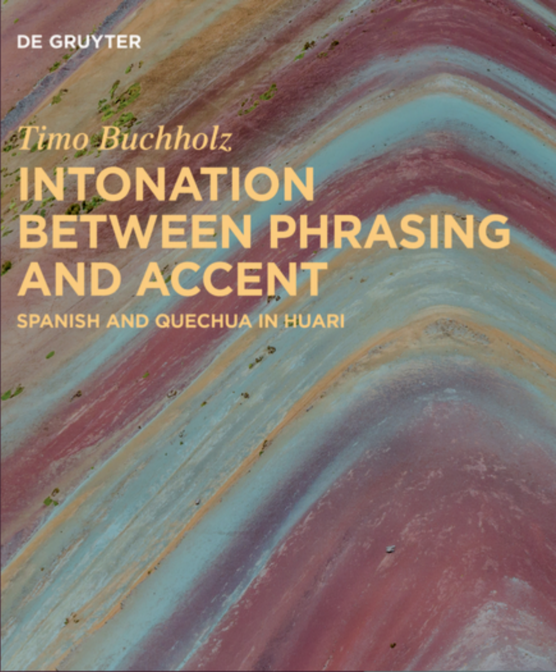
DE GRUYTER



Timo Buchholz

INTONATION BETWEEN PHRASING AND ACCENT

SPANISH AND QUECHUA IN HUARI

LINGUISTICA LATINOAMERICANA

Timo Buchholz

Intonation between phrasing and accent

Linguistica Latinoamericana

Editado por / Edited by
Dermeval da Hora Oliveira, Carlos Garatea Grau, Uli Reich

Volumen / Volume 7

Timo Buchholz

Intonation between phrasing and accent

Spanish and Quechua in Huari

DE GRUYTER

Slightly revised version of a thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (Dr. phil) at Freie Universität Berlin in August 2021 and defended in April 2022.

Die Publikation wurde ermöglicht durch eine Ko-Finanzierung für Open-Access-Monografien und -Sammelbände der Freien Universität Berlin.

ISBN 978-3-11-130364-2
e-ISBN (PDF) 978-3-11-130459-5
e-ISBN (EPUB) 978-3-11-130624-7
ISSN 2628-3875
DOI https://doi.org/10.1515/9783111304595

Library of Congress Control Number: 2023942615

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.dnb.de.

Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Acknowledgements

This book is a slightly revised version of my doctoral thesis that I defended in 2022 at the Free University of Berlin. It arose from a research project on bilingual prosody between Spanish and Quechua that my supervisor Uli Reich invited me to think about with him in 2014. Its eventual funding by the DFG allowed Raúl Bendezu Araujo, him, and me to undertake field trips to Peru and to devote several years to studying the speech of the amazing people we met in Huari.

Those people, whose speech provides the database for this book, generously gave us their time, humour, and friendship. My first thanks go to them.

Besides giving me the opportunity to participate in this adventure, I want to thank Uli for his support, for always encouraging me to question established wisdoms and for providing me with a strong idea of what qualities I would like to pursue as a researcher myself.

Raúl and I became friends on this journey together, solved Quechua puzzles together, and got called gringos together by our Huari friends.

Jan Fliessbach discussed nearly every page that made it into this thesis and many more that didn't, and he has always been there for me.

Judith Meinschaefer always asked the best and trickiest questions that made me think about my ideas and linguistics in general in a new way.

I would also like to thank Maria del Mar Vanrell for teaching me practical things about prosody, and everyone at the colloquium for Romance linguistics as well as many others for countless enjoyable discussions that helped me develop and shape my ideas.

Gabriel Barreto and Leonel Menacho shared their expertise on Quechua with us, provided us with invaluable ins into local society and were very patient with us.

This book and the thesis it is based on has been years in the making and in this process I have benefitted from discussions, advice and support by more people than I could possibly list here by name. My thanks go to all of you. My parents and my friends bore with me throughout this time and picked me up when it was all becoming too much. To the extent that this work has been a successful endeavour, you all contributed immeasurably to it. However, I firmly claim sole ownership of all remaining shortcomings.

https://doi.org/10.1515/9783111304595-202

Contents

List of figures

https://doi.org/10.1515/9783111304595-204

List of tables

https://doi.org/10.1515/9783111304595-205

List of symbols and abbreviations

¬	logical negation
{x, y}	the set with the members x, y
[	*in the object language line in examples:* start of speaker overlap
	in the tone line in examples: left boundary of a high-level prosodic phrase
]	*in the object language line in examples*: end of speaker overlap
	in the tone line in examples: right boundary of a high-level prosodic phrase
x >> y	x is higher ranked than y (OT)
.	*in a context like CV.CV*: syllable boundary
!	the element after ! is downstepped relative to the preceding element
¡	the element after ¡ is upstepped relative to the preceding element
#	*in a context like CVC#CV*: word boundary
# *Example*	the example marked with # is unfelicitous (at odds with / inappropriate in the context given)
* ***Example***	the example marked with * is not well-formed/ungrammatical (not possible in the language given, independent of context)
AM	autosegmental-metrical
AP	*prosody*: accentual phrase
	syntax: adjective phrase
C	*as part of syllable structure:* consonant
	as part of prosodic structure: clitic group
CG	common ground
EEQT	Eastern European Question Tune
Ft	metrical foot
f0	fundamental frequency
f1 – f5	first – fifth formant frequency
H	high tone
h	heavy syllable
IP	intonation phrase
IPA	international phonetic association
IS	information structure
iP	intermediate phrase
ι	high-level phrase (iP or IP)
L	low tone
l	light syllable
ms	millisecond
μ	mora
NP	noun phrase (syntax)
OCP	obligatory contour principle
OT	optimality theory
ω	prosodic word
ω'	metrically strongest prosodic word in a phrase
PhP	phonological phrase
PP	prepositional phrase (syntax)
Pwd	prosodic word

https://doi.org/10.1515/9783111304595-206

ϕ	phonological phrase
QUD	question under discussion
R^2	coefficient of determination (proportion of variation in the dependent variable predictable from the independent variable(s))
SLH	strict layer hypothesis
Sp_ToBI	Spanish tone and break indices
s	strong (metrical relation)
st	semitone
Σ	metrical foot
σ	syllable
σ'	metrically strongest syllable
T	a tone (either low or high)
T*	a tone (T) that is associated with a strong metrical position (pitch accent)
T-	a tone (T) that is a phrase accent or iP-boundary tone
T%	a tone (T) that is an IP-boundary tone
TBU	tone-bearing unit
ToBI	tone and break indices
TS	minimal tone sequence
V	vowel
VP	verb phrase (syntax)
w	weak (metrical relation)

Huari Quechua morphological glosses

suffix	gloss[1]	meaning
-:	1	first person (preceding vowel is lengthened)
-shaa ~ -shaq	1.FUT	first person, future tense
-ma	1.OBJ	first person object
-ntsi ~ -ntsik	1.PL.INCL	first person inclusive, plural
-shun	1.PL.INCL.FUT	first person inclusive plural, future tense
-shayki	1.SUB>2.OBJ.FUT	first person subject acts on second person object, future tense
-nki ~ -yki	2	second person
-n	3	third person
-nqa	3.FUT	first person, future tense
-shunki	3.SBJ>2.OBJ	third person subject acts on second person object
-shu … -nki	3.SBJ>2.OBJ	third person subject acts on second person object, past tense (in verbal constructions in the past tense, the sequence –shu-nki can be split up, e.g. *mutsa-shu-rqa-nki* kiss-3.SBJ>2.OBJ-PST-3.SBJ>2.OBJ 's/he kissed you')
-piq ~ -pita	ABL	ablative (spatial and temporal)
-pis ~ -si	ADD	additive ('also', 'even')
-q	AG	agentive
-m ~ -mi / -chaa	ASS	assertion
-sapa	AUGM	augmentative
-paa ~ -paq	BEN	benefactive (nominal)
-pu	BEN	benefactive (verbal)
-tsi / -raykur	CAUS	causative
-mu	CIS	cislocative
-chi	CONJ	conjectural

1 These are the Quechua glosses developed in the DFG-funded project "Zweisprachige Prosodie: Metrik, Rhythmus und Intonation in mehrsprachigen Kontaktsituationen" (PI Uli Reich, Freie Universität Berlin). See https://www.geisteswissenschaften.fu-berlin.de/we05/institut/mitarbeiter/reich/forschung/DFG-projekt-zweisprachige-Prosodie/Manuals/Glossen_HP_20190626.pdf and section 2.1. "~" indicates variation between phonologically similar forms, "/" indicates variation in form but with the same meaning as far as could be determined. If two forms have the same gloss but are listed in separate lines, their meaning is similar enough to warrant the same gloss, but not enough is known about the semantics of (one of) the forms in question to assign two different glosses.

https://doi.org/10.1515/9783111304595-207

(continued)

suffix	gloss	meaning
-tsun ~ -tsuraa	CONJ	conjectural
-raq ~ -raa ~ -ran ~ -ra	CONT	continuative (‘still’)
-man	DEST	destination, goal
-taa ~ -taq ~ -tan	DETVAR	variable determination (“contrastive”)
-rku ~ -rka	DIR	directional (*kawallu-man lluqa-rku-n* ‘s/he gets on the horse’)
-rpu ~ -rpa	DIR	downwards directional (*hita-rpu-n* ‘s/he throws it down-wards’)
-yku ~ -yka	DIR	inwards directional (*chura-yku-n* ‘s/he puts (it) inside’)
-rqu ~ -rqa	DIR	outwards directional (*aywa-rqu-n* ‘he goes out’)
-na	DISC	discontinuative (‘now’, ‘already’)
-ni	FON	euphonic (prevents segmental sequences prohibited by phonotactics)
-pa	GEN	genitive
-y	INF	infinitive (used also in imperatives)
-wan	INST	instrumental-comitative
-ntin	INTEG	integrative
-man	IRR	irrealis (conditional)
-shwan	IRR.1PL	irrealis (conditional), first person plural
-pa / -ri / -ska ~ -ski	ITER	iterative
-lla	LIM	limitative (used also as diminutive)
-chaw	LOC	locative
-ku ~ -ka / -kaa	MID	middle voice
-taaku	NEG	negator
-tsu	NEG	negator
-na / -nqa /	NMLZ	nominalizer
-ta	OBJ	object
-kuna	PL	plural (nominal)
-yaa ~ -ya	PL	plural (verbal)
-yuq	POSS	possessive (*qillay-yuq* money-POSS ‘someone who has money’)
-yka	PROG	progressive aspect

(continued)

suffix	gloss	meaning
-naa ~ -naq	PST.REP	reportative past tense
-rqa ~ -ra ~-rqu	PST	past tense
-q	PST.HAB	habitual past
-shqa ~ -shaa ~-sh	PRTCP	past participle
-ku	Q	polar question
-naku	RECP	reciprocal (*mutsa-naku-yaa-n* ‘they kiss each other’)’
-sh ~ -shi	REP	reportative
-naw	SIMIL	similarity
-pti	SUBDIFF	subordinator, different subjects in matrix and subord. clause
-shpa / -r	SUBID	subordinator, coreferential subjects in matrix and subord. clause
-naaqa	SUPL	superlative degree comparison
-yaq	TERM	terminative (‘until’, ‘up to’)
-qa	TOP	topic (focal)
-paku / -ykacha	UNSPEC	non-specific reference
-ya	VBLZ.INTR	verbalizer (intransitive)
-tsaa ~ -tsa	VBLZ.TR	verbalizer (transitive)

stem	gloss	meaning
ka	COP	copula
tsay / washa	DEM.DIST	distal demonstrative
kay	DEM.PROX	proximal demonstrative
mana	NEG	negator (particle)
na	PSSP	*passe-partout* morpheme (placeholder, stand-in for any stem)

Glosses from other languages cited in this work

affix	gloss[2]	meaning	language
-no	GEN	genitive	Japanese
go-	HON	honorific	Japanese
-ga	NOM	nominative	Japanese
-wa	TOP	topic	Japanese
-ari	DAT.SG	dative singular	Northern Bizcayan Basque
ˈ-ari	DAT.PL	dative plural (preaccenting)	Northern Bizcayan Basque
-an	GEN.SG	genitive singular	Northern Bizcayan Basque
ˈ-an	GEN.PL	genitive plural (preaccenting	Northern Bizcayan Basque
-I	ACC	accusative	Turkish
-ya	DAT	dative	Turkish
ˈ-ymış	EVID	evidential (preaccenting)	Turkish
-ˈIyor	IMPERF	imperfective	Turkish
-lar	PL	plural	Turkish
-an	REL	relativizer	Turkish

2 For the origin of these glosses, see references in the text.

https://doi.org/10.1515/9783111304595-208

1 Introduction

This thesis describes the intonation of two varieties of Spanish and Quechua that are spoken by the same bilingual speaker community in Huari, Peru. Spanish and Quechua are generally seen to be typologically quite different and genetically unrelated, but in large parts of Latin America they have had a common history of being spoken by the same people, in some cases for centuries. In Huari, both Quechua and Spanish are spoken by nearly everyone and used daily in interaction. This means that for a lot of daily communicative functions, Huari speakers have (at least) two ways of putting them into language at their disposal. The thesis (and the research project out of which it has emerged) takes as its basis the assumption that using equivalent communicative interactions as a third of comparison provides a unique opportunity for studying how one way of putting things into language differs from another. This allows us to see what kind of constraints are specific to only one of those ways (linguistic variants or systems), and which are perhaps shared. Doing this across different types of communicative situations even allows us to see what kinds of interactive functions (meanings) are encoded in what way in these systems.

Intonation is particularly interesting in this regard, since in many languages including Spanish it has been found to convey a variety of interactional meanings that have to do with how a discourse progresses, how knowledge is negotiated by its participants and what their subjective takes on it are. For Quechua on the other hand, many similar meanings have been found to be encoded via morphology, yet the question whether this means that intonation/prosody plays no role for the expression of such meanings in Quechua has mostly not been addressed. In order to be able to say something about what kinds of meaning the prosody of a language encodes, its prosodic system first has to be described. Yet for this, it is important to know what meanings there are present in the interaction. This last part is luckily broadly dealt with via the general methodology of comparing the same kinds of interactions across both languages. Thus this thesis proposes to describe the intonational systems of Huari Quechua and Huari Spanish (both previously undescribed) from observing their uses by the same speakers in similar interactions, and in this way to learn something about how the interactional meanings that can be found to play a role are encoded. This fills a research gap in that no detailed intonational and prosodic descriptions exist of varieties of Quechua or Spanish that are geographically or typologically close to those of Huari, and almost no studies that relate prosody to discourse meaning in a pair of varieties of these languages spoken in the same speaker community.

https://doi.org/10.1515/9783111304595-001

Therefore the leading questions, subject to further expansion and refinement (in chapter 4), are:

(1) What are the relevant properties of the intonational systems of Huari Spanish and Huari Quechua?

(2) How and what kinds of interactional/discourse meaning do they encode?

(3) Which of these properties are specific to one language, and which are perhaps common to both?

The third question has a subquestion that will resurface from time to time throughout the thesis, namely whether certain kinds of intonational systems lend themselves better to the encoding of meaning than others.

To answer these questions, the thesis is structured as follows. Chapter 2 gives a short introduction to Huari and the general method. Chapter 3 provides a theoretical background for the analytical chapters, introducing the models of intonational and pragmatic analysis used and their assumptions, as well as setting out the dimensions along which intonational systems have crosslinguistically been shown to vary. This is done because the prosody of Quechua is largely undescribed and some of the existing descriptions are potentially influenced by the perceptual filters of linguists reared on European languages. Since I am also a European linguist (and many of my readers will be, too), I try to work against this bias by initially mapping out what options are known to be possible and thus to expand the expectational horizon. Based on the theoretical discussion in chapter 3, chapter 4 then provides further refined research questions. Answers to them are then sought in the analytical chapters on Huari Spanish (chapter 5) and Huari Quechua (chapter 6) intonation. Finally, chapter 7 first summarizes the results from the analytical chapters and then concludes the thesis on a discussion of how some of the results from the two languages can be brought together.

Before heading into the fray, I should address the elephant in the room. This thesis is not about language contact as such, even though its object are two genetically unrelated language varieties spoken by the same bilingual speakers. I subscribe to the view that language contact is ubiquitous even in speaker communities traditionally considered monolingual, and thus a fundamental component of variation, which the thesis is very much about (cf. Mufwene 2001; Enfield 2014; Otheguy et al. 2015; Höder 2018). I will however make no claim about the origin of variable features that turn out to be shared between the two languages. It seems to me that the logic of assuming that a feature observed in Huari Spanish and Huari Quechua “comes from” Quechua “into” Spanish because it is not known from other descriptions of Spanish is faulty in the circumstances of this project. It might well have been in existence in Huari Spanish for several generations, in which case it is simply part of Spanish for the speakers, or actually part of their repertoire across both languages. It might also have "come" from an as yet undescribed neighbouring variety of Spanish. Language contact only really becomes a meaningful concept once larger populations are considered diachronically. This thesis is a very localized and synchronic endeavour. My focus is on describing features of the prosodic systems in both languages and then comparing them, plausibilizing what can be thought of as common and convergent, or specific and divergent, in this speaker community. In any case, a detailed and empirically grounded description of what multilingual speakers can be seen to actually do at a certain point in time and space may serve as a data point for future studies that are interested in the possibility of the propagation of features across longer spatial and temporal distances.

2 Data & general methods

This chapter provides an overview over methods and procedures used in this thesis, both for data collection and analysis. It also serves to situate the data in a sociolinguistic context via a brief description of the region in the Peruvian Andes they come from, and some sociodemographic key characteristics of their bilingual speakers.

