, which is made up of a number of interleaved “facts” to be checked against the DM.

As to the implementation of this particular case of anaphora resolution: it is performed in the Discourse Anaphora module which checks both disjointness and activates an inferential process to look for the entity that satisfies the meaning constraints specified in the call. It is important to notice that in this call the head of the Referential Expression to be matched must correspond to a given linguistic element “father-in-law”. In particular, the meaning of the expression father-in-law is made to correspond to a name or property who has a daughter which has associated the constant semantic identifier Id, who is married to an individual entity Id2, who also has a name or property associated in the DM, and either he was a Secondary Topic and

the father-in-law's name was asserted as Main Topic in the previous utterance, or the opposite applies. Then gender is checked against the DM, to ensure we are not taking the wrong partner of the marriage, and finally disjointness is checked.

### **Case 11 – Selectional restrictions membership match**

1. Solo verso il 1950 decise di accettare un posto nella organizzazione della Fiera di Verona.
2. Lo nominarono delegato cioè una specie di funzionario viaggiante...
3. Questo era un compito che...

**TRIGGER:** The empty subject pronoun with plural number. Then the predicative nominal “delegato” and the predicative indefinite nominal “compito”.

**INFERENCES TO BE DRAWN:** There is a collective noun which can be the antecedent even though its grammatical number is singular: this is taken care by the agreement procedure of the anaphora resolution module at discourse level, which sets the number of nouns classified as “institutions” at a selectional restrictions level as plural if needed.

In the same vein is captured the cospecifying relation of the social role “delegato” to the previous “posto” a generic name for an occupation. The same process is at stake with the predicative indefinite NP “un compito”, computed as an “activity” at selectional restrictions lexical level.

### **Case 12 – Selectional restrictions membership match**

1. Lo confermò nell'incarico alla Fiera.

**TRIGGER:** The definite NP “incarico”.

**INFERENCES TO BE DRAWN:** This NP is physically removed from the source reference by quite a number of utterances; the DM is searched for and the matching element is again the selectional restriction, “activity” assigned to both “incarico” e “posto” or “delegato” the antecedent to be picked up. In this way an inferential chain is created between “posto” and “delegato”, “compito”, “collaborazione”, “incarico”, which are all related to Avveduti's activity.

### **Case 13 – Implication**

1. L'ex ufficiale del Novara Cavalleria gli era simpatico.

**TRIGGER:** The presence of a singular definite NP with a prefix modifier “ex”, and of a free external pronoun “gli”.

**INFERENCES TO BE DRAWN:** In the case of the pronoun “gli” this is correctly assigned as coreferent to the Main Topic, being the best candidate and the state of the automaton “Continue”.

However, the definite NP requires a complex chain of inferences: in fact Avveduti was mentioned at the beginning of the story as being an officer, but then as resigning



from this activity to become a lawyer. The prefix “ex” is understood by the system as requiring a given property to be no longer applicable to a certain individual: in other words, the property of being an officer should exist in the Model but there should also be an explicit indication that it no longer applies. This is derived by the presence of a second fact about the same role “ufficiale” which is temporally subsequent to the first fact and asserts that the property has polarity “0”, see above for the relevant portion of the DM.

#### Case 14 – Recency Constraint

1. Lo confermò nello incarico alla fiera.
2. Avveduti funzionava benissimo come segretario particolare.
3. Sapeva mobilitare prefetti e questori.
4. Tutti gli invidiavano il suo segretario particolare.

**TRIGGER:** The presence of a free external pronoun “gli” and a definite NP “segretario particolare” which is obviative in reference with the pronominal “gli”.

**INFERENCES TO BE DRAWN:** There is a singular definite NP “il segretario particolare” which is coreferent with Avveduti, being a property already assigned to him in the previous portion of text and derivable from the DM; thus the pronoun “gli” cannot possibly be referred to the same antecedent. A search is made in the stack of Main Topics to find the most recent one, which is different from the current one, and “Trabucchi” is found, a character already asserted in the DM as being an instance of a “man”: in this way the masculine reference of “gli” is correctly recovered.

## 6. REFERENCE AND GENERAL NOUNS

As for human entities, we want to differentiate generic from defining descriptions or properties; in turn these descriptions, which might be exemplified by “consumer” or “producer” in our case, are roles attached to default properties or general nouns like “person” or “man” which are computed as generic entities. A side effect of the special use of referential expressions in legal texts is their lack of referential content: in other words, legal texts do not refer extensionally to entities in the world but deal with generic entities or with classes of individuals; proper nouns may only refer to institutions and as such they can be treated as collective entities or sets; common nouns may deal with abstract concepts belonging to the vocabulary of law or to concrete individuals or objects which could be found in the world but needn’t do so. Usually, when common nouns are introduced in a directive or bill they are individuated by a specific definition: this is particularly so, when those nouns are the addressee or the main topic to which the norm applies.

The general noun “person” should be referentially empty in legal contexts: consider its lack of gender specification and its inherent generic nature if compared to common nouns related to social roles. It is heavily used to refer either to actual or

to potential addressees: in particular, we might be talking about “producers” and “consumer”, and then go on talking about “the injured person” or “two or more persons” which can cause damage. In fact, “person” can be used both with modifiers and without them, as a definite NP or as an indefinite one. When it is used as a definite NP, it requires some modifier to be properly coreferred; no bare NPs can be created in association with such a generic head.

While common nouns may refer either to individuals or to classes, singular definite NPs with a generic head cannot be computed as individuals but only as classes and this must be inferred from the modifier or adjunct that specifies it adequately.

Plural definite NPs with generic heads are computed as classes, and can be used to corefer to other singular generic nouns or to specific nouns which they share default properties with.

Differently from what happens in narrative texts, legal texts introduce indefinite NPs not to assert the existence of some individual or set but to refer to classes or to generic entities.

Now consider the problem of general nouns: person, man, individual etc. They constitute the most common and most frequent noun in our texts and this is due to their referential properties: they can pick the maximum set of human entities available when used with no attribute or modifier; they can also be used to corefer to single classes of individuals by adequate modification. In other words, the concept “person” will be present at different levels of generality in the DM and it will be instantiated with a different semantic identifier according to its “dattrs” (see Woods and Brachman). Reference to one or the other of the instantiations of “person” will depend upon the way in which the discourse entity is described, in other terms on its “dattrs”, i.e. attributes, properties, parts, constituents, features, relations and so on. Notice that in our DM every entity has a description in terms of situation semantics, with a polarity and a spatiotemporal anchorage in terms of indices for main spatial and main temporal location.

Let’s now go back to general nouns: in any text, “person” should be introduced effectively only whenever a number of possible specific entities which might be subsumed by “person” already exist in the DM. In the texts we analyzed, this is what happens: in particular, the Council Directive we used and implemented which is discussed in mode detail at the end of this paper, has the following textual structure:

1. introduce main topics and addressee of its contents, which are “producers” and “consumers” living in the “Member States”; the directive concerns the protection of consumes and the producers’ liability in relation to injury or damage which might result from defective products. Importers of products who present themselves as producers are also regarded as such.
2. subsequently, both main topics are coreferred by means of the general noun “person”.

On first appearance of common nouns which might be subsumed by “person”, the system checks whether there is any such entity in the DM: in the affirmative case it

simply inherits its identifier and in the negative case, a generic entity is asserted in the DM. This entity is not an individual nor a set but has the following properties:

- it may subsume other generic entities of the same kind;
- it may be used to infer the nature of references to individuals by means of “person” and some specific attribute or property.

In particular, in the following text we find “several persons liable” and “injured person”. The first referring expression is computed as “producer” and the second one as “consumer”. In order to get this interpretation, which is the one intended in the Directive, an inferential process must be carried out on the basis of external knowledge of the world. However, the trigger to start this process is constituted by the attribute “injured” which is computed in the property checker algorithm as a possible new property to be associated to some entity already existing in the DM. The search starts from a referring expression that has some attribute or modifier which is a property not yet asserted for the corresponding entity. Consider the case of “injured person”, the algorithm by looking into the DM recovers the identifier associated to “consumer” on the basis of the inference that he is a “person” which is expressed in the following list of facts:

```
class(_, Id)
class(_, Ind)
fact(_, isa, [_:Ind, class:person], 1, _, _)
fact(_, ist_of, [_:Ind, class:man], 1, _, _)
fact(_, role, [consumer, Id, Ind], 1, _, _)
fact(_, isa, [_:Id, class:consumer], 1, _, _)
in(_, Id, Ind)
```

When the Id of “consumer” is recovered from the DM by the property checker that looks for correspondences between literal predicates and their relations in the DM, the “injured” attribute and the Id is passed on to the following inference engine that collects knowledge of the world associated to the trigger and checks to see whether it applies to the current identifier and its properties in the DM.

```
infer_process(Trigger, Id):-
infer_trig(Trigger, Props),
infer_rels(Id, Props).
```

This is done by means of an inferential process that takes as input external knowledge of the world, where we deposit information related to implicit knowledge, mutual knowledge and specialized information which could be part of a T-box component in a knowledge base understanding system. In those inheritance networks (see Woods, 1978), each concept that the system understands is represented in a network of concepts, which have to cope with the problem of internal recursion - i.e. of a concept defined in terms of another concept. Our taxonomy for this domain is constituted by a small list of facts which might be regarded on a par with an axiom made up of a

certain number of logical implications and the portion that interests us now is represented as follows:

*infer\_trig(injured, [cause(damage, 0) & has(protection, 1) & liable(Id, 0)]).*  
*infer\_trig(liable, [cause(damage, 1) & has(liability, 1)]).*

Each term is made up of a trigger, “injured”, “liable” which are the properties by means of which the general noun “person” is modified; in turn each trigger is associated with a set of relations and properties which have as argument either a class predicate and a polarity, or a variable and a polarity.

In the text, we learn that “injured person” is only used to corefer to “consumer”; we also know that in order to be interpreted as “injured” a person has not to have caused the damage, nor to be “liable” for it, and finally be the one that is given “protection”. By definition, then, a person is liable in case “he caused damage” and in case “he has liability”.

## 7. THE BASIC ALGORITHM

In our system, the problem of assigning properties to entities in narrative genre is carried out in two main phases: first only referring expressions (ref\_exs) asserted as best candidates for topichood and ranked as Main, Secondary, Expected or Potential Topics by the Topic Hierarchy Mechanism are considered. This might or might not apply to legal genre, on the basis of the assumption that there is no plot nor main character to focus on. However, legal texts contain pronominals which refer in the previous discourse segment and cannot be done away with. In case no such coreferent is found, the remaining DM is searched for. Generally speaking, there are four possible ways in which the problem might be solved:

### CASE A

there is already at least one such entity in the Discourse Model (DM) with the same class restrictions which could be used to corefer or cospecify the current ref\_ex disregarding its possible contextually determined properties and considering only the Initial Description (ID);

### CASE B

the entity to be picked up as possible coreferent or cospecifier must be semantically equal: i.e. we consider number of the current ref\_ex and if plural, check whether the entity has the same cardinality or is asserted as a class; and if singular look only for individuals;

### CASE C

the entity and the ref\_ex must have the same attributes, roles or other properties assigned to it in the DM and present in the current ref\_ex;

## CASE D

works as case C and in addition it looks only for properties associated to entities asserted as Topics of discourse in the previous text; in case no such property is present, a new entity is entered in the DM;

It is easy to notice that, case D is the most restrictive but also the most genre bound procedure: in fact, it is only useful in case the text is a narrative one.

The first question is discovering whether the current ref\_ex is already present in the DM, i.e. if it has already been explicitly mentioned. If the problem at hand were that of matching the current linguistic form of the predicate or concept with those present in the discourse model (DM) this would be easily solved. However this procedure is clearly insufficient and leads to mistaken matches in case cardinality, definiteness, or simply the modifiers of the current linguistic use of a certain entity present in the DM does not coincide semantically with it.

The overall framework is further complicated by the presence of generic entities and by the fact that what we are dealing with are usually entities which might be easily subsumed by them or constitute a specific subclass. A “consumer” or a “producer” is clearly such a subclass in case the general noun “person” has already been introduced in the DM. However, problems will arise when the contrary applies, i.e. whenever a general noun is introduced with specific properties added by adjunction or modification: in this case it would both represent a subclass of a generic entity and a coreferent of a some class of individual or role.

In Situation Semantics where reality is represented in Situations which are collections of Facts: in turn facts are made up of Infons which information units characterised as follows:

Infon(Index,  
Relation(Property),  
List of Arguments - with Semantic Roles,  
Polarity - 1 affirmative, 0 negation,  
Temporal Location Index; Location Index)

In addition Arguments have each a semantic identifier which is unique in the Discourse Model and is used to individuate the entity uniquely. Also propositional facts have semantic identifiers assigned thus constituting second level ontological objects. They may be “quantified” over by temporal representations but also by discourse level operators, like subordinating conjunctions. Negation on the contrary is expressed in each fact. All entities and their properties are asserted in the DM with the relations in which they are involved; in turn the relations may have modifiers - sentence level adjuncts and entities may also have modifiers or attributes. Each entity has a polarity and a couple of spatiotemporal indices which are linked to main temporal and spatial locations if any exists; else they are linked to presumed time reference derived from tense and aspect computation. Entities are mapped into semantic individual with

the following ontology: on first occurrence of a referring expression it is asserted as an INDIVIDUAL if it is a definite or indefinite expression; it is asserted as a CLASS if it is quantified (depending on quantifier type) or has no determiner. Special individuals are ENTs which are associated to discourse level anaphora which bind relations and their arguments. Finally, we have LOCs for main locations, both spatial and temporal. If it has a cardinality determined by a number, it is plural or it is quantified (depending on quantifier type) it is asserted as a SET and the cardinality is simply inferred in case of naked plural, i.e. in case of collective nominal expression it is set to 100, otherwise to 5. On second occurrence of the same nominal head the semantic index is recovered from the history list and the system checks whether it is the same referring expression:

- in case it is definite or indefinite with a predicative role and no attributes nor modifiers nothing is done;
- in case it has different number – singular and the one present in the DM is a set or a class nothing happens;
- in case it has attributes and modifiers which are different and the one present in the DM has none, nothing happens;
- in case it is quantified expression and has no cardinality, and the one present in the DM is a set or a class, again nothing happens.

In all other cases a new entity is asserted in the DM which however is also computed as being included in (a superset of) or by (a subset of) the previous entity.

## 8/. DM AND INFERENCES: SOME EXAMPLES

We will show here below the DM relatively only to three entities: consumer, producer and their general noun “person” subsuming both. In utterance 4 with first appearance of “consumer and producer” the system generates a generic entity “person” which subsumes both: the inclusion relation is registered by the fact in( $\_$ , A, B) where A is the semantic identifier or initial description of the superset or superobject, and B is the subsumed entity:

[Whereas approximation of the laws of the ‘Member’ ‘States’ concerning the liability of the producer for damage caused by the defectiveness of his products is necessary, because the existing divergences may entail a differing degree of protection of the consumer, against damage caused by a defective product to his health or property.]

ent(infon102, id22)

fact(infon103, isa, [arg:id22, class:person], 1, univ, univ)

class(infon104, id23)

class(infon105, id24)

fact(infon106, cause, [causer:id5, arg:id24], 1, univ, univ)

fact(infon107, inst\_of, [ind:id24, class:abstract\_state], 1, univ, univ)

fact(infon108, isa, [ind:id24, class:damage], 1, univ, univ)

fact(infon109, against, [arg:id23, malefactive:id24], 1, univ, univ)

```

fact(infon110, inst_of, [ind:id23, class:man], 1, univ, univ)
fact(infon111, isa, [ind:id23, class:consumer], 1, univ, univ)
in(infon112, id23, id22)
fact(infon113, role, [consumer, id23, id22], 1, univ, univ)
fact(infon114, cause, [arg:id4, damage:id24], 1, univ, univ)
class(infon115, id25)
fact(infon116, inst_of, [ind:id25, class:man], 1, univ, univ)
fact(infon117, isa, [ind:id25, class:producer], 1, univ, univ)
in(infon118, id25, id22)
fact(infon119, role, [producer, id25, id22], 1, univ, univ)
class(infon133, id29)
fact(infon134, differing, [ind:id29], 1, univ, univ)
class(infon135, id30)
fact(infon136, has, [experiencer:id23, arg:id30], 1, univ, univ)
fact(infon137, inst_of, [ind:id30, class:abstract_state], 1, univ, univ)
fact(infon138, isa, [ind:id30, class:protection], 1, univ, univ)
fact(infon139, of, [arg:id29, specif:id30], 1, univ, univ)
fact(infon140, inst_of, [ind:id29, class:[measure]], 1, univ, univ)
fact(infon141, isa, [ind:id29, class:degree], 1, univ, univ)

```

In utterance 6 we learn that the producer should be made liable at certain conditions, and this is registered as an additional fact about producers:

```

[Whereas protection of the consumer requires that all the producers involved in the production process should be made liable in so far as their finished product, component part or any raw material supplied by them was defective.]
fact(infon248, liable, [nil:id25], 1, univ, univ)

```

In utterance 8 the text introduces an undefined set of “several persons” which are liable for a certain damage and are related to protection of the consumer in the same context. The system assigns to “injured person” the same identifier as the consumer, creates a new class of entities “person” which are liable and cause “damage”, included in the same superset of producers:

```

[Whereas, in situations where several persons are liable for the same damage the protection of the consumer requires that the injured person should be able to claim full compensation from any one of them.]
fact(infon323, injured, [ind:id23], 1, univ, univ)
class(infon324, id73)
fact(infon325, full, [ind:id73], 1, univ, univ)
fact(infon326, inst_of, [ind:id73, class:legal], 1, univ, univ)
fact(infon327, isa, [ind:id73, class:compensation], 1, univ, univ)
fact(infon328, cause, [arg:id25, damage:id24], 1, univ, univ)

```

```

class(infon329, id74)
fact(infon330, liable, [nil:id74], 1, univ, univ)
fact(infon331, cause, [arg:id74, damage:id24], 1, univ, univ)
in(infon332, id25, id74)
fact(infon333, inst_of, [ind:id74, class:man], 1, univ, univ)
fact(infon334, isa, [ind:id74, class:person], 1, univ, univ)
class(infon335, id75)
fact(infon336, inst_of, [ind:id75, class:abstract_state], 1, univ, univ)
fact(infon337, isa, [ind:id75, class:situation], 1, univ, univ)
fact(id76, claim, [experiencer:id74, theme_aff:id73, source:id74], 1, tes(finfl_aq4), univ)

```

In utterance 10 liability for damage is charged on the producer: the system looks for a similar predication in the knowledge base and finds the one related to infon248, which was asserted in utterance 6 above:

```

[The producer shall be liable for damage caused by a defect in his product.]
fact(infon382, cause, [arg:id25, damage:id24], 1, univ, univ)
class(infon383, id84)
fact(infon384, in, [arg:id84, locative:id5], 1, univ, univ)
fact(infon385, inst_of, [ind:id84, class:legal], 1, univ, univ)
fact(infon386, isa, [ind:id84, class:defect], 1, univ, univ)
fact(infon387, cause, [causer:id84, arg:id24], 1, univ, univ)
fact(id85, be, [prop:infon248], 1, tes(f3_dd07), univ)

```

In utterance 11 we find “the injured person” again, and the system picks up id23 associated to consumer. Notice the computation of the meaning for “relationship” which has “causal” as modifier: the system understands it as a relation and finds infon387 which is associated to a fact in the knowledge base asserted in the previous utterance and saying that there is a “cause” relation between “defect” and “damage”. This was expressed in terms of semantic roles, i.e. the defect is interpreted as a “causer” and its argument is id24, the “damage”. In the new utterance, this is linguistically formulated in terms of “relationship”, where there is a semantic marker “between” which expresses a relation, and the relation has two arguments:

```

[The injured person shall be required to prove the damage, the defect and the causal relationship
between defect and damage.]
ind(infon423, id95)
class(infon424, id96)
fact(infon425, inst_of, [ind:id96, class:abstract_state], 1, univ, univ)
fact(infon426, isa, [ind:id96, class:defect_damage], 1, univ, univ)
fact(infon427, between, [arg:id95, relation:id96], 1, univ, univ)
fact(infon428, cause, [arg:id95, relation:infon387], 1, univ, univ)
fact(infon429, inst_of, [ind:id95, class:abstract_state], 1, univ, univ)

```



```
fact(infon430, isa, [ind:id95, class:relationship], 1, univ, univ)
fact(infon433, [defect, damage, relationship], [arg:id97], 1, univ, univ)
fact(id98, prove, [actor:id23, prop:infon433], 1, tes(fin1_dd07), univ)
fact(id100, require, [actor:id23, prop:id98], 1, tes(f4_dd07), univ)
```

In utterance 12 the same predication is present, “be liable” and the same infon248 is picked up:

```
[Where, as a result of the provisions of this directive, two or more persons are liable for the same damage.]
fact(id104, be, [prop:infon248], 1, tes(f49_dd08), univ)
```

In utterance 13 a pronoun is introduced intersentially to corefer to the same persons, associated to the class of producers:

```
[They shall be liable jointly and severally, without prejudice to the provisions of national law concerning the rights of contribution or recourse.]
fact(id109, be, [prop:infon248], 1, tes(f3_dd10), univ)
```

In utterance 14, we find a reference to the superset of persons, the one introduced as generic entity, since the system does not find any hint in the utterance by means of which “a person” could be interpreted as belonging either to the class of consumers or to the class of producers:

```
[A product is defective when it does not provide the safety which a person is entitled to expect, taking all circumstances into account.]
fact(id116, expect, [experiencer:id22, theme_unaff:id111], 1, tes(fin1_dd10), univ)
fact(id118, entitle, [theme:id22, prop:id116], 1, tes(f6_dd10), univ)
```

Finally, in utterance 17, a new set of “person” is introduced which “imports a product” and according to the interpretation assigned by the utterance to this set it is included in the superset of the class of producers:

```
[Without prejudice to the liability of the producer any person who imports into the community a product for sale, hire or any form of distribution in the course of his business, shall be deemed to be a producer within the meaning of this directive.]
class(infon631, id132)
in(infon633, id25, id132)
fact(infon634, inst_of, [ind:id132, class:man], 1, univ, univ)
fact(infon635, isa, [ind:id132, class:person], 1, univ, univ)
fact(id138, import, [agent:id132, theme_aff:id7], 1, tes(f3_dd17), univ)
```

### 8.1 *Spatiotemporal locations, scenario and inferential processes*

As already mentioned above, some inferential process is triggered by a definite singular NP which causes the system to activate a search in the external knowledge of the world. This search determines the conditions for a scenario effect or a part\_of relation or other such semantic inferential processes. In particular, the description of these relation is assumed to be very simple and straightforward. What really matters is the way in which the system computes the inclusion relations that we list here below:

```
included(porta,casa).
included(finestra,casa).
included(chiavistello,porta).
included(corner,restaurant).
included(waiter,restaurant).
included(menu,restaurant).
```

As to the first three inclusion relations, they are used to introduce a part\_of relation, as the texts of the three little pigs commented further will show. The remaining three inclusion relations are taken by the system as a restriction on the spatial location of the singular definite NP. Then in case it is classified as a social role, a further inference causes the semantic module to assert a property which is a role in that location, or scenery, as is the case with the following text, where the final sentence has a SUBJECT pronominals which is bound to John and not to the waiter, because the latter is understood as a role (id11) in the restaurant (id2).

```
John went into a restaurant.
loc(infon3, id1, [arg:main_tloc, arg:tr(f5_r01)])
loc(infon4, id2, [arg:main_sloc, arg:restaurant])
ind(infon5, id3)
fact(infon6, inst_of, [ind:id3, class:man], 1, univ, univ)
fact(infon7, name, [john, id3], 1, univ, univ)
fact(infon9, isa, [arg:id2, arg:restaurant], 1, id1, id2)
fact(id4, go, [agente:id3, locat:id2], 1, tes(f5_r01), id2)
```

```
There was a table in the corner.
ind(infon21, id6)
ind(infon22, id7)
fact(infon23, inst_of, [ind:id7, class:thing], 1, univ, univ)
fact(infon24, isa, [ind:id7, class:corner], 1, id1, id2)
fact(infon25, in, [arg:id6, locativo:id7], 1, id1, id2)
fact(infon26, isa, [ind:id6, class:table], 1, id1, id2)
fact(infon27, inst_of, [ind:id6, class:thing], 1, univ, univ)
fact(infon29, isa, [arg:id8, arg:corner], 1, id1, id2)
fact(infon30, in, [arg:id6, locativo:id8], 1, id1, id2)
```

fact(id9, there\_be, [tema\_nonaff:id6], 1, tes(f4\_free\_r02), id2)

The waiter took the order.

ind(infon42, id11)

fact(infon43, inst\_of, [ind:id11, class:[social\_role]], 1, univ, univ)

fact(infon44, isa, [ind:id11, class:waiter], 1, id1, id2)

fact(infon45, role, [waiter, id2, id11], 1, id1, id2)

fact(infon47, isa, [arg:id12, arg:order], 1, id1, id2)

fact(id13, take, [agente:id11, tema\_aff:id12], 1, tes(f5\_r03), id2)

The air was nice and clean.

loc(infon59, id15, [arg:main\_tloc, arg:tes(f5\_r03)])

fact(infon60, isa, [arg:id16, arg:air], 1, id15, id2)

fact(infon61, [clean, nice], [arg:id16], 1, id15, id2)

fact(id17, be, [prop:infon61], 1, tes(f5\_r04), id2)

He began to read his book.

fact(infon76, isa, [arg:id19, arg:book], 1, id15, id2)

fact(infon77, poss, [arg:id19, poss:id3], 1, id15, id2)

fact(id20, read, [agente:id3, tema\_aff:id19], 1, tes(fin1\_free\_a1), id2)

fact(infon78, isa, [arg:id20, arg:pr], 1, tes(fin1\_free\_a1), id2)

fact(infon79, isa, [arg:id21, arg:tloc], 1, tes(fin1\_free\_a1), id2)

fact(infon80, pres, [arg:id21], 1, tes(fin1\_free\_a1), id2)

fact(id22, begin, [agente:id3, prop:id20], 1, tes(f3\_free\_a1), id2)

## 9. THE STORIES

We shall now discuss in more detail what the Interpreter does in each utterance of the texts analyzed. However we shall not include information from the DM which is irrelevant to the problem we are now tackling: in particular, we shall not comment on DM representations related to utterance which do not add any new entity to the model nor do require any special inferential process to be understood. In each story we have divided up relations from entities, in turn relations may show up as either fact, sit(uation), and conc(ept). Locations are used by the semantic interpreter to assign the Main Spatial Location Index to each term. Time intervals are used to assign the Main Temporal Location index. Facts and sits characterize objects or entities in the model.

### 9.1 Story 1

C'erano una volta tre fratelli porcellini che vivevano felici nella campagna.

loc(infon3, id1, [arg:main\_tloc, arg:volta])

loc(infon4, id2, [arg:main\_sloc, arg:campagna])

```

set(infon5, id3)
card(infon6, id3, 3)
fact(infon7, fratello, [ind:id3], 1, id1, id2)
fact(infon8, isa, [ind:id3, class:porcellino], 1, id1, id2)
fact(infon9, inst_of, [ind:id3, class:animale_cibo], 1, univ, univ)
fact(infon13, isa, [arg:id2, arg:campagna], 1, id1, id2)
fact(id4, vivere, [actor:id3, locativo:id2], 1, tes(f6_po1), id2)
fact(infon18, felice, [arg:id4], 1, tes(f6_po1), id2)
fact(id6, esserci, [tema_nonaff:id3], 1, tes(f7_po1), id2)
fact(infon23, isa, [arg:id1, arg:volta], 1, tes(f7_po1), id2)
fact(infon24, nil, [arg:id6, non_punct:id1], 1, tes(f7_po1), id2)

```

In the first utterance we have a set of three little pigs and a property “fratello”/ brother which are asserted as facts, two main relations “essere”/be and “vivere”/live which are also facts and a number of sits associated to these facts. All the entities have positive polarity and are located in space and time.

Nello stesso luogo però viveva anche un terribile lupo che si nutriva proprio di porcellini grassi e teneri.

```

loc(infon36, id8, [arg:main_tloc, arg:tes(f7_po1)])
ind(infon37, id9)
fact(infon38, terribile, [ind:id9], 1, id8, id2)
fact(infon39, isa, [ind:id9, class:lupo], 1, id8, id2)
fact(infon40, inst_of, [ind:id9, class:animale_feroce], 1, univ, univ)
class(infon41, id10)
fact(infon42, grasso, [ind:id10], 1, id8, id2)
fact(infon43, tenero, [ind:id10], 1, id8, id2)
in(infon44, id3, id10)
fact(infon45, inst_of, [ind:id10, class:animale_cibo], 1, univ, univ)
fact(infon46, isa, [ind:id10, class:porcellino], 1, id8, id2)
fact(id11, nutrire, [esperiente:id9, tema_nonaff:id9, cibo:id10], 1, tes(f5_po2), id2)
fact(infon48, isa, [arg:id13, arg:luogo], 1, id8, id2)
fact(id14, vivere, [actor:id9, locativo:id13], 1, tes(f6_po2), id2)
fact(infon53, pero_, [arg:id14], 1, tes(f6_po2), id2)
fact(infon54, anche, [arg:id14], 1, tes(f6_po2), id2)

```

In the second utterance the spatial location remains the same because the locative adjunct is constructed with a deictic “stesso”/same; the main temporal location also has the same index. The model is updated with a new individual, the wolf. The relations show how the static interpreter computes the argument “porcellini” for the predicate “nutrire”. This is taken as a sit owing to the fact that it is a plural naked NP, in other words, an intensional object. This NP will be exploited for inferencing the meaning of a metaphor, “pasto”/meal, as corresponding to little pigs, being the meal of the wolf.

Questi allora, per proteggersi dal lupo, decisero di costruirsi ciascuno una casetta.

```
set(infon79, id16)
card(infon80, id16, 3)
fact(infon81, piccolo, [ind:id16], 1, id8, id2)
sit(infon82, isa, [ind:id16, class:casa], 1, id8, id2)
fact(infon83, inst_of, [ind:id16, class:cosa], 1, univ, univ)
sit(id17, costruire_si, [agente:id3, tema_eff:id16], 1, tes(fin4_po3), id2)
fact(id19, decidere, [actor:id3, prop:id17], 1, tes(f5_po3), id2)
fact(infon97, allora, [arg:id19], 1, tes(f5_po3), id2)
fact(id21, proteggere, [agente:id3, esperiente:id3, malef:id9], 1, tes(fin1_po3), id2)
fact(infon103, per, [arg:id19, purpose:id21], 1, tes(f5_po3), id2)
```

When we get to third utterance a deictic pronoun is used to indicate the persistence of the same individual in the discourse; there is an indefinite NP which is computed as a set of sits by means of quantifier scoping detected in logical form, and owing to the fact that the NP is itself contained in a proposition which is computed as a sit, not a fact, with the relation “costruire”. The cardinality of the set of houses is derived from the abstract possessor built in the NP via the benefactive enclitic “si”, and the scope assignment wrought on it by the floating quantifier “ciascuno”/each. This in turn is computed in *f*-structure as an open adjunct which is controlled by the subject NP “questi”/these, a deictic whose resolvent is the set of three little pigs already present in the model.

Il maggiore, Jimmi, che era saggio, lavorava di buona lena e costruì la sua casetta con solidi mattoni e cemento.

```
loc(infon128, id23, [arg:main_tloc, arg:tes(f8_po4)])
ind(infon129, id24)
fact(infon130, name, [jimmi, id24], 1, id23, id2)
fact(infon131, saggio, [nil:id24], 1, id23, id2)
in(infon132, id24, id3)
ind(infon133, id25)
in(infon134, id25, id16)
fact(infon135, isa, [ind:id25, class:casa], 1, id23, id2)
fact(infon136, poss, [jimmi, id24, id25], 1, id23, id2)
fact(infon137, isa, [ind:id25, class:mattone], 1, id23, id2)
fact(infon138, isa, [ind:id25, class:cemento], 1, id23, id2)
fact(id26, lavorare, [agente:id24], 1, tes(f6_po4), id2)
fact(infon145, isa, [arg:id28, arg:lena], 1, tes(f6_po4), id2)
fact(infon146, di, [arg:id26, nil:id28], 1, tes(f6_po4), id2)
fact(id30, costruire, [agente:id24, tema_eff:id25], 1, tes(f8_po4), id2)
fact(infon156, [cemento, mattone], [arg:id29], 1, tes(f8_po4), id2)
fact(infon157, con, [arg:id30, nil:id25, prop:infon156], 1, tes(f8_po4), id2)
```

At this point, a name is associated to the elder of the three little pigs and other properties are predicated. We also know he is the possessor of a house made of bricks and cement, both universally located properties of the house. However, the house is now introduced as a definite NP and not as an indefinite NP owing to the fact that the discourse strategy already presented a set of houses, as a sit in the previous portion of text: thus a sit has now supported a fact, the two houses being now computed as a subset of the previous set made up of three members.