2.1 Data collection

The speech data analyzed in this work was collected by Raúl Bendezú Araujo, Uli Reich, and the author during two field trips to the central Andean town of Huari in Ancash, Peru in August – October, 2015, and April – June, 2017. The field trips were part of a research project funded by the *Deutsche Forschungsgemeinschaft (DFG)* with the title *Zweisprachige Prosodie: Metrik, Rhythmus und Intonation zwischen Spanisch und Quechua* ("Bilingual prosody: meter, rhythm, and intonation between Spanish and Quechua", PI Uli Reich).[3] Potential participants for the recordings were contacted via a friend-of-a-friend-approach, through initial contact with Leonel Menacho López in Huaraz and Gabriel Barreto Echiparra in Huari, both of whom acted also as native speaker consultants for the production of the experimental materials and in general. Their contributions were invaluable to the success of the research project as a whole. Speakers participated of their own free will, and gave written consensus to be recorded and have their data published anonymously. They were remunerated financially for their time. Recordings took place in various localities in Huari. Efforts were made to record in silent, closed environments and to reduce background noise during recordings as much as possible. However, the surrounding soundscape of a rural Andean town could not always be excluded. The data used in this thesis comes from those recordings that were the best both in terms of recording quality and naturalness of production, i.e. from speakers that spoke freely and comfortably, and enjoyed participating in the tasks. Recordings were made using a mounted Røde NT-1A condenser microphone and a Marantz

3 The DFG-project number is 274614727. Its website is at: https://www.geisteswissenschaften.fu-berlin.de/we05/institut/mitarbeiter/reich/forschung/DFG-projekt-zweisprachige-Prosodie/index.html. It contains links to those data that are already available online, descriptions of the experimental tasks, and a full list of contributors to the project. The project extends the same elicitation methods to other language pairs in Latin America, including Guaraní-Spanish in Asunción (Paraguay) and Nheengatú-Portuguese in Saõ Gabriel da Cachoeira (Brazil), for which data is in the process of being published on the website.

https://doi.org/10.1515/9783111304595-002

PMD 670 audio recorder (in 2015) or a Zoom H4N audio recorder (in 2017) in 44.1 KHz (BPM). The Quechua data recorded in 2015 were transcribed and translated by bilingual students and Leonel Menacho at the Universidad Nacional Santiago Antúnez de Mayolo (UNASAM) in Huaraz, and their morphology annotated by students at the Pontificia Universidad Católica del Perú (PUCP) in Lima and Leonel Menacho. They were paid for their work. Transcriptions and annotations were checked for consistency by Raúl Bendezú Araujo and the author, in accordance with the glossing conventions devised by them, Uli Reich, and Leonel Menacho, based on Parker (1976) and Max Planck Institute for Evolutionary Anthropology (2015). The full list of glosses can be found at the beginning on this thesis. They are also the glosses used in each Huari Quechua example in this study. The Spanish data were transcribed and translated by students at Freie Universität Berlin, Raúl Bendezú Araujo, and the author.

2.2 Brief description of Huari and its linguistic situation

Figure 1: Map of Huari within Ancash (left), and within Peru (right). © OpenStreetMap.

Huari is a small town of about 5000 inhabitants, the capital of the province of Huari in the department of Ancash, whose capital is Huaraz. It lies on the eastern slopes of the cordillera blanca range of the central Andes in the Conchucos region (cf. Figure 1). Until 2016, the main road into Conchucos was not fully tarmacked (Wikipedia 2021b), limiting access to the region. The area is one of the regions in Peru with the highest share of the population giving Quechua as their first language (cf. the left map in Figure 2). In Huari province, 72.7% of the population stated in

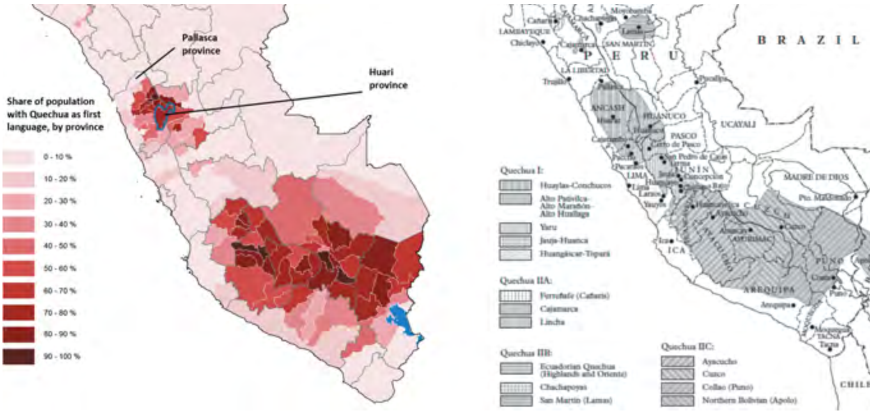


Figure 2: Maps of the share of the population that names Quechua as first language per province in central and southern Peru (left)[4] and of the distribution of the Quechuan languages in the same area (right, from Adelaar & Muysken 2004: 184).

2017 that Quechua is the language in which they learned to speak, while 24.7% said that it is Spanish; out of a total population of about 60.000 (Instituto Nacional de Estadística e Informática 2018). The variety of Quechua spoken in Huari belongs to the branch of Quechuan languages[5] that is variously called Quechua I, Quechua B, or central Quechua. I will refer to it as Quechua I when necessary. Quechua I varieties are spoken in the central Peruvian Andes, mainly in the mountainous parts of the departments Ancash, Huánuco, Junín, Pasco, and Lima (Adelaar & Muysken 2004: 185, cf. also the map on the right in Figure 2). The other main branch of Quechuan is called Quechua II or A, and its varieties are spoken in southern Peru including Cuzco, Bolivia, and a small enclave in Argentina, as well as in Ecuador, areas of Amazonian Colombia and parts of northern Peru. The Quechua I branch has a far greater internal diversity on a much smaller area, which is why the area where its varieties are spoken has been suggested as the Quechua homeland (Adelaar & Muysken 2004: 181). The Quechua varieties in Conchucos both differ from those spoken in Huaylas on the other side of the cordillera blanca and show a considerable internal (phonological) diversity (Parker 1976: 28). Diversification is plausibly due to the mountainous terrain facilitating the maintenance of very local

4 Map adapted from Wikipedia (2021a) based on data from Instituto Nacional de Estadística e Informática (2018).

5 All in all, 8–10 million people are estimated to speak languages belonging to the Quechuan family in an area extending from Putumayo in Southern Colombia throughout Ecuador and Peru and into parts of Bolivia, with small numbers of speakers also in Argentina and Chile, and internal diversity in the family comparable to that in the Romance or Slavic families (Adelaar & Muysken 2004: 168).

linguistic traditions. In the northern part of the Conchucos region, Sihuas Quechua and Corongo Quechua have been claimed and partially described as separate varieties (Parker 1976: 28; Hintz 2000; Hintz & Hintz 2017), while the Quechua spoken in the southern part including Huari province has been called South Conchucos Quechua but also described with some further internal variation (Hintz 2006, 2007; Hintz 2014; Hintz & Hintz 2017). For practical purposes, I will call the data sample analyzed in this work simply Huari Quechua and Huari Spanish by virtue of it having been recorded in Huari by speakers living in Huari town. The analysis itself will show how much variation still exists in this sample,[6] but I will remain agnostic about how much difference there might have to be in order to say that it consists of separate varieties or whether this is even a useful measure. Parker (1976) is a grammar of Quechua varieties that include Huari Quechua. More general grammatical descriptions of Quechua can be found in Cerrón-Palomino (1987); Adelaar & Muysken (2004). These works have a strong focus on morphosyntax and segmental (morpho-)phonology. Apart from brief impressionistic descriptions in Parker (1976) and comparatively short studies of the South Conchucos Quechua stress system (Stewart 1984; Hintz 2006), as well as our own previous work on acoustic correlates (Buchholz & Reich 2018), the prosody of both Huari Quechua and Huari Spanish is undescribed. See section 3.3.3 for a more detailed review of what is (not) known about the prosodic characteristics of Quechua in Conchucos.

Regarding Spanish, to my knowledge almost no studies from any linguistic subdiscipline exist of a variety from a region close to Conchucos. Andrade Ciudad (2016) is a study of northern Peruvian Spanish varieties (not their prosody) that includes data from the northernmost tip of Ancash, but specifically from Pallasca province (Andrade Ciudad 2016: 32), where Quechua bilingualism is actually very low (0–10% Quechua as first language, cf. both maps in Figure 2). Andrade Ciudad (2021) lists only one further small work on any variety of Spanish from the entire department of Ancash apart from Andrade Ciudad (2016). He concludes that what has come to be called "Andean Spanish" in the literature is a label for a somewhat heterogeneous cluster of varieties and almost exclusively drawn from studies of bilingual southern Peruvian regions, especially Cuzco (Andrade Ciudad 2021: 128–133).[7] One of the

6 In terms of prosody this will be made explicit, but the examples in the work also demonstrate in passing, as it were, that variation exists at various levels. In Huari Spanish this concerns e.g. the realization of the 3rd person singular present indicative form of *estar*, which is *está* for most speakers, but seemingly regularly *tá* for some (cf. note 177). In Quechua there is e.g. some variation in pronunciation, as between *tsuqllu* [t͡suxʎu] and *chuqllu* [t͡ʃuxʎu] 'corncob' (cf. Figure 126 vs. Figure 158).

7 He relates this to a broader academic tendency to favour the southern Andes and the Cuzco region as object of study in various disciplines (Andrade Ciudad 2021: 133). A similar imbalance most salient phonological features described for “Andean Spanish” is not attested for the monolingual Northern Peruvian varieties studied by Andrade Ciudad (2016: 224–225), the variant realization of high and middle vowels called *motoseo*, in which e.g. *mesa* ‘table’ is said to be produced as a homophone of *misa* ‘church mass’, or *poro* ‘leek’ of *puro ‘pure’*, and vice versa. This phenomenon is usually connected to bilingualism with Quechua, which only has three vowels /a i u/ or /a ɪ ʊ/ instead of the Spanish five /a i u e o/.[8] Since Huari is a fully bilingual area but the Quechua spoken is very different from the Quechua II varieties in regions on which the description of “Andean Spanish” has mostly been based (mutual intelligibility is usually said not to exist at least between the two main branches of Quechua, cf. Parker 1976: 24, see also Adelaar & Muysken 2004: 188–191 for a list of divergent features), neither what has been described for southern “Andean Spanish” nor for the monolingual northern varieties by Andrade Ciudad (2016) can simply be assumed to hold for Huari Spanish. This applies also to the prosodic descriptions of Cuzco Spanish by O’Rourke (2005, 2012), for which also see sections 3.4.4 and 3.5.1.

In Huari town, our own observations as well as statements by a number of people interviewed suggest that the overwhelming number of inhabitants are bilingual and use both languages in daily interaction, yet we also met a few individuals who were functionally monolingual in either Spanish or Quechua.[9] Short interviews were conducted with all experimental participants. Apart from eliciting basic sociolinguistic characteristics (see next section), participants were also asked which language they would normally use in different communicative situations, e.g. to speak with one’s family, with friends, with unknown people, with authorities, in the market, etc. The responses from 39 bilingual participants that were

---

seems to prevail in studies on Quechua, although a number of detailed descriptions of Quechua I varieties exist (cf. Adelaar & Muysken 2004: 193–194). Quechua I varieties are certainly underrepresented in terms of instrumental prosodic studies compared to Cuzco Quechua (see sections 3.3.3 and 3.7.3.2).

8 Guion (2003) investigates quality of front and back vowels in bilingual speakers of Spanish and Ecuadorian Quichua and monolingual Spanish speakers and paints a more nuanced picture: some speakers do not have separate realizations for /i/ and /e/, and /u/ and /o/, respectively, but others do. Some even produce three front vowels, two ([i] and [e]) for Spanish, and a distinct [ɪ] for Quechua, with the most differential system most often found with the earliest bilinguals. Pérez Silva et al. (2008) present similar results on Cuzco Spanish and Quechua: less proficient bilinguals produce Spanish mid and high vowels with more overlap in F1 than Spanish monolinguals, but more proficient bilinguals have a similar vowel space to monolinguals. They make the important sociolinguistic point that the characterization that Quechua speakers “confuse” the vowels (e.g. produce [misa] meaning /mesa/ and vice versa) is a baseless racist prejudice against the indigenous population.

9 Data by one monolingual Quechua speaker is contained in Bendezú Araujo et al. (2019).

interviewed in 2015 are given in Figure 3. These are of course self-reports, so based on the assessments of the speakers themselves, and therefore shaped by individual biases and value judgements. They thus give a measure of the subjective perception of language prevalence in a given situation rather than an objective measure of it. Participants were all bilingual and selected via a friend-of-friend-approach, so the results are not representative for the population of Huari as a whole. However, they still suggest tendencies regarding a possible functional differentiation according to communicative domains between Huari Quechua and Spanish for bilingual speakers.

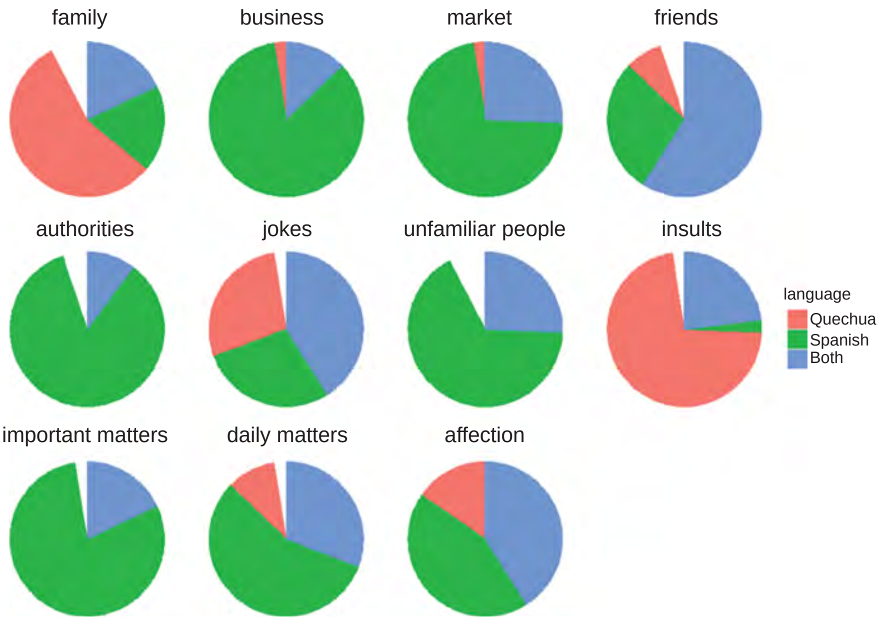


Figure 3: Responses from 39 bilingual speakers in Huari town regarding which language they would normally use in different communicative situations. White space indicates where not all 39 participants responded to a question.[10]

Figure 3 suggests that the use of Spanish dominates overall, in 7 of the 11 communicative situations asked about. It is especially prevalent in situations of social distance between speakers ("business", "authorities", "unfamiliar people"), commerce ("business", "market"), or to discuss important issues. Quechua, in contrast,

10 One speaker gave "English or French" as response to which languages they would use with unfamiliar people.

is dominant only in situations characterised by social proximity or expressive needs (“family”, “insults”), but not all of them (cf. “affection”). It is notable that in a number of situations of daily life (“friends”, “jokes”, “daily matters”, “affection”), a large share of the participants (a third or more) stated that they use both languages interchangeably. In fact, both languages were given as response by at least a quarter of all participants for 6 of the 11 situations, suggesting a functional equivalence between the two languages in these domains for them. In general, the results suggest that at the moment of recording, both Quechua and Spanish are healthy and that speakers are indeed in general functionally bilingual across a number of domains. However, it also seems likely that Spanish is slowly encroaching upon more of the communicative domains occupied by Quechua.

2.3 Short sociolinguistic profile of the speakers

This section gives an overview over some basic sociolinguistic characteristics only of those speakers whose data is analysed in this work. Overall, data from 27 speakers (9f, 18m, median age 21 years) was used. A short interview was conducted with each participant to obtain the sociolinguistic data. All information is thus based on self-description. At the time of recording, all speakers had their place of residence in Huari town. All speakers are bilingual[11] and fully literal in Spanish but do not read or write Quechua. Table 1 gives the identifier code for each speaker that is also used in each example in this work (column 1), together with their age at the time of recording (column 2), their sex (column 3), place of birth (column 4), level of education (column 9) and occupation at the time of recording (column 10). It also provides data about which of the two languages they first acquired (column 5), age of acquisition of the second language (column 6) and the response to two questions aimed at approximating a notion of which language the speakers are more dominant in or have a preference for.[12] They were *¿qué lengua te/le parece más fácil?* ‘which language is easier for you?’ (column 7) and *¿qué lengua utiliza(s) más?* ‘which language do you use more?’ (column 8). In columns 5, 7, and 8, „Q“ stands for Quechua, „Sp“ for Spanish, and „+“ for both. Not all speakers gave the

11 Speaker QP44 very rarely speaks Quechua, but that seemed partially due to a social stance oriented towards a life outside the limits of Huari rather than lack of linguistic competence.

12 This is of course a bare-bones-approximation. A thorough assessment of the degree of bilingualism of the Huari speakers both individually and as a community is not possible on the basis of the sociolinguistic data presented here and it is also not the goal of this study. See Schmeißer et al. (2015); Treffers-Daller (2019) for a discussion of the complexities involved in defining and assessing language dominance in bilinguals.

same response to both questions, suggesting that they in fact capture slightly different aspects. It is notable that even though the first language for 20 of the 27 speakers is Quechua (about 75%, just as in the overall population according to the 2017 census), its share is far lower in response to both of these questions, with only 2 stating that it is the language they use more, and 6 that it is easier. In both cases, "both" is the most frequent response (14 both, 2 Quechua, 11 Spanish for „most used"; 10 both, 6 Quechua, 10 Spanish for „easier"), again overall indicating a fairly solid bilingualism both in these speakers and the community they come from. Clearly there is also some speaker-specific variation in the responses given here. It will however be seen in the analytical chapters that it is not a straightforward predictor of the prosodic variation observed there. The rightmost column in Table 1 gives the types of experiment corpora that were used in this study from each speaker. They are described in the following section. Bolder horizontal lines in the table above and below the data from two speakers indicate that they were dialogue partners in the dialogical tasks. All participants who were partners in tasks are either close relatives, spouses or friends, and volunteered for the experiments together, except for SG15 & QF16, who volunteered individually and were partnered by us.

2.4 Description of the experimental tasks

This section describes the communicative experimental tasks whose recordings are the production data forming the database for this work. For the dialogical tasks, speakers participated in pairs that stayed the same throughout all the tasks. Speaker pairs are marked by bolder horizontal lines in Table 1 in the previous section. Participants could choose which language to start in and first performed all of the tasks in one language and then in the other, with breaks whenever they chose. They were told to treat the tasks like communicative games and the overwhelming feedback they gave was that they very much enjoyed themselves, many saying they would like to come back and "play more games".

None of the tasks required reading, since speakers are not literate in Quechua (but all are in Spanish). Verbal materials used in the tasks were thoroughly checked and approved by native speakers, Leonel Menacho and Gabriel Barreto in the case of Quechua, and Raúl Bendezú Araujo in the case of Spanish. They appeared in the form of oral recordings in the tasks, spoken by Gabriel Barreto in Quechua and Raúl Bendezú Araujo and Gabriel Barreto in Spanish. The other type of material consisted in pictures of objects that occurred in different functions throughout the tasks. They were chosen so that the names of the objects depicted are controlled metrically, i.e. they systematically vary heavy syllables ((C)VC or

Table 1: Basic sociolinguistic characteristics of the speakers whose speech data is used in this study.

Speaker	Age[13]	Sex	place of birth[14]	L1	age of acquisition L2 (years)	easier L	most used L	education[15]	occupation	corpora[16]
NQ01	34	f	Huari	Sp	0–4	+	+	sec. ed.	housewife	Elqud
TP03	34	m	Huari	+	–	+	+	sec. ed.	municipal administrator	Conc, Maptask, Cuento, Elqud
KP04	29	m	Huari	+	–	+	+	sec. ed.	mechanic	Conc, Maptask, Cuento
QZ13	26	m	San Luis, CFF	Q	9–12	+	+	sec. ed.	food production technician	Conc, Maptask, Cuento, Elqud
OZ14	27	m	San Luis, CFF	Q	5–8	Sp	Sp	sec. ed.	food production technician	Conc, Maptask, Cuento, Elqud
SG15	36	f	Huari	Q	9–12	+	Sp	sec. ed. (3rd year)	cook	Conc, Maptask, Cuento, Elqud
QF16	50	m	Huari	Sp	0–4	Sp	Sp	sec. ed.	primary school teacher	Conc, Maptask, Cuento
LJ22	26	m	Huachis	Q	9–12	Sp	+	sec. ed.	technician	Elqud
AZ23	24	f	San Luis, CFF	Q	5–8	Sp	+	sec. ed.	university student	Conc, Maptask, Cuento, Quién
ZZ24	21	m	San Luis, CFF	Q	5–8	+	+	sec. ed.	university student	Conc, Maptask, Cuento, Quién, Elqud
MS27	20	f	Chavín de Huantar	Q	9–12	Q	Sp	sec. ed.	university student	Conc, Maptask, Cuento, Quién
CF28	22	m	Chavín de Huantar	Q	9–12	Sp	Sp	sec. ed.	university student	Conc, Maptask, Cuento, Quién

13 Age at time of recording in 2015, or 2017 if only data from the task *Elqud* by that speaker is used.
14 Refers to a district within Huari province if only a single name is given and to a district within a different province in Ancash if two names of the form "x,y" are given. "CFF" and "AR" stand for Carlos Fermín Fitzcarrald and Antonio Raimondi, respectively, two neighbouring provinces.
15 "sec. ed.", standing for "secondary education", here means that secondary education (from ages 12–18) was completed.
16 See next section for an explanation of the tasks.

ZR29	19	f	Huari	Q	9–12	Sp	Sp	sec. ed.	university student	Conc, Maptask, Cuento, Quién
HA30	19	f	Huari	Q	5–8	+	Sp	sec. ed.	university student	Conc, Maptask, Cuento, Quién
XU31	21	m	Huari	Q	13–16	Sp	+	sec. ed.	university student	Conc, Maptask, Cuento, Quién
OA32	21	m	San Nicolás, CFF	Q	13–16	+	Q	sec. ed.	university student	Conc, Maptask, Cuento, Quién
XQ33	20	m	Llamellín, AR	Sp	9–12	Sp	Sp	sec. ed.	university student	Conc, Maptask, Cuento, Quién
LC34	21	m	Huari	Q	0–4	Sp	Sp	sec. ed.	university student	Conc, Maptask, Cuento, Quién
CJ35	21	m	Huari	Q	9–12	Q	Q	sec. ed.	university student	Conc, Maptask, Cuento, Quién
KA36	20	m	Huari	Q	5–8	Q	+	sec. ed.	university student	Conc, Maptask, Cuento, Quién
OV37	16	m	Huari	Q	0–4	Q	+	in sec. ed.	student	Quién
AC38	16	m	Huari	Q	9–12	Q	Sp	in sec. ed.	student	Quién
SO39	16	f	Huari	Q	5–8	Q	+	in sec. ed.	student	Conc, Maptask, Cuento, Quién
MD40	20	f	Yauya, CFF	Q	5–8	+	+	sec. ed.	university student	Conc, Maptask, Cuento, Quién
QP44	25	m	Huari	Q	0–4	Sp	Sp	sec. ed.	university student	Elqud
XJ45	18	m	Huari	+	–	+	+	sec. ed.	university student	Elqud
ZE55	40	f	Llamellín, AR	Sp	5–8	n.a.	+	univ. ed.	housewife	Elqud

(C)V:) in any position in bi- and trisyllabic words in Quechua. Almost the same set of pictures was used for Spanish, chosen so that the intended words varied stress position. Words and candidate pictures were checked for their aptness to elicit the intended words with Gabriel Barreto before being used in the tasks. The different combinations of heavy and light syllables in Quechua did not yield the same number of easily depictable words, as (4) shows. In the tasks, speakers also sometimes spontaneously chose other than the intended words for the objects on the pictures.

(4) Intended Quechua words depicted by images of objects in the tasks, with heavy (**h**) or light (**l**) syllables
  a. **h.h** *ach.kas*[17] 'lamb'; *is.may* H.H 'excrement'
  b. **h.h.l** *aq.tsall.ku* 'hair of the corncob'
  c. **h.l** *all.qu* 'dog'; *an.ka* 'eagle'; *hir.ka* 'mountain'; *tsiq.tsi* 'bat'; *tsuk.lla* 'hut'
  d. **h.l.h** *pam.pa.kuy* 'funeral'; *ull.tu.kuy* 'tadpole'
  e. **h.l.l** *pin.ku.llu* 'type of flute'; *as.wa.na* 'clay vessel'
  f. **l.h** *a.ñas* 'skunk'; *a.tuq* 'fox'; a.rash 'lizard'
  g. **l.h.h** *ka.puq.yuq* 'wealthy person'
  h. **l.h.l** *cha.kall.wa* 'jawbone'; *na.ran.ha* 'orange'
  i. **l.l** *qa.qa* 'rock'; *ha.cha* 'tree'; *ha.ka* 'guinea pig'; *tsu.ku* 'hat'
  j. **l.l.h** *ha.tu.saq* 'giant, enormous'
  k. **l.l.l** *i.lla.pa* / *ti.lla.ku* 'lightning'; *pi.tsa.na* 'broom'

17 For the orthographic transcription of Quechua, I use the official alphabet for Peruvian Quechua varieties as set down by the *Resolución Ministerial* 12-18-85-ED from November, 1985 (cf. Cerrón-Palomino 2008: 70–71). The phonological values for its characters correspond largely to the ones in the IPA alphabet (International Phonetic Asssociation 2005), except in the cases of <ch>, <ll>, <ñ>, and <y>, that are adopted from Spanish orthography and therefore correspond to /t͡ʃ/, /ʎ/, /ɲ/, and /j/, respectively. It uses only three vowel graphs <a>, <i>, <u> corresponding to the three vowels of Quechua. I follow this use except in the case of Spanish loanwords in which I sometimes use five vowels to make the word orthographically more recognisable to readers familiar with Spanish. The consonantal phonemes of Huari Quechua are /p b t d k g q t͡s t͡ʃ f s ʃ h m n ɲ r l ʎ w j/, with the voiced occlusives only occurring in loans from Spanish (adapting slightly from Parker 1976: 37–44). Quechua I varieties do not have the glottalized and aspirated occlusives of Cuzco Quechua (cf. section 3.2.1). Regarding actual phonetic realization, which is quite variable, each figure with a pitch track and spectrogramme of Quechua speech in this work contains a tier with a syllable-separated phonetic transcription following IPA alphabet conventions.

(5) Intended Spanish words depicted by images of objects in the tasks
  a. Proparoxytones *águila* 'eagle'; *murciélago* 'bat'; *relámpago* 'lightning'; *mandíbula* 'jawbone'
  b. Paroxytones *escoba* 'broom'; *gigante* 'giant'; *roca* 'rock'; *árbol* 'tree'; *choza* 'hut'; *naranja* 'orange'; *zorro* 'fox'; *zorrillo* 'skunk'; *lagartija* 'lizard'; *olla* 'pot'; *renacuajo* 'tadpole'; *perro* 'dog'; *montaña* 'mountain'; *choza* 'hut'; *cordero* 'lamb'
  c. Oxytones *maní* 'peanut'; *funeral* 'funeral'

Beyond those imposed by the instructions at the beginning of each task and the materials, there were no restrictions by the experimenters on the communicative behaviour of the participants during the tasks. Besides a role as game master in some of the tasks, the experimenters in general only interfered when participants had clearly misunderstood some of the instructions, or asked for help. In particular, no interference was made when speakers used other than the intended words for the game objects, or when they switched between languages. None of the tasks involved scripted speech, and all of the speech data used in this work can therefore be classified as at least semi-spontaneous. Raúl Bendezú Araujo and the author were present as experimenters in all of the tasks whose data is used in this work; with Raúl Bendezú speaking Quechua sufficiently to act as game master in some of them. The following describes the individual tasks. All the data from dialogical tasks (*Conc, Maptask, Cuento, Quien*) were recorded in 2015. Data from the pseudo-dialogical task *Elqud* was recorded in 2017.

Conc. Short for *concurso* "guessing game", and similar to the children's game *Memory*. Twelve picture cards chosen according to the metrical control were laid out in a grid on the table between the participants, face up. The participants had half a minute to memorize where each picture was, then the cards were turned face down. The speakers then took turns saying where which picture was, one at a time. They were not allowed to touch the setup or to use gestures, so that they had to use spoken language. After a guess, the game master confirmed that he had understood correctly and then turned over the card the speaker had indicated. If they had been right, they scored a point, the card was put aside, and the speaker was allowed to continue. If not, the card was turned face down again and it was the other speaker's turn. The game was over once all cards had been successfully identified.

Maptask. Map tasks are an established type of elicitation game (cf. e.g. Anderson et al. 1991; Grice & Savino 1997, 2003). In our version, the two participants were each given a map with pictures of objects whose names were controlled metrically. Only one participant's map had a path drawn between the objects. The task consisted in communicating so that the participant without the path was able to draw the correct path also on their map. Participants were not allowed to look at each other's maps or to use gestures. The two maps were also different in that some of the objects were at different locations in them. The speakers were not told about this initially, so that solving the task necessarily included resolving the communicative conflict that would arise from the mixed-up locations. The game was over once the speakers had communicated the correct path successfully to their satisfaction.

Cuento. Spanish for "story", this task consisted of a version of the game *chinese whispers*. One of the participants left the room, and the other was played a recording of a short story containing many of the words chosen according to the metrical control. The stories were invented by the three experimenters and revised and read aloud by Gabriel Barreto for the Quechua recordings, and by Raúl Bendezú Araujo for the Spanish recordings. The participant could ask to have the recording replayed until they were confident that they remembered it. Then the other participant reentered the recording room, and the first participant told the story to them. After hearing the story and being free to ask anything about it, the second participant then told the story to the experimenters, while the first participant could help or correct them.

Quién. Spanish for "who", this is a task based on the game *who am I?*. The experimenters prepared four cards with names of people that they knew both participants to know (very famous people or people known personally to all involved), and one of the participants drew one of the cards in secret. The second participant then had to find out about the identity of the person on the card using any type of question, except for directly asking who it is via a wh-question, but with polar guessing questions like "Is it person x?" allowed. The first participant had to answer truthfully and to the best of their knowledge, without giving too much information away that wasn't directly asked for. The game was over once the person on the card had been correctly guessed or the second participant gave up.

Elqud. This is the only task whose data is used here that did not involve two participants, but only simulates utterances in a dialogical context. It uses a combination of visual and audio stimuli. Images or short animations were prepared that showed short situations or interactions between entities and objects. For each such situation, a corresponding audio was recorded (checked for acceptability and appropriateness and spoken for the recordings in both languages by Gabriel Barreto) that described the situation. In each of the experimental items, as opposed to the fillers, there was an intended discrepancy between audio and visual stimulus. The discrepancy was intended to correspond to different elements in the experimental items, e.g. a noun phrase corresponding to one of the figures in the visual stimulus ("the black dog"), a noun or adjective corresponding to an attribute of one of the figures ("the **white** dog"), a verb corresponding to an action or relation between the figures ("the dog **ate** the bone"), or also more complex parts of the audio stimulus corresponding to similarly complex relations in the visual stimulus, such as double-topic constructions ("the **white** dog plays with the **blue** ball, the **black** dog plays with the **red** ball"). Participants were shown a visual stimulus and heard the corresponding audio stimulus. They were asked to treat the audio stimulus as a contribution by a dialogue partner about the situation shown by the visual stimulus, and to respond to this contribution, either by correcting it (in the experimental items), or by agreeing to it (in the fillers). They were asked to respond as expansively as possible in general (verbalizing as many of the elements shown), but otherwise not constrained in the form of their response, and no intervention was made when they responded elliptically or contrafactually. In total, 69 stimulus pairs (including 17 fillers) each were created for Quechua and Spanish, and speakers were shown them in two pseudo-randomized sequences, first in one language, then the other.

The data from the three dialogical tasks *Conc*, *Maptask,* and *Cuento* forms the empirical backbone of the analyses throughout this study. Data from *Quien* (in both languages) and *Elqud* (only Spanish) is used where indicated, in particular in sections 6.1.6. and 5.2., respectively.

2.5 Partial availability of the speech data used in the analysis

The *Conc*, *Maptask*, *Cuento*, and *Quien* speech data from speaker pairs TP03 & KP04, QZ13 & OZ14, AZ23 & ZZ24, ZR29 & HA30, XU31 & OA32, and XQ33 & LC34 in Quechua and Spanish has been published with annotation files and metadata, as Bendezú Araujo et al. (2019) and Bendezú Araujo et al. (2021), respectively. It is freely available online[18] for any interested party to be used non-commercially with attribution (creative commons licence CC BY-NC-SA 4.0).

2.6 General method

Overall, this study uses a mixture of qualitative and quantitative analysis. In all sections, the analytical argument is based on an in-depth knowledge of the data acquired through several years of study and by the author's having been present during all of the recordings. In some sections, the results are only exemplified using individual examples. Other sections also employ quantification where possible, and occasionally also inferential statistical analysis.

18 At https://refubium.fu-berlin.de/handle/fub188/25747 and https://refubium.fu-berlin.de/handle/fub188/29497, respectively. Note that *Conc* is called *Memoria* in the published data.

2.7 Acoustical analysis, visualization and statistics

Acoustical analyses of the data were conducted in *praat* (Boersma & Weenink 2020). Scripts for the automatic extraction of measurements were written by the author, partially based on routines taught by Rafèu Sichel-Bazin. For the generation of pitch track, oscillogramme, and spectrogramme images used for the examples, a slightly customized version of the script *Create Pictures* (Elvira-García 2017) was used with a voicing threshold of 0.6 if not otherwise noted. Occasionally, pitch movement due to consonantal microprosody or background noise was manually removed. Statistical analyses were conducted in *R* (R Core Team 2020) using *Rstudio* (RStudio 2009–2020). Plots from *R* were produced using *ggplot2* (Wickham 2016). Further *R* packages are cited in those sections for whose analysis they were used. Individual sections also contain further specific methodological descriptions.

2.8 A short note on examples

Each Huari Spanish and Quechua example with a pitch track as well as each longer context example appearing in this study has a corresponding audio file hosted online via the Open Science Framework (Foster & Deardorff 2017) and published open access under creative commons license CC BY-NC-SA 4.0, at https://doi.org/10.17605/OSF.IO/MAH8Z (Buchholz 2021a). In electronic versions of this document that are hyperlink-compatible, the audio file is accessible via the weblink in a separate footnote for each pitch track or example.

Appendix **A** also lists the stable URLs for all audio files so that they can be accessed in document formats without hyperlink compatibility. In the text, examples are given a code, which is to be read as follows: *speaker_task_language_start time*, e.g. KP04_Conc_Q_1653 meaning that the example is by speaker KP04, from the *Conc* task in Quechua, starting at 165.3 seconds from the beginning of the recording of the task.

3 Theoretical background and literature review

In this chapter, I lay out the theoretical groundwork necessary for the analytical chapters. An empirical investigation into the prosody of Huari Spanish and Quechua necessitates not only the establishment of theoretical and analytical terminology, but also an understanding of the range and limits of prosodic variability.

3.1 Introduction

In the following I will introduce and develop fundamental concepts (sections 3.2 and 3.3), putting particular emphasis on the range of typological variation observed in the areas of stress, accent, and prosodic phrasing (section 3.4), and the relation between tones, prosodic structure, and segmental material (section 3.5). We will see that this variation goes beyond the descriptive coverage of labels such as "culminative stress", "delimitative phrasing", or "segmental anchoring". Such a detailed exploration of the ranges of prosodic variability is relevant because only from a comprehensive viewpoint is it possible to characterize the prosodic phenomena we will observe in both Spanish and Quechua, and to bring them into an appropriate perspective towards each other. Not doing so would run the risk of studying Conchucos Quechua, about whose prosody and intonation very little is known, with Spanish as an analytical foil, simply because so much more is already known about the prosody of Spanish varieties. With such a widened typological perspective, on the other hand, it will hopefully be better possible to analyze both local languages, Quechua and Spanish, on their own terms and to locate them, with their individual variability and areas of overlap and divergence, on the variation space mapped out by what is known about prosodic typology. Over the course of this chapter, the discussion of the relevant prosodic concepts therefore will often be based not only on what is already known about Spanish, but also about typologically more distant languages as comparison. Quechua will also feature wherever possible, but unfortunately, for the most part to mark gaps in our knowledge.

Finally, the phenomena observed in the areas of pitch peak and accent distribution, as well as of (scaling-based) cues for prosodic phrasing in the analyses of Huari Quechua and Spanish (in particular sections 5.1.2, 5.1.3, 5.2, 6.2.3.3, and 6.4), make it necessary to include the issues of possible recursiveness in prosody and the prosodic cueing of information structure into the analysis. The theoretical groundwork for these later analytical decisions is also laid in this chapter, in sections 3.6 and 3.7, respectively. Their analysis is intended to contribute to suggestions about how the prosodic variation space can be conceptualized cross-linguistically.

https://doi.org/10.1515/9783111304595-003

From the second half of the 20th century onwards, it has been increasingly recognized that speech sounds exhibit systematic behaviour that can only be described by going beyond individual segments. The study of these aspects of speech is therefore sometimes called *suprasegmental* phonetics and phonology. Linguistic sound systems organize individual speech sounds into larger units via the concerted manipulation of the parameters[19] pitch, loudness, spectral quality of vowels and consonants, and duration, and they allow for the signaling of meanings independent of those conveyed by the segmental string in particular by the manipulation of pitch. All aspects of this system concerned with the grouping and composition of units are together called *prosody*. The systematic manipulation of pitch for postlexical linguistic functions[20] in particular is called *intonation* (cf. Ladd 2008: 5–7). Prosody and intonation are intimately connected, since prosody defines the domains on which an intonational event must occur in order to fulfill a specific function, and the two terms are often used nearly interchangeably. The restriction to postlexical functions distinguishes intonation from the use of pitch to distinguish lexical contrasts in tonal languages.