Gli altri, Timmy e Tommy, pigri e oziosi se la sbrigarono in fretta costruendo le loro casette con la paglia e con pezzetti di legno.

```

set(infon171, id32)
card(infon172, id32, 2)
ind(infon173, id33)
in(infon174, id33, id32)
fact(infon175, name, [timmy, id33], 1, id23, id2)
ind(infon176, id34)
in(infon177, id34, id32)
fact(infon178, name, [tommy, id34], 1, univ, univ)
in(infon179, id32, id3)
fact(infon180, ozioso, [ind:id32], 1, id23, id2)
fact(infon181, pigro, [ind:id32], 1, id23, id2)
fact(infon182, poss, [porcellino, id32, id35], 1, id23, id2)
set(infon183, id35)
card(infon184, id35, 2)
fact(infon185, piccolo, [ind:id35], 1, id23, id2)
fact(infon186, isa, [ind:id35, class:paglia], 1, id23, id2)
fact(infon187, isa, [ind:id35, class:legno], 1, id23, id2)
in(infon188, id35, id16)
fact(infon189, isa, [ind:id35, class:casa], 1, id23, id2)
fact(infon190, isa, [arg:id36, arg:fretta], 1, id23, id2)
fact(id37, sbrigarci, [agente:id32, modale:id36], 1, tes(f5_po5), id2)
fact(infon197, [legno, paglia], [arg:id39], 1, tes(f5_po5), id2)
fact(id40, costruire, [agente:id32, tema_eff:id35], 1, tes(fgerund1047_po5), id2)
fact(infon203, [legno, paglia], [arg:id39], 1, tes(fgerund1047_po5), id2)
fact(infon204, con, [arg:id40, nil:id35, prop:infon203], 1, tes(fgerund1047_po5), id2)
fact(infon205, coincide, [arg:id37, arg:id40], 1, tes(f5_po5), id2)

```

In this utterance, the subset of two little pigs is coreferred by a nominal substitute, “gli altri”/the others, which we must compute as two in number, and names are associated to these two little pigs. They are assigned evaluative properties, and they are individuated as possessors of two little houses, which however are introduced as previously happened, with a possessive and no cardinality. The possessor relation is built accordingly on the subset of two houses by accessing one of the

most recent properties being assigned to the set of possessors, which in this case is “ozioso”/lazy.

Notice that the possession is not specified for each individual house but for the set, and the following text should make this fact persistent, even though we will see that this is not the case. An open adjunct is then used to predicate a property of these house, the fact that they have been built of “paglia”/straw: this property is computed by accessing the controller of the open adjunct, i.e. the houses.

I due porcellini pigri passavano le loro giornate suonando e cantando una canzone che diceva:  
chi ha paura del lupo cattivo.

```
loc(infon219, id42, [arg:main_tloc, arg:tes(f5_po5)])
ind(infon220, id43)
fact(infon221, isa, [ind:id43, class:canzone], 1, id42, id2)
fact(infon222, inst_of, [ind:id43, class:[astratto, informa]], 1, univ, univ)
fact(infon225, isa, [arg:id44, arg:giornata], 1, id42, id2)
fact(infon226, poss, [arg:id44, poss:id32], 1, id42, id2)
fact(id45, suonare, [esperiente:id32], 1, tes(fgerund1051_po6), id2)
fact(id47, passare, [tema_nonaff:id32, tema_nonaff:id44, gerund:id45], 1, tes(f8_po6), id2)
fact(infon240, focus, [arg:id49], 1, id42, id2)
fact(infon241, isa, [arg:id49, arg:chi], 1, id42, id2)
fact(infon242, paura, [nil:id49], 1, id42, id2)
sit(id50, avere, [arg:id49, prop:infon242], 1, tes(f6_po6), id2)
fact(id52, dire, [agente:id43, propq:id50], 1, tes(f7_po6), id2)
fact(id54, cantare, [esperiente:id32, tema_nonaff:id43], 1, tes(fgerund1052_po6), id2)
fact(id47, passare, [tema_nonaff:id32, tema_nonaff:id44, gerund:id45], 1, tes(f8_po6), id2)
```

At this point, the set of two little pigs is specifically indicated. However we make coreference work in the model due to the previous inference on “the others” which made us build a set of little pigs with cardinality two. Notice that we have two class entities, which translate the implicit locative argument of “passare”/spend, and the subject wh- word of “avere”/have: the restriction on the class is constituted by a semantic feature. However, if it is true that the use of definite article points to the existence of some entity in the model that matches the required reference, it might also be the case that the appropriate reference must be recovered from the presence of the attribute “pigri”/lazy. This property has been associate to one set in the previous DM and this corresponds to the most recently built set of little pigs, with id32 semantic constant associated.

Ma ecco che improvvisamente il lupo apparve alle loro spalle.

```
fact(infon286, inst_of, [ind:id56, class:cosa], 1, univ, univ)
fact(infon287, isa, [ind:id56, class:spalla], 1, id42, id2)
fact(infon288, part_of, [spalla, id56, id32], 1, id42, id2)
fact(id56, apparire, [actor:id9], 1, tes(f1_po7), id2)
```

fact(infon289, a, [arg:id56, nil:id35], 1, tes(f1\_po7), id2)

No entity is introduced in this and the following utterance, which is not included here.

Questo intanto si leccava già i baffi pensando al suo prossimo pasto così invitante e saporito.

```
loc(infon332, id66, [arg:main_tloc, arg:tes(f6_po8)])
ind(infon333, id67)
in(infon334, id67, id32)
sit(infon335, isa, [ind:id67, class:pasto], 1, id66, id2)
sit(infon335, kind_of, [ind:id32, class:pasto], 1, id66, id2)
fact(id70, leccare_baffo, [esperiente:id9], 1, tes(f5_po9), id2)
fact(id72, pensare, [actor:id9, tema:id32], 1, tes(fgerund1064_po9), id2)
fact(infon346, isa, [arg:id72, arg:st], 1, tes(fgerund1064_po9), id2)
fact(infon350, coincide, [arg:id70, arg:id72], 1, tes(f5_po9), id2)
```

In this utterance “pasto” has been computed as an individual included in the set of two little pigs by means of the semantic features associated to it [edible,event], where “edible” is a feature associated to the semantic role “food” and this in turn appears in a relation “think/pensare” where the subject/experiencer is the wolf, and the complement, the metaphoric intensional “food”, refers to the “little pigs”. The knowledge base is tapped in order to compute fully the inference from selectional restrictions and a further sit is added by means of the “kind\_of” relation. Note that the main sentence is constituted by an idiomatic expression whose predicate “leccare”/lick has an OBJECT NP which is computed as a FORM. This causes the relation to be treated as a sit rather than a fact with a single argument. Also notice that the predicate has been reinterpreted from “leccare” into the semantic concept “leccare\_baffo” by means of the same procedure that has transformed “costruire” into “costruire\_si” in a previous utterance. Here below we give the full f-structure of the current utterance.

```
index:f4
pred:leccare
lex_form:[np/subj/agent/[animate,human], np/obj/theme_aff/[object, body_part], pp/obj2/
benef/a/ [animate, human], idioms/obj/form/[baffo], idioms/obj2/form/[si]]
mood:ind
tense:imp
cat:subjective
subj/experiencer:index:sn1
  cat:[animate, human]
  pred:questo:[pers:3,gen:mas ,num:sing,case:nom]
  spec:def:+
  tab_ref:[+ ref, + pro, - ana, + class]
  antecedent:external
```



interpretation:specific  
 obj/form:index:sn148  
   cat:[object, body\_part]  
   pred:bafo:[pers:3,gen:mas ,num:plur]  
   spec:def:+  
   subj/poss:index:sn149  
     cat:[object, body\_part]  
     pred:pPro  
     spec:def:+  
     controller:sn146  
     tab\_ref:[+ ref, + pro, + ana, - me]  
   tab\_ref:[+ ref, - pro, - ana, + class]  
 obj2/form:index:sn146  
   cat:[ ]  
   pred:si:[pers:3,gen:\_ ,num:\_,case:dat]  
   spec:def:+  
   tab\_ref:[- ref, - pro, + ana, + me, - subj]  
   antecedent:sn1  
   interpretation:specific  
 adjs:adv:pred\_adv:intanto  
   tipo:temp  
   duraz:non\_punt  
   config:[td=tr]  
 adv:pred\_adv:gia\_  
   tipo:temp  
   duraz:non\_punt  
   config:[td=tr]  
 adj/gerund:index:fgerund6  
   pred:pensare  
   lex\_form:[np/subj/actor/[human,animate], pp/obl/theme/a/[event| \_]]  
   mood:ger  
   tense:pres  
   cat:mental\_activ  
   subj/actor:index:sn153  
     cat:[human, animate]  
     pred:pPro  
     tab\_ref:[+ ref, + pro, + ana, - me]  
     antecedent:sn1  
     interpretation:specific  
   obl/theme:index:sn151  
     cat:[event, object]  
     pred:pasto:[pers:3,gen:mas ,num:sing]  
     spec:def:+

```

subj/poss:index:sn152
  cat:[]
  pred:suo:[pers:3,gen:mas ,num:sing]
  spec:def:+
  tab_ref:[+ ref, + pro, - ana, + me]
  antecedent:sn1
  interpretation:specific
mods:mod_r:index:saa14
  cat:[evaluative]
  coord:head:coordination
    coord:index:saa12
      cat:[evaluative]
      pred:invitante:[gen:mas ,num:sing]
      adjs:int:cosi_
    coord:index:saa13
      cat:[evaluative]
      pred:saporito:[gen:mas ,num:sing]
  mod_a:index:sa6
    cat:[tempor]
    pred:prossimo:[gen:mas ,num:sing]
  tab_ref:[+ ref, - pro, - ana, + class]
aspect:activity
aspect:state

```

Finalmente i porcellini riuscirono a raggiungere la loro casetta e vi si chiusero dentro sbarrando la porta.

```

loc(infon369, id74, [arg:main_tloc, arg:tes(f8_po10)])
ind(infon370, id76)
fact(infon371, isa, [ind:id76, class:porta], 1, id74, id35)
fact(infon372, part_of, [casa, id35, id76], 1, id74, id35)
sit(id77, raggiungere, [actor:id32, goal:id35], 1, tes(fin1_po10), id35)
fact(id79, riuscire, [agente:id32, prop:id77], 1, tes(f6_po10), id35)
fact(id81, chiudere, [agente:id32, tema_aff:id32], 1, tes(f8_po10), id35)
fact(infon389, isa, [arg:id83, arg:dentro], 1, tes(f8_po10), id35)
fact(infon390, in, [arg:id81, locativo:id83], 1, tes(f8_po10), id35)
fact(id84, sbarrare, [agente:id32, tema_aff:id76], 1, tes(fgerund1072_po10), id35)
fact(infon396, coincide, [arg:id81, arg:id84], 1, tes(f8_po10), id35)

```

No new entity is introduced in this utterance. However, notice that the house being talked about is the set of two houses already existing in the world, even though the actual NP used in this utterance is a singular one. This coreference is worked out on the basis of definiteness and the possessive which is bound to the subject NP. This NP is clearly ambiguous between a distributive and a collective reading: in the

second case, only one house should appear, and since no single individual house exists in the DM belonging to the two lazy little pigs a new entity should have been added. However, the system prefers a distributive reading and looks for inds or sets belonging to a possessor if there is one, and picks up the entity in the DM which fits the current description.

Dalla finestra cominciarono a deridere il lupo cantando la solita canzoncina: chi ha paura del lupo cattivo.

ind(infon417, id86)

fact(infon418, isa, [ind:id86, class:finestra], 1, id74, id35)

fact(infon419, part\_of, [casa, id35, id86], 1, id74, id35)

sit(id87, deridere, [actor:id32, tema\_nonaff:id9], 1, tes(finfl\_po11), id35)

fact(id89, cominciare, [agente:id32, prop:id87], 1, tes(f5\_po11), id35)

fact(infon433, da, [arg:id89, nil:id32, loc\_origin:id86], 1, tes(f5\_po11), id35)

fact(infon437, focus, [arg:id91], 1, tes(f5\_po11), id35)

fact(infon438, isa, [arg:id91, arg:chi], 1, tes(f5\_po11), id35)

fact(infon439, paura, [nil:id91], 1, tes(f5\_po11), id35)

fact(id92, avere, [arg:id91, prop:infon439], 1, tes(f5\_po11), id35)

fact(id94, cantare, [esperiente:id32, tema\_nonaff:id43], 1, tes(fgerund1076\_po11), id35)

fact(infon448, coincide, [arg:id89, arg:id94], 1, tes(f5\_po11), id35)

In this utterance we have a property associated to the individual “canzone”/song, which is interpreted as its subject, “paura”/fear: owing to its definiteness, this song is taken to be the same as the one mentioned previously in discourse. No new entity is introduced in the following utterance.

Il lupo stava intanto pensando al modo di penetrare nella casa.

loc(infon468, id96, [arg:main\_tloc, arg:tes(f5\_po11)])

sit(id97, penetrare, [actor:id9, locativo:id35], 1, tes(finfl\_po12), id2)

fact(id99, pensare, [actor:id9, prop:id97], 1, tes(f5\_po12), id2)

In this utterance the story becomes inconsistent: the reference to the set of two houses where the two little pigs are hiding has now become a reference to “la casa”/the house, singular definite. There are two possibilities now: either the system might look for individual satisfying the description and find the house belonging to the elder brother, Jimmy; or it might take the most recent set satisfying the description. And this is what it does.

Esso si mise ad osservare attentamente la casetta e notò che non era davvero molto solida.

loc(infon497, id101, [arg:main\_tloc, arg:tes(f8\_po13)])

sit(id102, osservare, [esperiente:id9, tema\_nonaff:id35], 1, tes(finfl\_po13), id2)

sit(infon501, isa, [arg:id102, arg:st], 1, tes(finfl\_po13), id2)

sit(infon505, attentamente, [arg:id102], 1, tes(finfl\_po13), id2)

```
fact(id104, mettersi, [actor:id9, prop:id102], 1, tes(f5_po13), id2)
fact(infon510, solido, [nil:id35], 0, id101, id2)
fact(id106, essere, [prop:infon510], 0, tes(f7_po13), id2)
fact(id108, notare, [esperiente:id9, prop:id106], 1, tes(f8_po13), id2)
```

The attribute “solido”/solid is now predicated rightly of the two little houses. Moreover, in the following utterance, the (set) of (two) little house(s) gets destroyed by the wolf’s powerful blow. No new entity is being introduced in the following utterance.

```
Spaventatissimi i due porcellini corsero a perdifiato verso la casetta del fratello.
fact(id117, correre, [agente:id32, loc_direct:id25], 1, tes(f5_po15), id2)
fact(infon575, isa, [arg:id119, arg:perdifiato], 1, tes(f5_po15), id2)
fact(infon576, a, [arg:id117, nil:id119], 1, tes(f5_po15), id2)
fact(infon577, spaventatissimo, [arg:id117], 1, tes(f5_po15), id2)
included(tr(f5_po15), id110)
```

In order to understand that the little house being referred to is the one belonging to Jimmy, the system looks for an individual, which is a house, and has a possessor, which is again an individual and is a brother: it captures the relative identifier, id24. However, the Main Location has not changed yet, owing to the fact that a semantic role like Loc\_direct, i.e. a directional location, is not taken as a valid main location.

```
“Presto, fratellino, aprici! Abbiamo il lupo alle calcagna”.
loc(infon584, id120, [arg:main_tloc, arg:tes(f5_po15)])
fact(infon585, inst_of, [ind:id24, class:animale], 1, univ, univ)
fact(id121, avere, [esperiente:id32, tema_bound:id9], 1, tes(f4_po16), id2)
fact(id123, aprire, [agente:id24, goal:id32], 1, tes(f2_po16), id2)
```

The appearance of direct speech within the story requires the system to compute the reference of first and second person pronouns to previous participants in the story, which however were introduced by third person. Further problems are constituted by the reference to a singular little brother as “fratellino”, which should be understood as the wise little pig Jimmy: this is computed by the existence of an individual in the world associated to Jimmy, as opposed to the presence of a set associated to the remaining two little pigs, as discussed above.

```
Fecero appena in tempo ad entrare e tirare il chiavistello.
ind(infon602, id125)
fact(infon603, isa, [ind:id125, class:chiavistello], 1, id120, id25)
fact(infon604, part_of, [casa, id25, id125], 1, id120, id25)
fact(id126, entrare, [agente:id32], 1, tes(fin1_po17), id25)
fact(id128, fare_tempo, [agente:id32, prop:id126], 1, tes(f9_po17), id25)
```

fact(id130, tirare, [agente:id32, tema\_aff:id125], 1, tes(fin4\_po17), id25)  
fact(id128, fare\_tempo, [agente:id32, prop:id126], 1, tes(f9\_po17), id25)

It is interesting to notice the appearance of the *part\_of* relation as the singular definite NP “il chiavistello” is computed as being a part of the house: the inference is based on a constraint based on the spatial location of the situation the definite NP is contained. Also notice that semantic interpretation transforms a coordinate structure into a double relation holding in the DM; the main predicate has also undergone a semantic transformation, from “fare” into “fare\_tempo”. The same happens in the following utterance where the singular definite NP “pranzetto” is computed by an inference as coreferring to any of the little pigs, since it corefers with “pasto” above.

Il lupo stava già arrivando deciso a non rinunciare al suo pranzetto.  
fact(id132, arrivare, [tema\_aff:id9], 1, tes(f5\_po18), id2)  
sit(id135, rinunciare, [actor:id9, tema\_aff:id67], 1, tes(fin1\_po18), id2)  
fact(infon655, not, [prop:id135], 1, tes(f5\_po18), id2)  
fact(infon656, deciso, [arg:id132, nil:id9, prop:infon655], 1, tes(f5\_po18), id2)

In the previous utterance no new individual was introduced and the wolf was the main character being in focus. In the following utterance [non ci fu niente da fare], an impersonal sentence, the wolf is still taken to be the focussed character binding the big PRO subject of the infinitival “fare”.

I tre porcellini si sentivano al sicuro nella solida casetta di mattoni.  
loc(infon815, id165, [arg:main\_tloc, arg:tes(f6\_po22)])  
fact(infon817, sicuro, [nil:id3], 1, id165, id25)  
fact(id166, sentirsi, [esperiente:id3, prop:infon817], 1, tes(f6\_po23), id25)  
fact(infon818, isa, [arg:id167, arg:mattone], 1, tes(f6\_po23), id25)  
fact(infon819, di, [arg:id25, matter:id167], 1, tes(f6\_po23), id25)  
fact(infon820, in, [arg:id165, nil:id25], 1, tes(f6\_po23), id25)

Notice that at this point of the story coreference with the three little pigs is reestablished explicitly by indicating cardinality: in this way the system is able to extract the right ID. Problems arise with the definite singular NP “la solida casetta di mattoni”: of course we know that has been built “con”/with, by means of “mattoni”/bricks and “cemento”/cement. An inference is fired in order to ascertain whether these two properties can have the same meaning.

Riconoscenti i due porcellini oziosi promisero al fratello che da quel giorno anche essi avrebbero lavorato sodo.  
sit(id169, lavorare, [agente:id32], 1, tes(f246\_po24), id25)  
sit(infon841, anche, [arg:id169], 1, tes(f246\_po24), id25)

```

sit(infon842, sodo, [arg:id169], 1, tes(f246_po24), id25)
fact(infon843, isa, [arg:id171, arg:giorno], 1, tes(f246_po24), id25)
sit(infon844, da, [arg:id169, nil:id171], 1, tes(f246_po24), id25)
fact(id172, promettere, [actor:id32, tema_nonaff:id24, propint:id169], 1, tes(f247_po24), id25)
fact(infon849, riconoscente, [arg:id172], 1, tes(f247_po24), id25)

```

However, the final utterance of the story mentions the set of lazy little pigs again by indicating explicitly its cardinality. As we already said, at the end of the computation the discourse model algorithm automatically generates the rank list of the entities of the model or world of the story according to the relevance score: we compute this score by taking into account the topicality of each entity, i.e. the type of Topic it represents in the story and the times it is so regarded by the system. There is a separate entry for each SET, INDIVIDUAL or CLASS where hierarchically all properties and relations the story has ascribed to it are inherited.

## 9.2 Story 2

As we said in the introduction, this story is just a small excerpt from a newspaper article which deals with politics, and in particular with a man called Avveduti who was implied in a number of political scandals. The first sentence of the story has been added in order to provide the rest of it with some context, which otherwise would have been lacking. The portion of the text that we chose is remarkable as far as textual cohesion mechanisms are concerned: the use of pronouns is very extended and the discourse module captures all the relevant passages in the right way. The model representation is enriched here as to relations, in that it assigns a restriction to them in terms of event or state. This information is important for the temporal interpretation which however is not represented in the book.

Fino a tre anni fa Franco Avveduti non si era mai immischiato col mondo della pubblica amministrazione.

```

loc(infon3, id1, [arg:main_tloc, arg:anno])
ind(infon4, id2)
fact(infon5, surname, [franco, id2], 1, id1, univ)
fact(infon6, inst_of, [ind:id2, class:uomo], 1, univ, univ)
fact(infon7, name, [avveduti, id2], 1, univ, univ)
fact(infon22, isa, [arg:id3, arg:mondo], 1, id1, univ)
fact(infon23, isa, [arg:id4, arg:amministrazione], 1, id1, univ)
fact(infon24, pubblico, [arg:id4], 1, id1, univ)
fact(infon25, di, [arg:id3, specific:id4], 1, id1, univ)
fact(id5, immischiare, [agente:id2, tema_aff:id2, comitat:id3], 0, tes(f3_avv01), univ)
fact(infon30, isa, [arg:id1, arg:anno], 1, tes(f3_avv01), univ)
fact(infon31, fino_fa, [arg:id5, nil:id1], 1, tes(f3_avv01), univ)

```

Come burocrate, era un immigrato che veniva dal di fuori.

```
loc(infon42, id7, [arg:main_tloc, arg:tes(f3_avv01)])
loc(infon43, id8, [arg:main_sloc, arg:fuori])
ind(infon44, id9)
fact(infon45, burocrate, [ind:id9], 1, id7, id8)
fact(infon46, role, [immigrato, id9, id2], 1, id7, id8)
fact(infon47, inst_of, [ind:id9, class:[ind, real, soc, role]], 1, univ, univ)
fact(infon48, isa, [ind:id9, class:immigrato], 1, id7, id8)
fact(infon49, immigrato, [nil:id2], 1, id7, id8)
class(infon50, id10)
fact(infon51, inst_of, [ind:id10, class:uomo], 1, univ, univ)
fact(infon52, isa, [ind:id10, class:uomo], 1, univ, univ)
fact(id11, essere, [prop:infon49], 1, tes(f4_avv02), id8)
```

Figlio di buona famiglia, a venti anni decise di iscriversi alla Accademia Militare di Cavalleria.

```
loc(infon71, id13, [arg:main_tloc, arg:anno])
loc(infon72, id14, [arg:main_sloc, arg:accademia])
fact(infon74, isa, [arg:id14, arg:accademia], 1, id13, id14)
fact(infon75, militare, [arg:id14], 1, id13, id14)
fact(infon76, isa, [arg:id15, arg:cavalleria], 1, id13, id14)
fact(infon77, di, [arg:id14, specific:id15], 1, id13, id14)
fact(id16, iscrivere, [actor:id2, tema_unaff:id2, goal:id14], 1, tes(finfl_free_avv03), id14)
fact(id18, decidere, [actor:id2, prop:id16], 1, tes(f3_free_avv03), id14)
```

[era un buon allievo con ottime qualifiche]

```
loc(infon97, id20, [arg:main_tloc, arg:tes(f3_free_avv03)])
fact(infon98, allievo, [nil:id2], 1, id20, id14)
fact(infon99, buon, [ind:id2], 1, id20, id14)
fact(id21, essere, [prop:infon98], 1, tes(f3_avv04), id14)
```

Più tardi fu un ufficiale di successo.

```
loc(infon113, id23, [arg:main_tloc, arg:piu_tardi])
ind(infon114, id24)
ind(infon115, id25)
fact(infon116, inst_of, [ind:id25, class:[ind, abst, soc, state]], 1, univ, univ)
fact(infon117, isa, [ind:id25, class:successo], 1, id23, id14)
fact(infon118, di, [arg:id24, specif:id25], 1, id23, id14)
fact(infon119, role, [ufficiale, id24, id2], 1, id23, id14)
fact(infon120, inst_of, [ind:id24, class:militare], 1, univ, univ)
fact(infon121, isa, [ind:id24, class:ufficiale], 1, id23, id14)
fact(infon122, ufficiale, [nil:id2], 1, id23, id14)
fact(id26, essere, [prop:infon122], 1, tes(f3_avv05), id14)
```

At this point of the story we have a certain number of properties and relation being predicated of the main character Franco Avveduti, who has been assigned the semantic constant identifier id2. It is important now to notice that in the latest utterance we learn that he was also an officer. This new property however is indirectly negated in the following utterance, where we are told that Avveduti “si dimise”/resigned “dall’esercito”/from the army. No indication is directly made to the role of officer which must be indirectly inferred. A number of interesting inferences are realized in the following utterances and are all based on the collective noun “esercito”/army. They are all discussed in the text above.

```
Poi nel 1945 Avveduti si dimise dall' esercito.
loc(infon136, id28, [arg:main_tloc, arg:tes(f3_free_avv03)])
loc(infon137, id29, [arg:main_sloc, arg:esercito])
set(infon138, id30)
card(infon139, id30, >100)
fact(infon140, inst_of, [ind:id30, class:militare], 1, univ, univ)
fact(infon141, isa, [ind:id30, class:esercito], 1, id28, id29)
in(infon142, id24, id30)
fact(infon143, role, [ufficiale, id24, id2], 0, id28, id29)
fact(id31, dimettere, [actor:id2, actee:id2, locativo:id30], 1, tes(f4_avv06), id29)
fact(infon150, isa, [arg:id33, arg:1945], 1, tes(f4_avv06), id29)
fact(infon151, in, [arg:id31, nil:id33], 1, tes(f4_avv06), id29)
fact(infon152, adj, [temp_seq:id31], 1, id28, id29)
fact(infon153, poi, [prop:infon152], 1, id28, id29)
```

In the following utterance “i militari” a plural definite NP is equalled to the “army” seen that the latter is made up of “militari”, as the default class inclusion clearly testifies. So no new entity is asserted in the DM.

```
I militari lo avevano deluso.
loc(infon163, id34, [arg:main_tloc, arg:tes(f4_avv06)])
fact(id35, deludere, [causer_emot:id30, esperiente:id2], 1, tes(f3_avv07), id29)
```

In the following utterance, however a metonymic coreference is realized: “divisa”/uniform is made to refer to the “army”, where such clothes are used by “militari”. The solution we chose is thus to fire a part\_of assertion holding between “uniform” and “army”.

```
Avveduti, deposta la divisa, si iscrisse alla università.
class(infon173, id37)
fact(infon174, inst_of, [ind:id37, class:cosa], 1, univ, univ)
fact(infon175, isa, [ind:id37, class:divisa], 1, id34, id29)
fact(infon176, part_of, [divisa, id37, id30], 1, id34, id29)
```



fact(infon178, isa, [arg:id38, arg:universita\_], 1, id34, id29)  
fact(id39, iscrivere, [actor:id2, tema\_unaff:id2, goal:id38], 1, tes(f3\_avv08), id29)

Nel 1947 era laureato e nel 1948 era procuratore legale.

ind(infon201, id43)  
fact(infon202, role, [laureato, id43, id2], 1, id34, id29)  
fact(infon203, inst\_of, [ind:id43, class:[ind, real, soc, role]], 1, univ, univ)  
fact(infon204, isa, [ind:id43, class:laureato], 1, id34, id29)  
fact(infon205, laureato, [nil:id2], 1, id34, id29)  
ind(infon206, id44)  
fact(infon207, legale, [ind:id44], 1, id34, id29)  
fact(infon208, role, [procuratore, id44, id2], 1, id34, id29)  
fact(infon209, inst\_of, [ind:id44, class:[ind, real, soc, role]], 1, univ, univ)  
fact(infon210, isa, [ind:id44, class:procuratore], 1, id34, id29)  
fact(infon211, procuratore, [nil:id2], 1, id34, id29)  
fact(id45, essere, [prop:infon205], 1, tes(f3\_avv09), id29)  
fact(infon218, isa, [arg:id47, arg:1947], 1, tes(f3\_avv09), id29)  
fact(infon219, in, [arg:id45, nil:id47], 1, tes(f3\_avv09), id29)  
fact(id48, essere, [prop:infon211], 1, tes(f5\_avv09), id29)  
fact(infon225, isa, [arg:id50, arg:1948], 1, tes(f5\_avv09), id29)  
fact(infon226, in, [arg:id48, nil:id50], 1, tes(f5\_avv09), id29)

New properties are introduced in the previous and the following utterance. However, the most interesting ones are the latter properties because they allow the system to realize the meaning postulate associated to “father-in-law”. In the following utterance we are told that Avveduti met and married Paola, daughter of Antonio Alberti, important senator belonging to the political party of the Christian Democrats. All these information are needed in the following text to derive the coreference for the definite singular NP “il suocero”/the father-in-law.

Intanto a Verona aveva conosciuto Paola, figlia di Antonio Alberti, potente senatore democristiano, e la aveva sposata.

loc(infon239, id51, [arg:main\_sloc, arg:verona])  
ind(infon240, id52)  
ind(infon241, id53)  
fact(infon242, surname, [antonio, id53], 1, id34, id51)  
fact(infon243, inst\_of, [ind:id53, class:uomo], 1, univ, univ)  
fact(infon244, name, [alberti, id53], 1, univ, univ)  
fact(infon245, relat, [figlia, id52, id53], 1, id34, id51)  
fact(infon246, inst\_of, [ind:id52, class:[ind, real, nat, relat]], 1, univ, univ)  
fact(infon247, isa, [ind:id52, class:figlia], 1, id34, id51)  
fact(infon248, inst\_of, [ind:id52, class:[ind, real, nat, hum]], 1, univ, univ)  
fact(infon249, name, [paola, id52], 1, univ, univ)

```

fact(infon252, isa, [ind:id53, class:senatore], 1, id34, id51)
fact(infon253, democristiano, [ind:id53], 1, id34, id51)
fact(infon254, potente, [ind:id53], 1, id34, id51)
ind(infon257, id55)
fact(infon258, inst_of, [ind:id55, class:[luogo]], 1, univ, univ)
fact(infon259, name, [verona, id55], 1, univ, univ)
fact(id56, conoscere, [esperiente:id2, tema_emot:id52], 1, tes(f3_avv10), id51)
fact(infon268, a, [arg:id56, nil:id55], 1, tes(f3_avv10), id51)
fact(id58, sposare, [agente:id2, tema_aff:id52], 1, tes(f5_avv10), id51)

```

A number of utterances follow which however do not modify the DM. The following utterance of the text introduces “suocero”/father-in-law which is computed as such by means of inferential processes based on previous knowledge available in the DM.