19 The names given here are strictly speaking those of the perceptual/psychoacoustic correlates of physical properties of the acoustic signal. Pitch is the perceptual correlate of fundamental frequency, or F0, loudness that of intensity (energy amplitude), vowel quality that of formant frequencies above (overtones of) F0, consonant quality a mixture of that, aperiodic signals in the spectrum and timing measures such as voice onset time (VOT), and duration that of quantity (Ladefoged 2005; Fastl & Zwicker 2007; Ladefoged & Johnson 2011). The relationship between acoustic and psychoacoustic correlates is not straightforward in human perception, see e.g. Traunmüller 1990, 2017; Dawson et al. 2017; Fahey & Diehl 1996; Kuang & Liberman 2016; Ladd et al. 2013. For practical purposes in intonation research, the labels are often used interchangeably (Ladd 2008: 5).

20 Ladd (2008) stresses the distinction between *linguistic* and *paralinguistic* uses of pitch and other prosodic parameters. Paralinguistic uses signal emotional and attitudinal states of the speaker, and are characterized by a gradual and iconic relationship between what is signaled and the signal itself: paralinguistically, loudness signals arousal or emotional involvement; the louder, the more emotionally involved. Linguistic uses, on the other hand, are supposed to convey aspects of meaning that are discrete and categorical and be expressed by formal means that are also essentially discrete. It has long been recognized that intonation straddles the boundary between these domains (Bolinger 1978: 475 famously calling it a "half-tamed savage"), in particular because pitch variation is by nature gradual and continuous, instead of discrete and categorical. The idea of *paralinguistics* is furthermore problematic because many meanings thought to belong to it, such as propositional speaker attitudes, are not only expressed by quite discrete morphosyntactic means in several languages (e.g. German discourse particles, morphology for expressing *mirativity*, etc.), but can also be shown to be semantically well-defined, and expressable by equally well-defined intonational cues (Fliessbach 2023). Ladd (2008: 34–42) maintains the distinction conceptually, but allows for the integration of gradual form into phonology.

3.2 Prosodic units and their hierarchy

Prosody research has proposed the following hierarchy of units apart from segments: segments group together to form onsets, nuclei and codas of syllables (σ). Nucleus and coda together are called the rhyme. A rhythmic unit, the mora (μ), is often positioned below the syllable, but is better seen as orthogonal to it. One or several syllables together form feet (F or Σ), which in turn form prosodic or phonological words (ω). Prosodic words are not isomorphic with lexical or morphological words, but it is a reasonable heuristic to assume that they broadly map onto each other. In Spanish, some morphological words such as adverbs formed by attaching *–mente* to an adjectival root (e.g *rápidamente* "quickly", *nuevamente* "newly, again") are often produced as two prosodic words, as can be seen from them realizing two stresses, and even pitch accents, one on the stressed syllable of the root, and another on the penult of *–mente*. On the other hand, clitics do not affect stress and pitch accent placement on the words they attach to (e.g. *te lo traduzco* "I translate it for you", *cocinárselo* "to cook it for oneself"). Thus on the criterion of being the domain of stress assignment, they form a single prosodic word on more than a single morphological one. For Quechua, this is largely unexplored. Given that Quechua is agglutinating and allowing for quite long polysyllabic words consisting of a root plus a number of suffixes, observations about multiple stresses or accents on a single word in several varieties (e.g. in Parker 1976; Adelaar 1977; Hintz 2006) indicate that some morphological words are produced as more than one prosodic word. Non-isomorphy between the morphological word and the prosodic word as domain of stress assignment is assumed by Stewart (1984: 207–209) for Conchucos Quechua,[21] and by Levinsohn (1976: 20) for Inga Quechua (Quechua IIB, Southern Colombia), with the domain of stress assignment sometimes larger, sometimes smaller, than the morphological word.

Above the prosodic word, a number of units can be roughly classed into two phrasal groups: a smaller phrasal unit has been identified under the name of Accentual Phrase (AP), Minor Phrase, phonological phrase (PhP or ϕ), or intermediate phrase (iP). A larger unit has been called Major Phrase or Intonational Phrase (IP or ι), sometimes said to be equal to or yet below, the utterance (U or υ). Their attestation differs substantially between languages as well as between descriptive and theoretical approaches. Table 2, taken from Frota (2012: 257), gathers proposed hierarchies of units into three groups based on different approaches in the litera-

21 Conchucos Quechua includes Huari Quechua. When giving information about a different Quechua variety, I indicate the classificational group the variety belongs to (cf. section 2.2) as well as its geographic region.

Table 2: Different prosodic hierarchies proposed in the literature, from Frota (2012: 257).

a. Rule-based	b. Intonation-based	c. Prominence-based
Intonational Phrase (IP)	IP	Nuclear Accent
Phonological / Major Phrase	Intermediate Phrase	
Clitic Phrase / Minor Phrase / Prosodic Word Group	Accentual Phrase	Accent
Prosodic Word (PW)	(PW)	Stress
Foot	Foot	Full Vowel
Syllable	Syllable	Syllable
Mora	Mora	

ture. Early proposals in the rule-based generative tradition (Selkirk 1984; Nespor & Vogel 2007) assume that the prosodic hierarchy is universal across all languages. Later approaches (Selkirk 1996, 2011; Ito & Mester 2007, 2012; Féry 2017) maintain this claim but reduce the inventory by making all domains recursive in a departure from earlier views.

3.2.1 Different kinds of evidence for the existence of prosodic units in languages

Empirically speaking, a prosodic unit should only be said to exist in a language if there is tangible evidence for it, i.e. if phonological or phonetic processes can be shown to make reference to it, or if it is perceptually or articulatorily robust. In many languages, prosodic domains constrain how segments are distributed. This is most frequently seen with syllables, which impose restrictions on segment distribution in nearly all languages, based on preferences that enhance contrast and rhythmicity (Vennemann 1988). Thus, e.g. the sequence /sp/ is not ilicit *per se* but must always straddle a boundary between two syllables in Spanish, because codas may not contain more than one consonant and in onsets the two segments do not create a sufficiently steep sonority gradient. But such restrictions can also act on larger domains: either an aspirated or a glottalized consonant, never both, may optionally occur only once per word (on the initial stop in the root and never in the suffixes) in Cuzco Quechua (Quechua IIC, southern Peru; Quechua I varieties do not have these consonants), unlike their simple counterparts, which are not constrained in this way (cf. (6); Parker 1997: 2; Cusihuamán 2001: 34). This restriction of once-per-domain is an example of a *culminative* property by which the relevance of a word-level unit can be argued, here for Cuzco Quechua.

(6) Distribution of glottalized and aspirated consonants in Cuzco Quechua: culminative but optional
  a. t'anta "bread"
  b. thanta "old, used up"
  c. tayta "father, old man"
  d. *tant'a
  e. *t'antha

Hyman (2006: 229) lists other properties besides culminativity that can be used to define prosodic word domains in a language. Phonological processes specific to a domain can e.g. affect the segmental makeup by allowing assimilation processes only within a domain, but blocking it at its edges. For s-aspiration in Spanish, the domain of application differs between varieties:

(7) /s/-aspiration in Spanish varieties (adapted from Strycharczuk & Kohlberger 2016)

	Andalusian / Honduras Spanish (Hualde 1991; Kaisse 1999)	Rio Negro Argentinian Spanish (Kaisse 1999)	Chinato Spanish (Hualde 1991)	Buenos Aires Spanish (Kaisse 1999)
dieces "tens"	[die.seh]	[die.seh]	[die. ðeh]	[die.ses]
desigual "unequal"	[de.hi.gual]	[de.si.gual]	[de.ði.gual]	[de.si.gual]
dos palas "two shovels"	[doh.pa.lah]	[doh.palah]	[doh.pa.lah]	[doh.pa.las]
dos alas "two wings"	[do.ha.lah]	[do.ha.lah]	[do.ða.lah]	[do.sa.las]

Data such as in (7) is often interpreted to show that in all varieties, the minimal condition for aspiration of /s/ is for /s/ to be in coda position. However, what counts as a coda position is influenced by another process, usually called "resyllabification", which causes postvocalic consonants to be produced as an onset when followed by a vowel, even if a word boundary intervenes. The syllable boundaries, symbolized by the dot (.), that are given in (7), are intended to be those after this "resyllabification" has occurred. To explain (7), Hualde (1991) and Kaisse (1999) propose that aspiration and resyllabification occur in different orders between the varieties (see there for details). In Buenos Aires Spanish, but not in the other varieties, aspiration does not occur before a pause (Kaisse 1999: 206–207), i.e. before a phrase boundary (Strycharczuk & Kohlberger (2016: 2)). Thus, /s/-aspiration interacts with the boundaries of up to three different domains: syllables, prosodic words and (phonological) phrases. Resyllabification also does not occur across pauses, but its complex workings lead Cardinaletti & Repetti (2009) to propose a new prosodic constituent, the “phrasal syllable level”, as its domain. In the next section we will see what the pitfalls of proposing prosodic constituents based on individually observed phenomena can be.

3.2.2 Universal aspects of the prosodic hierarchy

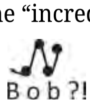
Assuming universality of the units of the prosodic hierarchy would at this point mean postulating the phrasal syllable level also for all other languages, even if no processes ever take it as their domain apart from Spanish resyllabification. However, on the criterion of demonstrably being the domain of at least one phonological process in all languages, it turns out that few if any of the proposed units are really universal (cf. also Grijzenhout & Kabak 2009). Even the syllable, maybe the most universally accepted of these units, is perhaps not the domain of any phonological processes in at least one language, Gokana (Niger-Congo, cf. Hyman 2011, 2015 for a discussion). The prosodic word has been argued not to be universal based on several languages, including Vietnamese and Limbu (Sino-Tibetan, cf. Bickel et al. 2009; Schiering et al. 2010). In contrast, Himmelmann et al. (2018) present quite robust evidence that a unit corresponding to the intonational phrase can be identified in perception consistently even in languages the listeners are unfamiliar with. Its length also seems to average at 1.5 seconds in some data on English, French, and German reported on by Jun (2005d: 443), suggesting that there is at least some amount of overlap among observing linguists regarding what constitutes an IP crosslinguistically.

Arguments made about segmental prosodic processes are often made based on symbolic, categorical data such as given in (7), but increasingly, instrumental (acoustic and articulatory) findings are also brought to bear on them. They have resulted in similar observations of gradual variability in segmental realizations depending on prosodic domains across a variety of languages and are in general known as the prosodic strengthening of domain edges. The two edges of domains do not show the same effects: very broadly speaking, domain-initial strengthening often seems to make consonants at the beginning of larger prosodic domains such as utterances or IPs more consonant-like, judging from articulatory measurements such as increased linguopalatal contact and acoustic measures such as increased voice onset time (Fougeron & Keating 1997; Onaka 2003; Keating et al. 2004; Keating 2006; Cho & Keating 2009), while vowels enhance those features that make them more contrastive against other vowels (Georgeton & Fougeron 2014). On the other hand, final lengthening, as the name suggests, is an effect of increased duration at the end of phrasal domains that can affect both individual segments as well as syllables (Beckman & Edwards 1990; Rao 2007, 2010; Fletcher 2010; Petrone et al. 2017). The durational measurements are sometimes able to distinguish between positions defined with respect not just to a single, but several, prosodic domains (e.g. Strycharczuk & Kohlberger (2016: 7–9) on /s/-realization in Peninsular Spanish). Although both of these phenomena have been observed across a variety of languages and are thought to be "phonetic" markers of prosodic structure rather than (language-specifically) "phonological" ones (cf. Vaissière 1983; Keating 2006; Cho & Keating 2009: 466), individual studies also show differences both in terms of the precise nature and strength of the effects observed as well as the domain at which they occur between languages, and also between different information structural conditions (cf. Cruttenden 1997: 33; Fletcher 2010: 529–532). Differences between individual speakers should also not be discounted. For example, Strycharczuk & Kohlberger (2016: 9) note several degrees of durational domain-sensitivity in /s/-realization among their speakers, ranging from differential realization for each of the categories to total insensitivity across them.

3.3 Intonation in the autosegmental-metrical model

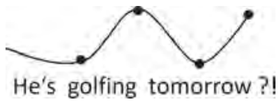
For the purposes of this work the most relevant phenomena identifying domains of the prosodic hierarchy are tonal ones. In the model of intonational phonology adopted here, the autosegmental-metrical (AM) model of intonation (Pierrehumbert 1980; Ladd 2008), tones are represented on a tonal tier that is autonomous from the segmental tier,[22] and both are independently assigned to a specific level and position in the prosodic structure. This autonomy of tones and segments and their mediation via the prosodic structure (called "tune-text-association") is vital for making sense of how tonal contours relate to segmental strings of different length and morphosyntactic complexity:

(8) The "incredulity" contour on utterances of different length and complexity

a. 

22 „Autosegmental" does not only mean that tones and segments are on separate tiers, although this was one of the original applications of autosegmental phonology (Goldsmith 1976). Different classes of segmental features are also taken to be located on separate tiers (e.g. McCarthy 1986; Hellmuth 2013; Venditti et al. 2008: 458–459).

b. 

c. 

In (8), the "incredulity contour"[23] (Ward & Hirschberg 1985; Hirschberg & Ward 1992), represented by the schematic rise-fall-rise above the text, is produced on three different utterances. In (8)a, it is realized on the monosyllabic name *Bob*, appropriate to a context where Bob has just been suggested as the answer to a pending question by someone else but prior to that was deemed by the speaker not to be a likely candidate for the question (e.g. who might cook a complex meal for six in the evening if the only culinary action the speaker has ever witnessed Bob performing was to bake a frozen pizza, and to burn it). In (8)b, it is produced on the complex noun phrase *the long-lost heir to the throne*, felicitous in a context e.g. where the speaker has just been told that Princess Peach, the royal scion, has made a public appearance and the speaker up to that point had believed the princess's whereabouts to be unknown. In (8)c, the same contour is found on the intransitive sentence *he's golfing tomorrow*, e.g. in a context where the speaker knows that an important parliamentary debate is taking place the next day and has just been told that the president will be golfing at the time. The domain for such a contour – not just this one, but all comparable ones – is taken to be the Intonational Phrase (Hirschberg & Ward 1992: 242). This is one aspect why AM assumes tune-text-association to happen via the prosodic structure: the important point is that whether the contour is felicitous depends not on the morphosyntactic makeup of the utterance, nor its length (although it must certainly consist of at least one sufficiently sonorous sound), but only upon the appropriate context conditions and that it be realized on one intonational phrase, in the correct form.

The correct form, however, depends on more than just proportionally adapting the tonal movement to the length of the text, spreading the rises and falls evenly across it: in both (8)b and (8)c, the high and low points in the contour are again

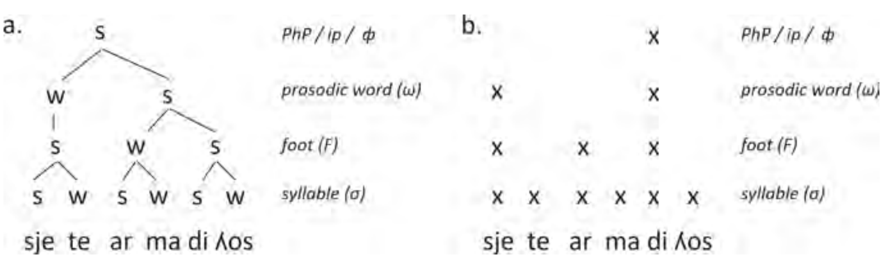
23 Also discussed under various labels in Liberman & Sag (1974); Liberman (1975); Marneffe & Tonhauser (2019) and elsewhere. Descriptions of its meaning are multifaceted. Hirschberg & Ward (1992) find evidence that increased pitch excursion shifts listeners' perception towards an "incredulity" reading of ambiguous sentences, while perception is shifted towards an "uncertainty" reading when pitch excursion is comparatively decreased. The same contour has also been found to evoke more negative scalar implicatures in listeners when used as an indirect answer to a polar question than a neutral declarative contour (Marneffe & Tonhauser 2019). I will use "incredulity" as a shorthand for the meanings associated with it.

located relative to specific positions in the prosodic structure. AM assumes just two tonal primitives, a high (H) and a low (L) tone. The rise-fall-rise contour under discussion here is usually taken to be made up of a sequence of four tones, LHLH. These are responsible for the pitch movement in the tune: only where a tone is specified is pitch actively manipulated; between tonally specified locations, any tonal movement is due to interpolation (Pierrehumbert 1980: 52). The tones in the phonological representation are not quite directly reflected in the actual pitch contour: they cause tonal targets to exist in the phonetic implementation, which is responsible for realizing the tonal contour together with the segmental string in speech production by, amongst other things, assigning pitch values relative to the overall pitch range of the utterance: high tonal targets, due to high tones, get assigned relatively high pitch values, low tonal targets, due to low tones, get assigned relatively low pitch values. In the LHLH contour, the position of the final H tone is determined simply by the right edge of the intonation phrase: it is as rightmost in the phrase as it can be. Such tones are called boundary tones. For the adequate specification of the other tones' position in the contour, the edges of prosodic units are not enough. That specification must make reference to the metrical structure of the utterance.

3.3.1 Stress and metrical structure

Metrical structure (Liberman & Prince 1977; Hayes 1995; Ladd 2008; Calhoun 2010b) assigns strength relations (prominence) throughout the prosodic structure. At each level of prosodic structure, all the constituents of the level below that are dominated by a constituent at that level, are assigned a strength relation such that only one of them is strong (s), and the others are weak (w).

(9) Metrical structure for the phrase *siete armadillos* "seven armadillos"



In (9), the metrical structure for the phrase *siete armadillos* "seven armadillos" is given in branching tree notation (a) and grid notation (b). The two notations are effectively equivalent for our purposes. In both, prominence is built from the ground up. Crucially, it is relative: at the level of the foot, the two syllables [ar] and [di] in *armadillos* are still equal in strength, but at the level of the prosodic word, the foot which contains [ar] is weak relative to the strong one containing [di], which is how the fact that [di] is the stressed syllable in *armadillos* is represented. Strong nodes at one level must always be founded on a strong node at the level below, so that strength propagates upwards. In this way, structures such as (9) can represent three facts at one. Firstly, that [sje] and [di] are the stressed syllables in their respective words. Secondly, that [di] is the strongest syllable in the phrase. And finally, that *armadillos* is stronger than *siete* in it. Even though it might seem so from this example, metrical structures are not maximally binary branching, but n-ary branching in principle, in order to allow the assignment of exactly one strong position for level x-1 at each constituent of level x dominating it (Nespor & Vogel 2007: 7). The metrical structure also assigns prominence at levels higher than the iP/PhP: in the utterance *Juan encontró siete armadillos* "Juan met seven armadillos", it would also put *siete armadillos* in a strength relation with the rest of the utterance. How it would do that however would depend on how the utterance is phrased and on its information structure (cf. section 3.7).

3.3.2 Word stress crosslinguistically

The strongest position at the word level is particularly important. It is usually called "word stress", "lexical stress" or simply "stress", although this is a somewhat confusing terminology, since "stress" also signifies the particular way in which many (European) languages configure their prosody in order to mark this position both phonologically and phonetically. This typically includes increased duration and intensity on stressed syllables in comparison to unstressed ones (Fry 1955 for English, Ortega-Llebaria & Prieto 2007, 2011 for Spanish, Gordon & Roettger 2017 for a broad crosslinguistic overview), but the evidence for such claims has to overcome a considerable number of methodological pitfalls (cf. Roettger & Gordon 2017). The most relevant of those is to separate stress from pitch accent. For Spanish in particular, Ortega-Llebaria & Prieto (2007, 2011) find that stressed syllables are longer and louder than unstressed ones even when not pitch accented. Acoustic cues to prominence are statistically robust in many languages, but in individual instances, they can often be missing or misleading. Yet experimental subjects securely identify prominent positions, both at the word level and above also under less than ideal cues and even against cues, up to a point (Terken & Hermes 2000; Bishop 2012; Cole et al. 2019, cf. also the works cited in Calhoun 2010b: 4–5), and the expectation of a prominent position to occur increases attention to detail even when cues are absent (Zheng & Pierrehumbert 2010). All of this indicates that metrical structure should primarily be thought of in relation to this generation of expectations about prominence (Reich & Rohrmeier 2014), which can then be exploited for interpretative information structural effects (Ladd 2008; Calhoun 2010b; Bishop 2012), rather than with regards to its acoustic correlates.

Language-specific stress-related phonological phenomena include e.g. the historical alternation in Spanish in the morphological paradigms of many words whereby /o/ and /e/ occur when a syllable is unstressed, but the diphthongized or pre-glided counterparts /we/ and /je/, respectively, occur when it is stressed, e.g. *apostár – apuésto* "to bet – I bet", *tenér – tiéne* "to have – s/he has" (accent indicates stressed syllable). Historically, there is also a tendency to gradually reduce material following the stressed syllable (syncope), resulting in increasingly few word forms where the stressed position is followed by more than one further syllable. This is a particular instance of the much more general observation that stressed syllables in many languages are most resistant to reduction processes, both historically and in production, relatable also to the fact that more peripheral and sonorous vowels tend to occur in stressed position, while unstressed positions often have a more restricted set of less sonorous vowels in many languages (Crosswhite 2004). For an overview of stress-related phonological processes in various Romance languages, see Meinschaefer (in press). In contrast, there are languages that do not share in any or most of the properties associated with "stress" but still have a comparable unique syllabic position at the word level. The stereotypical example is Japanese, where about 45% of the words in the lexicon (Kubozono 2008: 167) have such a unique lexically specified position which can be anywhere in the word, but it is not marked by increased duration or intensity, or any other phonetic or phonological process except a characteristic pitch movement (Beckman 1986; Venditti et al. 2008). Japanese is usually said to have a lexical (pitch) accent, which is somewhat confusing terminologically, because *pitch accent* is what the postlexically assigned tones linked to prominent syllables are called in AM. Leaving aside the issue of the phonetic and phonological correlate for the moment, word stress has two defining properties that directly derive from the nature of the metrical structure, *culminativity* and *obligatoriness* (cf. Liberman & Prince 1977: 263; Hayes 1995: 29; Hyman 2014: 60). Culminativity is the property that only exactly one syllable can be stressed (= assigned highest prominence) in a word; as we have already seen, this is a property of metrical structure that holds at each level. Obligatoriness means that every lexical content word in a language must be stressable. This is a corollary of every utterance having a metrical structure that assigns strength relations at each level. Having more than half of the lexicon consist of words that are not accented is clearly a different matter than having a closed and relatively small set of function words that cannot be stressed: these can often taken to be clitics and therefore never form prosodic words on their own (see Hualde 2007, 2009 for stress on function words in Spanish). Japanese then does not have word stress under this definition quite independently of its acoustic correlates, and in the analysis of our own data we will see that Huari Quechua shares some, but not all of the properties that lead to this conclusion. That does not mean that Japanese does not have a metrical structure, it just means that the position of its lexical pitch accents does not necessarily have something to do with it. This is borne out by the fact that accented words are not in any sense more prominent in Japanese than unaccented ones. High vowels bearing the H tone of the lexical pitch accent are often reduced (Venditti et al. 2008: 480–481), which would be unexpected in a stress system where the stressed syllable is the most prominent at the word level. The Japanese lexical pitch accent is simply a lexical specification that has the additional property of being culminative at the word level, similar to the distribution of aspirated and glottalized plosives in Cuzco Quechua (cf. (6), see also Hyman 2006: 238–239). Yet Japanese, and other languages that do not have word stress, clearly do have prominence above the prosodic word, so that one word is more prominent than the others in a phrase (cf. Venditti et al. 2008 on Japanese, Roettger 2017: 135 on Tashlhiyt Berber). One theoretical solution to account for this would be to assume that the relevant prosodic domain, the prosodic word, is not headed (Gussenhoven 2018: 398 on Ambonese Malay), which effectively means that metrical structure is not assigned at its level. Another option could be to assume that it is assigned a metrical structure, but this finds no, or only very little, expression in the language under discussion. The last option might be less abstract than it sounds: Hyman (2014: 57–58, 61) develops properties and functions that stress systems may have and shows that typologically, languages do not just cluster around having either all or none of these, but seem to occupy many of the available positions in between.

(10) Properties of a highly stress-oriented phonological system (Hyman 2014: 57–58)
   a. Stress location is not reducible to simple first or last syllable (which could simply represent a boundary phenomenon).
   b. Stressed syllables show positional prominence effects:
      i. Consonant, vowel, and tone contrasts are greater on stressed syllables.
      ii. Segments are strengthened in stressed syllables (e.g. Cs become aspirated or geminated, Vs become lengthened, diphthongized)
   c. Unstressed syllables show positional non-prominence effects:
      i. Consonant, vowel, and tone contrasts are fewer on unstressed syllables.
      ii. Segments are weakened in unstressed syllables (e.g. Cs become lenited, Vs become reduced).

d. Stress shows cyclic effects (including non-echo secondary stresses).
e. Stress shows rhythmic effects lexically/post-lexically (cf. the English 'rhythm rule').
f. Lexical stresses interact at the post-lexical level, e.g. compounding/phrasal stress.
g. Lexical stress provides the designated terminal elements for the assignment of intonational tones ('pitch-accents').
h. Other arguments that every syllable is in a metrical constituent which can be globally referenced.

The properties in (10) all apply in the case of English, but as Hyman (2014) points out, other languages "care" far less about word stress than English. Spanish could be said to be less interested in word stress at least with regards to properties (10) b, c, and probably e. Thus it is also quite conceivable that a language might rank very low on these dimensions, with a stress system that serves only very few of these functions and leaves barely any trace, so to speak. Some languages have also been argued to have a stratified lexicon, with one set of lexical items exhibiting one configuration of properties, and a second set another (Hyman 2006: 228). We will return to a discussion about languages without stress in section 3.5.3. How the position of word stress is determined in different languages is subject to a vast research body (cf. van der Hulst 1999b; Goedemans et al. 2010 for surveys). Some stress systems seem very simple in their regularity, e.g. stress always occurs on the initial syllable of a word in Hungarian (if its expression is not blocked by postlexical processes, Siptár & Törkenczy 2007: 21–23). Other languages require much more complex metrical algorithms that can differ at least in whether they are quantity-sensitive (sensitive to syllable weight, e.g. as expressed in moras), whether feet in it are left- or right-headed (trochaic or iambic), whether stress is oriented towards the left or the right edge of the word, and whether they systematically ignore constituents at one of the edges (extrametricality, cf. Hayes 1995: especially 54–61, 71–74). Together with additional morphophonological constraints e.g. on how certain classes of affixes interact with stress, this can then lead to systems that at first sight seem quite opaque, such as that of English, but they also allow for the exploitation of stress contrasts for minimal pairs (e.g. English *óbject* (noun) vs. *objéct* (verb), German *Ténor* "thrust of an idea" vs. *Tenór* "tenor voice", Spanish *páso* "step / pass.1SG" vs. *pasó* "pass.3SG.INDEF"), which is of course impossible when stress is always assigned to the same position in a word. There are also proposals that the primary word stress in many languages is simply specified in the lexicon, but interacting with the metrical structure (cf. van der Hulst 1999a: 73–75). In Spanish, word stress always falls on either of the three final syllables in a content word, unless it is followed by two clitic pronouns (e.g. *cuéntamelo* "tell it to me"), but within this three-syllable window, the difference between stress on the antepenult (proparoxytonic), on the penult (paroxytonic) or on the ultima/final syllable (oxytonic) can encode lexical and grammatical contrasts, as seen above. Among these three options, paroxytonic words are by far the most common overall (almost 75%), oxytonic words are only more common in the subgroup of words ending in a consonant other than /s/, and proparoxytonic words make up just a little less than 6% of a corpus of the 4289 most frequent Spanish words, according to Eddington (2000: 96). Stress assignment in Spanish has been analyzed algorithmically (Harris 1983, 1987; Roca 1997, 1999, 2006; Lipski 1997, among many others) as well as via an exemplar-based lexical approach (Eddington 2000).

Stress assignment can be subject to considerable variation within a language (cf. e.g Behnstedt & Woidich 1985a, 1985b on substantial regular differences in regional varieties of Egyptian Arabic). Lexical items with low use frequency seem often to show more variability in their stress placement, which would indicate that stress position is at least partly lexicalized, but the relationship between the two seems to be quite indirect in many cases, at least in English (cf. Tokar 2017: 19, 21–22). Since stress derivation is not the focus of this study, we will simply treat stress position as a given property of prosodic words in Spanish, part of the knowledge of speakers, and also mostly ignore the level of feet in representations henceforth. For now, we resolve to call the culminative, obligatory highest prominence assigned by metrical structure at the level of the prosodic word "word stress" or "lexical stress", without necessarily implying the concomitant bundle of phonological and phonetic properties attached to stress in many European languages. "Stress accent", on the other hand, will be used when necessary to distinguish it from other kinds of lexical accent like the Japanese one. For the treatment of intonation, word stress is crucial in many languages: the position specified by it provides the other anchoring site, besides constituent edges, for tones in AM. Those tones assigned to positions designated as prominent by the metrical structure are called pitch accents. Before returning to the discussion about how intonational contours are related to prosodic landmarks crosslinguistically and in AM, we will look at what is known about metrical structure and word stress in Quechua.

3.3.3 What is (not) known about stress in Quechua

For Quechua, the facts about word stress are far less clear than in many European languages and making efforts at resolving them is one of the contributions of the present work. In overviews based almost entirely on impressionistic data, word stress is said to regularly fall on the penult, or on the initial syllable, in some of the Quechua I varieties (Cerrón-Palomino 1987: 128–129; Adelaar & Muysken 2004: 207–

208; Wetzels & Meira 2010: 352–353). Focusing on those central Peruvian Quechua I varieties closest to Huari Quechua, for the Quechua varieties spoken in the department of Ancash, Parker (1976: 57–60) describes a complex, partly weight-sensitive system: broadly speaking, stress (which he equates with *intensidad* "intensity") regularly falls on the penult, as in other varieties. In exclamatives, it seems to be final, but he tentatively proposes that this is due to a particular exclamative intonation (cf. also Cusihuamán 2001: 79–81, who similary describes the difference between declaratives and exlamatives in terms of intonation for Cuzco Quechua). In particular for the varieties spoken in the Callejón de Huaylas, the following details are given: in phrase-final words, a non-final heavy syllable ((C)VC or (C)V:) receives stress; or the initial syllable, if no heavy syllable is present. In words in a non-final position in the phrase, the initial syllable is stressed, or, in slow speech, one of the heavy syllables, if present, preferentially the prefinal one (Parker 1976: 58–59). Parker (1976: 59–60) also explicitly remarks on the inadequacy of applying notions of stress from Spanish to the Quechua he describes, because he observes a disjunction of phonetics cues: in a word consisting of three light ((C)V) syllables in final position in the phrase, the highest intensity is heard initially, but the highest pitch on the penult. A particularly exceptional case is also described: syllable structure does not allow closed syllables with a long vowel ((C)V:C) in this variety; when they occur morphologically, the vowel is shortened. For some exceptional lexical items, they do however surface word-finally, and Parker (1976: 56) states that they then also cause irregular oxytonic stress. For Conchucos Quechua, Stewart (1984) paints a picture similarly full of complex exceptions and optionality. In general, she also describes word stress to fall on the penult. Unlike Parker, she treats only syllables containing a long vowel as heavy, at least in all words that do not consist of three syllables (Stewart 1984: 195). Quantity-sensitivity acts variably, and differently on words of different length: in words of more than three syllables, a final heavy syllable following a light penult can but need not attract stress, but in bisyllabic ones, it normally does not (Stewart 1984: 199, 201–203). No explanation for this optionality is offered. For trisyllabic words, the situation becomes more complicated. Only in them, closed syllables ((C)VC) also count as heavy. They attract stress in a positional hierarchy: if the penult is heavy, it is stressed. If it is not, then the initial syllable is stressed if heavy. If that is also not the case, then the final syllable may receive stress if heavy (Stewart 1984: 205–206). In all words, when both final syllables are heavy, then the penult is stressed. Words where all three syllables are light, such as *yaku-ta* (water-OBJ) "water (obj.)" or *hacha-man* (tree-DIR) "towards the tree", are variably stressed either on the initial syllable or the penult (Stewart 1984: 198–199). This last case sounds similar to what Parker describes for such trisyllabic words (see above), but unlike Parker, Stewart does not differentiate between intensity and pitch. It is imaginable that some of the unexplained optionality, based presumably on what she as analyzing linguist perceives as prominent, might be resolved when considering phonetic cues individually.

Unlike the previous studies, Hintz (2006) uses instrumental acoustic data in her study on stress in South Conchucos Quechua. She does not find evidence for weight-sensitivity, even though the varieties studied by her and Stewart (1984) overlap.[24] Based on data from four speakers, she finds that stress is word-initial in spontaneous connected speech, with a secondary stress on the penult. In isolated words and under conditions of "emphasis", the situation reverses, and primary stress is now located on the penult, with secondary stress on the initial syllable. Her perceptions are partially confirmed by judgments by three of the same speakers, who were asked to identify prominent positions on a subset of the recorded words. Both in her own judgment and that of the native speakers, variation occurred in general, but in particular in questions and exclamatives, where stress was found to have moved to the final syllable of the word (Hintz 2006: 488). Identical word forms containing the suffix *–shun* (1.PL.INCL.FUT), such as *aywakushun* "we will go/let's go" were found to have different stress patterns depending on whether the speaker intended to express agreement or a request (Hintz 2006: 490). Vowels in a voiceless environment are occasionally observed to be voiceless, with a large majority of them (>80%) occurring word-finally (Hintz 2006: 498, cf. Delforge 2011 for a similar phenomenon in Cuzco Quechua). To explain part of the remaining variation in stress perception, Hintz (2006: 499–500) proposes to view such devoiced syllables as "optionally extrametrical", i.e. optionally not entering into the stress assignment algorithm, giving a number of suffixes where she has found this to be the case. Interestingly, Stewart (1984: 209) also describes extrametricality, but directly relates it to certain morphemes, which do not enter into the stress assignment algorithm, again explicitly optionally. The list of candidate syllables men-

24 "South Conchucos Quechua" is a label for Quechua I varieties spoken in the Southern part of the Conchucos valley, including Huari province and town, by about 250.000 speakers according to Hintz (2006: 478) Her speakers are from Huaripampa, a small community outside of the town of San Marcos, about an hour's drive by car away from the town of Huari. "Conchucos Quechua", the label used by Stewart (1984), seems to be more comprehensive, but she does not indicate where precisely her data is from. In Stewart (1987: 5–8), she explains however that "Conchucos Quechua" is spoken in an area bounded to the north by the town of Pomabamba, and to the south by that of San Marcos, by about 200.000 speakers, which is largely coterminous to the area described by Hintz (2006) to fall under the label "South Conchucos Quechua". A large part of the data in Stewart (1987) comes from a community close to Pomabamba, which is a car drive of 3–5 hours away from the town of Huari. If the description in Stewart (1987) also applies to Stewart (1984), then her data is thus from the northern edge of the region in which "South Conchucos Quechua" is spoken, Hintz' data is from its southern edge, and the data on which the present study is based is from a more centrally southern part of it. It is not known how much prosodic variation exists within this area.

tioned by the two authors mostly does not overlap, only –*ta* (OBJ) and –*qa* (TOP) are mentioned by both. Primary, secondary and unstressed syllables were found to be significantly different according to the cues of F0, intensity, and duration in Hintz' analysis. However, only for one speaker was there actually a difference between all three conditions, via pitch height; both pitch height and intensity were different in the data of all speakers between stressed and unstressed syllables; duration was the least reliable cue (Hintz 2006: 505–506). This last observation is somewhat expected: as vowel length is phonemic in Quechua I varieties, it is perhaps less likely to be used additionally as a cue to stress.

As already pointed out by Parker (1976) and also by Hintz (2006), many of these observations would benefit from a treatment that separates phonetic cues, intonational phenomena, and word stress. In our own Huari Quechua data, we found a situation where our non-native perceptions were very heterogeneous. In multiple recordings of the same lexical items, impressions of highest prominence did not agree between recordings and hearers; in many words, prominence could be heard on different syllables depending on which cue we focused on. We therefore decided to refrain from relying too much on our own perceptions which are apparently biased from exposure to languages with other prosodic systems. In a study based on the speech of 2 speakers, Buchholz & Reich (2018) investigated whether acoustic cues (pitch height, pitch range, duration and intensity), both individually and taken together, served to make syllable positions stand out from their environment. We did not find consistent evidence across words of different length that either the penult or the initial syllable in the word was cued to stand out relative to the others, except for an indication that the penult seemed to have slightly higher intensity values overall (Buchholz & Reich 2018: 147, 155). CVC-syllables in the word penult also were found to be somewhat lengthened in this position, but not CV-syllables in the same position (Buchholz & Reich 2018: 153–155). On the other hand, we found that pitch height formed a distinctive pattern on phrases, operationalized as material between pauses, such that in general a gradual rise from the beginning of the phrase was observed, until a more abrupt fall taking place across the last two syllables of the phrase (Buchholz & Reich 2018: 151–153).