Gli piaceva parlare del suocero come di una facile occasione mancata che chiunque altro avrebbe sfruttato ma che lui, Avveduti, preferiva lasciare perdere.

```

fact(infon379, isa, [ind:id53, class:suocero], 1, id1, nil)
fact(infon380, relat, [suocero, id53, id2], 1, id1, nil)
ind(infon383, id81)
fact(infon384, facile, [ind:id81], 1, id80, id14)
fact(infon385, mancato, [ind:id81], 1, id80, id14)
fact(infon386, isa, [ind:id81, class:occasione], 1, id80, id14)
fact(infon387, inst_of, [ind:id81, class:[evento]], 1, univ, univ)
class(infon388, id82)
fact(infon389, inst_of, [ind:id82, class:uomo], 1, univ, univ)
fact(infon390, isa, [ind:id82, class:uomo], 1, univ, univ)
fact(id83, sfruttare, [agente:id82, tema_aff:id81], 1, tes(f3_p01), id14)
fact(id87, perdere, [actor:id2, tema_nonaff:id81], 1, tes(fin_f3_p01), id14)
fact(id89, lasciare, [actor:id2, prop:id87], 1, tes(fin_f2_p01), id14)
fact(id91, preferire, [actor:id2, prop:id89], 1, tes(f4_p01), id14)
fact(id93, parlare, [actor:id2, subj_dis:id53], 1, tes(fin_f1_p01), id14)
fact(id95, piacere, [propint:id93, esperiente:id2], 1, tes(f5_p01), id14)

```

In the third utterance a new individual is created, which is a “posto”/position and is treated as a sit rather than a fact because it is the argument of a [-tense] predicate in turn an open complement of a higher predicate “decidere”/decide which produces opacity for its arguments.

Solo verso il 1950 decise di accettare un posto nella organizzazione della Fiera di Verona.

```

loc(infon448, id97, [arg:main_tloc, arg:1950])
loc(infon449, id98, [arg:main_sloc, arg:verona])
ind(infon450, id99)

```

```

ind(infon451, id100)
ind(infon452, id101)
ind(infon453, id102)
fact(infon454, inst_of, [ind:id102, class:[luogo]], 1, univ, univ)
fact(infon455, isa, [ind:id102, class:verona], 1, id97, id98)
fact(infon456, di, [arg:id101, specif:id102], 1, id97, id98)
fact(infon457, inst_of, [ind:id101, class:[coll, real, soc, event]], 1, univ, univ)
fact(infon458, isa, [ind:id101, class:fiera], 1, id97, id98)
fact(infon459, has, [arg:id100, tema:id101], 1, id97, id98)
fact(infon460, inst_of, [ind:id100, class:istituzione], 1, univ, univ)
fact(infon461, isa, [ind:id100, class:organizzazione], 1, id97, id98)
fact(infon462, in, [arg:id99, locativo:id100], 1, id97, id98)
sit(infon463, isa, [ind:id99, class:posto], 1, id97, id98)
fact(infon464, inst_of, [ind:id99, class:[attivit a, luogo, stato]], 1, univ, univ)
fact(id103, accettare, [actor:id2, tema_nonaff:id99], 1, tes(finf1_p02), id98)
fact(id105, decidere, [actor:id2, prop:id103], 1, tes(f3_p02), id98)
fact(infon480, verso, [arg:id105, nil:id107], 1, tes(f3_p02), id98)

```

In the fourth utterance there is a long list of predicates and arguments which however do not constitute legal IDs, being either plural or singular NPs with no co-referent in the world of the model nor in the external knowledge network. We only know now that a number of properties are associated to the main topic with a number of restrictions.

Lo nominarono delegato cio e una specie di funzionario viaggiante con incarichi diplomatici di tenere i rapporti con le delegazioni commerciali, curare i produttori stranieri, le grandi ditte, la stampa.

```

loc(infon493, id108, [arg:main_tloc, arg:tes(f3_p02)])
ind(infon494, id109)
fact(infon495, role, [delegato, id109, id2], 1, id108, id98)
fact(infon496, inst_of, [ind:id109, class:uomo], 1, univ, univ)
fact(infon497, isa, [ind:id109, class:delegato], 1, id108, id98)
fact(infon498, delegato, [nil:id2], 1, id108, id98)
class(infon499, id110)
fact(infon510, viaggiante, [ind:id110], 1, id108, id98)
fact(infon512, inst_of, [ind:id112, class:uomo], 1, univ, univ)
sit(infon513, isa, [ind:id112, class:funzionario], 1, id108, id98)
sit(infon514, kind_of, [ind:id109, class:funzionario], 1, id108, id98)
fact(id113, nominare, [agente:id100, prop:infon498], 1, tes(f4_avv15), id98)

```

No new ID is created in this utterance and in the following one. However, we should note the special function devoted to the indefinite NP “un compito/a task” in a predicative function, which is translated by the system as a property ascribed to

an entity coreferred by the deictic “this/questo” subject of the predication. The entity is recovered by means of the special semantical restriction associated to the noun “task”, which is classified as “activity” in the lexicon. Now the semantic interpreter understands an activity to be coreferable to a “social role”: and the one available in the previous portion of text is the “delegato/delegate”, which is picked up as coreferent for “this”.

Questo era un compito che corrispondeva bene alla sua vocazione e nel quale avveduti sapeva giostrare con, notevole agilità.

loc(infon530, id115, [arg:main\_tloc, arg:tes(ff6\_p03)])

ind(infon531, id116)

fact(infon532, questo, [nil:id116], 1, id115, id98)

fact(infon533, isa, [ind:id116, class:compito], 1, id115, id98)

fact(infon534, isa, [compito, id109], 1, id115, id98)

fact(infon535, isa, [arg:id117, arg:vocazione], 1, id115, id98)

fact(infon536, poss, [arg:id117, poss:id2], 1, id115, id98)

fact(id118, corrispondere, [tema\_bound:id116, tema\_nonaff:id117], 1, tes(ff5\_p04), id98)

fact(infon538, bene, [arg:id118], 1, tes(ff5\_p04), id98)

sit(id120, giostrare, [agente:id2, modale:id122], 1, tes(finff1\_p04), id98)

fact(infon540, isa, [arg:id122, arg:agilità], 1, tes(finff1\_p04), id98)

fact(infon541, notevole, [arg:id122], 1, tes(finff1\_p04), id98)

sit(infon542, con, [arg:id120, nil:id122], 1, tes(finff1\_p04), id98)

fact(id123, sapere, [actor:id2, prop:id120], 1, tes(ff6\_p04), id98)

fact(id125, essere, [prop:inffon532], 1, tes(ff7\_p04), id98)

In this utterance there is a new ID which is created and a number of individuals are also referred to by proper names. However no coreferent exists in the world so that they are simply dropped. In case we computed the whole of the story, that we would have known which person the proper name “Alberti” corefers to.

A Verona il collegio di Alberti lo aveva ereditato Trabucchi e col collegio aveva ereditato la presidenza della Fiera.

loc(infon585, id123, [arg:main\_tloc, arg:tes(ff6\_p05)])

ind(infon586, id124)

fact(infon587, inst\_of, [ind:id124, class:uomo], 1, univ, univ)

fact(infon588, name, [trabucchi, id124], 1, univ, univ)

fact(infon590, isa, [arg:id125, arg:collegio], 1, id123, id98)

fact(id126, ereditare, [receiver:id124, tema\_nonaff:id125], 1, tes(ff3\_p06), id98)

fact(infon596, a, [arg:id126, nil:id55], 1, tes(ff3\_p06), id98)

fact(infon597, isa, [arg:id129, arg:collegio], 1, id123, id98)

fact(infon598, con, [arg:id128, nil:id124, comitativo:id129], 1, id123, id98)

fact(infon599, isa, [arg:id130, arg:presidenza], 1, id123, id98)

fact(id131, ereditare, [receiver:id124, tema\_nonaff:id130], 1, tes(ff5\_p06), id98)

fact(infon604, isa, [arg:id129, arg:collegio], 1, tes(f5\_p06), id98)  
fact(infon605, con, [arg:id131, nil:id124, comitativo:id129], 1, tes(f5\_p06), id98)

No new entity is introduced in this utterance: rather, the story now has two protagonists which are continually coreferred by the use of pronouns.

Trabucchi continuò a valersi della collaborazione di Avveduti.  
fact(infon625, isa, [arg:id133, arg:collaborazione], 1, id123, id98)  
fact(id134, valersi, [esperiente:id124, tema\_nonaff:id133], 1, tes(finfl\_p07), id98)  
fact(id136, continuare, [agente:id124, prop:id134], 1, tes(f3\_p07), id98)

In this utterance a reference to the main topic, Avveduti, is realized by means of a compound name, “ex-ufficiale”/ex-officer, where the prefix indicates that the property of being an officer no longer holds for the referent, Avveduti. To compute the meaning of this noun, a complex inferential process should have been activated in case we had analysed the whole text: in particular we know at the beginning of the story that Avveduti gets enlisted as officer of cavalry. Then in the following text we also know that he decides to quit the army to become a lawyer. Reasoning processes would have analysed these events in such a way as to produce an interpretation for the prefix “ex”, indicating explicitly that the property denoted by the noun was previously predicated of the individual but that at a certain moment in time, preceding the actual time indicated by the discourse has been suspended.

L' ex ufficiale del Novara\_Cavalleria gli era simpatico.  
loc(infon643, id138, [arg:main\_tloc, arg:tes(f3\_p07)])  
fact(infon644, simpatico, [nil:id2], 1, id138, id98)  
fact(infon649, isa, [arg:id140, arg:novara\_cavalleria], 1, id138, id98)  
fact(infon650, di, [arg:id2, specif:id140], 1, id138, id98)  
fact(id141, essere, [prop:infon644], 1, tes(f3\_avv20), id98)

In this utterance and the following one no new ID is created.

La sua distinzione lo impressionava.  
fact(infon665, isa, [arg:id143, arg:distinzione], 1, id138, id98)  
fact(infon667, poss, [arg:id143, poss:id2], 1, id138, id98)  
fact(id144, impressionare, [causer\_emot:id143, tema\_emot:id124], 1, tes(f3\_p09), id98)

Consider now the inference required to maintain the same location: we are faced with a property, “fiera/fair” which is also a location, however we do not want to assert a new location since the fair we are talking about is just a place in Verona. Inferences are activated on the model on the basis of a specification role associated with the id of the fair.

Lo confermò nello incarico alla Fiera.

```
ind(infon681, id146)
fact(infon682, isa, [ind:id146, class:incarico], 1, id138, id98)
fact(infon345, isa, [incarico, id109], 1, id138, id98)
fact(id148, confermare, [esperiente:id124, prop:id2], 1, tes(f6_p10), id98)
```

In this utterance we know that Avveduti was working as special secretary of the ministry Trabucchi, something which did not appear in the previous text.

Avveduti funzionava benissimo come segretario particolare.

```
loc(infon697, id150, [arg:main_tloc, arg:tes(f4_p10)])
ind(infon698, id151)
fact(infon699, particolare, [ind:id151], 1, id150, id98)
fact(infon700, role, [segretario, id151, id2], 1, id150, id98)
fact(infon701, inst_of, [ind:id151, class:uomo], 1, univ, univ)
fact(infon702, isa, [ind:id151, class:segretario], 1, id150, id98)
fact(infon703, segretario, [nil:id2], 1, id150, id98)
fact(id153, funzionario, [prop:infon703], 1, tes(f3_p11), id98)
fact(infon714, bene, [arg:id153], 1, tes(f3_p11), id98)
```

In this utterance we are informed about Avveduti's skill to contact important people which however are introduced in the plural as bare generic NPs. In the model, the relation interpreter creates a set of individuals which includes both reference to classes. Also notice that default properties are augmented by the additional information that they can be referred as "uomo"/man.

Sapeva mobilitare prefetti e questori.

```
class(infon725, id155)
fact(infon728, isa, [ind:id155, class:questore], 1, id150, id98)
fact(infon730, inst_of, [ind:id155, class:uomo], 1, univ, univ)
fact(infon731, isa, [ind:id155, class:uomo], 1, univ, univ)
class(infon732, id156)
fact(infon735, isa, [ind:id156, class:prefetto], 1, id150, id98)
fact(infon737, inst_of, [ind:id156, class:uomo], 1, univ, univ)
fact(infon738, isa, [ind:id156, class:uomo], 1, univ, univ)
set(infon733, id159)
card(infon734, id159, >10)
in(infon736, id156, id159)
in(infon739, id155, id159)
fact(infon740, [questore, prefetto], [arg:id159], 1, id150, id98)
fact(id160, mobilitare, [agente:id2, prop:infon740], 1, tes(finfl_p12), id98)
fact(id162, sapere, [actor:id2, prop:id160], 1, tes(f3_p12), id98)
```

Finally, the last sentence is rather tricky because the reference to Avveduti is recovered via the property “segretario”/secretary which makes it possible to assign the right coreferent to the clitic pronoun “gli”/him and the possessor “suo”/his, i.e. Trabucchi.

Tutti gli invidiavano il suo segretario particolare.

class(infon757, id164)

card(infon758, id164, all)

fact(infon759, isa, [arg:id164, class:umano], 1,id150, id98)

fact(id165, invidiare, [esperiente:id164, tema\_nonaff:id2, actor:id124], 1, tes(f5\_p13), id98)

## 10. REFERENCES

- Abney A. (1996). *Part-of-Speech Tagging and Partial Parsing*. In K. Church et al. (eds). *Corpus-Based Methods in Language and Speech*. Dordrecht, Kluwer Academic Publishers.
- Bianchi D., Delmonte R. (1999). *Reasoning with A Discourse Model and Conceptual Representations*. Proceeding of VEXTAL, pp.401-411. Unipress.
- Bianchi D., Delmonte R. (1997). *Rappresentazioni concettuali nella comprensione di storie*. In *Atti Apprendimento Automatico e Linguaggio Naturale*, pp. 95-98. Torino.
- Bianchi D., Delmonte R. (1996). *Temporal Logic in Sentence and discourse*. Proceeding of SIMAI'96, pp. 226-228.
- Bianchi D., Delmonte R., Pianta E. (1993). *Understanding Stories in Different Languages with GETA\_RUN*. Proceeding of EC of ACL, p. 464. Utrecht.
- Bianchi D., Delmonte R., Pianta E., Sartori S. (1993). *Reference Resolution by Contextual Reasoning from Grammatical Representations with GETA\_RUN*. University of Venice Working Papers in Linguistics, Vol. 3, n. 1, pp. 31-68.
- Bresnan J. (2001). *Lexical-Functional Syntax*. Oxford, Blackwell.
- Delmonte R. (1993). *GETA\_RUN: A fully integrated system for Reference Resolution by Contextual Reasoning from Grammatical Representations*. ACL-93, p. 2. Columbus, Exhibitions and Demonstrations.
- Delmonte R. (1994). *Inferences and Discourse Structure*. In G. Ferrari (ed.). *Proceeding of IV Conference of the Italian Artificial Intelligence Association – AIIA*, pp. 11-14.
- Delmonte R. (1995a). *Lexical Representations: Syntax-Semantics interface and World Knowledge*. Notiziario AIIA (Associazione Italiana di Intelligenza Artificiale), pp. 11-16. Roma.
- Delmonte R. (1995b). *Reference Resolution and SpatioTemporal Locations. Presented at Eurolan '95, Second Euro-Romanian Summer School on Natural Language Processing, "Language and Perception – Representations and Processes"*. MS.
- Delmonte R. (1995c). *Understanding texts in different languages with Geta\_Run*. Proceeding of JADT'95, pp. 279-286.

- Delmonte R. (1996). *Contextual Reasoning and Inferential Processing*. Proceeding of SIMAI'96, pp. 229-231.
- Delmonte R. (1997). *Lexical Representations, Event Structure and Quantification*. Quaderni Patavini di Linguistica, Vol. 15, pp. 39-93.
- Delmonte R. (1997). *Rappresentazioni lessicali e linguistica computazionale*. In Atti SLI, Lessico e Grammatica. Teorie Linguistiche e applicazioni lessicografiche, pp. 431-462. Roma, Bulzoni.
- Delmonte R. (2000). *Generating from a Discourse Model*. Proceeding of MT-2000, pp. 25-1/10. Exeter (UK), BCS.
- Delmonte R. (2002). *GETARUN PARSER – A parser equipped with Quantifier Raising and Anaphoric Binding based on LFG*. Proceeding of LFG2002 Conference, pp. 130-153. Athens.
- Delmonte R. (2003a). *Getaruns: a hybrid system for summarization and question answering*. Proceeding of EAACL 2003 Workshop “NLP for Question Answering”, pp. 21-28. Budapest.
- Delmonte R. (2003b). *The Semantic Web Needs Anaphora Resolution*. Proceeding of Workshop ARQAS 2003, International Symposium on Reference Resolution and Its Applications to Q/A and Summarization, pp. 25-32. Venice, Ca' Foscari University.
- Delmonte R., Bianchi D. (1992). *Quantifiers in Discourse*. Proceeding of ALLC/ACH'92, pp. 107-114. Oxford (UK), OUP.
- Delmonte R., Bianchi D. (1994). *Computing Discourse Anaphora from Grammatical Representation*. In D. Ross, D. Brink (eds). Research in Humanities Computing 3. Oxford, Clarendon Press, pp. 179-199.
- Delmonte R., Bianchi D. (1999). *Determining Essential Properties of Linguistic Objects for Unrestricted Text Anaphora Resolution*. Proceeding of Workshop on Procedures in Discourse, pp. 10-24. Pisa.
- Delmonte R., Bianchi D. (2002). *From Deep to Partial Understanding with GETARUNS*. Proceeding of ROMAND 2002, pp. 57-71. Roma, Università Roma2.
- Delmonte R., Bianchi D., Pianta E. (1992). *GETA\_RUN – A General Text Analyzer with Reference Understanding*. Proceeding of 3rd Conference on Applied Natural Language Processing, Systems Demonstrations, pp. 9-10. Trento, ACL.
- Delmonte R., Dibattista D. (1995b). *Switching from Narrative to Legal Genre*. Working Papers in Linguistics, Vol. 5, n. 1, pp. 1-41. University of Venice.
- Delmonte R., Pianta E. (1994). *Discourse Structure and Linguistic Information*. ACH/ALLC '94, pp. 61-62. Paris, Consensus Ex Machina.
- Delmonte R., Pianta, E. (1996). *IMMORTALE – Analizzatore Morfologico, Tagger e Lemmatizzatore per l'Italiano*. In Atti del V Convegno AI\*IA “Cibernetica e Machine Learning”, pp. 19-22. Napoli.
- Dibattista D., Pianta E., Delmonte R. (1999). *Parsing and Interpreting Quantifiers with GETARUN*. Proceeding of VEXTAL, pp. 215-225. Unipress.
- Hirschman L. et al. (1999). *Deep Read: A reading comprehension system*. Proceeding of ACL '99. University of Maryland, College Park MD.



<http://csli-publications.stanford.edu/hand/miscpubsonline.html>.

Preiss J. (2003). *Using Grammatical Relations to Compare Parsers*. Proceeding of EACL.03, pp. 291-298. Budapest.

# Multilingualism and the Web

*Elisabeth Burr*

Universität Bremen

(elisabeth.burr@uni-bremen.de)

*Only a few years ago, the new technologies in general and the web in particular seemed to constitute a danger for most languages of the world. In the meantime, however, it would seem that quite the opposite is the case and that the new technologies favour multilingualism, the deconstruction of norms and of linguistic institutions, and the formation of hybrid languages. In fact, not only has the web itself become a multilingual environment but also minority languages and local dialects are migrating into this digital space. Furthermore, the (intentional) development of mixed languages even in environments dominated by writing would seem an indication of the weakening of the national and standard language paradigm, of the search for a we-code counterbalancing institutionalised they-codes and for a linguistic identity built on the multilingual knowledge of the speakers. In this contribution I arrive at the conclusion that the process from monolingualism to multilingualism and to hybridity as an important ingredient of digital culture a) presents linguistics with a unique possibility to study directly how the 'oral' is being transformed into the 'written', how norms evolve and spread, and from which tools depends the recognition of languages or dialects as modern communication systems, b) questions those schools of linguistics which have modelled language as a homogeneous entity, the ideal speaker-listener's competence as inherently monolingual, and have reduced linguistic variety to a common core, and c) constitutes a serious challenge for Humanities Computing and the developers of language technologies which, if taken up, will open up, finally, this new domain of research to the many Creoles and mixed languages spoken (and written) in the world.*

*Keywords:* multilingualism, new technologies, humanities computing

## 1. INTRODUCTION

Only a few years ago, the new technologies in general and the web in particular seemed to constitute a danger for most languages of the world. The parallels between:

- the invention of the printing press, which led to the normalisation of languages for economic reasons and caused the disappearance of language varieties and dialects, and the diffusion of the new technologies,

- the creation of hegemonic centralised European national powers, which forced the imposition of monolingualism, and the expanding hegemonic position of the US,
- the concept of *bon usage* or standard language much favoured by nation states and by their educational systems, and the concept of *centre* and *periphery* or homogeneous language communities favoured by modern linguistics, seemed, to indicate, in fact, that another detrimental process of linguistic homogenisation and normalisation was to be expected.

In the meantime, however, it would seem that quite the opposite is the case and that not monolingualism, homogenisation and normalisation are favoured by the new technologies but multilingualism, the deconstruction of norms and of linguistic institutions and the formation of hybrid languages.

## 2. FROM MONOLINGUALISM TO MULTILINGUALISM

The challenge which multilingualism constitutes in the digital world has not just been recognized by firms like Microsoft or Netscape which have developed and continue to develop language and culture sensitive versions of their browsers, but also by online marketing communications consultancies like the US based GlobalReach, whose aim it is to facilitate the access of transatlantic firms to the European market and who, as the following quotes from their web site show, are extremely conscious of the economic importance of languages in the digital market:

- Online marketing always takes place in the language of the target country, as any form of marketing does (GlobalReach, 2004d).
- People surf the Web in their own language: it is as much a part of their life in their own language as other media (TV, radio, printed publications, etc.) (GlobalReach, 2004e).
- Web marketing is global by nature, and requires a multi-lingual approach, since each market has to be addressed in its own language (GlobalReach, 2004e).

The rising market value of languages in the digital world is, however, only one of the aspects of the growing tendency towards multilingualism. Another aspect is the clear indication to be found on the web that the traditional view of multilingualism as representing a burden for humankind is substituted with the consciousness that multilingualism in itself constitutes a cultural value. Emblematic for this change in perspective is the story and the image of the Tower of Babel.

In fact, whereas the story of the Tower of Babel during the Middle Ages was used to explain linguistic diversity as being the result of the linguistic confusion God inflicted on humanity for questioning the divine hierarchies, and whereas later, during the phase of linguistic normalisation brought about by the advent of the printing

press and the creation of national states, the same story was regularly exploited by grammarians when justifying their proscription of variety and their establishment of a unique lexical and grammatical norm, in the last decade, on the contrary, Babel



and Pieter Bruegel's painting of the building of its tower has become one of the most frequently used symbols on the web when multilingualism, seen in terms of cultural richness, is at stake.

The building of the tower of Babel by Pieter Bruegel, 1563. Oil on oak panel, Kunsthistorisches Museum, Vienna (source: Lawrence 2004)

Furthermore, forecasts like

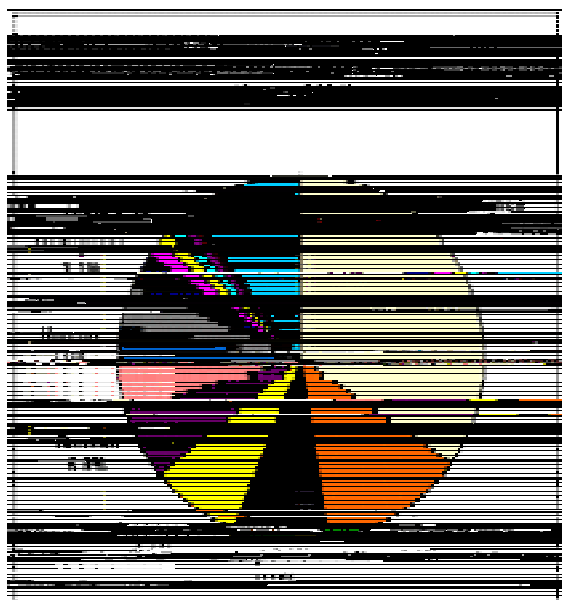
No speak English nella rete del 2000 (Mola, 1999)

La Babele delle lingue sopravvive in Rete (Retico, 2000)

have long become true and the web which, when invented in 1993 and when first known above all in the Anglo-American world and in regions like New Zealand, Australia and Canada was a truly monolingual, i.e. English resource, has itself become a multilingual environment.

## *2.1 National languages*

As the statistics of GlobalReach which has been systematically tracking the development of the online populations since 1996, show, in fact, since 1995 when the first non-English speaking countries connected to the Web, the percentage of the native English speaking online population has gone down continuously: from 84% in 1997 to 64.9% in 1999, to 47.5% in 2001, to 35.6 in September 2003. At present the relationship between native speakers of English and native speakers of other languages



on the Web seems to have stabilized at around 35.2% to 64.8% respectively. During the same period the percentage of English Web pages went down from about 100% to 68.4% (cfr. GlobalReach, 2004c)<sup>1</sup> and the number of languages on the Web has grown explosively.

(GlobalReach, 2004b)

## 2.2 *Minority languages and dialects*

Multilingualism on the Web does not mean, however, the presence of national languages alone. Instead, there are clear indications that living minority languages and dialects are also finding their way into the Web, a phenomenon which Italian newspapers thought to be worth reporting already quite a few years ago when articles like the following appeared:

- Altro che inglese: Internet ora parla napoletano (Ferraiulo, 1998)
- Proverbi e aforismi anche in dialetto (Mandò, 1998b)
- Minoranze linguistiche e dialetti in Rete (Mandò, 1998a)
- I dialetti italiani sopravvivono online (Benigni, 2000)

As newspaper articles as such are not a reliable enough proof that dialects and minority languages are really migrating into the Web, research into the presence of northern Italian dialects and minority languages in the digital world has been undertaken (cfr. Burr, 2004). This research has not only shown that there is such a migration but also that it happens on several different levels each of which is itself indicative for the value attributed to the linguistic entities by the individual people and / or the respective communities:

- 1) On a first level dialects appear in the Web exclusively in the form of their history and explanations of their particularities. Such pages are normally written in Italian (cfr. Associazione L.I.S.A., 2004).
- 2) Another way to give dialects a space in the digital world is to put collections of traditional texts like rhymes, popular sayings, proverbs in the respective dialect online together with an Italian translation (cfr. Giuse, 2003).
- 3) On a third level we can position sites which are technologically more advanced, i.e. where the multimedia possibilities offered by the Web are exploited in such a way that you can listen to the sayings and rhymes being spoken in the dialect (see for example the *filastrocche* in Bortoluzzi, Bolzan, 2000).
- 4) On a fourth level we find, instead, sites which do not just aim at the conservation of dialects as testimonies of a lost past or a dying culture such as the ones named so far, but which by exploiting the interactive possibilities offered by the Web make available tools like online dictionaries and forums or chat facilities for the active usage of dialects. One example is *dialettando.com*.

The most complex sites are found, however, at least at the moment, when it comes to minority languages. In fact, communities which traditionally have a strong feeling of cultural and linguistic identity and have long sought to gain recognition as autonomous entities have quickly recognized the possibilities offered by the new technologies for the affirmation of their identity and have thus been present on the Web already for quite some time now. In this case, too, we have to distinguish, however, between several types of communities.

There are first of all communities which exploit the possibilities offered by the new technologies more or less exclusively for the conservation of their cultural heritage. An example is the site of *Union Ladina del Cadore de Mèdo*, who, by presenting the wealth of historical and cultural information gained by ethnographic, sociological and linguistic research not in Ladinian, the traditional language of the region, but in Italian and by giving a voice exclusively to very old representatives of the Ladinian community, categorizes the minority language as belonging to the past. Another type of communities, on the contrary, with the help of the new technologies aims clearly at promoting its language as a modern means of information and communication. This is the case with *La Patrie dal Friûl*, a monthly journal of cultural and political information where Friulian is actively used. The importance which this community attributes to the present technological development for the affirmation of their own language in the modern world is furthermore shown by the fact that the site not only makes available an interactive Friulian-Italian dictionary but also an interactive multilingual dictionary, where the Italian, Slovene, German, English, Spanish and French correspondences of Friulian words, and vice versa, can be retrieved.

But the process of migration into the digital world does not stop with those communities who have always defended their identity. Instead, also communities who in the past have been much less active on a level which transcends their own locality are now discovering the possibilities of the web and are creating their own sites. In-

terestingly, newcomers like the Germanophone communities in northern Italy seem less governed by the wish to strengthen just their own identity, rather than by the intention to overcome at the same time the geographical borders which separate the individual communities. The site *Isole linguistiche* is a good example here.

These and many more indications that information and communication technologies do not really serve monolingualism and homogeneity but that they encourage, on the contrary, the recognition of the value of multilingualism and heterogeneity concern, however, only one aspect of the present development, i.e. the presence of more and more national, regional and minority languages, of individual language varieties and local dialects on the Web. Another aspect of the same development would seem to be the creation, affirmation and acceptance of mixed languages or hybrid varieties.

### 3. FROM MULTILINGUALISM TO HYBRIDITY

Lately, the crossing of language and dialect borders, the creation of multilingual spaces, the production of new identity patterns and the speaking of several languages at the same time, for which Jürgen Erfurt (2003: 6) has coined the term *Multisprech* (Multispeech) has more and more engaged the interest of socio- and applied linguists in Germany (cfr. Erfurt, 2003). Their studies are, however, mostly confined to spoken discourse in urban migration milieus. Yet, as other research has shown, phenomena of Multispeech are neither limited to urban migration milieus nor to the domains of spoken speech. Instead, it would seem that the new technologies are one of the factors which encourage the overcoming of certain norms and help the (intentional) development of mixed languages even in environments dominated by writing.

#### 3.1 SMS and Multispeech

The phenomenon and its inherent technological determination was first noted, as far as I can see, by Flavia Ursini (in print), when she studied two corpora of SMS written mainly by northern Italians. In these corpora, next to messages written wholly in the dialect of the region as in the following examples:

C1, 89: A MI PAR MI VA BEN, CASA TUA SOITA ORA.

C1, 95: oppo quoa ke te gao conta, pi gnente, ovio kendemo. Saeuti. M.

C2, 31: Son a B\*. Tren ritard. Un'oremèz daspetàr! Pazienz! Ma ce fredd! Sopporterem!

or messages where code-switching between northern Italian varieties and standard Italian takes place as in the following examples:

C1, 29: Dottore, si faccia vivo ogni tanto. So sempre mi che te gho da ciamare.

C1, 34: /.../ è da un bel pezzo che non ti fai sentire. Vien fora da che busa!

C2, 154: /.../ Mi tocca andare fino a M\*\*\*\*\* col trenino, e poi tornare da lì! Eccì! Desperassiùn! Eccì! Rafredùr! Eccì! Ciau!

C2, 198: Sper che tut vad ben (poveri bamb.). B\*\*\*\*\* mi ha detto di dirti di contattarlo. Ciao. G.

Ursini found, in fact, also quite a number of messages where the writers either mixed their own northern Italian dialect with central- and southern Italian dialects:

C1, 2: Non mi vhenire addire che anche questo sabato te ne vuoi stare acchasa a far finta di studiare!

C1, 7: non è checc'hai tèmpo di passà eddi vvenirmi a prèndere in macchina?

C2, 34: Ciò le foto! Sono 'ribbili (le mie). Le tue accusi accusi... (invidia...). Pa' i seminari te ciamme dopp!

C2, 107: Agge ditt'a 'o ccompiuto mio che lo duomilo è ggjà arrivate. Chillo m'ha ditto: "Fòttiti! A me nun me fregghi!" E io, che cce devo penzà? Ah, chiste tecnuologie!

or, and this is what interests most in our context, where a mixed language made up of different languages and dialects appears:

C1, 81: /.../ io non so gnancora perké dovaria studiar /.../ in + dovrei lavarmi i cavei. Fame savuar

C2, 62: Guten morgen! Son sul tren / per V\*\*\*\*\* / in un scompartiment / pien di Giap. Que ùrlan como napoletaneros! /.../

C2, 63: Sì, ce la fàrem! Je tèm! Moi superplus! /.../

C2, 65: (And your makinina? Como està? Y los peritos? Arriban or not arriban? Salùtamela, la makinina. Poarèta, qui sa como sofre...)

C2, 79: Ich bin bunn bang from zu art ort urt urt uqtt! Gutarbaiten! Nächtilüc henziertenrtar!

C2, 142: /.../ I have interest for una solla (très chère une, la ùnica querida)

C2, 570: The insalataro is open. I'm chez moi tra 10 minuti

That these phenomena have nothing to do with backwardness, uneducated speech or thoughtlessness is shown by the fact that the language used in the other messages written by the same people is the modern middle level Italian or *italiano dell'uso medio* and that in these messages its orthography is generally respected.