There are several indications that what has been described as word stress might partially consist of postlexical intonational phenomena. First, from a functional perspective it seems inefficient to have a system of stress assignment that is as complicated as described above yet does not fulfill a distinctive function: all descriptions agree that stress in Quechua is not lexically distinctive. Secondly, Parker (1976) and Cusihuamán (2001) state that the "stress shift" in exclamatives is likely an intonational phenomenon; the same argument could also be made for the "stress shifts" due to pragmatic conditions observed by Hintz (2006), namely that these are possibly due to different pitch accent and boundary tone configurations at a phrasal level. Boundary tones adjacent to a phrase edge might also better explain why voiceless syllables do not seem prominent, because such syllables are insufficient landing sites for tones. Thirdly, Stewart (1984: 207–209) gives evidence that the domain of stress assignment is not isomorphic with the morphological word (cf. (11)).

(11) Conchucos Quechua data points that necessitate a domain of stress assignment larger or smaller than the morphological word, according to Stewart (1984: 208)

a. áchikày mikùkuskínàa
achikay miku-ku-ski-naa
Name eat-MID-ITER-PST
"The Achikay [wicked old woman] ate it all up"

b. hákàakunáta
hakaa-kuna-ta
guinea.pig-PL-OBJ
"the guinea pigs (obj.)"

c. kèedanantsíkpaqkáqta
keeda-na-ntsik-paq ka-q-ta
stay-NMLZ-1.INCL-BEN COP-AG-OBJ
"that which is to stay for us"

d. taríntsikpístsu
tari-ntsik-pis-tsu
find-1.INCL-ADD-NEG
"we don't find it at all"

All the examples in (11), where the acute accent (´) marks primary and the grave accent (`) secondary stress, are attested but incompatible, according to Stewart (1984: 208), with her stress algorithm unless the domain of stress assignment is not the individual (morphological) word. In (11)a, the words *achikay*, the name of a wicked woman from folklore, and *mikukuskinaa* "s/he ate", are both full content words, yet if they were assigned stress independently according to Stewart's system, then the initial syllable of *mikukuskinaa* would have to have primary stress. In contrast, (11)b-d lead Stewart (1984: 208) to conclude that the morphemes *–kuna* (noun plural), *ka-* (copula),[25] and "frequently" *–pis* (additive meaning, "also" or "even")

25 Stewart (1984) treats *kaq* as a "definitivizer" and apparently as a dependent morpheme. As indicated in the glosses, I treat it as a combination of the copula verb *ka-* and the agentive nominalization *–q*, and as also morphologically independent because it can occur on its own (which

form their own domains of stress assignment. If we take a postlexical perspective, both types of cases can be seen as indicative of correlates of phrasing or postlexical pitch accenting rather than of lexical stress.

Finally, formulations such as that "stress can be distributed over several long syllables" (Adelaar & Muysken 2004: 207) also indicate that what has been described as stress in much of the impressionistically based literature on Quechua is probably to be understood as a broad label collecting a number of suprasegmental phenomena, rather than culminative word stress as defined in the previous section.[26] The results in Buchholz & Reich (2018) about an identifiable pitch trend (different from declination) across a phrase-like unit can also be seen as indication that at least pitch movement is sensitive less to stress as a word-level phenomenon, but to a unit above the prosodic word.

Hyman (2006: 246–247) cautions against interpreting any prosodic differences observed in an unfamiliar language via the lens of stress accent familiar to the analyzing linguist from European languages. Citing the example of Indonesian, he proposes the heuristic that "if word-stress is so hard to find, perhaps it is not there at all". Indonesian has received a number of conflicting stress-based accounts, yet van Zanten et al. (2003); van Zanten & Goedemans (2009) showed that what these accounts had taken to be correlates of word stress seems most likely to be pitch accenting independent of word stress, instead seeking proximity to a phrasal edge. Gordon (2014: 111–112) even estimates that perhaps a majority of word stress-based accounts especially of lesser-studied languages actually reflect such systems of phrasal, not lexical, prominence realized via pitch events that seek to occur close to phrasal edges. In this vein, as an alternative hypothesis to the complex and optionality-heavy stress-based accounts reviewed here, I will consider the possibility that Huari Quechua has no word stress, or at least only "cares" very little about it, following the characterization by Hyman (2014). In chapter 6, I will present data in evidence for this hypothesis and also provide an analysis of the intonational phenomena of Huari Quechua that only marginally needs to make reference to a lexical stress position.

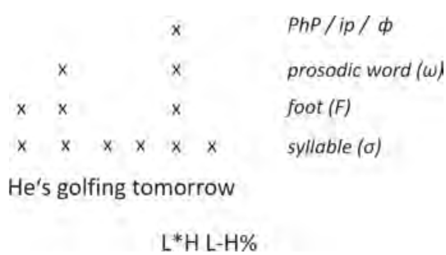
–*kuna* and –*pis* cannot). For an in-depth treatment of the functions of *ka-q* and its derived forms, see Bendezú Araujo (2021).

26 See also Wetzels & Meira (2010: 314–315), who state that the definitions for phenomena like stress or pitch accent in descriptions of suprasegmental phenomena in South American indigenous languages in general are often vague. They make it clear that much more research is needed before reliable generalizations can be made.

3.3.4 Pitch accents and arriving at phonological tones from pitch contours

Returning to the discussion of how to relate the pitch contour and the text to the tones along the example of the rise-fall-rise contour, we can now see that the position of the first rise in (8)a-c is clearly linked to the stressed syllable in the words *Bob*, *heir*, and *golfing*, respectively. Specifically, it is linked to the stressed syllable of that word which is most prominent in the entire IP. In some languages, among them English and German, it is only the most prominent word in a phrase whose stressed syllable has to be pitch accented. Because of this link to a prominent position, such pitch accents are sometimes called “prominence-lending” (e.g. Welby 2006: 364). They have also sometimes been confused with a direct correlate of stress, but this should be kept apart: syllables that are not pitch accented but stressed are still often longer and louder than unstressed syllables. In languages with stress, a stressed syllable is a necessary condition for prominence-lending pitch accents to occur, but not a sufficient one. In other languages, this does not have to be the case. Continuing with our example, we can see that it is the strongest position in the phrase that receives the pitch accent because in *he’s golfing tomorrow*, it could also occur on another word than where it does in (8)c, e.g. on *tomorrow* (see (12)).

(12) Metrical structure and rise-fall-rise contour on *he’s golfing tomorrow* with highest prominence on *tomorrow*



That would, however, indicate another context:[27] e.g. one in which the parliamentary debate is still taking place the next day, but the speaker has just asked when the

27 Note that in principle, the same thing could be done with 8b, e.g. moving the highest prominence and also the pitch accent to the stressed syllable on *throne*. However, that would imply a context in which long-lost heirs to different things are contrasted with each other. Stereotypically, long-lost heirs are often those to thrones (“throne” is by far the most frequent collocate to “heir to” in a four-word window to the right of “heir” in the >13 billion-word web corpus English Web 2015 as searched via Sketch Engine (Jakubíček et al. 2013; Kilgarriff et al. 2004)), so such a context is simply not highly likely from our knowledge of the world.

president will be going golfing next and has been told it will be tomorrow. In this context, (8)c would clearly be odder than (12). However, the other way around, (12) in the context given for (8)c would arguably be more acceptable, because there is a general preference for locating the highest prominence rightmost (cf. Ladd 2008: 252). This hints at the complex influence that metrical structure, pitch accenting and context have on the interpretative categories of focus and contrast, which will be treated in more detail in section 3.7. Note that the "incredulity" aspect of the meaning conveyed (and made plausible by the context) does not change, just the location at which alternatives to (8)c and (12) would have to differ from them, respectively, in order to be less incompatible with the speaker's expectations. That it can change, however, is evidence that it does not suffice to simply say that "the position of the first rise is linked to the stressed syllable" in a word, as we just did above. Instead, the linking relationship has to be established between the individual tones making up the rise, the L and the H. If the rise is timed so that the pitch trough preceding it (caused by the presence of the first L tone) extends throughout the stressed syllable of the most prominent word, then the "incredulity" reading obtains, while if most of the rise, sometimes including its peak, takes place on the stressed syllable, then a different reading obtains which is also contrastive but without the additional meaning of "incredulity", cf. (a) and b) in Figure 4.

That this difference in relative timing between text and tune creates a difference in perception that for the majority of speakers seems to be close to categorical[28] was first established by Pierrehumbert & Steele (1989) in a categorical perception and imitation task. Providing contexts that make one or the other realization more felicitous (Pierrehumbert & Steele 1989: 185) and thus demonstrating that

28 Note that of their five test subjects, one did not reproduce any difference in contours between the two conditions on average. Pierrehumbert & Steele (1989: 190) readily ascribe this to the L*H pitch accent not occurring in the speaker's tonal inventory. It's since been amply demonstrated that intonation can vary considerably also between experimental tasks, speaking styles and various social categories, even within speakers from a locally relatively restricted area, and also show effects of interlocutor accomodation (cf. Face 2003; Henriksen 2013; Romera & Elordieta 2013; Huttenlauch et al. 2016; Huttenlauch et al. 2018; Martín Butragueño & Mendoza 2018; Vanrell & Fernández Soriano 2018, to cite only a few recent works on Spanish). Furthermore, categorical perception experiments have been less successful in other instances (see also Gerrits & Schouten 2004 for a critical assessment of the paradigm). Overall the conclusion seems to solidify that for intonation, the relation between form and meaning is often distributional and probabilistic, rather than strictly categorical: realizations of two meaning categories often show a bimodal yet also clearly overlapping distribution (meaning that statistically, differences with clear trends do emerge, but two randomly chosen instances from each of the categories are relatively likely to be similar or even to counteract the trend); they can also differ substantially in their internal homogeneity, all of which seems to firmly locate variability at the heart of what intonation is (cf. Ladd 2008: 150–154, 2014; Cangemi & Grice 2016; Roettger 2017).

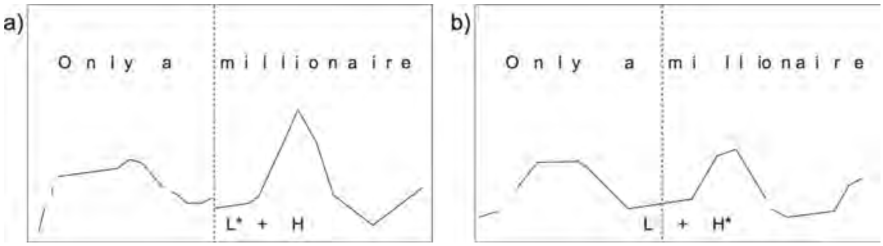


Figure 4: Two pitch track contours of *only a millionaire*, with the "incredulity" reading (a, L*H L- H%) and with a contrastive reading without "incredulity" (b, LH* L- H%), adapted[29] from Pierrehumbert & Steele (1989: 182).

the contours in Figure 4 can be intonational minimal pairs, they propose to encode their difference phonologically via specifying which of the two tones in the first rise is linked to the stressed syllable in *millionaire*. This link between the TBU of the stressed syllable and one of the tones of the pitch accent is called association, and it is marked in the phonological annotation by the asterisk after that tone (see Figure 4). Note that even in the contour in Figure 4 b), where the association holds between the stressed syllable and the H tone (LH*), the pitch peak only occurs quite late in [mil]. Association of a starred tone with a stressed syllable does not necessarily mean that the pitch peak always occurs within that syllable, but that the timing of its tonal event can be reliably defined relative to its position (Arvaniti et al. 2000, 2006; Ladd 2006, 2008). How the other types of tones in the contour are related to the segmental string will be treated in section 3.5. It is a reasonably good initial heuristic in AM-based intonation analysis to take pitch turning points (peaks, i.e. local maxima, and shoulders, points which form the left or right edges of high pitch plateaux, for high tones; troughs, local minima, and elbows, points which form the left or right edges of low stretches of pitch, for low tones) as indication for tonal targets. Yet there are many factors that impede this simple equation (cf. Ladd 2008: 134–138; Barnes et al. 2010). For one, it is well known that voiceless obstruents tend to locally raise F0 while voiced obstruents will lower it (so-called *consonantal microprosody* even though its effects can be quite large, cf. Ohala 1978; van Santen & Hirschberg 1994; Hanson 2009; Kirby & Ladd 2016; Ladd & Schmid 2018). These have to be discounted when trying to deduce the presence of tones from the pitch signal. Then, as we have seen, quite subtle but inarguable pragmatic contrasts can affect pitch in also quite subtle but perceptible ways. Phonological context can also

29 It is problematic to obtain open access reproduction rights for pitch tracks originally published under more restrictive copyright schemes. I therefore use schematized adaptations throughout in these cases.

play an important role, e.g. when the proximity of other tones or prosodic boundaries causes tonal targets to undershoot (*truncation*, which can vary between languages and prosodic positions, cf. Rathcke 2013, 2016; Cho & Flemming 2015). Thus, inferring tones from the pitch signal can only be done with some certainty once the relevant landmarks in the prosodic structure and the tonal inventory of the language are known. Both the relevant landmarks as well as the tonal inventory can vary considerably between languages.

3.4 Intonation systems typologically

In the following, we will explore how languages vary in their intonation systems. The goal is to arrive at a variation space mapped out by what is known about prosodic typology. This will help us avoid a treatment of Huari Quechua and Huari Spanish only in terms of what is known about other Spanish varieties or other European languages.

3.4.1 Domains of tone assignment

The very broad typological division of languages into "tone" and "intonation" languages[30] is partially based on the domain of tone assignment. Tone languages assign tones at the level of the syllable or the prosodic word, but more crucially, different combinations of segmental strings together with tonal contours can encode distinctive lexical meanings in them. In intonation languages, with which we are here mostly concerned, tones never carry lexical meaning; instead they can convey meanings that are orthogonal to the lexical meaning. In this sense, they are always *postlexical* in these languages. However, while this distinction is useful in itself, it would be misleading to characterize languages as only belonging to either category. See e.g. the studies in Downing & Rialland (2017) on how intonation interacts with tone in African tone languages. Languages that are not tonal according to this

30 This division is used here only descriptively. The prosodies of many languages pose problems for this dichotomic typology, in particular the so-called "lexical pitch accent languages" among which Japanese, Basque and Turkish have sometimes been counted; but also many African tone languages, or languages like Egyptian Arabic or West Greenlandic, which assign tones to a unit that looks like a word but without conveying lexical meaning. Hyman (2014) proposes a prosodic typology that does not "pigeon-hole" languages, but instead allows to class them along several dimensions via a number of criteria that do not necessarily have to cluster. See also Hyman (2009, 2014, 2018); Beckman & Venditti (2010) for extensive discussion.

definition (and their descriptions) are also sometimes described as assigning tones also at the level of the prosodic word (Egyptian Arabic according to Hellmuth 2005, 2007; West Greenlandic according to Arnhold 2014). It is not entirely clear, however, whether the unit identified as the prosodic word in these descriptions is really distinct from a small phrasal domain such as the AP or PhP.[31] The tonal movements resulting from the placing of (combinations of) tones at prosodic constituent edges help to identify these constituents. However, this identification is not just such a straightforward matter. In the following section we will briefly consider how pitch cues can signal prosodic constituency beyond a delimitative and culminative placement of phonological tones.

3.4.2 Tonal scaling and other gradual cues for prosodic constituency

Besides marking the edges of prosodic units with tones, thus delimiting these units and making the domain they belong to identifiable, languages also use further tonal cues for the delimitation of prosodic units. They are related to tonal scaling, the height of a tonal event in relation to another or a reference point. Each human speaker has a typical vocal range for speech, the interval on the acoustic spectrum on which they are most comfortable speaking and at which they have the greatest control over pitch modulations, differing according to gender and individual. Within this range, pitch register, the global height of the pitch of an utterance, can still differ according to social context or emotional state of the speaker (Ladd 2008). Pitch registers can differ either in pitch level, if both the lowest and the highest points of a contour are lowered or heightened together, or in pitch span, if the lowest and highest points are moved closer towards or further away from each other. Within a given register, speakers have a lower and an upper baseline, in relation to which individual tonal events are scaled; this local scaling is called pitch excursion. The baselines normally fall gradually over the course of an utterance, this is called declination and likely due to decreasing subglottal pressure over an exhalation (Pompino-Marschall 2009: 246–247). It has been shown for some languages (Pierrehumbert 1980; Liberman & Pierrehumbert 1984 for American English, Prieto et al. 1996 for Mexican Spanish) that at the very end of

31 In particular, Arnhold (2014: 221–222) argues for the prosodic word as the domain of tone assignment mainly on the grounds of it being identical to a morphological or syntactic word. However, since individual words can also be realized without tones (Arnhold 2014: 222), the tones could also be analysed as belonging to a phrasal category that is usually, but not always, isomorphic with a morphosyntactic word (which can also be quite long in West Greenlandic). Solving this argument will most likely always depend on one's theoretical stakes.

some utterances, pitch events take place at a lower local level than projected from the declining lower baseline; this has been called *final lowering* and is one of the ways in which pitch scaling can be used to cue the edges of prosodic units. Pitch reset, the return of the pitch level to the initial baseline, is the second. Pitch reset often occurs between larger chunks of speech and is proposed by Himmelmann et al. (2018: 239–240) to be a universal cue, together with pauses, for what they call the "phonetic IP", a universally present speech unit which serves as the basis for the language-specific development of "phonological IPs". The third is systematic manipulation of the pitch level according to prosodic constituency. We will return to this last issue in section 3.6. The suggestion by Himmelmann et al. (2018) that there is both a "phonetic" and a "phonological" IP in some languages is targeted at the open question whether the different types of prosodic cues, such as segmental processes, phonetic cues such as initial strengthening and final lengthening, tonal cues based on the location of phonological tones, and tonal cues based on pitch scaling and pauses, actually align in signaling the *same* prosodic units. This is an ongoing discussion. It is evident that the consequent assignment of segmental phonological processes to prosodic domains has led to a proliferation of them. Prosodic domains also seem to differ in the way they manifest: e.g., unlike other prosodic domains, there is quite robust articulatory evidence for the syllable even though it is perhaps not the domain of phonological processes in all languages (Hyman 2015). Pauses, sometimes seen as the quintessential signal that a part of speech has ended, have been shown to also occur in hesitations or as indicators of increased cognitive load, instead of just at the end of larger prosodic groups and not even always there (Cruttenden 1997: 30–32; Frota et al. 2007; Seifart et al. 2018). Presumably, such points will all have to be considered in order to eventually resolve the issue of how many prosodic hierarchies there are and how they are related. However, since solving this problem is far beyond the scope of this work, I will here follow Frota (2012: 257–258) in her assessment that there is good evidence for a convergence at least between the segmental processes and the tonal cues to prosodic structure, as well as pauses, at least on average. I will take this as an optimistic methodological heuristic, undergoing constant re-evaluation. It is also quite evident that languages differ in the specific weight they assign these various cues. In the following section we will leave these murky waters and consider the differing types of tonal inventories across languages.