Ursini's conclusion that such a mixing of languages is indicative of a search for a language which is as far removed as possible from the standard language and comes as near as possible to the pole of informality (cfr. Ursini in print: 453) certainly captures an important aspect of this type of technology-mediated writing. I would, however, go further and interpret this type of linguistic practice as an indication of the weakening of the national and standard language paradigm, of the search for a *we-code* which counterbalances the *they-code* represented for most people in Italy by Italian and for a linguistic identity built on the multilingual knowledge of the speakers. Such a hypothesis can, in fact, be supported by research on other hybrid languages and above all on *Nouchi*.



### 3.2 Nouchi

Nouchi, according to Sabine Kube (2003) was originally created by youngsters in Abidjan (Côte d’Ivoire). In this multi-ethnic and multi-lingual city, French, the official language of the country, functions also as the *lingua franca*. The reason for this situation seems to be that given the very complex linguistic situation, none of the original languages managed to take on this function. This does, however, not mean that French is also attributed an identity-building function. Instead, this function is realised by Nouchi, a hybrid language composed of elements of French, of elements of the African languages present in Côte d’Ivoire and of neologisms invented by the people of Côte d’Ivoire themselves. Nouchi, the usage of which is no longer confined to oral domains but has become the language of music groups, artists and satirical weeklies, seems to constitute, in fact, the *we-code* which opposes the *they-code* represented by either the standard French language which is taught in the schools or by the French *lingua franca*. It is even said that Nouchi could well become the expression of Côte d’Ivoire’s identity altogether in the future (cf. Kube, 2003: 131-137).

### 3.3 Europanto

Similar things have been said more than once about Europanto, a mixed language ‘created’, or better, brought to light in 1996 by Diego Marani, an Italian translator who works in Brussels:

“Is this a new artificial language, an alternative to Esperanto? Or could it be the language of the future?” (Marani, 1999)

«L’Europanto: la langue des citoyens européens?» (Vandendooren without year)

«Europanto - Könnte das la lingua majeure for Europa para the nouvelle millennium sein?» (Karismo, 1999)

Notwithstanding the fact that on *The Linguist List* Europanto is seen to be closely related with mixed (artificial) languages like Esperanto, Interlingua, Brithenig, Ceqli, Klingon, Lojban, Orcish, Ido, Quenya, Sindarin, Volapük, Jakelimoto (cfr. Linguist List, 2004), Europanto cannot really be included in this list, because it was not created for literary or film characters as Brithening, Lojban or Orcish, nor was it created by linguists so to speak from above, like Esperanto, Ido or even Interlingua, which is presented in the following way:

“Interlingua es un lingua international facile e de aspecto natural elaborate per linguistas professional como un denominator commun del linguas le plus difundite in le mundo in le dominios del scientia, cultura, commercio, etc. Un texto in interlingua es immediatamente intelligibile a millones de personas in tote le mundo, sin necessitate de studio previe”. (UMI, 2004)

Europanto, on the contrary, is as Marani puts it, developing in a natural way from the bottom of the magma of European multilingualism (cfr. Marani, 1999). An example of a text in Europanto written by Marani is the following:

G7 INFORMATIOGEZEL PILOTAPROJECTO

Diego Marani 09.04.1997

“Eine globalo kaufpunto por Piccola und Media Entrepresas”

Que would happen if, wenn Du open your computero, finde eine message in esta lingua? No est Englando, no est Germano, no est Espano, no est Franzo, no est keine known lingua aber Du understande! Wat happen zo! Habe your computero eine virus caught? Habe Du sudden BSE gedevolved? No, Du esse lezendo la neue europese lingua: de Europanto! Europanto ist uno melangio van de meer importantes Europese linguas mit also eine poquito van andere europese linguas, sommige Latinus, sommige old grec. Qui know ten moins zwei europese linguas kan Europanto undergrepen. From nu avanti, Du need keine mas foreignas linguas studie und Du kan mit el entiero mundo communicate danke al Europanto. Du no believe? Ich zal aan you demonstre brefly describendo en Europanto el pilota projecto „Eine globalo kaufpunto por Piccola und Media Entrepresas“. El but del projecto ist de facilitate PME in der electronicommerz. De aczione in tres themas axed esse: 1. Informatio Network por PME; 2. PME Exigentias: juridica, institutionale, technical; 3. Internazionale testbeds

Europanto, furthermore, is certainly not any longer merely ‘a language of Belgium’ or restricted to the members of the European Institutions as the Ethnologue defines it (ethnologue.com, 2004), nor is it used exclusively by Marani for his newspaper columns, novels or translations. In fact, in the meantime quite a few Internet-sites of Europanto have been created which are dedicated either to the collection of texts written in Europanto such as the site *Europanto*, which claims to possess “Mas than 100 textes” (Anonymous, 2004), or to the usage of Europanto, like the forum created in 2001 by *Neuropeans* where people talk about Europanto in Europanto (neuropeans.com, 2004).

#### 4. RELEVANCE FOR LINGUISTICS AND HUMANITIES COMPUTING

Whether Europanto constitutes a similar case to Nouchi in Côte d’Ivoire remains to be seen. The same goes for the question regarding the links which exist between Europanto and (European) identity, as this can only be answered by systematic research into the character of Europanto and the arguments used in online forums in favour of using and spreading it<sup>2</sup>.

What is already now clear, however, is that the whole process from monolingualism to multilingualism and to hybridity, which seems to be an important ingredient

of digital culture, raises many questions and opens up new research fields, above all for linguistics. Digital culture is, after all, heavily dominated by writing and by specific tools. With the migration of languages, dialects and hybrid varieties which have not been written before (or at least not according to a unique norm) into domains of digital culture (SMS, the Web, etc.) where no academy or other institution exerts control or establishes binding norms, linguistics is presented with a unique possibility to study directly how the 'oral' is being transformed into the 'written', how norms evolve and spread, which tools are considered to be vital if a language or dialect is to be recognised as being part of the modern world and how these tools are being created.

Mixed languages and their creation and acceptance by speakers, furthermore, question thoroughly those schools of linguistics which in conformance with the idea of a standard or national language have modelled language as a homogeneous entity, the ideal speaker-listener's competence as inherently monolingual, and have reduced linguistic variety to a common core. It becomes clear, in fact, that such models and theories have hindered linguists from seeing the multilingual nature of human communities, of the linguistic knowledge of the speakers and of their linguistic behaviour. By looking at migration in urban environments, instead, by studying the role new technologies play in freeing people from the control of print and from state norms, by examining the free choices taken by multilingual people and by following up what less main stream linguists like Hymes (1972), Wandruszka (1979) and Coseriu (1988) had to say about the multilingual nature of humans and their communities we will open the way for a quite different conception of languages, of linguistic knowledge and of speaking.

Last but not least, mixed languages and / or hybrid varieties raise a number of questions for Humanities Computing. One question is, for example, how to handle such language texts when we build corpora and above all how to treat the elements of the different languages of which they are composed when we mark up such texts? Such texts are after all not made up just of elements which can clearly be attributed to one or the other language, as in the example of code-switching given below:

“Drittens, and not the least, nous le croyons probable, daß die milieux multiculturels avec leur richesse et créativité linguistiques will grow in importance and strength, and thereby achieve the respect necessary für sprachliche Mannigfaltigkeit to become the standard of society in Europe” (Jørgensen, Kristiansen, 2000: 167)

but also of elements which are being created spontaneously or by taking into account certain characteristic traits of certain languages as in the following quote:

It was ein ernormer plaisir ihre Europanto™-website zu découvrir (grâce to the New York Times del Web). Saviez-vous que el existiert un phenomemon della same genus an ein niveau Transpacifcano? Ici à Los Angeles (place déjà at least bilingue), parmi los immigrants, como Sinistas, Koreanos, Japonicos,

sans mention Latinoamericanos, une sprache incorporant divers elements vom grammatik y vocabular both from Europa et d'Asien has evoliiert. Avec mes freunden, jo speak oft upgemixt Englisch with Chinese wörtorder and etwa Japonais verba, ou bien vice versa. C'en est the same chez youth in Taiwan, who ajoutent auch two dialecta differenta della lingua sinica.

Natürlich sind you Europeans beaucoup mieux organizado, mit wesbites u.s.w. - and j'applaud ihre accomplissements en esta demesne. But les 'Ricains (though j'admets io non esta) y los Asiaticos might well vous rattraper, sooner or tard.

Finalmente, suggestare un challenge pour les informaticos: ist es possible to développer un spell-checker Europanto™?

En tout K, grazie füe ihre site,

B.R., 25/03/98

And what do mixed languages – which are perhaps much more common than linguistics has realised up to now – mean not only for grammar and dictionary building but for speech recognition as a whole? Can the linguistic knowledge used by the speakers when they produce such mixed language texts be exploited for the processing of natural languages? What does this knowledge tell us about languages and speaking? What sort of tools do we need if we want to study it? And how about a spell-checker and other tools for Europanto and other mixed languages like Nouchi and the many Creoles spoken (and written) in many parts of the world? It will be interesting to see whether this challenge will be taken up in the near future.

## 5. REFERENCES

Anonymous (2004). Europanto.

<http://www.contagions.com/europanto/europanto.html>

Associazione L.I.S.A. (2004). *Regioni d'Italia*. Italia Donna. Il portale delle donne

<http://www.italiadonna.it/public/percorsi/02014/indicedialetto.htm>

B.R. (25/03/98). "Europanto™ Réactions"

<http://www.europanto.contagions.com/react.html>.

Begnini A. (2000). *I dialetti italiani sopravvivono online*. La Repubblica 12.08.2000

<http://www.repubblica.it/online/societa/dialetti/dialetti/dialetti.html>.

Bortoluzzi R., Bolzan M. (eds) (2000). *Vecio parlar* <http://digilander.libero.it/vecioparlar/index.htm>.

Burr E. (2004). *Dialetti e lingue minoritarie delle montagne in Rete*. in G. Marcato (ed.): *I dialetti e la montagna*. Atti del Convegno internazionale di Studio Sappada/Plodn (Belluno), pp. 351-357. Padova, Unipress.

Coseriu E. (1988). *Sprachkompetenz: Grundzüge der Theorie des Sprechens* (= UTB 1481). Tübingen, Francke.

Dialettando.com (2004). <http://www.dialettando.com/>

- Erfurt J. (2002) „*Multisprech*“: *Migration und Hybridisierung und ihre Folgen für die Sprachwissenschaft*. In J. Erfurt (ed.) *Multisprech: Hybridität, Variation, Identität*. Osnabrücker Beiträge zur Sprachtheorie 65, pp. 5-33.
- Ethnologue.com (2004). *Europanto: a language of Belgium* [http://www.ethnologue.com/show\\_language.asp?code=EUR](http://www.ethnologue.com/show_language.asp?code=EUR).
- Ferraiuolo L. (1998). *Altro che inglese: Internet ora parla napoletano*, La Repubblica, 23.01.1998  
<http://www.repubblica.it/online/tecnologie/write/napoletano/napoletano.html>.
- Giuse N. (2003). “Il canton del dialett” [barlassina.it](http://www.barlassina.it/)  
<http://www.barlassina.it/>.
- GlobalReach (2004a). *Evolution of non-English speaking online population* GlobalReach  
<http://global-reach.biz/globstats/evol.html>.
- GlobalReach (2004b). *Global Internet Statistics (by Language)*  
<http://www.global-reach.biz/globstats/index.php3>.
- Global Reach (2004c). *Global Internet Statistics: Sources & References*  
<http://global-reach.biz/globstats/refs.php3>.
- GlobalReach (2004d). *Global online marketing, multilingual Web promotion and Website translation will develop your international sales*  
<http://global-reach.biz/>.
- GlobalReach (ed.) (2004e). *A guide to Web sites around the world, arranged by language in topic-specific directories*  
<http://glreach.com/gbc/index.php3>.
- Hymes Dell H. (1972). *On Communicative Competence*. In: J. B. Pride, J. Holmes (eds). *Sociolinguistics*, pp. 269-293. Harmondsworth, Penguin.
- Isole linguistiche <<http://www.isolelinguistiche.it/prima.htm>> [18.11.2004]
- Jørgensen N., Kristiansen T. (2000). *La tâche de la sociolinguistique au 21e siècle – Ein europäisches soziolinguistisches Mehrsprachigkeitsmanifest*. *Sociolinguistica* 14, pp. 164-167.
- Karismo A. (1999). *Europanto – Könnte das la lingua majeure for Europa para the nouvelle millennium sein?* <http://www2.helsinginsanomat.fi/english/archive/thisweek/41041999.html>.
- Kube S. (2003). *Das Nouchi in Abidjan: Vom Argot der Straßenkinder zur zukünftigen Nationalsprache der Côte d'Ivoire?* In J. Erfurt (ed.). *Multisprech: Hybridität, Variation, Identität*, pp. 131-153. Osnabrücker Beiträge zur Sprachtheorie.
- La Patrie dal Friül. *Sfuei mensil di informazion* (2004)  
<http://www.friul.net/>.
- Lawrence M. (ed.) (2004). *The Multilingual, multicultural online journal and community of arts and ideas*  
<http://www.towerofbabel.com>.
- Linguist List (2004). *Search for a language* by The Linguist List <http://linguistlist.org/forms/langs/find-a-language-or-family.html/>.
- Marani D. (1999). *Europanto. From productive process to language*  
<http://www.mediamatic.net/article-200.5711.html>

- [http://www.repubblica.it/online/forum\\_rete/sititop/2811/2811.html](http://www.repubblica.it/online/forum_rete/sititop/2811/2811.html)
- Mandò M. (1998a). *Proverbi e aforismi anche in dialetto*. La Repubblica, 11.04.1998  
[http://www.repubblica.it/online/forum\\_rete/sititop/1104/1104.html](http://www.repubblica.it/online/forum_rete/sititop/1104/1104.html).
- Mandò M. (1998b). *Minoranze linguistiche e dialetti in Rete*, La Repubblica 28.11.1998.
- Mola G. (1999). *No speak English nella Rete del 2000*. La Repubblica 17.06.1999  
<http://www.repubblica.it/online/internet/lingue/lingue/lingue.html>.
- Neuropeans.com (2004). *Europanto the language that connects Europeans* <http://www.neuropeans.com/communication/forum/category1/topic/FORUM7>.
- Nouchi.com. *La Culture Africaine avec le Sourire* (2004)  
<http://nouchi.com>.
- Retico A. (2000). *La Babele delle lingue sopravvive in Rete*. La Repubblica 25.10.2000  
[http://www.repubblica.it/online/cultura\\_scienze/lingue/lingue/lingue.html](http://www.repubblica.it/online/cultura_scienze/lingue/lingue/lingue.html).
- [UMI] Le Union Mundial pro Interlingua (2004). Union Mundial pro Interlingua  
<http://www.interlingua.com/>.
- Union Ladina del Cadore de Mèdo (2003)  
<http://www.ladinia.org/index.html>.
- Ursini F. (in print). *Tra scritto e parlato: I 'messaggi brevi' tra telefoni cellulari* in E. Burr E. (ed.): *Tradizione & Innovazione. Il parlato: teoria – corpora – linguistica dei corpora*. In Atti del VI Convegno Internazionale della SILFI, pp. 443-455. Firenze, Cesati.
- Vandendooren J. (without year). *L'Europanto: la langue des citoyens européens?* [http://www.espace-citoyen.net/article.php3?id\\_article=470](http://www.espace-citoyen.net/article.php3?id_article=470).
- Wandruszka M.(1979). *Die Mehrsprachigkeit des Menschen*. München/Zürich: R. Piper.

#### 4. NOTES

- <sup>1</sup> That the number of English Web pages is still relatively high is due to the fact, that in a lot of countries, the first language to appear next to the national language on the Web is English.
- <sup>2</sup> As such research can be carried out systematically only on the basis of a corpus, at the moment of writing a marked-up and TEI conformant corpus composed of the messages which appear in the above presented forum is being created. This corpus will be integrated later with other types of texts written in Europanto.



# Multiculturalità e rete: voci migranti in Italia

Luisa Carrer

Università di Edimburgo

(l.carrer@ed.ac.uk)

*Immigration is a permanent feature of Italian system in the present phase and therefore it deserves the development of new editorial and communication policies, that go beyond traditional broadcasting policies. The paper will consider the potential of e-publishing for both multi-cultural communication and for educational application. Present on line resources show three main fields of application: news; multi-cultural representation of migration, comprising literary expression, educational field, specifically for teaching Italian. The digital space stimulates literary production and communication at a transnational level. The article states a set of recommendations concerning: a) the need to consider the literature of migrants within linguistic and literary curricula; b) the need of to train teachers to digital technology for educational purpose; c) the need to fully exploit the potential of digital writing by migrants.*

*Keywords:* e-publishing, Multilingualism, Literature of migrants

## 1. PREMESSE

Presento i risultati di una ricerca ancora giovane, incoraggiata dai temi centrali a questo convegno e motivata dal mio interesse per i periodici culturali telematici, in particolar modo quelli dedicati alla letteratura della migrazione. Suo scopo è indagare in quale misura e con quali modalità il fenomeno immigratorio sia riscontrabile nel panorama mediale italiano contemporaneo e quale funzione svolga l'editoria digitale in tema di multiculturalità, difesa della diversità e promozione della letteratura della migrazione.

Non vi sono studi, a tutt'oggi, che trattino in modo sistematico il ruolo della tecnologia nella vita, anche artistica, dei cittadini italiani di origine immigrata e che censiscano criticamente le risorse elettroniche dedicate a multiculturalismo e integrazione valutandone l'impatto sull'utenza. Il presente vuole essere un primo passo in tal direzione. A partire da due considerazioni preliminari.

Il fenomeno dell'immigrazione in Italia inizia ad essere palese nei primi anni ottanta ed assume connotazioni di rilievo a partire dai primi anni novanta del secolo scorso. Al 31 dicembre 1991 gli immigrati regolarmente soggiornanti in Italia erano poco meno di 650.000. Dieci anni dopo, alla fine del 2001, la stima era più che duplicata e contava oltre 1.300.000 presenze (Tab. 1). Negli ultimi anni l'aumento



della popolazione immigrata è legato in gran parte ai ricongiungimenti familiari (31,7%) e alla conseguente influenza quantitativa dei minori, fattori che denotano la tendenza ad un insediamento stabile.

Ripartizione territoriale	1991		2001	
	numero	%	Numero	%
Nord-Ovest	181.359	27,9	444.876	32,6
Nord-Est	127.636	19,7	328.488	24,1
Centro	220.913	34,0	396.834	29,1
Sud	65.572	10,1	133.263	9,8
Isole	53.455	8,2	59.169	4,3
Totale	648.935	100,0	1.362.630	100,0

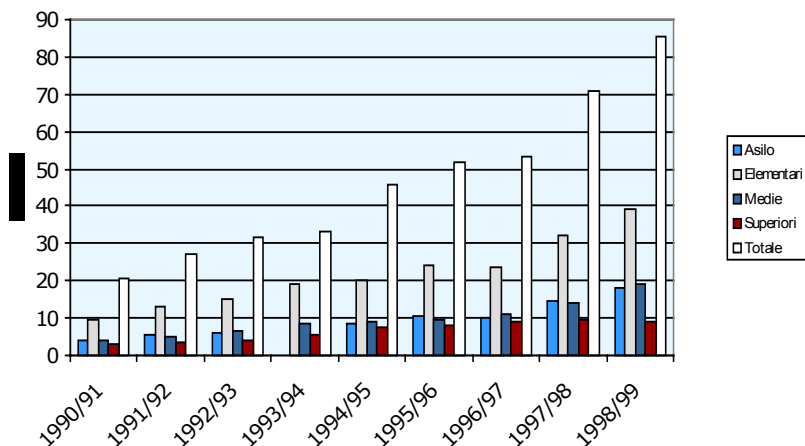
Tab. 1 Permessi di soggiorno per ripartizione territoriale 1991 e 2001. (Fonte: Caritas su dati del Ministero degli Interni <<http://www.caritasroma.it/immigrazione>>)

Al principio del 2004, il panorama si struttura in una presenza straniera di almeno due milioni e mezzo di soggiornanti regolari. La Romania è il primo gruppo, mentre il Marocco e l'Albania seguono distaccati di qualche decina di migliaia di unità: tutti e tre i paesi si collocano, comunque, al di sopra delle 250.000 unità<sup>1</sup>. Se si considera poi che anche le rappresentanze a seguire, seppure con numeri ridotti, hanno una buona consistenza<sup>2</sup>, si coglie il significato del policentrismo dell'immigrazione italiana, fatto del tutto caratterizzante, tanto più se a confronto con i fenomeni migratori nei paesi dell'Europa del Nord. Si coglie inoltre la complessità delle problematiche da gestire ma, allo stesso tempo, la ricchezza a disposizione per un adeguato progetto di convivenza.

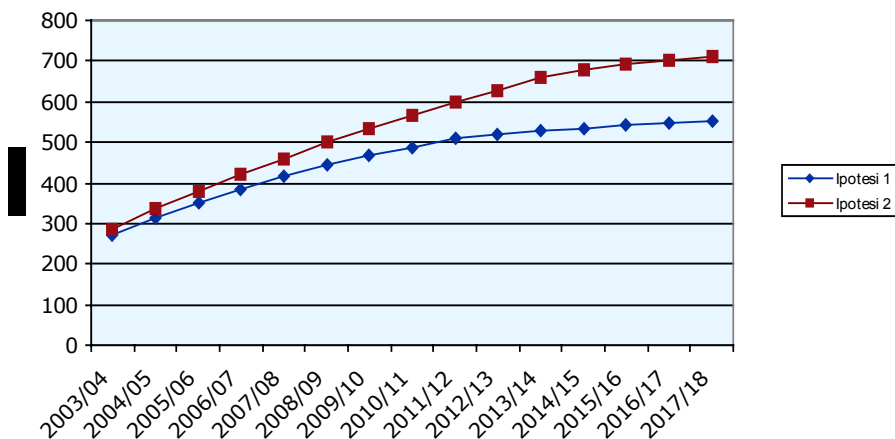
Il continuo incremento del flusso migratorio in Italia avviene in un contesto locale di grave declino demografico. Si ipotizza un calo della popolazione italiana pari al 5,3% nei prossimi vent'anni, con la fetta degli under 20 destinata a ridursi del 32,3%<sup>3</sup>. L'incidenza degli studenti stranieri sul totale della popolazione scolastica nazionale, pur essendo attestata su un valore ancora piuttosto esiguo (2,3% del totale pari a 191.767 unità), inferiore all'incidenza che i soggiornanti stranieri hanno sulla popolazione italiana nel suo complesso, mostra un incremento annuo sempre più consistente (+23,3% dall'anno scolastico 2000/01). Nel corso degli ultimi cinque anni, gli studenti stranieri sono addirittura più che triplicati e in meno di venti anni la loro consistenza è aumentata di trenta volte (erano appena 6.104 nell'anno scolastico 1983/84) (Graf. 1).

Jonathan Chaloff porta a riflettere su come i ragazzi immigrati al di sotto dei venti anni d'età verranno a costituire il 7,7% circa della popolazione totale prima del 2020, con la conseguenza di una forte presenza interculturale all'interno del sistema educativo italiano<sup>4</sup>. Una più recente proiezione ministeriale<sup>5</sup> sull'evolversi del numero degli alunni con cittadinanza non italiana dal corrente anno scolastico fino al 2017/18, conferma tale ipotesi, se pur, bisogna dire, con maggiore cautela: è presumibile,

infatti, che progressivamente verranno a stabilizzarsi alcuni o tutti i fattori che oggi costituiscono forti spinte alla crescita (Graf. 2)



Graf. 1 Alunni con cittadinanza non italiana nelle scuole italiane 1990-1999. (Fonte: Chaloff, CHIP, Censis 1999)



Graf. 2 Ipotesi di evoluzione del numero di alunni con cittadinanza non italiana nei prossimi quindici anni: l'ipotesi 1 prospetta un flusso costante in entrata di 150.000 immigrati l'anno (4,5% del totale nel 2017/18) mentre l'ipotesi 2 ne suppone uno di 200.000 unità (6% nel 2017/18). (Fonte: Ministero dell'Istruzione, dell'Università e della Ricerca 2002)

Seconda premessa. Un quadro normativo comunitario sull'editoria elettronica fu la principale richiesta della proposta di risoluzione che il Parlamento Europeo discusse nel febbraio 2001. Il primissimo punto della relazione, firmata dall'eurodeputata britannica Barbara O'Toole, sottolineava "l'importanza di incoraggiare la creazione letteraria e linguistica, di mantenere e migliorare la diversità culturale e linguistica in Europa, di salvaguardare l'accesso universale alla parola scritta e di consolidare la conoscenza e la cittadinanza"<sup>6</sup>. Perché in Europa l'editoria elettronica venisse a rappresentare una garanzia nella difesa del multilinguismo e della multiculturalità, il documento esortava la Commissione "a promuovere il contenuto europeo creativo nell'editoria elettronica in tutte le lingue [...] e a sostenere un gran numero di piccole imprese di pubblicazione elettronica al fine di mantenere la diversità culturale e linguistica in modo da creare un polo alternativo ai grandi gruppi mediatici"<sup>7</sup>. Invitava inoltre la Commissione a garantire, una volta che l'editoria elettronica si fosse affermata nell'Unione, la messa a disposizione di fondi per promuovere la pubblicazione di libri in formato elettronico, e per sostenere gli sviluppi tecnologici per la tutela dell'autenticità e dei diritti d'autore.

In ambito italiano, peraltro, fu contemporanea alla proposta di risoluzione l'entrata in vigore della nuova legge sull'editoria (62/2001) che ampliava la definizione di 'prodotto editoriale' fino a comprendervi, significativamente, anche quello realizzato "su supporto informatico, destinato alla pubblicazione o, comunque, alla diffusione di informazioni presso il pubblico"<sup>8</sup>. La legge 62/2001 dunque, con assoluta novità ed in linea con le raccomandazioni comunitarie, regolava la trasposizione dell'attività editoriale dal supporto cartaceo al campo multimediale.

## 2. ARGOMENTAZIONI

Date le premesse, avvio la ricerca da due assunti principali:

- a) la consapevolezza della portata non più emergenziale bensì oramai strutturale dell'immigrazione all'interno della società italiana deve comportare un maggiore investimento nei servizi tanto sociali quanto culturali. Multiculturalismo deve essere sinonimo di politiche ed iniziative rivolte a incorporare il riconoscimento della differenza etnica, culturale, linguistica e religiosa all'interno del sociale<sup>9</sup>. La sfida è principalmente legata all'educazione linguistica<sup>10</sup> ed interculturale, divenute una priorità nella preparazione e nella pratica didattica attuale;
- b) una maturazione dei mezzi di comunicazione di massa e delle politiche editoriali – dell'editoria piccola, e spesso già coraggiosa, e soprattutto della grande – sembra a questo punto dovuta, non solo per la promozione del contenuto creativo multilingue tradizionalmente europeo, ma anche in favore di argomentazioni che riconoscano gli immigrati quali nuovi cittadini dell'Unione, li rendano partecipi di una rete informativa plurilingue, ne assecondino il potenziale creativo.

Di fatto, l'immigrazione in Italia e la sua rappresentazione mediale sono state oggetto di numerose ricerche negli ultimi anni: ne sono esempi *Tuning into Diversity*<sup>11</sup>,

progetto finanziato dall'Unione Europea realizzato nel 2002 dal Censis (Centro Studi Investimenti Sociali) cui ha contribuito il Cospe (Cooperazione per lo Sviluppo dei Paesi Emergenti); ancora del Censis è la ricerca *L'immagine degli immigrati nei media del 2002*<sup>12</sup>, sostenuta dall'iniziativa comunitaria Equal e dal Ministero del Lavoro e della Politiche Sociali; Extracommunication<sup>13</sup>, monitor su informazione e immigrazione attuato nel 2003 sempre in ambito Equal e curato dal Dipartimento di Sociologia e Comunicazione dell'Università degli Studi "La Sapienza" di Roma; On-line/More Colour in the Media, programma d'azione comunitaria contro le discriminazioni realizzato con il sostegno finanziario della Comunità Europea. Nel novembre 2003, il programma ha lanciato la "Settimana Europea di Monitoraggio dei Media"<sup>14</sup>.

L'attenta considerazione dei risultati di queste indagini è fondamentale per il corretto inquadramento del ruolo che l'editoria digitale italiana riveste in materia di immigrazione<sup>15</sup>.

Partendo dal presupposto che l'informazione sia la risorsa basilare per assicurare a ciascuno una prima forma di inclusione sociale, *Tuning into Diversity* ha analizzato quale immagine di migrante, rifugiato o richiedente asilo avessero trasmesso i media italiani negli ultimi anni. Alcune peculiari deficienze del sistema mediale italiano – la debole presenza della dimensione locale e la scarsa rilevanza riservata all'informazione di servizio, ad esempio<sup>16</sup> – sono apparse influire grandemente sui percorsi di informazione rivolti agli immigrati. Cionondimeno, la ricerca del Censis ha evidenziato la presenza di 16 emittenti TV e 44 stazioni radio che hanno o hanno avuto in tempi recenti almeno un'esperienza di iniziativa multilinguistica nel loro palinsesto e 31 testate editoriali dedicate agli immigrati. I primi tentativi risalivano all'inizio degli anni novanta (11,1%) e la massima attivazione si verificava nella seconda metà di quegli anni, anche se molte erano le iniziative nate nell'anno dell'indagine stessa.

Le esperienze censite assommavano a 117 prodotti (programmi radio 70, televisivi 16 e prodotti editoriali 31), di cui 87 ancora in corso nel 2002 e 30 conclusi. Erano prodotti estremamente eterogenei: la stessa espressione 'trasmissioni per immigrati' poteva significare, di volta in volta, inserimento nel palinsesto di un programma pensato per un certo tipo di *target*; uso di un'altra o molte altre lingue (la Tab. 2 indica come il plurilinguismo fosse sicuramente promosso, in percentuale pari all'italiano); collaborazione redazionale di immigrati; programmazione di musica etnica; presenza di notizie o programmi sull'immigrazione.

Lingua utilizzata	TV e radio	Stampa	Totale	
	v.a.	v.a.	v.a.	%
Solo italiano	24	11	35	29,91
Solo lingua straniera	19	5	24	20,51
Bilingue	12	10	22	18,8
Plurilingue	29	5	34	29,05
Non rilevabile	2	—	2	1,73
Totale	86	31	117	100,0

Tab. 2 Lingue utilizzate dalle iniziative radiotelevisive ed editoriali rivolte agli immigrati, secondo la tipologia delle iniziative (v.a. e val. %). (Fonte: Cospe-Tuning into Diversity 2002)

Il 49,6% delle esperienze censite sussisteva da un periodo compreso da uno a cinque anni e l'11,1% superava la soglia dei cinque anni. Si trattava evidentemente di una fenomenologia tutt'altro che volatile, decisa a evolvere verso forme di stabilizzazione. Lo strumento privilegiato era la radio, affiancato dall'esteso utilizzo della carta stampata, due mezzi a costo limitato ed elevata flessibilità. Meno presenti i programmi televisivi, e il dato non stupisce se si considera la particolare strutturazione del sistema televisivo italiano.

Un punto assai limitante, tuttavia, veniva rilevato: il bacino d'utenza (radio e TV) e di diffusione (carta stampata) attestava la prevalenza di iniziative a carattere prettamente locale (23% iniziative cittadine e locali, 34% iniziative provinciali). Stabilire chi era l'immigrato tipo' rappresentato in televisione e quali erano i programmi e le trasmissioni che trattavano di immigrazione o in cui comparivano persone straniere o immigrate è stata ulteriore preoccupazione del Censis nell'ambito della ricerca *L'immagine degli immigrati nei media*. L'indagine ha riguardato l'intera programmazione televisiva delle tre reti Rai, Mediaset e Telemontecarlo (poi La 7) per sei settimane nel corso di due anni (2001 e 2002). All'interno dei telegiornali, oltre l'80% delle notizie che riferivano di persone di origine straniera rientrava nella rubrica di cronaca (e cronaca nera per la quasi metà dei casi), il 16% in quella di politica interna, il 2,1% in cultura e società, e l'1% nella pagina estera. Esiguità, scarsità di approfondimento e monotematicità stereotipata sembravano caratterizzare anche l'attenzione di rotocalchi e rubriche. È questo il secondo punto dolente dell'offerta mediale italiana: la continua associazione al mondo criminale non fa che alimentare la tendenza a sovrapporre immigrazione e delinquenza e suggerisce una netta contrapposizione di mondi. L'opportunità di esprimersi su ciò che li riguarda direttamente – preconditione, direi, per una corretta rappresentazione – risultava pure limitata: nel 64% dei casi gli immigrati venivano solo citati, per il 2,1% erano consultati in qualità di esperti, e nell'1% soltanto dei casi censiti erano ospitati in una trasmissione. Al settembre 2003, *Extracommunication* rilevava a sua volta la generale faziosità dei quotidiani nazionali censiti in relazione a struttura e contenuti delle notizie sull'immigrazione. La rilevazione si è estesa per sei settimane fra il dicembre 2002 e il febbraio 2003 ed ha interessato otto quotidiani nazionali con sezione locale "Roma", laddove presente. La forma dominante era quella delle brevi di cronaca, spesso riguardanti episodi di microcriminalità urbana che si qualificavano come una costante (56,7% delle notizie per sezione dei quotidiani censiti) nell'immagine mediale della presenza straniera in Italia (Tab. 3).

Tipologia notizie	Sezioni nazionali	Sezioni locali	Totale
Crimini e terrorismo	40,8	69,6	56,7
Sbarchi e regolarizzazioni	35,6	7,8	20,3
Economia, lavoro e sanità	11,6	9,2	10,3
Religione, cultura e integrazione	8,4	10,5	9,5
Altro	3,6	2,9	3,2
Totale	100,0	100,0	100,0

Tab. 3 Argomento trattato per sezione del quotidiano. (Fonte: Extracommunication 2003)

Tale pratica era talmente diffusa che più del 50% del pubblico intervistato – 400 cittadini romani di diciotto anni e più, interpellati fra il marzo e il giugno 2003 – esprimeva insoddisfazione rispetto alla propria conoscenza della realtà dei fatti e al comportamento delle testate: un atteggiamento critico da parte dell'utenza che segnala un'importante, netta dissonanza tra l'offerta erogata dai media e le aspettative dei destinatari.

Dalle ricognizioni di Censis, Cospe ed Etnequal – i cui risultati sono pienamente confermati dal recentissimo rapporto sulla già menzionata “Settimana Europea di Monitoraggio dei Media”<sup>17</sup> – emergono due questioni cruciali. Da un lato, come sottolineavo or ora, affiorano la coscienza e l'insoddisfazione del pubblico nei confronti del sistema informativo corrente in materia di immigrazione; d'altro lato, manipolazione e faziosità dell'informazione sembrano di gran lunga predominare, con la conseguente assenza, a livello nazionale diffuso, della voce diretta di coloro che di quell'informazione sono i protagonisti.

Pur date le raccomandazioni della Commissione Europea sulle nuove frontiere dell'editoria, le indagini considerate mancano di esaminare le iniziative nate in ambiente digitale e dedicate al multilinguismo ed alle culture dell'immigrazione. Dati freschissimi su italiani, scuola e Internet portano a riflettere sull'opportunità di fronteggiare tale mancanza. Vediamoli.

Secondo l'ultima indagine dell'Istituto per gli Studi sulla Pubblica Opinione (ISPO) – realizzata per il Rapporto 2004 dell'Osservatorio Associazione Italiana Editori (AIE) sull'editoria digitale<sup>18</sup> – fra i 470 milioni di utenti *non*-anglofoni in cui il numero globale della popolazione ‘internauta’ si sarebbe espanso a partire dagli anni novanta (i due terzi circa del totale)<sup>19</sup>, ben 21 sarebbero italiani, pari al 43% della popolazione e all'83% di quanti utilizzano regolarmente il computer. Quasi la metà del totale (49%) sarebbe addirittura disposta a pagare per accedere a servizi a valore aggiunto (incentivo alla ricchezza dei contenuti, ad aggiornamenti, alla velocità di accesso). Ancora, il 17% degli italiani – perlopiù giovani e studenti – sceglierebbe in prevalenza le tecnologie per accedere a vari tipi di contenuto editoriale: comparando i dati dell'indagine ISPO 2002 con questa del 2004 emerge infatti una crescita nel numero di visitatori di almeno un sito internet a soggetto informativo nell'ultimo semestre, da 14,7 a 17,8 milioni rispettivamente. Sul fronte educativo, sebbene siano ancora una minoranza (37%) gli insegnanti che si valgono regolarmente del computer nella fase preparatoria delle lezioni, vi sono stati incrementi notevoli dal punto di vista della formazione: il 42% di quanti non erano in grado di lavorare al computer nel gennaio 2003 ha imparato a farlo; il 53% di quanti non sapevano navigare in Internet ha fatto altrettanto; e il 55% dei docenti intervistati che già utilizzavano i mezzi informatici dichiara di aver migliorato le proprie capacità. È la scuola stessa il motore del cambiamento, dunque, con l'introduzione di corsi ministeriali *on-line* per docenti da sommare a quelli organizzati dagli istituti scolastici a livello locale.

Con un tale bacino di ricezione, in costante incremento, il potenziale delle risorse digitali italiane su integrazione culturale e multilinguismo mi parrebbe inequivocabile, nel campo dell'informazione come in quello dell'interculturalità, verso un'alternativa quali-quantitativa efficace ai canali – lacunosi, abbiamo visto – di televisione, radio e stampa. Procediamo allora allo spoglio critico della risorse più rappresentative.

### 3. RASSEGNA RISORSE<sup>20</sup>

Ho riscontrato un triplice orientamento in ambiente elettronico: spiccano, in primo luogo, i servizi informativi su iniziative, normative, statistiche ed associazionismo liberamente accessibili e diretti agli immigrati stessi come agli operatori del settore. Alcuni esempi:

- *Atlante immigrazione* <<http://www.provincia.torino.it/xatlante/>>. È un progetto del Dipartimento Solidarietà Sociale della Provincia di Torino operante dal 1997. Si rivolge principalmente agli addetti ai lavori con l'obiettivo di fornire una banca dati ricca di informazioni su legislazione nazionale, competenze istituzionali, enti locali (cosa fanno e cosa si propongono), servizi, organismi internazionali, luoghi del sapere, il fenomeno immigrazione. L'accesso è gratuito, previa iscrizione. Il sito è continuamente aggiornato, è composto da più di 1500 pagine e conta più di 2000 iscritti;
- *Centro Interculturale della Città di Torino* <<http://www.comune.torino.it/cultura/interculturale/>>. Attivo dal 1996, il Centro si propone come luogo di incontro, scambio e formazione. Il sito ruota intorno alle tematiche dell'economia sostenibile, diritti delle minoranze, emigrazione e religioni, su cui offre ricchi approfondimenti critici. Comprende uno spazio dedicato ai servizi per l'intercultura nella città di Torino, in Italia e nel mondo della scuola, fornendo informazioni, analisi, contatti, bibliografie. Mette a disposizione un valido glossario a chiarimento dei principali termini dell'intercultura. Il sito comprende, inoltre, i dati relativi alle attività di formazione ed ai laboratori organizzati dal Centro, nonché il rimando al catalogo *on-line* della sua biblioteca – interrogabile secondo sezione tematica – e ad altre biblioteche selezionate;
- *Cestim on-line* <<http://www.cestim.org/>>. A cura del Centro Studi Immigrazione di Verona. Disponibile in italiano e in inglese, è un sito di documentazione sui fenomeni migratori organizzato in *Sezioni* (dedicate alla cultura con rimandi a biblioteche, riviste, novità librarie, ecc.) e in *Argomenti*, con *link* di approfondimento su tematiche specifiche quali "Intercultura", "Islam", "Mass Media". Per una prospettiva comparata, vi sono rimandi a documentazioni multilingui riguardanti i fenomeni migratori nel resto d'Europa e del mondo;
- *Città Multi-etnica* <<http://www3.iperbole.bologna.it/immigra/index.php>>. Servizio Immigrati, Profughi e Nomadi del Comune di Bologna attivo dall'aprile 1995. È un portale di informazioni su normative, statistiche, associazionismo, servizi per salute, casa e lavoro, mediazione culturale. Dedicata una sezione a scuola e università con schede informative e riferimenti normativi. Dal 2002 è anche un periodico di informazione, liberamente consultabile e scaricabile;
- *Migra* <<http://www.migranews.net/>>. Agenzia *on-line* d'informazione degli immigrati, uno dei risultati cardine dell'iniziativa comunitaria Equal dedicata all'immagine dell'immigrato nei media, nella società civile e nel mondo del lavoro, gestita in Italia dal Ministero del Lavoro e delle Politiche Sociali e promossa dall'Organizzazione Internazionale per le Migrazioni (OIM), dalla

Caritas di Roma e dall'Archivio dell'Immigrazione. Attiva dal principio del 2003, l'Agenzia propone *MigraNews*, notiziario costantemente aggiornato con approfondimenti tematici, inchieste, dossier fotografici "che si propone di fornire un'informazione non episodica, approfondita, rispettosa delle differenze culturali, non riduttiva, non sensazionalista. Un'informazione dalla parte degli immigrati, non in senso ideologico: per offrire il loro punto di vista, puntando sulla loro capacità di fare informazione, di essere parte dell'opinione pubblica, di fare cultura". La redazione vanta la collaborazione di circa settanta corrispondenti di origine immigrata su tutto il territorio nazionale;

- *Migr@re* <<http://www.migrare.it/migrare/migrare.nsf/?Open>>. Viaggio nel mondo dell'immigrazione per cittadini e operatori (Città di Reggio Emilia). Raccoglie i contenuti in quattro nodi principali: *Leggi, percorsi, diritti e doveri*: con ragguagli dettagliati su permessi, casa, lavoro, salute, scuola e integrazione; *Tante culture, una città*: servizi, associazioni e sostegno nell'ambito di Reggio Emilia; *Appunti e appuntamenti*: aggiornato con notizie, novità e giorni da ricordare; *Libri, riviste e siti*: rubrica di cultura con rimandi a ricerche, materiali, liste biblio e webliografiche;
- *Roma – Intercultura* <<http://www.roma-intercultura.it/>>. A cura del Centro di Informazione e Documentazione su Immigrazione e Intercultura (CIDII), il servizio è operativo dal dicembre 2000 ed è promosso dal Comune di Roma come spazio dedicato alla conoscenza della realtà multietnica e multiculturale cittadina e alla realizzazione di iniziative interculturali. Il sito mette a disposizione di insegnanti e studenti i cataloghi *on-line* della biblioteca e mediateca del Centro; la consultazione degli indici (anche in lingua inglese) della rivista *SE*, provvisti di *abstract*; il *download* rapido di documenti con dati statistici, rapporti ricerche, atti convegni, manuali e guide; segnalazioni aggiornate di incontri di informazione e formazione a tema e novità librarie;
- *Stranieri in Italia* <<http://www.stranieriinitalia.it/>>. Portale ricchissimo provvisto di annunci, forum, informazioni e guide su lavoro, regolarizzazioni e leggi in cinque lingue. Contiene un *Vademecum* per l'immigrazione in otto lingue, fra cui arabo, ucraino, russo, romeno. Fornisce i *link* ai portali dedicati in modo specifico ai romeni e ai brasiliani soggiornanti in Italia – rispettivamente in rumeno e portoghese – e, in francese e inglese, agli africani in Italia. Interessante è pure il servizio di rimandi diretti a testate giornalistiche e stazioni radio *on-line* di tutto il mondo.

Se da questa prima sezione risaltano gli intenti informativi e di servizio per una corretta, alternativa e multilingue rappresentazione dei fenomeni migratori, indipendente da leggi di palinsesto, il secondo orientamento della rete guarda più alle risorse pedagogiche e ad autorevoli percorsi formativi per docenti in tema di educazione interculturale e didattica dell'italiano come L2. Di ricerca e sviluppo si occupano:

- *Approccio Lingua Italiana Allievi Stranieri* (ALIAS) <<http://venus.unive.it/aliasve/index.php>>. Il progetto, nato nel 1998/99 dalla collaborazione tra



l'Università Ca' Foscari e il Ministero della Pubblica Istruzione (MPI), si occupa delle tematiche e delle problematiche legate all'insegnamento dell'italiano come lingua seconda. Fra i suoi obiettivi sono la ricerca; la formazione e l'aggiornamento a distanza e in presenza dei docenti di allievi stranieri e di chi desidera approfondire le tematiche relative all'italiano L2; la creazione di materiali e strumenti didattici e critici multimediali per la formazione e per la didattica, interamente scaricabili; l'assistenza *on-line* ai docenti sui problemi specifici legati alla materia; la realizzazione di bibliografie ragionate di testi e di materiali didattici per l'insegnamento dell'italiano L2 e per l'educazione interculturale; la progettazione di programmi personalizzati per l'autoformazione e la promozione di sperimentazioni didattiche e di progetti di ricerca-azione nel campo della didattica delle lingue seconde e dell'educazione interculturale;

- *Centro COME* <<http://www.centrocome.it/homemenu.htm>>. Il Centro COME è un servizio della cooperativa sociale *Farsi Prossimo*, promossa dalla Caritas Ambrosiana, attiva dal 1994. Il sito si propone di realizzare un intervento di sostegno e di consulenza a distanza per insegnanti attraverso servizi informativi riguardanti iniziative, seminari, corsi d'aggiornamento sui temi dell'accoglienza, dell'insegnamento dell'italiano L2, dell'educazione interculturale; aggiornamenti bibliografici e webliografie; chiarimenti in tema di normativa sull'inserimento scolastico degli alunni stranieri; offerta di materiali didattici via internet (scaricabili in formato word e pdf) o posta ordinaria e di modelli di progetti organizzativi e didattici sperimentati in varie città; consulenza alla progettazione di azioni e interventi e risposte mirate riguardanti percorsi didattici specifici o casi particolari d'inserimento; forum degli insegnanti;
- *COSPE On Web* <<http://www.cospe.it/onweb/web.htm>>. Sito a cura della Ong Cooperazione per lo Sviluppo dei Paesi Emergenti, con sezione dedicata alle iniziative on-line sulle problematiche interculturali. Fra i progetti di maggiore interesse sono CRO.CU.S. (Cross Cultural Satellite Service) <<http://www.crocusproject.net/>> che dal marzo 2001 offre servizi a distanza per scuole plurilingui e interculturali (comprende alcune sezioni ad accesso libero ed una serie di corsi di lingua e altri servizi accessibili solo agli abbonati); *Me Too-Anch'io* <<http://www.socrates-me-too.org/>> che mette a disposizione materiali didattici elaborati dalle équipes Cospe con l'obiettivo di promuovere la diffusione del plurilinguismo – sia della seconda lingua, sia delle varie lingue materne – fra gli allievi figli di lavoratori migranti, nelle scuole di ogni ordine e grado, e dunque favorire pari opportunità di successo scolastico agli allievi di origine straniera; *ParlaMondo* <<http://www.rete.toscana.it/sett/polsoc/parlamondo/index.htm>> progetto per le pari opportunità nella comunicazione, nell'informazione e nei servizi ai cittadini: pubblica ricerche, promuove servizi, e diffonde le offerte informative di TV, radio e stampa rivolte in modo specifico a cittadini stranieri nell'ambito della Regione Toscana;
- *CRES* <<http://www.manitese.it/cres/cres.htm>>. Sito del Centro Ricerche Educazione allo Sviluppo, associazione professionale costituita da esperti e

insegnanti di ogni ordine di scuole (elementare, media inferiore e superiore), riconosciuta dal Ministero della Pubblica Istruzione e curante l'attività di educazione allo sviluppo di *Mani Tese*. Offre sussidi didattici per l'educazione all'intercultura, corsi per insegnanti su richiesta, e *Strumenti Cres*, quadrimestrale di didattica interculturale;

- *Educazione Interculturale*

<<http://www.educational.rai.it/corsiformazione/intercultura/default.htm>>. Il sito di *Rai Educational* è strutturato a partire dai tredici nodi della educazione interculturale così come individuati nel kit didattico *Educazione interculturale nella scuola dell'autonomia* pubblicato dal Ministero della Pubblica Istruzione (MPI 2000). È rivolto a docenti ed operatori del settore educativo e sociale e, in un'ottica collaborativa e cooperativa, intende offrire una documentazione completa ed aggiornata – liberamente scaricabile – sia dal punto di vista teorico che della concreta mediazione educativa e didattica. Ogni nodo, a partire da una breve presentazione che ne delinea l'orizzonte, mette a disposizione una molteplicità di articoli d'approfondimento curati da esperti – fra le tematiche vi sono, ad esempio, “Identità, differenze e nuova cittadinanza”, “Dimensione europea dell'educazione”, “Multimedialità e linguaggi”, “Italiano come L2”. Dai nodi è poi raggiungibile la documentazione relativa a oltre mille progetti didattici, educativi e di intervento culturale già conclusi e validati ed inseriti in un data base ricercabile per parole chiave;

- *Interground* <<http://www.interground.it/>>. Alla base di *Interground* sta il concetto che “scuola e media sono i due grandi sistemi di rappresentazione-interpretazione della realtà e di socializzazione-controllo dei comportamenti”. Per “giocare e studiare con l'intercultura”, *Interground* offre una rete di metodologie frutto di esperienze concrete e materiali didattici realizzati dalle scuole elementari e medie della Puglia nell'ambito di un progetto di ricerca pluriennale svolto in collaborazione con il Ministero della Pubblica Istruzione e l'Unione Europea. I materiali sono organizzati in 39 unità didattiche, 1400 pagine web, 90 file testo, 60 sonori e 20 video – ricchi di interazioni, spunti, approfondimenti, esercizi, giochi e linee guida – facenti capo a sette curricula: “Lingua”, “Religioni”, “Storia”, “Musica”, “Scienze”, “Matematica”, “Media”. Davvero originale la metafora della ‘metropolitana dell'intercultura’, su cui il progetto è incentrato e da cui trae il nome: la rete dell'*interground* è composta di sette linee (i sette curricula di studio) ed ogni fermata corrisponde ad una o più unità didattiche del curriculum corrispondente. La metafora della metropolitana, visualizzata attraverso ricostruzioni in grafica 3D, consente di evidenziare, attraverso gli incroci tra le linee, l'interdisciplinarietà dell'educazione interculturale. Si accede per iscrizione gratuita all'*Interground Club*, al momento di quasi 700 membri fra studenti e docenti;

- *Io Parlo Italiano* <<http://www.educational.rai.it/ioparloitaliano/>>. Progetto promosso dal Dipartimento per gli Affari Sociali, della Presidenza del Consiglio dei Ministri in collaborazione con il Ministero della Pubblica Istruzione, Rai Educational e i Centri territoriali per l'educazione permanente degli adulti (2001). Le quaranta lezioni del corso sono suddivise in due sessioni introduttive

per i tutor, quattro moduli per gli studenti e quattro lezioni conclusive dedicate alle prove di certificazione. Il materiale è visualizzabile in sei finestre, all'interno delle quali è possibile selezionare i singoli moduli ed accedere così ad una scheda informativa che ne organizza gli elementi. La documentazione completa suddivisa per area è a disposizione per il *download* in formato compresso;

- *I.S.MU.* <<http://www.ismu.org/pag.asp?page=132>>. Sito della Fondazione Cariplo per le Iniziative e lo Studio sulla Multietnicità – Regione Lombardia. Struttura di servizio, rivolta principalmente agli enti e alle istituzioni pubbliche, al mondo del volontariato, agli insegnanti, ai singoli operatori e agli studiosi. Disponibile in italiano e inglese. L'attività della Fondazione, costituitasi nel 2001, si articola in quattro aree principali, di cui il sito è specchio: la documentazione ed il monitoraggio; la promozione di ricerche e studi; la formazione, che comprende anche attività di supporto, rivolta a enti pubblici e privati; l'informazione, attraverso l'organizzazione di seminari, convegni, pubblicazioni. Pare evidente, a partire dal 2000, il sostanziale incentivo di internet sul fronte educativo linguistico-letterario. A questo strettamente connesso, è il terzo indirizzo tematico della rete, mirato ad incentivare il potenziale creativo ed a promuovere le iniziative legate alla letteratura italiana della migrazione – produzione letteraria di scrittori stranieri presenti a vario titolo in Italia i quali, per scelta, si esprimono nella lingua del paese ospitante<sup>21</sup>. Dal repertorio di manuali di orientamento e di formazione, passiamo, dunque, all'analisi di risorse che sono espressione già autonoma di identità personale e collettiva. A ragione di un chiaro salto qualitativo e di intenti, e per un mio interesse specifico ai temi letterari, propongo di seguito una valutazione particolarmente dettagliata di cinque esempi significativi.
- *BASILI* <<http://www.disp.let.uniroma1.it/basili2001/>> (1997). Banca dati Scrittori Immigrati in Lingua Italiana realizzata da Armando Gnisci, nel 2001 potenziata e abbinata alla rivista semestrale *Kúma-Creolizzare l'Europa* <<http://www.disp.let.uniroma1.it/kuma/kuma.html>> (n. 1, maggio 2001) con la quale costituisce tuttora uno strumento informativo e critico unico nel genere. In oltre 400 schede, *BASILI* raccoglie i dati bibliografici della letteratura pubblicata in italiano *da e su* scrittori immigrati in Italia: comprende saggi, articoli, recensioni, come pure opere di scrittori la cui migrazione non rientra nel modello principe dell'emigrazione di massa che ha portato in Italia gran parte degli scrittori lì archiviati, ma che fa parte di un fenomeno molto più longevo ed articolato che riguarda i letterati autoesiliatisi, avventuratisi in Italia per diverse ragioni – politiche, innanzitutto – o lì trasferitisi deliberatamente, scegliendo di scrivere e pubblicare in italiano. La redazione rende esplicito l'intento che ispira i fronti di indagine e raccolta di *BASILI*, e vale a dire tessere una rete complessa che possa offrire, in futuro, elementi di riflessione sul rapporto che lega le diverse forme della scrittura alle altrettanto varie tipologie della migrazione.

L'interfaccia della banca dati è essenziale ed efficace. Dalla pagina iniziale risultano chiari i rimandi alle modalità di ricerca e alle informazioni relative alla ragion d'essere ed ai

responsabili di *BASILI*. Tutti i campi di ricerca del *database* sono in relazione logica *AND*. È possibile inviare interrogazioni secondo scrittore ed opera letteraria, secondo critico ed opera critica o secondo tesi universitarie – importante, a riguardo, la considerazione di *BASILI* per i lavori dei giovani studiosi. L'utente nuovo alla scrittura della migrazione potrà inserire all'interno dei vari campi un qualsivoglia termine che verrà utilizzato come chiave di ricerca: la parola sarà confrontata con i dati del campo in qualunque posizione presente. Una volta ottenuti i dati di interesse, selezionando, a seconda del caso, il nome dello scrittore, del critico o il titolo della tesi sarà possibile visualizzare informazioni bio-bibliografiche e critiche complete ed eventualmente selezionare ulteriori rimandi interni.



Fig. 1 *Kúma*. Sezione *Poesia, Il peso delle parole*, di M. Gomes. Testo e traduzione a fronte.

La forma *BASILI&Kúma* si è manifestata, pionieristicamente, “in un momento in cui la coscienza che qualcuno ha avuto ed ha di partecipare ad un movimento epocale di *creolizzazione* della cultura europea comincia a diventare finalmente critica e attiva”<sup>22</sup> e di contaminazioni, sincretismo, ibridazioni e ibridità inedite la rivista (anche cartacea) *Kúma* – ‘parola’ nella lingua bámbara dell’Africa occidentale subsahariana – è senz’altro portatrice. L’interfaccia richiama immediatamente per colori e stile il fratello maggiore *BASILI*, in gradevole linea di continuità. Una testata più descrittiva potrebbe forse favorire il lettore alla prima visita nell’individuazione immediata dei contenuti e degli scopi del sito. La *homepage* offre immediato accesso ai contenuti del numero corrente, mentre dalla barra di navigazione di sinistra è possibile muoversi – non sempre agilmente, si deve dire, a causa dei necessari continui spostamenti verticali – all’interno dei testi pubblicati in precedenza. *Sezioni* ordina in categorie i testi di narrativa, poesia, critica e commento. I generi spaziano dalla letteratura alla musica al cinema e al teatro mantenendo come sfondo i temi della migrazione. Non è del tutto chiaro il criterio di ordinamento dei rimandi all’interno delle sezioni, non essendo alfabetico o cronologico e variando da titolo dell’opera a nome dell’autore. Ciò impedisce la rapida ricostruzione dei numeri arretrati e può disorientare il lettore esperto

alla ricerca di riferimenti precisi. *Kúma* offre la versione originale di un testo accanto alla sua traduzione (vedi Fig. 1) nonché notizie bio-bibliografiche sugli autori. Ad arricchimento del sito, la sezione *Rubriche* aggiorna su novità editoriali, iniziative per l'intercultura, *link* utili per la formazione a livello universitario sui temi della migrazione: il nodo *Decolonizziamoci*, in particolare, offre materiale teorico a supporto e chiarimento della poetica stessa di *Kúma* e del suo Direttore Armando Gnisci:

la decolonizzazione della quale parliamo non va confusa e sovrapposta, però, al “post-colonialismo”, come spesso avviene nei discorsi della cultura contemporanea. [...] Post-colonialismo [...] vuol dire: “a partire dall'avvento del colonialismo”. Il post-colonialismo inizia dal momento in cui un popolo-potere si impossessa di un altro, lo occupa, lo domina e lo sfrutta per il proprio profitto. Decolonizzazione vuol dire, invece, liberarsi dalla malattia del colonialismo venendone via. [...] de-colonizzazione è lotta per la dignità, pedagogia dell'indignazione, grido del risarcimento, creolizzazione e salvaguardia della differenza [...] Di essa sappiamo qualcosa attraverso i poeti e i musicisti. Noi tutti che stiamo intorno all'albero delle parole e siamo pronti a trasformarci<sup>23</sup>.

Chiare le informazioni relative alla redazione e ai curatori del sito – con relativi recapiti – in coda alla homepage. Gli interessati hanno la possibilità di iscriversi ad una mailing list per ricevere aggiornamenti.

- *El Ghibli. Rivista online di letteratura della migrazione* <<http://www.el-ghibli.provincia.bologna.it/>> (n. 0, giugno 2003; 12.859 hits al 10/5/2004). Diretta da Pap Khoumi è una rivista interamente gestita da scrittori migranti e dedicata al tema del viaggio della migrazione. *El Ghibli* – è questo il nome di un vento sahariano – è giunto al quarto numero e vanta una media di oltre mille visite mensili (erano 747 al principio di dicembre 2003) ed un conteggio quasi raddoppiato di *hits* in poco più di due mesi (febbraio-aprile 2004). Si leggeva nella lettera d'invito alla lettura – disponibile in quattro lingue – che il Comitato Editoriale pubblicò in occasione del numero inaugurale:

Da qualche tempo gli scrittori stranieri che vivono in Italia vogliono con più forza porre all'attenzione dei lettori italiani le loro opere. [...] Tuttavia, l'attenzione riservata loro dal mondo della cultura, dell'editoria e della stampa è minima e, nonostante questa produzione letteraria possa ritenersi in espansione, sono pochi gli italiani che hanno letto le loro opere<sup>24</sup>.

Di qui nasceva l'idea di una rivista *on-line* trimestrale, fatta e diretta da stranieri, che coniugasse la leggibilità tipica della stampa e navigabilità web.

Il sito di *El-Ghibli* è ospite dello spazio informatico della Provincia di Bologna, e pur è gestito del tutto autonomamente: ciò ne consente la realizzazione tramite l'uso esclusivo di *software libero*, a garanzia della completa accessibilità ai codici, della gratuità, della stabilità a lungo termine. Alberto Maurizi, 'anima' informatica della rivista, ha potuto informarmi su alcune scelte tecniche alla base del progetto: codifica in html valido (e senza javascript), fruibile anche privo di stile per quei *browser* che non supportano correttamente i fogli css; tecnica *tableless*; e applicazione di css all'html per la rappresentazione grafica ne sono esempio, mirati a favorire l'accesso al maggior numero di persone indipendentemente dalla disponibilità dei mezzi.



- El Ghibli
- Chi siamo
- Il manifesto
- Archivio
- numeri precedenti
- Documenti
- Dicono di noi
- Contatti
- Scrivi alla Redazione
- Seleziona un tema
- Per ricevere aggiornamenti
- Supporto tecnico
- Notizie
- Links
- Cerca nel sito

Anno 0, Numero 3 - marzo 2004

<p><b>editoriale</b> pap khouma</p> <p><b>racconti e poesie</b> iglaso soego tahar larbi helene paraskeva elisa lidani dario edgar gonzales helene oliviera</p> <p><b>stanza degli ospiti</b> rico terzi gruppo SE Bologna saverio fattori</p>	<p><b>edizionale de almeida</b> pereira</p> <p><b>parole dal mondo</b> denise dahamei pablo araujo teodoro lirio</p> <p><b>generazione che sale</b> progetto tea punto verde roma amarachi auzie zein zefrit</p> <p><b>recensioni</b> armonica dell'emigrazione di sara vitali arta' spon'illi, cronaca di una vita in silenzio</p>	<p><b>il ritorno di amanda</b> cinzia atturra e le altre</p> <p>marcoca poesia araba ozzi</p> <p><b>interventi</b> miquel vaie de almeida alexandre portali gebre tyasui gorlu jeshua massareri</p> <p><b>supplemento</b> incontro con jermia shkarova</p>
--	---	--

Fig. 2 *El Ghibli*. Pagina iniziale (n. 3, marzo 2004)

L'interfaccia, stilisticamente affine alla pagina stampata, colpisce subito per eleganza e sobrietà. La *homepage* – aggiornata di recente: il sito è attualmente in fase di profonda ristrutturazione – rende immediatamente visibile l'indice del numero corrente mentre la barra di navigazione è disposta sulla destra in linea con la testata (vedi Fig. 2). I contenuti sono organizzati in sezioni: *Racconti e poesie* ospita la produzione letteraria di scrittori immigrati in Italia che hanno scelto l'italiano come lingua d'espressione; *Stanza degli ospiti* dà spazio a scrittori stanziali – i cosiddetti viaggiatori immobili – di ogni nazionalità; *Parole dal mondo* ospita la produzione letteraria di migranti non italiani, in altri paesi del mondo; *Generazione che sale* pubblica i migliori scritti dei ragazzi delle scuole sul tema della rivista – un'iniziativa, quest'ultima, che mi pare di grande valore ai fini della sensibilizzazione e del coinvolgimento dei più giovani nei temi della pluriculturalità. All'interno di ciascuna sezione i testi sono presentati, dove opportuno, sia in originale che in traduzione italiana, sono corredati da *Note biografiche* riguardanti gli autori e sono leggibili anche a piena pagina grazie ad una comoda *Versione lettura*. Vi sono poi una rubrica dedicata alla recensione e all'analisi storico-critica della produzione degli scrittori migranti (*Recensioni*), un'area di commento e testimonianza su questioni politico-culturali (*Interventi*) e, a partire dal numero 2, un *Supplemento* monografico dedicato all'incontro con un personaggio della cultura. È fra i programmi futuri di *El-Ghibli* – resomi noto da Gabriella Ghermandi, coordinatrice e promotrice

del progetto, in un recentissimo scambio via posta elettronica – l'incremento dei testi in traduzione, anche dall'italiano, a rafforzamento così del proprio carattere multilingue.

La barra di navigazione, sempre presente, propone innanzitutto un'occasione di incontro con i redattori, presentati con foto e notizie bio-bibliografiche. Seguono il manifesto poetico della rivista e l'*Archivio* dei numeri precedenti, i cui contenuti sono presentati in struttura identica, organizzata in sezioni, a quella del numero corrente: un elemento di consistenza che facilita l'orientamento e la navigabilità. *Contatti* offre informazioni ad aspiranti collaboratori e, per chi volesse ricevere aggiornamenti, la possibilità di sottoscrivere la *mailing list* della rivista. Ancora di utilità e interesse sono le rubriche *Notizie*, con novità su concorsi, premi letterari, segnalazioni bibliografiche, manifestazioni culturali, e *Links*, pagina di risorse elettroniche consigliate su letteratura migrante, intercultura, sviluppo sostenibile, informazione alternativa. Un motore di ricerca, al momento disattivato, sembra essere una delle priorità della ristrutturazione in corso. Il piano di ampliamento del sito prevederebbe anche una rubrica 'innovazione-tradizione' sulla letteratura orale con brevi filmati e *banner* di narrazioni, interviste e simili originali aggiunte multimediali, per il momento non ancora concretizzate – mi comunica Gabriella Ghermandi – per questioni prettamente legate ai fondi. Ritornerò in coda sull'auspicabile sperimentazione di interattività e multimedialità da parte delle scritture della migrazione.