3.4.3 Tonal inventories

Research based on the AM model, by now the dominant approach, has produced intonational analyses of a considerable number of languages and varieties: see Jun

(2005c, 2014a) and the works therein for a typologically broad overview, Frota & Prieto (2015) and the works therein for an overview over intonational variation in Romance languages, and Prieto & Roseano (2009–2013, 2010) as well as the works therein and in particular Sosa (1999, 2003); Beckman et al. (2002); Estebas Vilaplana & Prieto (2008); Hualde & Prieto (2015) on intonation in Spanish varieties. Jun (2005a) classes languages into whether they assign tones to prominent positions and edges of prosodic domains, or to edges only, and to my knowledge it still holds that no language has been described as not assigning tones to the edge(s) of at least one prosodic domain. However, languages differ considerably with regards to the paradigmatic tonal inventory that they employ according to available descriptions. AM only assumes two underlying tones, a high (H) and a low (L) one, but for many languages it has been argued that bi- or even tritonal (e.g. Gabriel et al. 2010 for Argentinian and García 2016 for Peruvian Amazonian Spanish) complexes are assigned to a single position under certain conditions. As an aid for facilitating comparison between data, a transcription system has been developed based on the AM description of American English by Pierrehumbert (1980), called Tones and Break Indices (ToBI, Beckman & Hirschberg 1994; Beckman et al. 2005). ToBI conventions for a wide range of languages have been developed that all use a related set of symbols. An asterisk following a tone (e.g. H*, L*, a "starred" tone) signifies that the tone is a pitch accent and associated with a tone-bearing unit (TBU) in a prominent position (a stressed or accented syllable). TBUs are usually taken to be either the rhymes of syllables, or moras, depending on the language (cf. Gussenhoven 2004, 2018). Languages sometimes make more fine-grained differentiations than that, e.g. only allowing consonants of sufficient sonority as TBUs (cf. the situation in Tashlhiyt Berber according to Roettger 2017). In bi-or tritonal pitch accents the tones are either connected with a plus sign (+) or simply written together (the convention followed here), with the tone associated with the prominent position marked by the asterisk (e.g. LH*/L+H*, L*H/L*+H, the tones not followed by * are also called "unstarred"). Unstarred tones preceding the starred tone in a more than monotonal pitch accent are called leading tones, those following are called trailing tones. Boundary tones of the smaller phrasal level (ip or AP) are marked by a minus sign (e.g. H-, L-), those of the higher phrasal level (IP) with a percent sign (e.g. H%, L%). Placing the minus or the percent sign after the tone symbol indicates a boundary tone at the right edge, placing it before it one at the left edge of a phrase (e.g. -H, %H vs. H-, H%). The specific status of starred and unstarred tones in pitch accents as well as boundary tones and their phonetics and phonology will be discussed below, in section 3.5. There are further symbol conventions used in some language-specific ToBI schemes, the most frequent of which are ! for downstep and ¡ for upstep (e.g. !H, ¡H), intended to mark a tone whose excursion is markedly lower (downstepped) or higher (upstepped) relative to its surroundings.

SpToBI, the ToBI system developed for Spanish (Beckman et al. 2002; Face & Prieto 2007; Estebas Vilaplana & Prieto 2008; Aguilar et al. 2009; Hualde & Prieto 2015) also specifically marks delayed rising peaks, i.e. pitch accents occurring in prenuclear or prefinal position in which the peak is realized after the stressed syllable but the H tone is taken to associated with it, using the "smaller than" symbol between the L and the H tone (L<H* / L+<H*). Whether tonal scaling (down/upstep) is best encoded paradigmatically at the level of individual tones in all cases and the independent status of the delayed rising pitch accent in Spanish are both somewhat contested issues. The former will be discussed in section 3.6, while we discuss the latter in the next section.

3.4.4 Delayed peaks in Spanish and the notion of the nuclear accent

Regarding the issue of rising pitch accents with a delayed (posttonic) peak, it is uncontroversial that in many peninsular varieties of Spanish, prenuclear/prefinal rises are generally[32] delayed in the way described, but in nuclear / final position in the utterance, they are not (see Hualde 2002: 103 for a list, since then further enlarged, of works concurring with this finding). Other varieties, e.g. Cuzco Spanish, do not show this behaviour as strongly (O'Rourke 2005). In those that do, however, a pitch accent that is not delayed on a prefinal word causes the word to be interpreted as being narrowly focused; equally, oxytonic words (stressed on the final syllable) do not usually delay the peak into the following word (cf. Hualde 2002: 103–105; Face & Prieto 2007). A rather neat unifying explanation is proposed by Hualde (2002: 106), based on findings by Nibert (1999) that intermediate phrase boundaries follow the non-delayed accents in prefinal position: the delayed realization of the peak due to the H tone of the pitch accent is blocked by the presence of the boundary tone realized on the last syllable of the word, whether the word is in ip-final or IP-final position. A concomitant assumption is that focused constituents are always directly followed by the right edge of an ip or IP. This is the view espoused by Gabriel (2007), based on further empirical evidence and formulated in optimality-theoretic (OT) terms.[33] We will also follow this view here and

32 There are some curious exceptions: Prieto & Torreira (2007: 475) state that the prenuclear peak can regularly occur late in the accented syllable, instead of in the postaccentual one, on a first accent in a phrase "when the first accent belongs to an utterance-initial phrase which contains two accents".

33 Even though he proposes that ip- or IP-boundaries follow focused constituents, Gabriel (2007: 191) formulates doubts that it is really the presence of the ip boundary tone which causes the peak to be realized earlier, based on the observation that proparoxytones in narrow focus also realize can then also define nuclear accents, with Ladd (2008: 134), as the final and only obligatory accent in the intermediate phrase. We already saw that in English or German, the nuclear accent, the pitch accent assigned to the strongest prominence in the phrase, is often the only one. This is mostly not the case in Spanish, where words in prenuclear position are also regularly pitch accented (Hualde & Prieto 2015: 358). However, using the reasoning and evidence given above, we will take the nuclear accent in the Spanish case to be not only the rightmost one (which it is also in English, all following ones being deaccented, Pierrehumbert 1980: 37), but also on the final word in an ip.[34] That is to say, an ip-boundary is inserted after the nuclear accent in Spanish if it is not already final in the IP. The special status of the nuclear accent is also corroborated by the fact that in many languages, fewer pitch accents are available in prenuclear position than in nuclear position (cf. Ladd 2008: 286).[35] This is reflected in the Spanish ToBI system, which only recognizes two different prenuclear pitch accents (L*H and L<H* (which we here take to be a variant of LH*)) in prenuclear position, but five in nuclear position (cf. Hualde & Prieto 2015). We will return to the issue of focus and nuclear accent further below (section 3.7.3).

the peak within the stressed syllable. However, Prieto et al. (1995: 438–439) find that at least for one of their two Mexican Spanish speakers, the presence of a following phrase boundary clearly reduces peak delay also on proparoxytones, showing that tonal crowding effects can persist also at distances larger than single syllables. Their other speaker does not show this effect, highlighting the importance of considering individual variation in these discussions.

34 This is not the same as saying that *focus* is always rightmost in a *sentence* or in an *utterance* in Spanish, the categorical claim by Zubizarreta (1998) which has since then been amply shown to be unsubstantiated (Face 2001; Hualde 2002; Gabriel 2007; Muntendam 2010; Hoot 2012, 2016; Uth 2014; Vanrell & Fernández Soriano 2018; Dufter & Gabriel 2016 and the works cited therein). The nuclear accent is here defined prosodically as a statement about pitch accent location and metrical structure in an intermediate phrase. In the conception espoused here, this only relates probabilistically to the interpretative category of focus (Calhoun 2010b). The optimality-theoretic constraints that relate focus to prominence and phrasing are also formulated as violable in Gabriel (2007: 235, 278).

35 As an example for a case where this does not hold as strongly, cf. Gussenhoven (2005: 121) who states that standard Dutch has a large choice of prenuclear pitch accents. However, in a sequence of several prenuclear accents within an IP, they tend to be of the same type, thus still largely conforming to the assertion by Ladd (2008: 286) that the type of prenuclear accent within a single “tune is – or may be – a single linguistic choice”.

3.4.5 Differing degrees of combinatorial freedom in the tonal make-up of prosodic constituents

Intonational descriptions of languages also differ with regards to the (number of) phrasal domains at which tones are assigned and whether edge tones occur at the right or the left edge of a prosodic constituent. For example, German as analysed in Grice et al. (2005) has (monotonal) tones assigned to the right edges of intermediate phrases (iP; L-, H-, !H-) and to the left (%H) as well as right edges of intonational phrases (IP; L%, H%, ¡H%, not all of which are attested in free combination with the ip tones), as well as six monotonal and bitonal pitch accents (L*, H*, LH*, L*H, HL*, H!H*) available for assignment at prominent positions. Unangan (Eastern Aleut), according to Taff (1999), makes do with a single monotonal pitch accent (H*), three monotonal (H-, L-, ¡H-) and one bitonal (LH-) ip-boundary tones and two monotonal (L% and H%) IP-boundary tones; all of the boundary tones occur at the right edge of their respective phrases. Tokyo Japanese assigns two tones (%LH-) at the left and one tone (L%) at the right edge of the Accentual Phrase (AP), a single bitonal pitch accent (H*L) on lexically accented syllables and provides a choice between four (H%, LH%, HL%, HLH%) boundary tone combinations (mostly) at the right edge of the IP, according to Venditti (2005); Venditti et al. (2008); Igarashi (2015).

"Peninsular" Spanish, in the description by Hualde & Prieto (2015), has seven monotonal and bitonal pitch accents (L*H, L<H* (these two only prenuclear), L*, H*, LH*, HL*, L¡H*), three monotonal boundary tones at the right edge of the ip (L-, H-, !H-), and six monotonal and bitonal boundary tones at the right edge of the IP (L%, H%, !H%, LH%, L!H%, HL%). In the only available descriptions of the (declarative) intonation of a Quechua variety in AM terms, O'Rourke (2009) tentatively proposes that Cuzco Quechua assigns an –L tone at the left edge and optionally a H- or L- tone at the right edge of the ip, and has an inventory of two pitch accents, L*H, and LH*, the first of which occurs phrase-medially and the second phrase-finally. O'Rourke (2005: 182–201), discussing interrogative intonation, does not add a further pitch accent. IP-finally, L% is attested most often, also in questions, but H% does occasionally occur. As stated before, the intonation of both Huari Quechua and Spanish is as yet unexplored. Thus the range of variation in tonal inventories is quite large. A further relevant aspect pertains to the combinatorial freedom of these tones in the different languages. While in Spanish and German, there are several options to choose from at each point in the prosodic structure that a tone could be assigned to, in Japanese, an AP will always begin with a rise, either because of the AP-initial %LH- tones or because of a combination of AP-initial L plus the high tone of the lexical pitch accent H*L if the first or second syllable in the AP is lexically accented, in which case the AP-initial H- is superseded by the lexical H (Venditti 2005). In addition, the only pitch accent is H*L. That is to say, there is less combina-

torial freedom in the Japanese system compared to the Spanish or German one (cf. Igarashi 2015: 559–560). The Japanese case is comparable to other languages like French (Jun & Fougeron 2000, 2002) or Korean (Jun 1993, 2005b) that have also been described as having an AP with quite fixed initial and final boundary tones, and to the Cuzco Quechua case in the description by O'Rourke (2009). Quite obviously, less optionality at a given structural point means fewer possibilities for exploiting these options to encode meaning differences, but on the other hand, it means that the unit that is delimited by these tones (here, the AP) is made more easily recognizable, because its edges are signaled by a less variable set of cues. Paradigmatic variability at a given point in the prosodic structure therefore stands in a tradeoff relationship with the delimitation of larger prosodic units.

From the point of view of information transmission, this is a tradeoff between having many paradigmatically variable cues at each point, each signaling some differential information, on the one hand, and having cues converge on only a few units that are signaled so that listeners may adjust their expectations to these units, and also recognize when deviations from this convergent pattern are exceptionally exploited for the coding of meaning, on the other. Redundant coding is part of what constitutes recognizable structure in languages, and it is especially necessary in the acoustic signal for human communication to be robust under noisy conditions (Shannon 1948). Paradigmatically variable prosodic cueing thus is capable of encoding more information at each structural point, but it runs a greater risk of information loss. By having a greater combinatorial freedom for tones, such cueing systems can also encode information syntagmatically at the level of the tones, but we will also see that it is less easy to encode information syntagmatically via the sequence of the larger units that are delimited by these tones because the make-up of the tonal cues that signal their edges is more varied. Paradigmatically fixed cueing, on the other hand, while being less free at each individual structural point, can invest more resources in ensuring lossless transmission and it is freer to signal information syntagmatically at the level of the units delimited by the tones, e.g. by varying the size of the few units that are robustly cued in relation to the morphosyntactic units that are contained in them (*phrasing*).

3.4.6 "Phrasing" and "Accenting" languages: An appropriate typology?

In fact, Korean, Japanese and French have occasionally been called "phrasing" languages (cf. Igarashi 2015, also called "edge-prominence languages", in the typology developed by Jun 2005d, 2014b), because they optimize the (accentual) phrase in this way and use phrasing to a large degree for the encoding of information structure. Languages like English, German, or Spanish, on the other hand, have been called “accenting” languages (“head-prominence languages” in Jun 2005d, 2014b), because information structure and other pragmatic meanings are mainly encoded via the placing of tonal accents and choice of tones. Venditti et al. (1996) show that indeed, very similar information structural effects are conveyed via (de-)phrasing in Japanese and Korean, on the one hand, and by (de-)accenting in English, on the other, which supports the point that these are divergent coding strategies but with similar purposes (cf. Ladd 2008: 279–280). While deaccentuation in “accenting” languages means that pitch accents are not realized on constituents where they would otherwise be expected, often following a focus, dephrasing in Japanese or Korean signifies that AP-tones are not realized following a focus, so that the last AP in an IP is the one beginning with the focused constituent. This is also called “prosodic subordination” in Venditti et al. (2008), which is an apt term also for the phenomenon observed in the “accenting” languages. They also note that in Japanese, “full” dephrasing, i.e. the total deletion of AP tones and lexical pitch accents does not always take place and that often instead, tonal movements on focused constituents are scaled higher, followed by a substantial reduction in the local excursion of subsequent movements, with the difference to “full” dephrasing being gradual (Venditti et al. 2008: 484–485).

The same is probably true for “accenting” languages, with a gradual continuum between “deaccentuation” and “compression” (Kügler & Féry 2017; Vanrell & Fernández Soriano 2018). In general, the dichotomic typology between “phrasing” and “accenting” languages is probably misleading, because on the one hand, it depends to a certain degree on the descriptive approach chosen: French e.g. has also been described in a more “accenting” framework (Post 2000, 2002), and under closer inspection, individual languages seem to nearly always occupy an intermediate position between the two theoretical endpoints (cf. Igarashi 2015: 561). It has also been argued that phrasing always plays a role for the encoding of information structural categories in nearly all languages, and that only in some, accenting is an additional optional means for it (Féry 2013), but Kügler & Calhoun (2020) point out that there are counterexamples to the universal validity of this claim. Furthermore, the dimensions of freedom of combinatoriality and “phrasing”-“accenting” must clearly be kept apart: Egyptian Arabic realizes pitch accents on the stressed syllables of prosodic words, and is thus an “accenting” language. However, unlike in German or Spanish, there is no space for accent choice: the only available pitch accent is LH*. Additionally, this pitch accent usually occurs on each prosodic word, no matter its information structural status (Hellmuth 2005, 2007; Chahal & Hellmuth 2014). That means that it also has few options to encode pragmatic meaning via accent choice (cf. Igarashi 2015: 561–562), and it tonally optimizes a recurring prosodic unit, like the “phrasing” languages. Madrid Spanish shares some, but not all of these characteristics: it is usually described, unlike German or English, as also realizing a pitch accent on each prosodic word, at least in prenuclear position (Hualde & Prieto 2015: 358), while postfocally, deaccentuation or compression does often take place (Hualde 2002; Face & Prieto 2007; Gabriel 2007; Torreira et al. 2014; Vanrell & Fernández Soriano 2018). In addition, as mentioned above, there are only two pitch accents available for words in prenuclear position, compared to the choice between five for the nuclear position (the last accent in a phrase, cf. Hualde & Prieto 2015). Thus Spanish seems to use tone distribution both to optimize words to a certain degree (by the recurrent realization of rising accents prenuclearly), while the placement of the nuclear accent in rightmost position in a phrase also serves as a delimiting cue for that phrasal category (cf. Kügler & Calhoun 2020 for a similar view), and it also maintains quite a large bit of combinatorial freedom through its tonal inventory. We can see some of the typological dimensions discussed exemplified in the utterances from Tokyo Japanese, Egyptian Arabic, Lima Spanish and German (Figure 5, Figure 6, Figure 7, Figure 8).

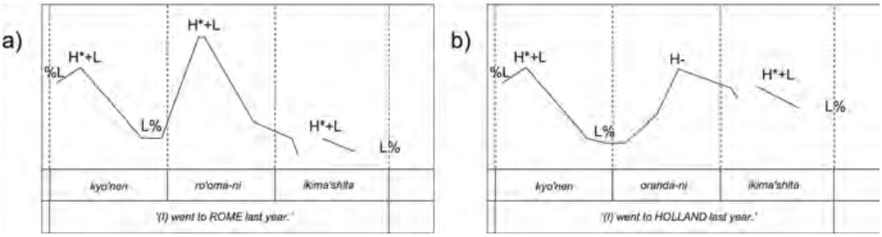


Figure 5: Two declarative Tokyo Japanese utterances with dephrasing, *kyo'nen ro'oma-ni/oranda-ni ikima'shita* "(I) went to Rome/Holland last year", adapted from Venditti et al. (2008: 485). The accent (') marks the location at which the fall from the lexical pitch accent takes place. Both utterances cue focus on the location, which in a) bears a lexical pitch accent (*ro'oma* "Rome") but in b) doesn't (*oranda* "Holland"). Note that the AP-initial H- does not get expressed separately if the mora bearing the pitch accent is initial or peninitial in the AP.