- *Il Gioco degli Specchi* <<http://www.ilgiocodeglispecchi.org/>>. È un progetto dei volontari dell'Associazione Trentina Accoglienza Stranieri, sostenuto dal Comune di Trento, dalla Provincia Autonoma e dall'Università di Trento. Propone la letteratura come luogo d'incontro con l'alterità, luogo di riconoscimento e di colloquio. Rivolge particolare attenzione alla letteratura dei migranti, degli immigrati e degli emigrati. Il sito dispone di un ampio *database* che permette interrogazioni per appuntamento (incontro con l'autore, didattica, conferenze, laboratori teatrali, ecc.), opera (secondo dati bibliografici ma anche secondo tematica e tipologia di testo), autore (italiani emigranti o immigrati in Italia), film o fumetti. Nel febbraio 2003 ha organizzato la prima edizione del *Festival di Letteratura Italostraniera* con letture pubbliche, incontri con autori, presentazioni di testi, spettacoli, conferenze, tavole rotonde, rappresentazioni teatrali, rassegne cinematografiche, mostre.
- *Sagarana on-line* <<http://www.sagarana.net/>> (n. 1, ottobre 2000; 850 visite giornaliere, 80.000 visite per numero, +30% di visite ogni numero). Diretta da Julio Monteiro Martins, *Sagarana* conta 15 numeri all'aprile 2004. Il titolo è neologismo di João Guimarães Rosa, a significare 'storia plurima' o 'saga infinita': *Sagarana* è una miscela trimestrale di saggistica, sia letteraria che di attualità, narrativa e poesia – organizzata nelle sezioni permanenti di *Narrativa*, *Saggi* e *Poesie* – e con una specifica rubrica (*Ibridazioni*) coordinata da Armando Gnisci e dedicata al Sud del mondo e agli scrittori migranti. Propone racconti spesso inediti, anche non recenti, di autori esordienti o ben noti che senza la mediazione della rivista il pubblico italiano non avrebbe occasione di leggere. Offre una letteratura che nasce "dall'incrocio meticcio dei generi e ne asseconda la mischia e la disseminazione"<sup>25</sup>. Le sezioni "Spazio Sagarana" e "Spazio Lettori" ospitano i lavori degli allievi della *Scuola Sagarana* di scrittura creativa, lettere, critiche, suggerimenti.

*Ibridazioni* è accessibile dalla *homepage* della rivista, con un clic sulla barra di navigazione a sinistra, che pure offre accesso rapido all'*Archivio* dei numeri precedenti, ad un valido *Indice Autori*, all'elenco redattori, a *Link* consigliati relativi a cultura e letteratura migrante e scrittura creativa, e all'opportunità di sottoscrivere il *Sagarana Group*. Una volta immersi nella consultazione dei numeri precedenti, bisogna dire, non è sempre agevole fare rientro al numero corrente: mancherebbe di fatto un rimando costante e chiaro al punto di partenza. *Ibridazioni* organizza i testi in *Racconti*, *Saggi* e *Poesie* (vedi Fig. 3), li presenta in modo senz'altro pacato ed elegante, facilmente leggibile, spesso abbinando scelte iconografiche di valore. Anche *Sagarana*, dove opportuno, correda i testi di traduzioni e schede informative sugli autori. Data la già elevata consistenza testo della rivista, un sistema di ricerca per parole chiave sarebbe senz'altro conveniente.



Fig. 3 Sagarana. Sezione Ibridazioni (n. 14, gennaio 2004): la piccola freccia precedente il titolo indica l'anteprima della pubblicazione

Di *Sagarana* si apprezzano l'originalità delle scelte, la profonda coerenza interna, la ricchezza di spunti: il risultato sono delle pagine di peso, non canoniche, che accettano la scommessa di sfruttare le potenzialità comunicative della parola in rete. Fu lo stesso Direttore Julio Monteiro Martins a progettare l'architettura originale del sito, poi affidata per la realizzazione a diverse aziende specialiste in *web design*.

Lo scorso dicembre Monteiro Martins si è reso disponibile per un'intervista via posta elettronica sulla nascita e lo sviluppo di *Sagarana*. Alla mia domanda sulla misura e le modalità in cui, secondo la sua opinione, il supporto digitale possa favorire il po-



tenziale creativo delle comunità immigrate in Italia e, inoltre, quanto abbia potuto e possa per alterare le tendenze del sistema informativo di TV, radio e stampa e dirigerle in senso opposto, Monteiro Martins rispondeva:

Senz'altro Internet ha un ruolo decisivo nell'equilibrio necessario riguardo a quest'immagine. È evidente a questo punto che la rete è diventata – magari a scapito dei suoi creatori – il grande, e forse unico vero sistema di informazione alternativa a quella “ufficiale”. [...] Questo accade soprattutto perché la produzione e l'emissione d'informazione attraverso Internet, almeno per ora, ha dei costi relativamente bassi, e quindi sopportati da quei gruppi sociali e culturali che non possono contare sulle abbondanti sponsorizzazioni dei loro detrattori. Se tutto questo sarà o meno in grado di alterare le tendenze neo-conservatrici, o per usare una felice espressione di Pasolini, il “fascismo consumistico” dei nostri tempi, con il suo carico di preconcetti e di razzismo, solo il tempo dirà.

Interrogato sul perché di una rivista elettronica, il Direttore menzionava subito la questione dei costi, ma soprattutto, più importante dei costi, la distribuzione, che è il nodo centrale dell'esistenza di una rivista culturale. Poi, una filosofia di inserimento: a) arrivare dappertutto, anche in altri paesi dove si studia la cultura italiana e dove vi sono lettori, soprattutto nelle università; b) arrivare gratis a tutti quelli che avranno accesso a un computer (e oggi questo accesso è libero nelle scuole e nelle biblioteche). È la democratizzazione assoluta di questo tipo di informazione; c) Aprirsi ai contributi dei lettori in un ambiente interattivo possibile solo attraverso le nuove tecnologie della comunicazione; d) Sperimentare nuove possibilità estetiche e organizzative di un sito internet letterario, ottimizzandone quindi il potenziale.

Ho indagato, inoltre, l'esperienza di Monteiro Martins sul fronte dell'accoglienza di *Sagarana*-rivista elettronica da parte della comunità dei collaboratori e dei lettori. Ho potuto accertarmi del fatto che la rivista è accolta con sicuro entusiasmo dai collaboratori, con interesse e curiosità maggiori da quelli più anziani, i quali, tuttavia, non mancano a tratti di esprimere diffidenza – per questioni di principio e tradizione o per cautela in termini di *copyright*. Il *feedback* da parte dei lettori è ottimo, “i giovani autori soprattutto capiscono subito che si tratta di una rivista seria,” continuava Monteiro Martins, “che avranno una visibilità enorme se pubblicano in essa, una visibilità che oggi come oggi nessuna rivista cartacea in Italia potrebbe offrirgli”. E concludeva: “Senza altro la letteratura della migrazione è quella che trae più beneficio – rispetto, diffusione, spazi di riflessione – dalla pubblicazione in rete”.

- *Voci dal silenzio* <<http://digilander.libero.it/vocidalsilenzio/>>. Nasce da alcune esperienze di lavoro sull'immigrazione realizzate nell'ambito della scuola e del volontariato dal Centro Informazione ed Educazione allo Sviluppo (CIES) di Ferrara, attivo dal 1989 nell'Educazione Interculturale. Ha l'obiettivo di far uscire dal silenzio e dall'anonimato le voci di donne e uomini immigrati attraverso la scrittura. Dalla *homepage* – che richiederebbe senz'altro una revisione grafica secondo gli accorgimenti del *web design* accessibile – è possibile raggiungere le pagine dedicate alle indicazioni bibliografiche su letteratura *dell'*immigrazione e *sull'*immigrazione, una sezione di interviste a scrittori migranti ed una finestra sulle culture della migrazione provvista di articoli di approfondimento. *Voci dal silenzio*, inoltre, ospita racconti e poesie di scrittori migranti anche alla prima esperienza.

## 4. RIFLESSIONI CONCLUSIVE

### 4.1 *Editoria e rete*

L'editoria digitale in Italia sembra rispondere letteralmente alle raccomandazioni della risoluzione parlamentare europea del 2001, ossia dà campo alla creazione letteraria, mantiene e migliora la diversità culturale e linguistica salvaguardandone l'accesso universale<sup>26</sup>. Soprattutto, riserva un'attenzione specifica e promuove la diffusione delle voci dei cittadini immigrati in modo paritario, economico, capillare, ciò che i mezzi tradizionali ancora non sanno fare.

Le riviste culturali telematiche considerate rivestono in tal senso un ruolo del tutto esclusivo in quanto vetrine periodicamente rinnovate e liberamente accessibili in cui tematiche radicali eticamente motivate e sperimentazioni di stile sono avanzati da scrittori esordienti o affermati a pari diritto. Grazie alle tecnologie dell'informazione e della comunicazione – che, si badi bene, insieme all'emergere dei nuovi mercati globali, sono il fattore che sta all'origine dei grandi flussi migratori recenti<sup>27</sup> – la loro letteratura – fino alla metà degli anni Novanta ancora letteratura di testimonianza a interesse sociologico e antropologico ed oggi, al contrario, oggetto d'attenzione di tipo essenzialmente letterario – si sta assicurando la dimensione pubblica e democratica della diffusione, confermata dai dati relativi alle utenze. Una rivincita, viene da chiedersi, su quelle stesse dinamiche internazionali invasive e *colonizzanti* che il mondo-Sud da sempre soffre? Probabile, cosciente o meno che lo sia. Resta il fatto che la rete globale avvicina, interconnette, intensifica la produzione e la comunicazione letteraria, la libera da (pre)giudizi nazionali, linguistici o culturali, ne riconfigura lo spazio sia concettualmente che esperienzialmente: una deterritorializzazione dell'analisi letteraria in favore di procedimenti teorici trans-territoriali che deve rimanere al centro delle riflessioni più generali sulla cultura contemporanea, globalizzata e globalizzante<sup>28</sup>.

### 4.2 *Scuola e rete*

Se è vero, come sostiene Gnisci, che la letteratura della migrazione è un fenomeno valorizzabile “solo da chi si sia educato a una prospettiva interculturale, non-nazionale [...] [che] nasce e cresce insieme a una poetica-politica di decolonizzazione e creolizzazione dell'Europa” e che trova nella migrazione e nelle sue scritture “l'occasione e l'incontro più concreto, decisivo e fruttuoso”<sup>29</sup>, la scuola, e l'insegnamento delle materie letterarie *in primis*, rivestono, ne va da sé, un'importanza sostanziale. La scuola deve saper cogliere gli stimoli che l'editoria elettronica sta proponendo e deve incorporarli nella pratica pedagogica. Abbiamo riscontrato dati recenti senz'altro positivi ed in crescita sul fronte dell'alfabetizzazione informatica anche dei docenti; tuttavia, un valido programma di formazione sull'applicazione delle tecnologie alla didattica è raccomandazione non differibile, mirata ad evitare il rischio di una integrazione auto-didatta improvvisata e solo parzialmente efficace.

### 4.3 *Lingua e letteratura*

L'introduzione dei temi della scrittura dei migranti nei programmi d'insegnamento linguistico-letterario potrebbe rivelarsi estremamente fertile e favorire nuove consapevolezze, culturali come linguistiche. Le riflessioni di Steven Kellman sul translinguismo ci aiutano a inquadrare meglio la questione. Kellman ricorda<sup>30</sup> l'impossibilità di dare per scontata anche una sola parola quando si scrive in una lingua straniera. In tal senso, gli scrittori translingui rappresenterebbero un esempio radicale di quella che i formalisti russi consideravano essere la qualità distintiva di ogni letteratura, e vale a dire "estraniare", rendere gli oggetti difficilmente riconoscibili, prolungare il processo estetico della percezione. Scrivere in una lingua straniera è, in qualche modo, distaccarsi e distaccare dall'espressione verbale e certo il lavoro dei translingui – come e più di quello dei colleghi madrelingua – lancia sfide continue al suo stesso medium. È ancora Kellman a sottolineare come gli scrittori migranti siano, di fatto, "the shock troops of modern literature, and those avant-garde movements that, like dadaism, surrealism, and futurism, have been most insistent on the inadequacy and treachery of conventional speech"<sup>31</sup> e sull'occorrenza di espandere i limiti del linguaggio. L'unicità, poi, del caso della letteratura migrante in italiano sta propriamente nell'elezione da parte degli autori "di una lingua senza un passato coloniale tale da ricondurla al filone delle letterature post-coloniali – con cui condivide comunque molti risultati espressivi – scelta al di là di qualsiasi imposizione implicita o esplicita, che viene conquistata e fatta propria con più difficoltà, più lentamente e laboriosamente, e proprio per questo rinnovata molto più radicalmente"<sup>32</sup>. Gli esiti di questa scelta stanno contribuendo ad aprire nuovi territori linguistici e identitari, accolti e diffusi nello spazio della rete insieme e parallelamente alle zone franche dell'inglese. Si tratta di scenari che contribuiranno a riconfigurare i parametri critici con cui si è giudicata e classificata sin qui la stessa letteratura italiana, ora, è evidente, costretta a ripensare alla propria ragion d'essere e ad assimilare, senza appiattare, le *altre* letterature innestate nella sua lingua.

### 4.4 *Decentralità e ipertesto*

La proprietà fondante lo spazio ed il testo digitali è, lo sappiamo, la decentralizzazione del design: l'ipertesto mette a disposizione i mezzi adatti a stabilire un infinito numero di centri, per espanderli ed alterarne le relazioni<sup>33</sup>. Ebbene, di decentralità o, se si vuole, di passaggi di centralità, è evidentemente piena l'esperienza di mobilità dello scrittore migrante. Lo abbiamo visto or ora del resto, per lo scrittore migrante la creazione di una nuova voce è anche creazione di una nuova identità, e di un nuovo territorio, interiore ed esteriore. La sfida continua, personale come artistica, dello scrittore migrante ai limiti semiotici del linguaggio troverebbe, credo, un ulteriore superamento nel pieno sfruttamento della fluidità della scrittura elettronica, per sua stessa natura tesa verso sempre nuove forme significanti. Mi riferisco qui alle ambiziose convinzioni dei teorici dell'ipertesto, secondo i quali la *cybertextuality* non sarebbe altro che un punto di arrivo (e di nuova partenza) di quella persistente ricerca di innovazione linguistica che segna la storia della cultura da oltre due secoli. Marie-Laure Ryan richia-

ma la ricerca di linguaggi totalizzanti e multimediali propria dell'opera ottocentesca e del teatro brechtiano; la spinta alla democratizzazione dell'arte del Dadaismo e del Surrealismo; l'ambita multisensorialità della poesia modernista; fino all'ideale dell'intertestualità universale del postmodernismo: tutti obiettivi 'incorporati' e realizzati nella struttura multireferenziale e nel carattere interattivo dell'ipertesto<sup>34</sup>. Vedo allora una non trascurabile affinità di condizione e tensioni fra autore migrante e parola digitale – creatore e portatore di contenuto – che potrebbe certamente evolversi in forme ancor più feconde e innovative se il primo non utilizzasse la seconda unicamente come canale di distribuzione bensì ne sperimentasse tutte le potenzialità espressive.

Attualmente, lo spazio digitale dedicato ai temi della migrazione sembra ricoprire la funzione base di supporto e vetrina di contenuto a basso costo, ovunque fruibile. All'interno delle riviste telematiche considerate – seppur si siano registrate spinte innovative per l'immediato futuro – le risorse tecniche ed ergonomiche dell'ipertesto risultano sfruttate solo parzialmente, in strutture ramificate relativamente complesse provviste di *link* soltanto unidirezionali. Per il momento sembrerebbe rimanere priorità il livello informativo, la diffusione, l'affermazione della propria voce, la ricerca di *feedback* – si pensi alla costante presenza di sezioni accurate di notizie, rimandi di approfondimento, di *mailing list* e spazi-forum e, d'altra parte, all'incertezza, a volte, degli strumenti di navigazione.

Bene, l'emancipazione dalla necessità del mezzo e la sfida creativa allo stesso – risorse materiali ed umane permettendo, questione non irrisoria – porterebbe a risultati senz'altro originali: penso all'introduzione dei media audio e video, ad una codifica testi più sofisticata che permetta interrogazioni altrettanto complesse, penso anche a sperimentazioni di *e-fiction* ed *e-poetry*, di narrativa e poesia interattiva. Una riflessione teorica più profonda su ipertesto e scritture migranti potrebbe persino rimuovere ulteriori barriere “di potere”<sup>35</sup>, questa volta dalle mani dell'autore in direzione del lettore, per ora fruitore sì, e interprete, ma anche potenziale partecipante della *mobilità* della scrittura migrante nel suo compiersi sullo schermo.

## 5. RIFERIMENTI BIBLIOGRAFICI

- Argento F. (2002). La letteratura della migrazione in Italia. In M. Calabrese, C. Peverati, P. Trabucco (eds). *Atti del Primo Convegno Nazionale Culture della Migrazione e Scrittori Migranti*. Ferrara  
<http://digilander.libero.it/vocidalsilenzio/laletteraturadellamigrazioneargento.htm>.
- Censis-Unione Europea (acd) (2002). *Tuning into Diversity. Immigrati e minoranze etniche nei media*. Roma, Fondazione Censis  
<http://www.stranieriinitalia.com/briguglio/immigrazione-e-asilo/2002/maggio/ricerca-censis-media.html>.
- Chaloff J. (1999). Current Research into Education for Immigrants in Italy. In collaborazione con Child Immigration Project (CHIP) e Fondazione Censis, Roma, Fourth International Metropolis Conference, Washington, DC, 9/12/, p. 2  
[http://www.international.metropolis.net/events/washington/chaloff\\_paper.doc](http://www.international.metropolis.net/events/washington/chaloff_paper.doc).

- Dahan M., Sheffer G. (2001). Ethnic Groups and Distance Shrinking Communication Technologies. *Nationalism & Ethnic Politics*, VII, 1, pp. 85-107: 105.
- Danet B., Herring S. C. (2003). Introduction: The Multilingual Internet. *Journal of Computer-Mediated Communication*, 9, 1  
<http://www.ascusc.org/jcmc/vol9/issue1/intro.html>.
- Fiormonte D. (2003). *Scrittura e filologia nell'era digitale*. Torino, Bollati Boringhieri.
- Fondazione Censis (acd) (2002). *L'immagine degli immigrati nei media: confinati dentro il ghetto della cronaca*. Roma  
<http://www.immagineimmigratitalia.it/rapportocensis.html>.
- Gnisci A. (2001a). Basili & Kúma. *Sagarana*, 4  
<http://www.sagarana.net/rivista/numero4/ibridazioni8.html>.
- Gnisci A. (2001b). Decolonizzazione. *Kúma*, 1  
<http://www.disp.let.uniroma1.it/kuma/rubriche/decolonizziamoci/Decolonizzazione.html>.
- Gnisci A. (2001c). La Sezione Ibridazioni. *Sagarana*, 2  
<http://www.sagarana.net/rivista/numero2/gnisci.html>.
- Gnisci A. (2003). *Creolizzare l'Europa. Letteratura e migrazione*. Roma, Meltemi.
- Grillo R. (2002). Immigration and the Politics of Recognizing the Difference in Italy. In R. Grillo, J. Pratt (eds). *The Politics of Recognizing the Difference: Multiculturalism Italian-style*, pp. 1-24: 10. Adershot, Ashgate.
- Kellman S. G. (2000). *The Translingual Imagination, Lincoln and London*, p. 29. Nebraska, University of Nebraska Press.
- King R., Andall J. (1999) The Geography and Economic Sociology of Recent Immigration to Italy. *Modern Italy*, IV, 2, pp. 135-158: 142.
- Landow G. (1992). Hypertext 2.0. *The Convergence of Contemporary Critical Theory and Technology*, p. 89. Baltimore, The Johns Hopkins University Press.
- Lecomte M. (2004). Una mappatura di cicatrici. *Sagarana*, 14  
<http://www.sagarana.net/rivista/numero14/ibridazioni3.html>.
- McGann J. (2001). *Radiant Textuality. Literature after the World Wide Web*, pp. 70-74. New York, Palgrave.
- Ministero dell'Istruzione, dell'Università e della Ricerca. (2002). *Alunni con cittadinanza non italiana*, pp. 22-24. Roma  
[http://www.edscuola.it/archivio/statistiche/nonit01\\_02.pdf](http://www.edscuola.it/archivio/statistiche/nonit01_02.pdf).
- Ryan M. L. (1999). Introduction. In M. L. Ryan (ed.). *Cyberspace Textuality: Computer Technology and Literary Theory*, pp. 1-28. Bloomington and Indianapolis, Indiana University Press.
- Tosi A. (2001). *Language and Society in a changing Italy*, pp. 35-39. Clavedon, Multilingual Matters.
- Zamora L. P. (2002) Comparative Literature in the Age of 'Globalization. CLCWeb: *Comparative Literature and Culture: A WWWeb Journal*, 4, 3  
<http://clcwebjournal.lib.purdue.edu/clcweb02-3/zamora02.html>.

## 6. NOTE

- <sup>1</sup> Ringrazio Paola Casella, Gabriella Ghermandi, Armando Gnisci, Julio Monteiro Martins ed Alberto Maurizi. A loro sono riconoscente per lo scambio fruttuoso di informazioni e idee e per la squisita disponibilità.
- <sup>2</sup> Dati Caritas, cfr. I flussi di immigrazione in Italia alla luce dell'ultima regolarizzazione, a cura di Caritas Italiana, Caritas di Roma, e Fondazione Migrantes su dati del Ministero dell'Interno <<http://www.caritasroma.it/notizie/comunicati/Comunicati/Scheda%20regolarizzazione.pdf>>.
- <sup>3</sup> King e Andall rilevano 45 nazionalità aventi almeno 5.000 rappresentanti ciascuna. Cfr. King R., Andall J. (1999).  
Chaloff (1999), p. 2.
- <sup>4</sup> Ibidem.
- <sup>5</sup> Ministero dell'Istruzione, dell'Università e della Ricerca (2002), pp. 22-24.
- <sup>6</sup> Risoluzione del Parlamento Europeo sulle nuove frontiere nella produzione libraria: editoria elettronica e stampa su richiesta (2000/2037(INI)), GU C267 del 21/9/2001, p. 84 <[http://europa.eu.int/eur-lex/it/archive/2001/c\\_26720010921it.html](http://europa.eu.int/eur-lex/it/archive/2001/c_26720010921it.html)>.
- <sup>7</sup> Ivi, p. 86.
- <sup>8</sup> Legge 7 marzo 2001, n. 62, art. 1, comma 1 <<http://www.camera.it/parlam/leggi/010621.htm>>. L'uscita della norma suscitò vari orientamenti interpretativi volti a determinare se la pubblicazione di siti Internet – o nei siti Internet – dovesse essere indiscriminatamente assoggettata agli adempimenti previsti per i prodotti a stampa (vedi legge 47/1948 sulla stampa). Il dibattito sembrò sostanzialmente risolto dal D.Lgs. 9 aprile 2003, n. 70 <<http://www.parlamento.it/parlam/leggi/deleghe/03070dl.htm>> che all'art. 7, comma 3, dispose: "La registrazione della testata editoriale telematica è obbligatoria esclusivamente per le attività per le quali i prestatori del servizio intendano avvalersi delle provvidenze previste dalla legge 7 marzo 2001, n. 61".
- <sup>9</sup> Grillo R. (2002), pp. 1-24: 10.
- <sup>10</sup> Per considerazioni approfondite su educazione linguistica e minoranze etniche in Italia, cfr. Tosi, A. (2001), pp. 35-39.
- <sup>11</sup> Censis-Unione Europea (acd) (2002). <<http://www.stranieriinitalia.com/briguglio/immigrazione-e-asilo/2002/maggio/ricerca-censis-media.html>>.
- <sup>12</sup> Fondazione Censis (acd) (2002) <<http://www.immagineimmigratitalia.it/rapportocensis.html>>.
- <sup>13</sup> Extracommunication. Iniziativa Comunitaria Equal, Progetto Etnequal Social Communication (IT-S-MDL-144) (2003). A cura del Dipartimento di Sociologia e Comunicazione, Università degli Studi "La Sapienza", Roma, 2003. <[www.etnequal.it/docs/sintesi\\_dati\\_Extracommunication%20ver%205.0.pdf](http://www.etnequal.it/docs/sintesi_dati_Extracommunication%20ver%205.0.pdf)>.
- <sup>14</sup> On-line/More Colour in the Media: The European Day of Media Monitoring. Media and minorities – from exclusion to active participation <<http://www.multicultural.net/edmm/index.htm>>. Per i risultati dell'indagine condotta in Italia, curati dal Cospe e pubblicati on-line nel marzo 2004, cfr. <[http://www.multicultural.net/edmm/cospe\\_report.pdf](http://www.multicultural.net/edmm/cospe_report.pdf)>.
- <sup>15</sup> In questa sede la programmazione televisiva satellitare non sarà considerata. Mi atterrò ai criteri d'indagine di Censis-Cospe ed Etnequal e, dunque, vaglierò solo i palinsesti di canali di informazione nazionale più largamente fruiti.
- <sup>16</sup> Cfr. Censis Unione Europea (acd) (2002), cit., parte III.
- <sup>17</sup> Cfr. nota 14.
- <sup>18</sup> Il Rapporto è del 30 marzo 2004, vd. <<http://www.aie.it/ufficiostampa/visualizza.asp?ID=751>>.
- <sup>19</sup> Cfr. Danet B., Herring S. C. (2003).
- <sup>20</sup> Gli esempi che fornirò di qui in avanti – in ordine alfabetico – non hanno pretese di esaustività quanto di rappresentatività. Il criterio di selezione adottato ha favorito ricchezza, autorevolezza ed aggiornamento dei dati, chiarezza di presentazione, livello di usabilità e navigabilità web. La data ultima di consultazione delle risorse risale al 1/5/2004. Per un sommario più ampio di siti dedicati a statistiche e rapporti relativi non solo all'Italia ed a risorse per la didattica, si veda Rai Educational:
- <sup>21</sup>

<<http://www.educational.rai.it/corsiformazione/interculturala/link/link01.htm#1>>; per un'ampia lista di risorse sulla letteratura della migrazione si veda <<http://www.cestim.org>>, sezione Letteratura della migrazione.

- <sup>22</sup> Per una breve storia della letteratura della migrazione in Italia si veda Argento F. (2002). <<http://digilander.libero.it/vocidalsilenzio/laletteraturadellamigrazioneargento.htm>>.
- <sup>23</sup> Momentaneamente, il rimando alla lettera d'invito al Numero 0 di El-Ghibli non risulta disponibile in rete.
- <sup>24</sup> Gnisci A. (2001b).
- <sup>25</sup> Gnisci A. (2001a).
- <sup>26</sup> Gnisci A. (2001c).
- <sup>27</sup> Dati regolarmente aggiornati su periodici e pubblicazioni elettroniche e, in genere, sull'attività in rete dell'editoria italiana, elettronica e commerciale, sono resi disponibili dall'Osservatorio sull'editoria elettronica italiana, attivato nel settembre 2003 da Casalini Libri Digital Division. Al 2/4/2004, cui risale l'ultimo monitoraggio, l'Osservatorio contava 688 periodici elettronici, curati da università, editori tradizionali, o singoli: lo spoglio evidenziava la minoranza dei periodici per i quali la versione elettronica fosse quella di riferimento; spesso si trattava ancora di un servizio per gli abbonati all'edizione cartacea; in aumento risultava la fruizione a pagamento, anche se la maggior parte delle iniziative manteneva l'accesso libero. Cfr. <<http://digital.casalini.it/osservatorio/>> (accesso via sottoscrizione annuale; periodo di prova gratuito disponibile).
- <sup>28</sup> Cfr. Zamora L. P. (2002).
- <sup>29</sup> Opportuno, a riguardo, il riferimento alla recente crescita dell'utilizzo delle tecnologie – in relazione a numero di soggetti e frequenza – da parte di gruppi etnici territorialmente concentrati o dispersi. Grazie ai nuovi media queste minoranze hanno la possibilità di incrementare la comunicazione fra membri, conservare le proprie tradizioni culturali, la propria storia e, in particolare, la lingua madre. L'attività di alcuni verte anche su questioni politiche di giustizia sociale e diritti per le minoranze con target di ascolto precisi quali Ong locali e globali: ed è certo uno dei risultati politici di queste network l'indebolimento del concetto di "nation-state", in favore di "ethnic intrastate and trans-state ethnic groups". Cfr. Dahan M., Sheffer G. (2001), pp. 85-107: 105.
- <sup>30</sup> Gnisci A. (2003), pp. 11-12.
- <sup>31</sup> Kellman S. G. (2000), p. 29.
- <sup>32</sup> Ivi, p. 31.
- <sup>33</sup> Lecomte, M. (2004). At <<http://www.sagarana.net/rivista/numero14/ibridazioni3.html>>. Si veda anche Gnisci A. (2003), cit.
- <sup>34</sup> Cfr. McGann J. (2001), pp. 70-74.
- <sup>35</sup> Cfr. Fiormonte D. (2003), in part. cap. 2; Ryan M. L. (1999), pp. 1-28.
- <sup>36</sup> Landow G. (1992), p. 89.

# Envolving to ubiquitous e-learning environments: a collaborative writing scenario\*

*Pedro P. Sánchez-Villalón, Manuel Ortega, Asunción Sánchez-Villalón*

Departamentos de Filología Inglesa e Informática

Universidad de Castilla La Mancha - EOI Ciudad Real

(ppsanch@fimo-cr.uclm.es, manuel.ortega@uclm.es, asuncion.sanchez@uclm.es)

*Following the European Community directives on language education, learning a language should focus on “developing skills such as the ability to communicate, to work in group, to identify and solve problems, and to make use of the new technologies for all this”. For this purpose, the European Union have in place an eLearning Programme which regards language learning as preparation for the active use of the language for communication, as social interaction. Following Vygotsky (1978), learning should be closer to socialization, where an individual internalizes language by collaborating in common activities and by sharing the means of communicating information: language knowledge, speech acts and socio-cultural conventions, as well as technical tools necessary to communicate. In this sense the Common European Framework of Reference (2001) has defined new basic skills (including ICT skills and foreign languages, among others) and more integrating techniques (referred to as Interaction, including group work, collaborative projects and real communication). We propose a Technology-Enhanced Language Learning (TELL) environment, with real-world computer-based devices and tools to access information and to communicate synchronously and asynchronously in a pervasive way, which provides the appropriate learning interaction. It is the AULA system: a ubiquitous collaborative elearning environment for the learning of collaborative writing as an active skill.*

*Keywords:* e-Learning, collaborative writing, Technology-Enhanced Language Learning

## 1. INTRODUCTION

Considering the European directives about the new learning environments required, we need more interactive techniques than the ones in the conventional classroom (European Commission, 2002). The so-called active skills (Speaking and Writing) are half-achieved in the conventional classroom. The widely established communicative approach seems quite effective as to the skill of Speaking: we practise pair work mainly in conversations and oral activities alike. But in Writing this is neglected. The



students of languages mainly write with an active communicative purpose when they do exams, but they rarely write true letters or texts for a specific audience.

The Writing skill has been used mainly as a means of evaluating the students' lexico-grammatical competence, a technique for individualized assessment, where the student is isolated, without access to information and unable to communicate with other students. This does not reflect the application of writing in the real world, where when we write formally we use technologically enhanced devices, we have access to information and we make use of communication tools. We can extrapolate this situation to the world of work, where it is necessary to write reports frequently collaborating with members of other departments.

Consequently, the learning environment has to evolve, reflecting this challenging change in our society. The first reaction (CBI and CALL) was to transform the conventional classroom into a computer language lab, full of desktop computers with digitized drills and simulated language activities. Although in the last decades the computer has firmly burst into language teaching improving the methodology application, the application of this technology (CMC) has been scarce in the field of writing. Loads of exercises and activities have been transferred from the textbooks onto CD-ROMs and over the Internet, mirroring printed materials in digital support. They took the advantage of the attractive motivation it implied, but the drilling technique remained the same, neglecting the use of the authentic communicative facility computers can provide. Moreover, the integration of computers in the traditional classroom has been difficult.

Not only the technological components used in the classroom should be close to those commonly used in real life for communication and to access information, but also the techniques used to develop communication should be similar. The separation of these two worlds is critical: the computer and the rest of the classroom. This seems to hinder the activity in the classroom. Why not integrating them? The paradigm of ubiquitous computing allows us to break down this borderline fading the computational environment into the classroom. In this way, the students interact with the system without really being aware. They are in their natural scenario where they interact: a classroom with classmates, the whiteboard and the teacher. Is there a better scenario than the real one enhanced with the use of the latest writing tool (the computer as a notebook), with access to digital libraries and to the largest source of information (the Internet), and at the same time with real communication tools (a chat facility)?

The emergence of "ubiquitous computing" (Weiser, 1991) with small computational devices, interactive whiteboards and "wearable computers" as electronic notebooks, will finally make the computer unnoticeably pervade, "disappear", in the learning environment. In addition to the computer technology, Web-based and networked communication establishes new ways of social interaction, which makes Collaborative Learning essential as the new educational paradigm in the development of writing skills. Thus, we can use the Internet (and any network) as the means of communication and we make use of mobile computing devices to facilitate the collaborative writing activities in a ubiquitous online learning environment (Ortega et al., 2001a, 2001b).

## 2. TECHNOLOGY-ENHANCED LANGUAGE LEARNING WITH AULA

So, it is not another load of materials and content which we present, but a system which enables the practice of writing as an integrating learning skill. We propose a system which facilitates the active use of language for communication, developing language skill-based collaboration like in the real world. This is implemented by means of a Technology-Enhanced Language Learning (TELL) environment (Warschauer, 1997; Warschauer & Healey, 1998), with real-world computer-based tools to access information and to communicate, which provides the appropriate interaction, with the features of accessibility, and ICT and language tools integration. The TELL environment enables communication from anywhere (the so-called Mobility or M-Learning), communication both in real time and at any time, and pervasive access to the system, to information and to ICT tools (Ubiquity).

In the TELL classroom we can have an electronic whiteboard (a screen with a projector (a beamer) or a mimio® whiteboard), e-notebooks (Portable PCs, TabletPCs or PDAs), and access to the Web or to an Intranet (Wireless Connection to a server). With these elements we can transform any traditional classroom into a ubiquitous TELL classroom.

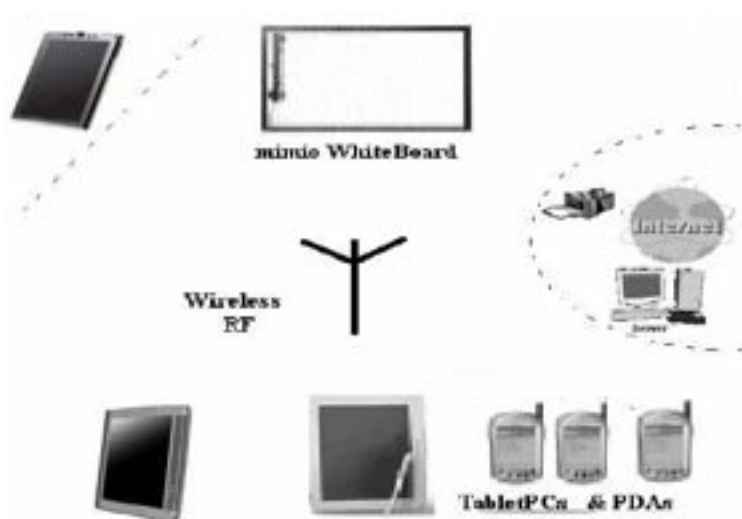


Fig. 1 AULA: a ubiquitous collaborative elearning environment

With this technology, we have established the AULA (A Ubiquitous Language Appliance) infrastructure, an online language learning environment for the learning of collaborative writing as an active skill. The technique consists in writing texts in group within assigned scenarios. This environment involves a pedagogical innovation: Writing a composition evolves from being a evaluation technique to being a learning tool based on a real social activity.

### 3. THE AULA SCENARIO-BASED WRITING PROCESS

We can now present a sample scenario: the AULA Newspaper, where the students are assigned the activity of writing news items for an online newspaper. The teacher guides the learners, who make use of a set of collaborative tools to communicate, to access information, to edit, to share the working space, to revise and correct. The writing activity follows the didactic methodology of the collaborative writing process, divided in three stages:



Fig. 2 The AULA Newspaper Scenario

#### *1st stage: Planning*

1. The teacher establishes a Scenario (e.g., a online newspaper, a travel brochure, film reviews, etc.), determining the conditions and pointing out the features of collaborative writings.
2. The participants propose titles within the scenario and create the documents. They generate ideas in notes by brainstorming or freewriting in the documents. For this purpose, access to information online is provided.

#### *2nd stage: Development*

1. At this stage, the discussion process takes place to select the most appropriate and relevant aspects and ideas proposed. The system provides synchronous and asynchronous tools to help the communication between the participants (a chat facility). Proposals are accepted or rejected under a chat-based discussion session.
2. Then, they write the accepted elements into text, with the aid of the language tools available: a grammar analysis tool and lexical options (dictionaries, a lexicon and translation tools). They write the notes (ideas) into sentences and organize them into paragraphs (aspects). The system proposes the structuring



take notes and write in what seems their small notebooks and frequently direct their attention to the whiteboard, where the teacher presents the information or where a wider view of the resulting writing process can be observed. This way the traditional elements do not disappear, but they evolve thanks to electronic devices, which are pervasively integrated in the learning environment of the classroom. The system is also accessible from outside the classroom.

Learning to write with AULA is based on scenarios, integrating the context, and a possible authentic audience for the purpose of the resulting writings. All in all, AULA makes the traditional learning evolve to the ubiquitous classroom, where students learn by sharing resources and information, by collaborating as a reflection of the future working life of the learner.

## 5. CONCLUSION

It is the application of the Technology-Enhanced Language Learning (TELL) through the online connection of diverse computational devices that seems appropriate for a really communicative environment. And the Internet is the widest connection. An online learning environment (e-Learning) offers the opportunities for authentic, communicative language activities. Under this perspective we have developed a system, AULA, an e-Learning skill-based environment for Collaborative Writing in an online synchronous and asynchronous scenario where ideas are shared, access to information is provided and electronic communication and edition tools are used. It integrates communication, information search and language tools to help the students write naturally in a collaborative way, as it is usual in the real world when writing formally. The system makes these tools permanently accessible to the student in a pervasive way. The chat facility allows both synchronous and asynchronous communication. The search tool gives access to information on the Web or in a local or networked drive. The language tools are based on an electronic dictionary that consists of a lexicogrammatical knowledge base of the target language (English in the prototype, easily adaptable to other European languages). AULA provides hyperlinked lexical information at every stage in the writing process, from the brainstorming phase to the final draft phase, and grammatical analysis at the final stage of revision. This accessibility enhances usability. The system includes a tracking mechanism, so that the teacher assesses every student's progress and participation in the writing activity.

Thus, the traditional learning scenario evolves to a ubiquitous elearning scenario which fades the classroom world and the real world by means of computational features.

The AULA system is at <http://chico.inf-cr.uclm.es/ppsv/aula.html>.

### *\*Acknowledgements*

This model is being developed thanks to the PBI-02-026 Project funded by the Junta de Comunidades de Castilla La Mancha and the Spanish Ministerio de Ciencia y Tecnología through the TIC2002-01387 Project, as well as thanks to the Junta de

Comunidades de Castilla La Mancha through the Joint Project between the Universidad de Castilla La Mancha and the Escuela Oficial de Idiomas de Ciudad Real.

## 6. REFERENCES

- <http://www.upgrade-cepis.org/issues/2001/5/up2-5Ortega.pdf><http://www.upgrade-cepis.org/issues/2001/5/up2-5Ortega.pdf>European Commission (2002). *Towards a knowledge-based Europe*.  
<http://europa.eu.int/comm/publications/booklets/move/36/en.pdf><http://europa.eu.int/comm/publications/booklets/move/36/en.pdf>
- <http://www.upgrade-cepis.org/issues/2001/5/up2-5Ortega.pdf>  
<http://www.upgrade-cepis.org/issues/2001/5/up2-5Ortega.pdf>
- Ortega M. et alii (2001a). Ubiquitous Computing and Collaboration: New paradigms in the classroom of the 21st Century. In M. Ortega, J. Bravo (eds). *Computers and Education: Towards an Interconnected Society*, pp. 261-273. Dordrecht, Kluwer Academic Publishers.
- Ortega, M. et alii (2001b). AULA: A Ubiquitous Language Teaching System. *Upgrade*, vol. 2, n. 5, pp. 17-22.
- Vygotsky L. S. (1978). *Mind in society: The development of higher psychological processes*, Cambridge MA, Harvard University Press.
- Warschauer M. (1997). Computer-mediated collaborative learning: Theory and practice. *Modern Language Journal*, Vol. 81, n. 3, pp. 470-481.
- Warschauer M., Healey D. (1998). Computers and language learning: An overview. *Language Teaching*, vol. 31, pp. 57-71.
- Weiser M. (1991). The computer for the twenty-first century. *Scientific American*, vol. 265, n. 3, pp. 94-104.



# About Writing in English for the Workplace

*Edward Tosques*

Università degli Studi di Firenze

(edward.tosques@unifi.it)

*The need to acquire written skills in today's Europe and in the globalized professional environment is in contrast to the general tendency to privilege spoken and image-based communication. However, the formal aspects which make writing a communication medium distinct from speaking is just as relevant as ever in any upper-level career endeavor. Interactive multimedia appear as a particularly appropriate vehicle for the acquisition of writing skills. The paper presents a corpus of computer-assisted learning materials for the acquisition of writing skills in English, implemented by the author during his teaching experience at the University of Florence. The framework responds to the general orientation of the Italian university reform. The organizational structure of Writing in English for the Workplace features a compendium of text-types related to four major areas (Mass Communications, Business, Research, Technology). Pull-down menus offer access to a range of style topics which include: Grammar, Rhetoric, Paragraph structure, Vocabulary building, Punctuation, Bibliography, Note-taking, Editing.*

*Keywords:* English L2, Writing skills, e-Learning

## 1. PREMISE

Writing in English for the Workplace is a corpus of computer-assisted learning materials for the acquisition of writing skills in English. The work, or that part of it completed to date, has grown out of my experience as a language instructor in the degree program in foreign languages of the Italian University system. As such, its immediate aim is to respond to the general orientation of the Italian university reform which has been in progress for the past several years, a reform which, despite setbacks and detours determined by the sometimes quirky shifts of government policy, has nevertheless maintained its course toward an educational preparation that is markedly skills-oriented, sensitive to what students actually need to learn to be able to compete effectively in today's Europe-wide and globalized professional environment.

My decision to concentrate on writing skills was dictated by the perception, shared by most educators and confirmed by statistics on learning proficiency, of a growing difficulty students have in dealing with the written word, in relation to the foreign languages they study as well as to their own native language, a phenomenon encountered with remarkable uniformity throughout the Western world. This trend of "impaired literacy" is often blamed on the shortcomings of teachers or educational



systems, but the root causes lie further upstream. It has been well-noted that communication in today's mass-media dominated society is primarily image-based, and that the role played by language, already ancillary, privileges the spoken word over the written. Previous ages were also distinguished by the primacy of pictographic or oral communication, but those pre-literary forms did not violate the principles of syntax; rather, they were the natural precursors of the codifications of syntax and grammar established with the invention of writing. In contrast, the "new Gutenberg revolution" of mass-media communication encourages a process of disjunction of the linear structure of discourse. Mass-media entertainment and advertising rely more and more on visual and word effects which are deliberately disarticulated, fragmented, exclamatory (rife with subliminal hints and subtexts craftily placed to sell the featured product) at the expense of consecutive and rational exposition. This studied discontinuity of discourse is further aggravated by the instantaneous access to multiple message sources and their haphazard juxtaposition (channel hopping, navigating on the internet), whose cumulative effect is to stunt one's ability to focus attention for very long on any one thing and elaborate it much beyond the phase of raw sensory data. No wonder that the generations nurtured on the mass-media have increasing difficulty in coping with the written word and especially with the act of writing.

How important is it in today's world to know how to write? To answer the question properly a distinction needs to be made between informal and formal writing. Almost everyone learns how to write, but such writing is confined almost entirely to the expedient shorthand supplements to speech communication, such as e-mail, SMS and chatline messages. This level of writing makes up a massive volume, quite possibly surpassing any other period of human history, but because of its essentially conversational nature, marked by a dialogue-type interplay that is barely less extemporaneous and elliptical than that of actual speech, it is conspicuously devoid of the formal aspects which make writing a communication medium distinct from speaking in its capacity to correlate syntactical units into complex, abstract, rationally ordered schema. The phenomenon manifests itself most directly in the learner's frequently blithe obliviousness in regard to the need to shift register and syntactical parameters where formal writing is called for. Even university-level students tend nowadays to make an appallingly spotty use of the basic elements of punctuation, capitalization and paragraph structure, not to mention the more recondite aspects of sentence structure and style, and when called to task correct themselves, as they are able, more to placate the teacher than from any real conviction about its appropriateness. This problem would be merely academic if formal writing occupied just a marginal place among the communication skills needed in the professional sphere, but the rub is that we live in a world whose primary commodity is as never before information, and that most of this information is transmitted, directly or indirectly, by means of the written word and according to very precise protocols. It has been noted that modern society is coming to be composed on the one hand of a mass of relatively inert, semi-literate consumers and on the other of an elite of technocrats, many of whom "manipulators of information" somewhere in the ratio of 80% to 20%. Educators, most of whom

work in the public sector and who therefore have a natural democratic leaning toward guaranteeing the greatest number of their students a decent scholastic preparation that will enable them to lead dignified and productive lives, can perhaps do little in the long-run to thwart the macroscopic effects of this “Orwellian” process, but they can work out minimalist strategies of counteraction. One of these, I think, is to relate the acquisition of writing skills to specific professional contexts, in as realistic and updated a manner as possible, in order to equip students with a technical-expressive knowhow crucial for succeeding in their careers and for improving the society in which they live. Another is to make instruction available beyond the physical and temporal limits of traditional educational settings by implementing the concept of “life-long learning”. A third, instrumental in accomplishing the other two, is to experiment and selectively exploit technical innovation.

## 2. ADVANTAGES OF E-LEARNING

These were the considerations that led me to adopt interactive multimedia as a particularly appropriate vehicle for the acquisition of writing skills. It seemed especially apt for learning how to write in English, in the light of the key role English plays as “world language” involved in all professional, cultural and economic spheres. Then, given the potentiality of interactive multimedia as an instrument for autonomous study, it seemed especially apt in the flexibility of its applications, whether in the context of traditional learning insitutions, professional training programs or individual self-study, capable of reaching persons of all ages and in all phases of their professional lives. Lastly, it seemed especially apt as an amplifier of the educational process, in its capacity to provide a vast number of learning hours beyond what is possible in traditional didactic settings, to differentiate the pace and paths of the learning process according to individual aptitudes and proficiency levels, to concentrate with patient persistence on details, the sum of which makes a significant difference in the achievement of learning goals, to provide immediate response to student input, thus accelerating and more effectively reinforcing the learning process, and to simulate through the use of mixed-media realistic workplace situations.

It must be specified that, apart from its captivating expositional means, the great strength and together the great limitation of interactive multimedia as a didactic tool lies in its capability of judging only relatively circumscribed ranges of right and wrong answers. Thus it can judge with extreme rapidity and precision a learner’s acquisition of the formal aspects of a study topic, but it cannot judge “indeterminate” content, i.e. content not clearly definable in terms of a right/wrong duality. This makes interactive multimedia inappropriate for dealing with the aesthetic, moral or philosophical implications of a piece of writing, for which a traditionally humanistic or otherwise subjective and all-embracing intervention of a flesh-and-blood teacher is necessary. But a “humanistic” evaluation of a piece of writing, if it is to have any real sense, must rest upon a solid grounding in the formal aspects of writing, since the abi-

lity to express oneself increases in direct proportion with a mastery of such aspects, what can be summed up in the traditional word “culture”. The dilemma of higher education today is a kind of dialogue at cross-purposes between an obstinately “lofty” institutional expectation and the dismal reality of low preparedness, a state of things which makes it difficult for students to exploit in full the academic experience. If we compare the different educational stages to letters of the alphabet, we might say that higher education or workplace requirements start at the letter “m”, while many students are chronically stuck at the letter “h”, with a yawning gap in between. Until we come to grips with this problem our institutions of higher learning will continue to turn out graduates with defective literacy and thus, paradoxically, contribute to lowering the cultural level of society. E-learning can do much to bridge this gap, to serve as an effective, low-cost, logistically agile remedial aid that can work in synergy with traditional teaching approaches and objectives. One computerized learning lab equipped with targeted quality educational software can produce learning results whose equivalent it would take a vast amount of human resources, time and physical space to achieve.

### 3. ORGANIZATIONAL STRUCTURE

The organizational structure of *Writing in English for the Workplace* features a compendium of text-types related to the following four major areas:

- Mass Communications
- Business
- Research
- Technology

For example, the unit on news reporting, which falls under the heading of Mass Communications, is subdivided into Headlines, Leads and Basic news story. Each of these sub-sections is composed of a tutorial and a set of exercises. Pull-down menus offer access to all the other text-types as well as to a range of style topics which include:

- Grammar
- Rhetoric
- Paragraph structure
- Vocabulary building
- Punctuation
- Bibliography
- Note-taking
- Editing

Each of these style topics consists of a set of tutorial “cards” with corresponding exercises and cross-references to related topics. The work has a comprehensive tracking structure which stores and evaluates student input data on two different levels:

- an institutional-oriented level which keeps a fully detailed record of the input-feedback dialogue, along with numerous variables useful for statistical purposes;
- a user-oriented level which keeps concise “score cards” consultable within each exercise in a version without the correct answers, and after each exercise in a version with the correct answers.

The work also has a comprehensive note-taking structure which sorts notes into different topic-related files consultable from any point in the program. In addition to on-screen consulting, users can print out both their score cards and notes. A culminating feature of each unit is role-play activity, which enables users to interact by means of film clips in virtual reality situations typical of the profession to which the text-type under study refers. For example, in the unit on news reporting, this role-play takes the form of a set of “Cub Reporter Assignments” in which the user assumes the identity of a novice reporter of a small-town newspaper and must gather information, often through interviews, and work up finished stories from his or her notes. Throughout *Writing in English for the Workplace* I have aimed at a high level of interactivity while allowing the user a maximum of navigational freedom. The various parts of the work can be used either separately or as a unified whole: for instance, any of the style factor modules, such as the one on grammar or the one on rhetoric, can be used independently of each other or of those on text types, or vice versa. However, in order to take full advantage of the program’s interrelated structure all the parts should be used together.

Despite its declared link to the professional sphere, *Writing in English for the Workplace* does not belong strictly to the category of didactic materials on “specialized English”. Its purpose is not to train future journalists or public relations operators or writers of reports to shareholders or of scientific research papers. In fact the various text types are presented in rudimentary form and with simple succinct explanations. Its purpose is rather to pave an initiatory way toward possible in-depth study of any of the sectors treated, or to habituate students *in general* to associate the act of writing with clearcut, profession-related protocols affecting choice of register, text length, standardized formulas, content structure, graphic layout, and even word-processing parameters, which have become such an intrinsic part of professional writing, so that further study of any sector will be quicker, easier and surer. In a sense, *Writing in English for the Workplace* proposes itself as an updated version of the Art of Rhetoric, in its original function as a set of practical guidelines for informing and/or persuading a given audience in a given social or professional context. Its aim is humble yet ambitious: to help students overcome their cultural deficits where writing is concerned, to get them from the letter “k” to the letter “m” and thus to give them a better chance to make it in today’s highly mobile and competitive world



# Un percorso didattico in rete per l'italiano lingua seconda

*Anna Bianchi*

Scuola Sec. 1° grado "Rizzo-Alessi" Genova  
Istituto Tecnologie Didattiche del C.N.R. Genova  
(anna.bianchi\_a@libero.it)

*A presentation of a teaching plan for teachers of Italian as a foreign language, implemented within a didactic archive on the net, inside the project "Netform2" (Ita.Ge.CNR.IT) by two junior high school teachers: Anna Bianchi (Genova), Fabio Mantegazza (Milano). The actors of the project "Netform2" are twenty teachers from all over Italy, already "expert" in the use of ITC tools in education, who have had a previous cooperative experience together within the NetForm-1 project (1998/1999), and a team of ITD-CNR researchers with different competences: system managers, community's tutors, educational technologists. The project is aimed to create, and to test, a grid suitable for both describing and putting in effect whatever kind of teaching plans, in different subjects, for different school levels, integrating ICT tools in educational practice, in order to build up an on-line archive. It will be available for teachers, newly trained in the technical use of ICT tools but till not expert in making a good educational use of them. The scheme has been tested by compiling it with a number of teaching plans. The one presented here, "Italian for beginners" tries to give an answer to the new and increasing problem challenging the Italian educational system: that of the teaching of Italian to foreign pupils, different as for countries, languages, age of immigration. This teaching plan can be examined under the heading "consultazione" at the url: <http://xgate.itd.ge.cnr.it:8080/test/presentazione.html> and can be presented referring to: it's addressees (teachers of Italian or of foreign languages); field of use (junior high school); aims (to provide teachers with tools to tackle the complex problem of language acquisition through examples of activities they can reproduce and extend); teaching methods and approaches.*

*Keywords:* Italian L2; ITC in education; Italian for beginners.

## 1. IL PROGETTO NETFORM

NetForm è un progetto di ricerca-azione realizzato dall'Istituto per le Tecnologie Didattiche del C.N.R. di Genova di concerto con il MIUR, che ha portato alla creazione di un database, disponibile *on-line*, contenente dei percorsi didattici. Esso vuole costituire un ambiente di apprendimento cooperativo, rivolto a insegnanti della scuola di base e secondaria, interessati all'impiego didattico delle TIC.

Negli ultimi anni il Ministero dell'Istruzione ha rivolto una crescente attenzione allo sviluppo dell'uso delle tecnologie informatiche da parte degli insegnanti, che ha portato a molteplici iniziative di formazione, a volte rivolte a grandi numeri di destinatari, con esiti scarsamente verificabili. Queste iniziative si sono generalmente centrate sull'acquisizione di competenze d'uso di strumenti informatici, ma altra cosa è essere in grado di farne un buon uso didattico.

L'idea centrale del progetto NetForm sta nell'utilizzare insegnanti che abbiano maturato una precedente esperienza nell'impiegare efficacemente le TIC nella didattica, come supporto di un processo di formazione all'interno di una comunità virtuale impegnata a cooperare a distanza.

Il progetto aveva avuto una prima fase, NetForm1, fra il 1998 e il '99, quando un gruppo di trenta insegnanti, individuati attraverso l'esame di prodotti didattici multimediali realizzati da scuole di ogni parte d'Italia, aveva lavorato insieme costituendo una comunità virtuale che aveva prodotto cooperativamente una "Quasi-guida" alla progettazione didattica centrata sulle TIC.

Venti di questi insegnanti hanno costituito, con un gruppo di ricercatori dell'I.T.D. di Genova, gli attori della seconda fase del progetto: NetForm 2, a partire dagli ultimi mesi del 2002.

L'idea centrale del progetto sta nell'uso delle TIC tanto per favorire la diffusione di esperienze didattiche già esistenti, quanto per consentire la comunicazione fra formatori e formandi, in un contesto in cui esse sono insieme oggetto e strumento della formazione.

## 2. IL "PERCORSO DIDATTICO"

Con questo nome si è voluto intendere contemporaneamente un itinerario di apprendimento, dal punto di vista dello studente, e un modulo di insegnamento, costituito da una successione di attività didattiche mirate al raggiungimento di specifici obiettivi didattici.

La fase più critica del progetto è stata quella di pianificazione della struttura del percorso, quando ogni docente ha dovuto mediare le proprie idee all'interno del gruppo sia riguardo gli intenti e le caratteristiche generali dei percorsi da realizzare, sia riguardo la struttura più adeguata per la loro presentazione *on-line*.

L'efficacia della struttura sarebbe poi stata messa alla prova attraverso la sua effettiva implementazione con una serie di percorsi didattici, diversi per argomento, destinatari, ampiezza.

Essi sono stati realizzati da piccoli gruppi (due o tre insegnanti) operanti a distanza in tempi definiti e coordinati da tecnici e tutor dell'ITD.

Concluso l'inserimento, i gruppi hanno operato una revisione reciproca dei percorsi, facendo riferimento, nella valutazione, a criteri discussi collettivamente e condivisi, costituiti da: accuratezza, pertinenza, chiarezza, completezza, efficacia dal punto di vista della motivazione, organizzazione.

Il data-base di percorsi risultante, che è attualmente visitabile alla voce "consultazione" nel sito <http://xgate.itd.ge.cnr.it:8080/test/presentazione.html>, potrà essere





Il percorso “Italiano per principianti”, attualmente consultabile all’URL: [http://xgate.itd.ge.cnr.it:8080/test/utenti/lista\\_percorsi.html?list=new](http://xgate.itd.ge.cnr.it:8080/test/utenti/lista_percorsi.html?list=new), presenta una struttura ad albero (Fig. 1), nella quale i rettangoli rappresentano le diverse attività, raggruppate in rami distinti in base alle abilità linguistiche prevalentemente prese in considerazione: abilità generali (fonetiche e grammaticali), competenza lessicale, abilità di lettura e scrittura, competenze comunicative (BICS), competenze linguistiche per lo studio (CALP).

I pulsanti superiori forniscono le informazioni generali che permettono di inquadrare il percorso a partire da un “problema didattico” e da un’ipotesi operativa descritta nei suoi presupposti teorici e metodologici.

Ogni “attività”, consultabile dispiegando i singoli pulsanti della struttura, fornisce in dettaglio: prerequisiti, obiettivi, risorse, attività preparatorie, tempi, modalità di lavoro e di documentazione, nonché una descrizione sequenziale dell’interazione docente/apprendenti. La struttura permette di fornire esempi di materiali e link a siti di interesse.

#### 4. RIFERIMENTI BIBLIOGRAFICI

- Benigno V. et al. (2002). *Recommendation on Quality Assurance and Selection of resources for the ETB network. Deliverable 3.2.*
- Benigno E., Olimpo G., Tavella M. (2000). *Multimedia in education: an experience in the propagation of teacher’s skills.* Proceeding of 16<sup>th</sup> World Computer Congress, pp. 289-295. China, Beijing.
- Benigno V., Sarti L. (2002). *ETB (European Treasury Browser): un sistema per la qualità delle risorse educative sul web.* Rivista di Tecnologie Didattiche, Vol. 26, n. 2, pp. 3-14.
- Altman G. (1997). *La scalata di Babele.* Milano, Feltrinelli.
- Bertocchi D. (2001). *La facilitazione nella comunicazione didattica nella classe plurilingue.* In E. Jafrancesco (ed.). Atti del X Convegno nazionale ILSA, pp.1-7. Firenze.
- Bettoni C. (2001). *Imparare un’altra lingua.* Bari, Laterza.
- Castellani M.C. (2000). *Agire nella classe plurilingue e multietnica.* In *Modulo di formazione Milia multimedia per l’italiano L2*, pp. 55-93. Genova, Sagep.
- Chini M. (1996). *Apprendere una seconda lingua: principi, fattori, strategie e problemi.* In E. Nigris, (ed.). *Educazione interculturale*, pp. 259-341. Milano, Bruno Mondadori.
- De Mauro T. (1980). *Guida all’uso delle parole.* Roma, Editori Riuniti.
- Diadori P. (2001). *Insegnare italiano a stranieri.* Firenze, Le Monnier.
- Favaro G. (2002). *Insegnare l’italiano agli alunni stranieri.* Milano, La nuova Italia.
- Lucisano P., Piemontese M. (1988). *GULPEASE: una formula per la predizione della difficoltà dei testi in lingua italiana.* Scuola e città, anno XXXIX, n. 3.
- Mastidoro N. (2003). *Leggibilità e lessico: il controllo con Èulogos CENSOR.* In A. Cattaneo (ed.). *Il cosmonauta. Guida per l’insegnante.* Milano, ELMEDI.
- Osborne D. (1999). *Teacher-Talk, A Sociolinguistic Variable.* English teachers forum on line, Vol. 37, n. 2

<http://exchanges.state.gov/forum/vols/vol37/no2/p10.html>

Pallotti G. (2000). *Favorire la comprensione dei testi scritti*. LEND, n. 3, pp. 28-35.

Serra Borneto C. (1998). *C'era una volta il metodo*. Roma, Carocci.

Vedovelli M. (2000). *Acquisizione e apprendimento linguistico*. In Modulo di formazione Milia multimedia per l'italiano L2, pp. 23-52. Genova, Sagep.

Vedovelli M. (2002). *Guida all'italiano per stranieri. La prospettiva del Quadro comune europeo per le lingue*. Roma, Carocci.



# Reflexiones en torno al uso de la prensa en línea para la enseñanza de lenguas desde la perspectiva del E/LE

Rosalie Sitman

Universidad de Tel Aviv

(rsitman@post.tau.ac.il)

*Most newspapers issued in Spanish are now available in real time on internet. This fact has encouraged language teachers all around the world to use newspapers for teaching Spanish. This is confirmed by a questionnaire distributed to teachers in Tel-Aviv University. They consider on line newspapers as a way to get real materials and an opportunity to get in contact with the different varieties of the Spanish Language. The language used is also quite close to real speech and therefore is suitable for use in a language lesson. Starting from the role played by the press in society the paper discusses practical and theoretical problems considering E/EL as a prototypical example. This may turn out to be significant for the use of on line newspapers in language teaching in general. Also new forms of communication in the digital scenario like weblogs are considered.*

*Keywords:* e-press; E/LE, Weblogs

## 1. PARA EMPEZAR..

El presente ensayo se enfoca desde la perspectiva de un consumidor de las ya no tan nuevas Tecnologías de la Información y la Comunicación (TIC), y está basado en la experiencia personal docente de un profesor de español como lengua extranjera (E/LE) en un país periférico, en el sentido de alejado del mundo hispanohablante. En realidad constituyen un eslabón más en la serie de reflexiones que he venido elaborando en distintos congresos y publicaciones profesionales en los últimos años acerca de la enseñanza del español asistida por Internet y, más específicamente, de las ventajas y los posibles peligros implícitos en la utilización de la prensa digital en el aula de lenguas. La pregunta clave, en ambos casos, es la misma: ¿por qué y para qué Internet?; ¿por qué y para qué la prensa en línea? En ambos casos, la respuesta es igualmente una y la misma: por necesidad – la urgencia íntima, vital, profesional de acercarnos e informarnos para conocernos y comunicarnos mejor – pero no por ello incondicionalmente y sin reservas; sobre todo, sin perder de vista el componente humano ni olvidar nuestra filosofía de la enseñanza y el aprendizaje. Más aún, si bien

estas reflexiones se circunscriben al caso del español, no así su relevancia, razón por la cual consideramos que, por extensión, podrían ser de utilidad a la hora de emplear la 'redprensa' para la enseñanza de lenguas en general en esta era de la información y de las comunicaciones.

El panorama ha cambiado muchísimo en los diez y tantos años transcurridos desde mi primera incursión en el ciberespacio cuando, impelida por la apetencia personal y profesional de romper el aislamiento geográfico y comunicarme con colegas y otros profesionales de la lengua española, di vida a ESPAN-L, la primera lista de distribución para profesores de E/LE. Hoy día Internet es ya un lugar común y los avances tecnológicos han modificado para siempre el cariz de la tarea educativa, sobre todo en lo referente a la educación a distancia: sabemos de aulas virtuales, cursos virtuales, tareas virtuales, materiales virtuales, prensa virtual... Con todo, debemos ser conscientes del doble peligro al que nos expone esta superabundancia virtual y cuidarnos de no caer en la tentación:

- a. de integrar Internet en la enseñanza por motivos equivocados (porque está de moda, por falta de tiempo, para no quedarnos atrás);
- b. de producir o utilizar más y más paquetes de materiales prefabricados, muchos de los cuales son de una potencialidad pedagógica discutible y – peor aún – cuya elaboración excluye *a priori* la espontaneidad de la comunicación, un elemento fundamental del aprendizaje de una lengua.

Del mismo modo que cada estudiante aprende a su ritmo y manera, cada profesor pone énfasis diferentes y adapta para sí lo que considera necesario de distintos métodos, de forma que un curso virtual, por más que lo intente, jamás conseguirá anticipar o reproducir estas variedades y combinaciones metodológicas individuales, menos aún los hilos invisibles del sutil entramado de la comunicación humana. ¿Qué estoy tratando de decir con todo esto? Simplemente que no debemos permitir que ni la ubicuidad ni la atracción del hipermedio nos encandilen y distraigan nuestra atención de lo que es realmente importante en el aprendizaje de una lengua y ése es el factor humano: la persona, el individuo, los seres detrás de las pantallas. Por lo tanto, a la hora de sopesar las posibles razones para utilizar Internet en un determinado momento, me parece instructivo medir hasta qué punto se justifica su inclusión en dichas circunstancias de acuerdo con los tres criterios que propuse en otra ocasión como patrón para evaluar la conveniencia, o no, de integrar Internet en el proceso de enseñanza: Internet como profesor; Internet como recurso; Internet como infraestructura para la comunicación. De esta manera, al obligarnos a reflexionar antes de actuar, esta guía nos ayuda a evitar el abuso o el mal uso del medio en el aula. En el caso específico que nos concierne aquí, el primer criterio quedaría descartado de entrada: la prensa digital jamás podrá reemplazar al profesor. En cuanto al segundo, como canal de información, sin duda los diarios electrónicos constituyen un recurso valioso para la adquisición de conocimientos, de tal modo que propiciarían el aprendizaje de una lengua. Con respecto al tercero, la prensa en línea como 'infraestructura para la comunicación' vendría a ser un instrumento idóneo para la enseñanza del E/LE, precisamente porque prioriza el

factor humano de que hablábamos antes y hace posible la comunicación espontánea y auténtica entre los estudiantes/profesores y otros practicantes, expertos o hablantes – nativos y no nativos – de la lengua meta en una situación comunicativa real. A continuación veremos que no sólo se justifica la utilización de la prensa digital para la enseñanza de E/LE, sino que las ventajas efectivamente sobrepasan con creces a las posibles desventajas.

## 2. ¿POR QUÉ Y PARA QUÉ?

Es un hecho que, gracias al avance de las nuevas tecnologías, hoy por hoy podemos leer en Internet las versiones electrónicas de los periódicos de casi todos los países hispanohablantes, casi al momento de su edición. Al mismo tiempo, el fácil acceso a estas ediciones digitales ha repercutido en una mayor utilización de la prensa en la clase de E/LE, sobre todo en el caso de los profesores que, como yo, trabajamos lejos de una comunidad de habla hispana. En un pequeño sondeo realizado entre mis colegas en Israel, todos respondieron que acuden a la prensa (mayoritariamente en línea) en busca de materiales auténticos y la oportunidad de entrar en contacto con las diversas variedades del español y las manifestaciones culturales de los distintos países donde se habla, se piensa, se siente y se sueña en este idioma<sup>1</sup>.

No cabe duda de que se trata de un género que, por su claridad, concisión, brevedad, actualidad y proximidad a la oralidad, entre otras muchas cualidades, es idóneo para su explotación en una clase de lengua. Siendo lengua y cultura dos realidades indisociables, la prensa, sea impresa o digital, acaba desempeñando un triple papel – representativo, normativo y educativo – que, más allá de su evidente utilidad en el aula, se conjuga fácilmente con los presupuestos del enfoque comunicativo y el enfoque por tareas, que tanta vigencia tienen actualmente. Por empezar, los textos periodísticos constituyen muestras auténticas de la lengua, producidas en una situación comunicativa real, que reproducen el español estándar o común (registro formal e informal) y por tanto son representativas de cómo los hablantes utilizan su lengua en la realidad, a diferencia de las representaciones artificiales de los materiales ‘prefabricados’ en cualquier soporte. A este respecto, la conjunción de los medios de comunicación convencionales y las redes de telecomunicaciones ha propiciado, en la forma de los diarios electrónicos, la creación de un espacio cibernético de encuentro, con enormes posibilidades para la expansión cuantitativa y cualitativa del idioma, en el que convergen las distintas expresiones del complejo mosaico que configura la realidad cultural del mundo hispanohablante. Se trata de un recurso que pone a disposición del profesor y del estudiante de español, a cualquier hora y desde cualquier lugar, grandes cantidades de información audiovisual, infográfica y textual actualizada, de diversa índole, posibilitando así su acercamiento a la realidad americana y peninsular por igual. La prescindencia de límites de tiempo y espacio y la intercomunicatividad del hipermedio no sólo facilitan la extensión del aprendizaje fuera de los confines del aula, sino que incluso pueden llegar a convertir la lectura

de la prensa digital en un elemento más de la rutina diaria, en lugar de una tarea forzada o artificial necesariamente restringida a 'la clase de español', cosa que es de fundamental importancia en los países 'extranjeros' con pocas o nulas posibilidades de contacto extracurricular con hispanohablantes.

En el plano de la lengua, el rompimiento de barreras físicas y políticas en las últimas décadas, gracias a la estupenda capacidad de difusión y comunicación de la red, ha conllevado una revalorización paralela de las distintas manifestaciones de la lengua española. En consecuencia, el creciente reconocimiento de la legitimidad de las distintas hablas ha ido desplazando al antiguo ideal académico de uniformidad del español impuesto desde la Península. Si bien los prejuicios y recelos perduran a ambos lados del océano, el cambio de actitud es notable en todos los ámbitos, según lo atestiguan las conclusiones dirimidas en los recientes congresos de la lengua (Zacatecas, Nueva York, Valladolid), en los que se priorizó la contribución positiva de las diferencias geográficas y las variedades dialectales a la riqueza de la lengua y la cultura españolas. Al mismo tiempo, la incorporación al ciberespacio de la comunidad hispanohablante en toda su diversidad no sólo ha contribuido a reforzar los vínculos de pertenencia a una comunidad cultural internacionalmente unida, sino que también ha azuzado el deseo de mantener y fortalecer el idioma común frente al acoso percibido del inglés, sobre todo en el área de las TIC. La actual eclosión de libros de estilo y manuales de redacción es un claro indicio de los esfuerzos invertidos por la industria periodística para elaborar un modelo normativo de criterios discursivos que sea válido para toda la prensa escrita en español, en consonancia con la creciente tendencia hacia la estandarización y la nivelación lingüística promovida por las exigencias comunicativas de la industria informática en la era de la globalización.

En tercer lugar, la prensa hispana, ya sea impresa o en su versión electrónica, tiene un gran valor didáctico en la enseñanza de L1 y L2, pues proporciona una rica fuente de documentos auténticos que el profesor puede aprovechar como texto para la elaboración de diversos materiales o como pretexto para la introducción o la práctica de distintos contenidos (gramaticales, léxicos, culturales), estrategias (de lectura y escritura) y destrezas (predecir, debatir, resumir), de acuerdo con sus objetivos y las necesidades de sus estudiantes. Huelga señalar que, por su naturaleza, estos textos son ideales para trabajar fenómenos como la organización discursiva, la tipología textual, los recursos metafóricos y las diferencias estilísticas. De tal manera que el uso de la prensa escrita en cualquier soporte potencia la adquisición de la competencia comunicativa en general, como también en cada una de las áreas de habilidad que la articulan --competencia gramatical, discursiva, estratégica, sociolingüística y sociocultural--, y además se presta a la elaboración de todo tipo de tareas; razones éstas que justifican de sobra su inclusión en el aula de E/LE.

Desde el punto de vista del estudiante, el trabajo con los diarios electrónicos puede ser beneficioso por varios motivos. En primer lugar, constituyen un entorno con el cual el estudiante está familiarizado en su propio idioma, reduciendo así la sensación de desconcierto que provoca el aprendizaje de una lengua desconocida y facilitando la comprensión a través de la presencia de palabras clave y de situaciones conocidas. Lue-

go, la posibilidad de acceder a la prensa digital en todo momento y, más importante, la satisfacción de poder comunicarse a través de los dispositivos interactivos (los foros, las cartas al director, las encuestas, etc.) y de recibir una retroalimentación real casi instantánea surten un efecto motivador de inestimable valor a la vez que promueven el autoaprendizaje. Eso sí, cabe recordar que tanto la motivación como las necesidades serán diferentes en el caso del estudiante que se halla estudiando en España, por ejemplo, y aquél que accede a las ediciones digitales desde un país no hispanohablante. El primero compra o coge algún diario español que encuentre por ahí, motu proprio, impulsado por la curiosidad de enterarse de lo que está pasando en el mundo o para consultar la cartelera de espectáculos, mientras que el segundo abrirá la versión electrónica de algún diario español o hispanoamericano, a instancias de su profesor, con el solo propósito de llevar a cabo alguna tarea específica que le haya sido encomendada; las noticias las leerá en la prensa local. Ineludiblemente, estas diferencias se manifestarán en el aula y, de alguna manera u otra, afectarán la dinámica de la clase.

### 3. UN ESPACIO DE ACERCAMIENTO

Al combinar las características de la prensa impresa con las ventajas de la malla mundial, la prensa en español en línea constituye un medio idóneo para llevar a cabo un viaje de (auto)descubrimiento de la lengua. La posibilidad de entrar en contacto con cualquiera de las variedades del español dentro del espacio aparentemente uniforme del periódico virtual desempeña un papel nivelador de suma importancia, tanto para los profesores (nativos o casi-nativos) como para los estudiantes (no nativos), especialmente aquéllos en un país alejado del mundo hispanohablante. Debemos puntualizar que normalmente un profesor peninsular nativo no tendrá mayor interés en recurrir a la prensa digital. Con todos los recursos humanos y materiales que pudiera desear a su disposición, se encuentra a gusto enseñando su variedad, en su entorno, y seguramente no siente mayor necesidad de familiarizarse con las variedades diferentes de la suya. ¿Para qué? El español se caracteriza por su homogeneidad, especialmente en los niveles más cultos, de manera que las diferencias – en su mayoría léxicas y tónicas y, en menor grado, morfosintácticas – no suelen entorpecer la comunicación entre los hablantes de un país y de otro. ¿O sí? Recordemos el impacto que causó en España la aparición del peregrino ‘chapapote’ en los reportajes sobre la marea negra que asoló las costas gallegas<sup>2</sup>.

A diferencia de España, en Israel, por ejemplo, la mayoría de los profesores son latinoamericanos. Por consiguiente, muchos se debaten en torno a qué variedad de la lengua han de enseñar, y más de uno se ve obligado por las circunstancias a enseñar una variedad que no es la suya. En estos casos, el acceso a las ediciones digitales de los diarios de los distintos países es de fundamental importancia como agente de acercamiento a la diversidad lingüístico-cultural hispana y ofrece una oportunidad singular para actualizarse y mantener viva la lengua, a pesar de la distancia, así como para conocer las variedades diferentes de la propia. Sea de donde sea, el profesor nativo sabrá



detectar los localismos y se familiarizará con ellos. Mas al irse habituando a las diversas manifestaciones del idioma a las que se ve expuesto, el profesor nativo en el extranjero pierde conciencia del momento en que incorpora a la suya un elemento lingüístico procedente de otra variedad y, de este modo, comienza a desarrollar una especie de ‘panespañol’. La inconsciente adopción de rasgos dialectales ajenos a nuestra variedad es un fenómeno bastante difundido entre los profesores nativos en el extranjero.

No así en el caso del profesor no nativo, cuya perspectiva, a estos efectos, se aproxima más a la del estudiante que a la de un profesor nativo: para ambos, todo lo desconocido es novedoso, sin tener conciencia del hecho de la variedad. Por tanto, uno y otro acudirán a la prensa en línea, pero ninguno será capaz de identificar las palabras desconocidas como pertenecientes a una variedad dialectal determinada, limitándose a buscar su significado en el diccionario. Dicho lo cual, debo acotar que esto está cambiando en Israel – y me imagino que también en otros países – donde el contacto con el mundo hispano a través de Internet, unido a un verdadero ‘boom’ de telenovelas (no culebrones) en la televisión, no sólo ha elevado el nivel umbral de los estudiantes, sino que ha contribuido patentemente a hacerlos conscientes de la diversidad lingüístico-cultural del español.

#### 4. CUIDADO CON LOS FALSOS AMIGOS...

Habiendo establecido la posible utilidad de la prensa digital para la enseñanza de E/LE en el plano metodológico, no es menos cierto que, en la práctica, su implementación puede depararnos más de una sorpresa, no siempre agradable, que acabe frustrando el aprendizaje en lugar de motivarlo. En la mayoría de los casos, estos obstáculos no suelen constituir vallas insalvables y pueden sortearse mediante la mediación del profesor o recurriendo a los libros de referencia. No obstante, no dejan de ser un imprevisto que interrumpe el flujo de la enseñanza y que el profesor ha de tener en cuenta a la hora de decidir utilizar la prensa en línea. A veces, incluso pueden causar incidentes embarazosos, sobre todo aquéllos derivados de lo que podríamos denominar ‘el habla de la ideología’.

Es un hecho que cada periódico, impreso o virtual, da igual, articula la realidad de una manera diferente, según a quién va dirigido y de acuerdo con una línea ideológica determinada. Según las reglas del juego, para que la noticia o información surta el efecto deseado, tiene que estar presentada de cierta manera: por un lado, deberá ajustarse a elementos predefinidos establecidos por la lengua; por el otro, estará sujeta a las manipulaciones imprevisibles efectuadas a fin de causar el mayor impacto sin desviarse de la línea editorial prevista. De este modo, el lector, que se supone ya conoce los códigos, automáticamente capta las claves que le permiten descifrar el mensaje. Leer un diario, entonces, puede ser una tarea simple, pero no así descifrar los códigos y tramas que lo componen; peor aún en el caso de una lengua que se está aprendiendo: ¿cómo puede saber un estudiante de E/LE que la decisión entre escribir “el Peñón” o “la Roca”, “Euskadi” o “País Vasco”, “Malvinas” o “Falklands”, “latino” o “hispano”, implica una posición política o ideológica?

Claramente, los elementos del ‘habla de la ideología’ – los presupuestos, los prejuicios, los estereotipos, la retórica, las cargas significativas – pueden ofuscar y frustrar a los no iniciados, e incluso convertirse en una verdadera traba para la comprensión. Neologismos con una clara connotación político-social, tales como “zapatismo”, “menemismo”, “fujimorazo”, pierden todo su efecto si el receptor desconoce la cultura política de la que surgieron. Más eficaz, el eufemismo sigue siendo un arma predilecta de ocultamiento y disimulo. Así, la (in)cultura de la violencia que nos ha tocado vivir está produciendo espeluznantes cosechas lingüísticas: el “Proceso” argentino, la “limpieza étnica” de Milosevic y las “eliminaciones puntuales” en Israel; donde lo censurado es ley, los códigos de significación reflejan el compromiso de la lengua con la realidad. Pero siempre y cuando la alusión sea compartida; y nuestro estudiante no siempre cuenta con las herramientas que le permitan captarla.

Algunas veces, el ‘habla de la ideología’ puede producir malentendidos o desencuentros a partir del choque inesperado entre el contexto emisor y el contexto receptor. Esto sucede en aquellos casos en los que el referente se trastoca, de tal modo que el sujeto (el estudiante/lector) se ve convertido en el objeto (el estudiante/protagonista de la noticia). Así, por ejemplo, los estudiantes israelíes (judíos, cristianos y árabes), que siguen con interés las noticias sobre Oriente Medio aparecidas en la prensa digital en español, muchas veces se sorprenden al encontrarse con una imagen o representación de sí mismos que no concuerda necesariamente con su autoimagen, ni con la imagen que cada cual tiene del otro. Otro problema es el de las connotaciones que se pierden o que cambian en la traducción de un idioma (y una cultura) a otro idioma (y otra cultura) y que, en consecuencia, afectan inadvertidamente tanto la transmisión como la percepción del mensaje. Un caso ilustrativo es el uso del término unívoco “Cisjordania” en la prensa hispanohablante para denominar lo que en la prensa israelí se traduce de una de dos maneras, según la posición ideológica del emisor: “territorios ocupados” (discurso opositor de izquierda) o “Judea y Samaria” (discurso de derecha). La diferencia es sutil, pero significativa, porque pone de manifiesto la existencia de distintas perspectivas en el horizonte discursivo israelí, aunque esto no siempre se trasluzca en las informaciones virtuales.

En algunas instancias, el cortocircuito se produce a raíz de carencias objetivas de conocimientos o de información, o debido a las características inherentes al medio. A esta categoría pertenecerían las dificultades atinentes a ciertos vacíos de información resultados del desconocimiento de lenguajes sectoriales (tales como el léxico de la economía, la jurisprudencia o la ciencia) o de tecnicismos especializados, sobre todo cuando son usados fuera de contexto y en conjunción con vocablos propios de un ámbito diferente (“una [universidad] de las de primera A en ciencias políticas y humanidades”). Esta tendencia a mezclar palabras y expresiones de distintos campos semánticos se ve exacerbada en el lenguaje deportivo, en el que es común encontrar deliciosos ‘cócteles’ confeccionados con términos extraídos del lenguaje mariner (al abordaje, ir a pique), del lenguaje militar (cañón, artillería), del lenguaje judicial (juzgado de guardia), del lenguaje culinario (a punto de caramelo), los cuales, aunque pintorescos, resultan ininteligibles para un simple y vulgar estudiante de E/LE, por muy aficionado que sea al deporte. Algo se-

mejante sucede con la riqueza y especificidad de los gentilicios localistas de que hacen gala los diarios españoles (abulense, ilerdense, donostiarra), así como con la presencia de voces de otras lenguas autonómicas, como el catalán o el ‘eusquera’ (generalitat, conseller, lendakari, abertzales), cuyo significado no siempre es fácil de dilucidar a través del contexto y que requieren cierto conocimiento previo del discurso político o identitario en España. ¡Del argot de los jóvenes y la jerga de la homosexualidad y la droga, mejor no hablar; menos aún de los ‘infortunios’ de la telefonía móvil!

En otras ocasiones, el obstáculo a la comprensión se debe a simples, pero no por ello menos frustrantes, errores grafemáticos de tipografía o de transliteración o a la ambigüedad producida por el mal uso de las comas, las comillas, la cursiva y demás signos de puntuación. Muchas veces, la confusión tiene su origen en ‘abusos’ típicos del lenguaje periodístico, tales como el uso indiscriminado o la lexicalización de siglas (no siempre reconocibles para el estudiante de E/LE acostumbrado a un alfabeto no latino), el deliberado empleo de extranjerismos en lugar de un perfecto equivalente español – pins (insignias), parking (estacionamiento/aparcamiento), panties (medias), gays (homosexuales) – o el uso incorrecto de voces españolas por contagio del inglés: argumento (discusión), evento (acto, acontecimiento), serio (grave), ignorar (desdeñar, ningunear)... Estos ‘falsos amigos’ son verdaderos enemigos y acaban desvirtuando el uso correcto y preciso de la lengua. Hablando de la contaminación del idioma, ¡el desconcierto infundido por el primer encuentro con un neologismo del calibre de “puenting” es simplemente indescriptible! En semejantes casos, aun teniendo plena conciencia de que el cambio es consustancial al hecho de una lengua viva, se aprecian las protestas que se están elevando en defensa del español.

Con frecuencia, las lagunas que surgen son de índole cultural más que lingüística. Tomemos el caso del fatídico 11 de septiembre, que ha quedado inscrito en el imaginario global como el día que cambió para siempre el paisaje urbano de Nueva York. Para los chilenos, no obstante, la fecha tiene un significado diferente, aunque no menos funesto: ellos asocian este día con el golpe militar que derrocó al presidente legítimo y cambió para siempre el paisaje humano del país. Mientras que para los catalanes es la Diada, que algunos festejan y otros conmemoran. Este fenómeno de ‘disociación’ es particularmente evidente en el mundo de la publicidad, que articula valores y símbolos compartidos por una comunidad de hablantes, los cuales pueden ser completamente desconocidos, e incluso ofensivos, para aquellos estudiantes procedentes de sociedades más conservadoras o fundamentalistas, cuyo bagaje cultural es radicalmente diferente. En estas ocasiones, el efecto se pierde y el mensaje no surte el efecto previsto por el emisor. Naturalmente, esto nos remite a cuestiones más amplias de interculturalidad y de representaciones del otro, que exceden los límites de esta comunicación.

Algo parecido ocurre con frecuencia cuando, ilusionados, intentamos utilizar chistes o tebeos, que nos parecen comiquísimos, sólo para estrellarnos contra la muralla de incompreensión reflejada en las caras impasibles de nuestros interlocutores en el aula. El humor, en tanto expresión de un pueblo en un determinado momento histórico, no se traslada fácilmente a otra lengua. Ni siquiera en el caso de la entrañable

Mafalda, la popular tira cómica argentina, en la que aun conociendo y entendiendo todas las palabras, se pierde mucho del significado, de la intención y de la ironía, si se desconocen las sutilezas del contexto sociopolítico argentino de aquellos años o ciertas peculiaridades de la variedad rioplatense<sup>3</sup>.

## 5. LOS 'BLOGS': ¿LAS VOCES DEL FUTURO?

En la sociedad de la información y de las grandes redes de comunicación, la realidad virtual ha trastocado los viejos límites conceptuales y arrasado con fronteras de todo orden. Así, la 'redprensa' nos permite consultar simultáneamente *La Opinión* de Los Ángeles, *El Universal* de Caracas, *El Mercurio* de Santiago, *La Nación* de Buenos Aires o cualquiera de los diarios de Madrid y cerciorarnos de que no hemos salido de la misma comunidad lingüística. Mucho más importante, también nos permite comunicarnos. Las implicaciones han sido enormes. Volviendo otra vez al fatídico 11 de septiembre de 2001: ¿cómo se explica que ese día una avalancha de internautas latinoamericanos haya descendido sobre las páginas-web de los diarios españoles? La primera razón es técnica: la mayoría de ellos dependía de servidores en los Estados Unidos, a los cuales, obviamente, no podían acceder. Por consiguiente, acudieron a los que, al menos lingüísticamente, les resultaban más próximos: los españoles. Pero, días después, la avalancha continuaba. Según lo atestiguan numerosas 'cartas a la redacción' electrónicas, buscaban una visión diferente de los hechos de la que les proporcionaban los medios estadounidenses. Más recientemente, el controvertido caso de Salam Pax en Irak parecería ejemplificar un fenómeno semejante.

Valgan estos ejemplos como ilustración de que, gracias a la prodigiosa infraestructura para la comunicación de Internet, indefectiblemente asistimos hoy a una redefinición de la tradicional razón de ser de la prensa, más allá de la información y la orientación, hacia la personalización y la comunicación. El rasgo más distintivo de la red es precisamente su capacidad para la ampliación de las relaciones sociales. Los grandes medios de comunicación convencionales –la televisión, la prensa impresa, la radio – son instrumentos de información que fluyen en una sola dirección y por lo tanto no tienen la capacidad de crear comunidades por sí mismos. Internet, por el contrario, ofrece la posibilidad de establecer una comunicación bidireccional y de este modo se van formando comunidades virtuales, cuyos vínculos son establecidos por los propios miembros, sin necesidad de estar juntos físicamente. El resultado es una especie de mundo virtual paralelo, pero constituido por personas reales que se conectan a través de sus ordenadores.

En este contexto, sin duda, los protagonistas en el escenario de la prensa digital son los numerosísimos *weblogs* o diarios personales de navegación y de publicación electrónicas. Comúnmente apodadas blogs, estas bitácoras constituyen, a un tiempo, una interesante herramienta para la difusión del conocimiento y un nuevo medio de comunicación global. Flexible por antonomasia, un weblog puede ser unipersonal, grupal o colectivo (hecho por un conjunto de editores) y también puede ser comuni-

tario, de tal forma que los lectores participan en la elaboración de los contenidos, en una suerte de tertulia o club de aficionados abierto al público y accesible a todos. A diferencia de la prensa convencional, al estar totalmente exentos de las exigencias de una línea editorial predeterminada o el temor a la censura, los *blogs* proporcionan un entorno interactivo donde los participantes pueden publicar su propio punto de vista sobre su(s) realidad(es).

Así es cómo van surgiendo versiones alternativas de un hecho concreto, tal el caso de Salam Pax. En otras palabras, los *blogs* recogen lo que piensa el ciudadano de a pie y, por ende, presentan la realidad de acuerdo a quien la cuenta.

Al mismo tiempo, las cualidades inherentes al hipertexto permiten conectarse fácilmente con otros navegantes de intereses afines que hayan escrito sobre lo mismo y, de este modo, se van formando subcomunidades dentro de una misma comunidad virtual más amplia: por ejemplo, una 'blogosfera' o índice de *blogs* en español.

Evidentemente nos hallamos ante una plataforma colaborativa, especie de aula global, con un enorme potencial didáctico para la enseñanza de lenguas. En primer lugar, el estudiante aprende a actuar en un sitio de dominio público, asimilando y adaptándose a la 'redetiqueta' del nuevo medio y a las convenciones del nuevo género discursivo. Al mismo tiempo, se trata de un recurso idóneo para desarrollar tanto la comprensión lectora como la expresión escrita. La necesidad de responder a algún comentario, tanto como el deseo de entablar un diálogo con otros cibernautas iniciando un nuevo hilo discursivo, obligan a publicar algún comentario propio en el blog a fin de comunicarse. Mas, a diferencia de la espontaneidad e inmediatez del correo electrónico, la naturaleza de la escritura en este caso será más deliberada y cuidada, más medida, puesto que se tiene plena conciencia de que el texto será leído por muchos otros.

No menos importante, los *blogs* favorecen el desarrollo de la competencia sociocultural, tanto a través de la comunicación establecida con otros internautas como por medio de la información dirimida de la multiplicidad de enlaces visitados. Al mismo tiempo, se prioriza el aprendizaje autónomo y el protagonismo del estudiante a partir del momento en que es este último quien determina el curso de su navegación por los distintos enlaces, según éstos piquen su curiosidad.

Por último, la combinación de la infraestructura comunicativa de Internet y el formato de los *blogs* propician el aprendizaje por tareas al prestarse para la realización de proyectos individuales o el trabajo en grupos.

La posibilidad de crear un diario personal propio o colectivo durante el curso, de utilizar herramientas reales para comunicarse en español con el profesor, con los compañeros, con otros navegantes de ideas afines en distintos países, sin duda constituye un aliciente muy atractivo para el aprendizaje y da mucho para pensar y para hacer en el camino hacia la tarea final. En el proceso, todos acaban beneficiándose de este tipo de colaboración. A fin de cuentas, ¿qué es enseñanza si no comunicación, interacción, inspiración, improvisación, producción, tolerancia, conocimiento lingüístico y sociocultural, colaboración...? Y los *blogs* son sólo un ejemplo del abanico de posibilidades didácticas que ofrece la redprensa para la enseñanza de lenguas sin fronteras.

## 6. PARA TERMINAR, Y PARA PENSAR..

La irrupción en el plano de la información de los medios electrónicos de prodigioso alcance, unida a la vertiginosa evolución de Internet, con su oferta realmente global y enciclopédica, está cambiando nuestra percepción de la prensa tanto como la naturaleza misma del medio. En la medida en que estos medios han acercado cada vez más a gente dispersa en el mundo entero, los hispanohablantes, repartidos por varios continentes, hemos visto ampliarse y fortalecerse nuestra unidad lingüística y cultural. La conciencia de que formamos parte de una comunidad cultural diseminada por el mundo, pero unida vigorosamente por la fidelidad al idioma común – el español – ese sentimiento de pertenencia es un valor que nos integra, que nos reconstruye, que robustece nuestra identidad en el período de globalización ‘imperial’ que estamos viviendo. Considerando la enorme influencia de los ‘media’ sobre la lengua común, el desafío, sobre todo para la prensa electrónica, consistirá en su capacidad para difundir, defender, renovar, enriquecer y preservar la cultura y la lengua españolas a partir del contacto con las otras lenguas y las otras culturas y a pesar de las embestidas de otras lenguas y otras culturas. Juntos y por separado, los medios de comunicación de habla hispana deberán asumir la doble responsabilidad de facilitar el descubrimiento de la diversidad y a la vez velar por el buen uso del castellano, marca de identidad de más de cuatrocientos millones de hablantes.

## 7. REFERENCIAS BLIOGRÁFICAS

- AA.VV. (1997). *Actas del Primer Congreso Internacional de la Lengua*. Zacatecas <<http://cvc.cervantes.es/obref/congresos/zacatecas/> >
- AA.VV. (2001). *Actas del Segundo Congreso Internacional de la Lengua Española*. Valladolid. <<http://cvc.cervantes.es/obref/congresos/valladolid/> >
- AA.VV. *Actas de los Congresos de ASELE*. (Málaga, Alcalá de Henares, Almagro, León).
- Alfaro Logorio C. (1998). El papel de la prensa en ELE. In *Actas del VIII Congreso Internacional de ASELE: El español como lengua extranjera: del pasado al futuro*, pp. 97-103. Alcalá de Henares, Universidad de Alcalá, Servicio de Publicaciones.
- Alonso Morales M. C., del Carmen M. (1993). Las perífrasis verbales en el español de la prensa. In Otesa Peydro M. et al. (eds). *Actas del tercer Congreso Nacional de ASELE. El español como lengua extranjera: De la teoría al aula*, pp. 291- 298. Málaga, Imagraf.
- Ballesta Pagán, J. La utilización didáctica de la prensa <http://dewey.uab.es/pmarques/EVTE/apliprensa.doc>.
- Barrós García P., de Molina Ridondo J. A. (1993). El uso de la lengua en la prensa deportiva: aportaciones para la enseñanza-aprendizaje del ELE. In *Actas del tercer Congreso Nacional de ASELE. El español como lengua extranjera: De la teoría al aula*, pp. 231-240. Málaga, Imagraf.

- Cruz Piñol M., Sitman Y R. (2000). Los diarios hispanos en la red: herramienta para el acercamiento cultural y lingüístico. *Cultura e intercultura en la enseñanza del español como lengua extranjera*  
<http://www.ub.es/filhis/culturele/diariosf.html>
- Cruz Piñol M., Sitman R. (1999). Tu español y mi español se encuentran en la red. Consideraciones sobre la prensa hispana 'en línea' en el marco de la enseñanza del E/LE. In A. Carcedo González (ed.). *Documentos del Español Actual*, pp. 89-103. Finlandia, Universidad de Turku.
- De la Heras A., Alonso M. R. (1991). *La prensa en la clase de lengua*. Madrid, Ediciones SM.
- Ferraz Martínez A. (1995). *El lenguaje de la publicidad*. Madrid, Arco Libros.
- Iglesias Casal I., Grande M. P. (2001). *¡A toda página!*. Madrid, Edinumen.
- Martinell Gifré E., Montalto Cessi D. (1999). Análisis contrastivo de la estructura gramatical de textos específicos en la prensa española e italiana. In *Actas del IX Congreso Internacional de ASELE: Español como lengua extranjera: Enfoque comunicativo y gramática*, pp. 659-761. Santiago de Compostela, Universidad de Santiago de Compostela, Instituto de Idiomas.
- Romero Gualda, M.V. (1994). *El español en los medios de comunicación*. Madrid, Arco Libros.
- Verdía E. (1996). 15 formas de trabajar la prensa. In *Actas del VI Congreso Internacional de ASELE: Actuales tendencias en la enseñanza del español como lengua extranjera*, pp. 413-418. León, Universidad de León.

## 8. NOTAS

- <sup>1</sup> En otro lugar ya hemos hablado de la contribución de la prensa en línea al intercambio entre las variedades de una lengua viva como el español: Cruz Piñol M. y R. (1999), pp. 89-103.
- <sup>2</sup> Ver Íñigo-Madrigo, L. "Chapapote. Historia de la palabra" <http://www.jamillan.com/chapa.htm>
- <sup>3</sup> Diccionario del habla de los argentinos, por la Academia Nacional Argentina de Letras.





Finito di stampare da  
Grafiche Cappelli - Osmannoro (FI)