They all involve instances of "prosodic subordination" – dephrasing or deaccentuation – in which the highest prominence in the IP occurs on a prefinal constituent and this is cued, to differing degrees, by increased pitch scaling on that constituent, and tonal compression up to deletion on following ones. This asymmetry in scaling effects an interpretation of focus on the most prominent constituent, in accordance with the elicitation contexts in which the other elements were given but the element corresponding to the most prominent one in the utterances was asked for.[36] Of all

36 Cf. Venditti et al. (2008: 484–486) for elicitation circumstances of the Japanese examples, Féry & Kügler (2008: 683–684) for those of the German one and Hellmuth (2006: 270–273) for those of

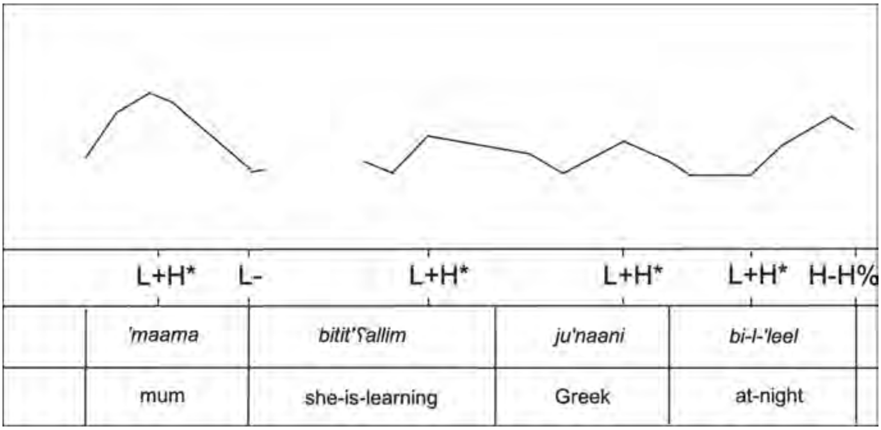


Figure 6: A declarative realization of *maama bititʕallim junaani bi-l-leel* "mum is learning Greek in the evening" by an Egyptian Arabic speaker, adapted from Chahal & Hellmuth (2014: 399). The accent (') is placed before the stressed syllable. The context is intended to yield a contrastive interpretation on *maama* "mum", with *ju'naani* "Greek" given, yet there is no deaccentuation, only some compression.

the examples, it is obvious that the Egyptian Arabic (Figure 6) one shows the least effect of this context, with the contrastively focused *maama* clearly receiving the highest scaled pitch accent, but the following given ones still robustly accented with the only available LH* pitch accent. We can also see a gradient, rather than a categorical difference between the Spanish and German examples: before the focused constituent *dem Rammler* in the German example (Figure 8), tonal movement is compressed but the constituent *der Hammel* is still pitch accented (the degree of pitch compression/scaling reduction here differs depending on whether the prefocal element is given or not, Féry & Kügler 2008), while following it, the verb is completely deaccented. In the second Spanish example (b) in Figure 7), the two prefocal pitch accents on *cuatro* and *policias* are clearly not compressed, both in comparison to the prefocal constituent in the German example, and to the postfocal pitch accents in the first Spanish example (a) in Figure 7). Those are compressed, but at least the first two, on *policias* and *arrestaron*, are still undoubtedly present and not deaccented. Full deaccentuation most likely takes place on the final word *sospechosos* in the second Spanish utterance. Note also the presence of the low

the Egyptian Arabic one. The Spanish examples are spoken by Raúl Bendezú Araujo and recorded by the author. They were purposely recorded to elicit differing intonation contours dependent on question context on the same sentence. This is of course not a very natural procedure, since in normal conversation, (null) pronominalization and elision of given constituents would likely result in different segmental strings between the two context conditions.

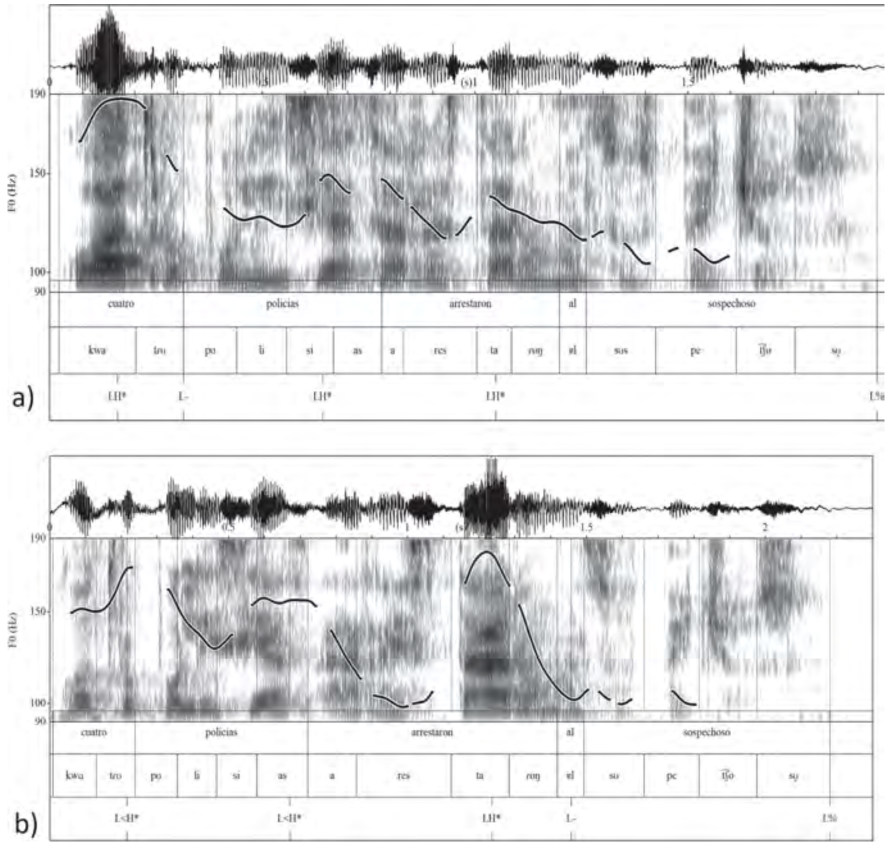


Figure 7: Two elicited declarative realizations of *cuatro policias arrestaron al sospechoso* "four police arrested the suspect" by a Spanish speaker from Lima. a) is an answer to the question "how many police arrested the suspect?" while b) answers "what did four police do to the suspect?".

ip-boundary tone (L-/Lϕ) directly following the focused constituents in both the German and Spanish examples, causing pitch to steeply drop after reaching the accent peak, and resulting in an extended final low stretch in the German and the second Spanish example. This is where we can make out the most striking difference to the Japanese examples (Figure 5). In both Japanese examples, the tonal movement on the AP bearing highest prominence is scaled high, which is comparable to what happens in the examples from the other languages. However, only in the first one, where the highest prominence is on the lexically pitch accented *ro'oma* "Rome", pitch then also drops to an overall low level, like in the German and Spanish examples – this is here caused by the increased scaling (relative lowering) on the L tone of the H*L lexical pitch accent. The lexical pitch accent on the verb

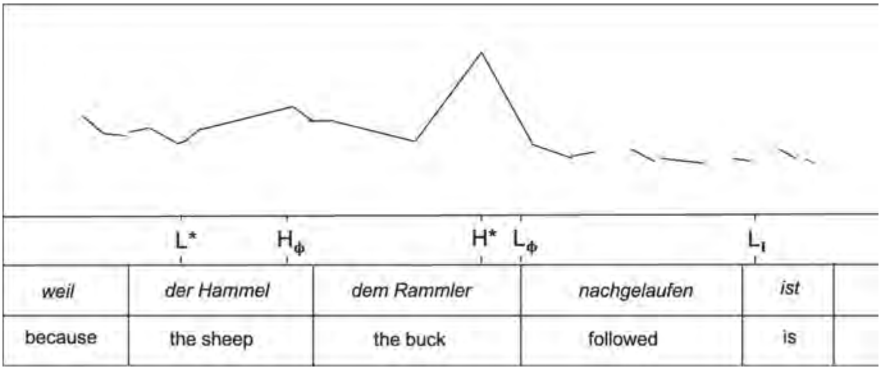


Figure 8: Declarative realization of *weil der Hammel dem Rammler nachgelaufen ist* „because the sheep ran after the buck“ by a German speaker, adapted from Féry (2017: 155). Focus is cued on *Rammler “buck”,* note the difference between prefocal compression and postfocal deaccentuation.

ikima'shita is either very reduced or deleted. In the second one, such a steep drop cannot be observed. Instead, pitch actually stays quite high after reaching the high tonal targer of the AP-initial LH- boundary tone, and only gradually sinks down to the IP-final L%, with very slight additional movement caused by the pitch accent on the verb, which is strongly compressed. Here we can see that dephrasing in this case really results in the absence of the tones belonging to the following phrase, and that the high pitch following it is not cueing prominence (cf. Venditti et al. 2008: 484–486).

Another important difference is the directedness of phrasing: in Japanese, the focused or most prominent constituent begins a new AP that then extends until the end of the ip or IP, whereas the analyses for German and Spanish place this constituent at the (right) end of an (intermediate) phrase which might have begun considerably earlier (further to the left). Note that these differences in directedness or prosodic headedness are probably again tendencies, rather than absolutes, as Beckman & Pierrehumbert (1986: 285) already point out. A further similarity could be seen in the tendency shown in Japanese, Spanish, and German to make the tonal movement on the focused constituent the last one in the IP before the final boundary tone, or at least the last one of comparable pitch scaling. Thus there are real differences in the realization of how prominence asymmetries are encoded intonationally in these four languages (Egyptian Arabic, Japanese, Spanish, and German), but it also seems that not all of them are actually categorical, and that in order to adequately analyze them, it is also necessary to pay considerable attention to gradient factors, such as pitch scaling. The description in O’Rourke (2005, 2009) suggests that Cuzco Quechua shares attributes with “accenting” languages in that it is said to realize pitch accents at prominent positions. However, it also shares some with "phrasing" languages in that it has both initial and final ip-level boundary tones, although in "phrasing" languages, the relevant domain is usually taken to be the AP. The relative scarcity of different pitch accents is also more reminiscent of "phrasing" languages. No evidence is found for dephrasing or deaccentuation. For Huari Quechua, comparable findings have not been made. My own analysis will show that prominence asymmetries are in fact encoded via deaccentuation/dephrasing, and also argue that both the Spanish and the Quechua data from Huari defies easy categorization as either "phrasing" or "accenting".

3.4.7 Restrictions on combinatorial freedom

Let us now return to a consideration of tonal inventories. Even in languages with a large inventory, combinatorial freedom seems to be actually more restricted than might be assumed. In this section, I will present evidence that argues that there are strong constraints on reducing the number of contours at a phrasal level, and that this aids both in the identification of contours and in consolidating this level as the domain at which intonational meaning is conveyed. The intonational finite-state grammar proposed by Pierrehumbert (1980) (see Figure 9b) in principle allows for any combination of pitch accent plus ip-boundary tone / phrase accent plus IP-boundary tone (in this order), as chosen from the inventory at each point, to be realized in American English. At each node in the sequence, there is a free choice between each of the options available, leading to 8 x 2 x 2 = 32 possible combinations of pitch accent plus ip and IP boundary tones for American English (leaving out the initial boundary tone), compared to 2 x 5 available combinations for Tokyo Japanese, according to the inventory in Figure 9a.

Empirical reality paints a somewhat different picture. Dainora (2001, 2002, 2006) shows, using a corpus of radio speech in American English comprising about 1200 IPs, that out of 100 possible sequences of prenuclear pitch accent plus nuclear pitch accent plus ip boundary tone plus IP boundary tone, only 44 are attested; that out of 20 nuclear sequences[37] – nuclear pitch accent plus ip and IP boundary tone – only 18 are attested and four of them (H* L- L%, H*L- H%, LH* L- H%, LH* L- L%, in order of decreasing frequency) account for nearly 80% of all attested nuclear sequences (Dainora 2006: 112–113). The identity of a pitch accent plus an

37 The pitch accents H*L and HL*, contained in Pierrehumbert's original inventory, are dropped in Dainora's investigation because they were originally introduced only to trigger downstep and are thus not taken to constitute independent categories, see Dainora (2001: 99). Downstepped H tones are also collapsed together with non-downstepped ones, because Dainora (2001: 46–69) finds that they come from the same distribution.

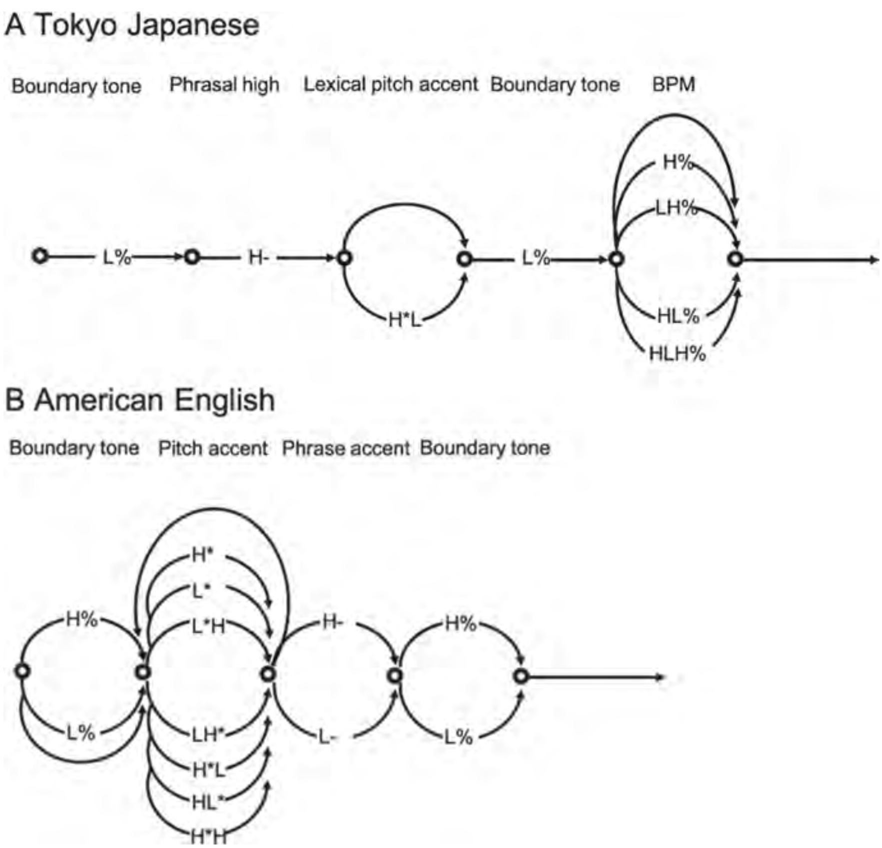


Figure 9: Comparative tonal finite-state grammars of Tokyo Japanese (A, above) and American English (B, below, originally from Pierrehumbert 1980: 29), adapted from Igarashi (2015: 560). The arrow pointing to the left in b) indicates that the pitch accent node can be repeated.

ip boundary tone is also a considerable predictor for the identity of an IP boundary tone, so much so that "in some cases the nature of the boundary tone is almost predetermined by other parts of the phrase" (Dainora 2002: 4). A further finding is that adjacent sequences of H tones are quite rare, irrespective of whether they are pitch accent or boundary tones: less than 16% of all attested contours contain such sequences (cf. table 1 in Dainora 2006: 113).[38] A similar observation, albeit

38 To bring this into perspective: even without assuming anything about what part of the nuclear contour a tone T belongs to, two tones in a sequence will have a probability of 0.5 * 0.5 = 0.25 of being HH, which is already more than the attested 16%. If we take the five pitch accents Dainora assumes as fundamental, atomic units, then 12/20 = 60% of all possible nuclear contours (pitch without data on the frequencies of individual contours, is made concerning Bengali by Hayes & Lahiri (1991). According to their analysis, Bengali has in nuclear position a pitch accent (L*, H*, LH*) followed by an optional iP-boundary tone (a PhP in their terminology, either H_P or L_P), plus an optional IP-boundary tone (H_I or L_I) and one obligatory IP-boundary tone (H_I or L_I).[39] Free combination would yield 54 possible tone sequences, but only eight of those are attested, each with a distinct intonational meaning (Hayes & Lahiri 1991: 72). Quite comparably, the summary of attested realizations in Peninsular Spanish of *bebe la limonada* "s/he is drinking the lemonade" in Hualde & Prieto (2015: 389), modestly said to list "some possible intonations" of that sentence, but based on a very comprehensive overview over the intonation research on Spanish in at least the last two decades by two of the leading experts in the field, only gives ten such combinations of prenuclear plus nuclear pitch accent and IP-boundary tone as well as an optional iP-boundary tone (see Table 3).

Table 3: Attested intonation contours of *bebe la limonada* "S/he is drinking the lemonade" in Peninsular Spanish, adapted from Hualde & Prieto (2015: 389).

	***be**be*	*la limo**na**da*	**Function**
a.	L<H*	LH* L%	*Statement or command*
b.	L<H*	L* L%	*Statement or command*
c.	LH* L-	L* L%	*Statement or command with emphasis on first word*
d.	L<H* H-	LH* L%	*Statement or command with emphasis on second word. First word is topic.*
e.	L<H*	LH* L!H%	*Statement of the obvious (see also echo-question expressing surprise)*
f.	L*H	L* H%	*Information-seeking question*
g.	L<H*	LH* HL%	*Confirmation-seeking question*
h.	L<H*	L¡H* L%	*Echo question (surprise etc.)*
i.	LH*	H* H%	*Quiz question*
j.	L<H*	HL* L%	*Insistent explanation / Insistent request*

This should be compared against the inventory, as stated above (cf. section 3.4.5), of 2 prenuclear and 5 nuclear pitch accents, six IP-boundary tones and three ip-boundary tones. The SpToBI-tradition has not followed the convention adhered to in the descriptions of American English or German that an IP-boundary tone must always be preceded by an ip-boundary tone (cf. Figure 9 and Pierrehumbert

accent + phrase accent + boundary tone) contain *at least* one sequence of an H followed by another H (some, like LH* H- H% or H*H H- H%, contain two or even three). From this expectation, the attested 16% clearly constitute a marked deviation.

39 This is the same as saying that IP boundary tones are either monotonal or bitonal, but generally obligatory, as done e.g. in analyses of Spanish.

1980; Beckman & Pierrehumbert 1986 for English, Grice et al. 2005: 68 for German). Instead, the ip boundary tone is often taken to occur somewhere before the nuclear pitch accent, but it is effectively optional (see Table 3). This means that the possible combinations must include four options for ip-boundary tones, the three tonal options (L-, H-, !H-, cf. section 3.4.5) plus the one that no tone occurs. Calculated like this, the attested combinations in Table 3 are only ten out of a possible 2 x 4 x 5 x 6 = 240 (taking the nuclear pitch accent inventory as not including L*H and L<H*, which are only attested as prenuclear from Table 3). Counting only possible nuclear contours would still yield 30 possible combinations, of which only eight are attested. It is doubtlessly the case that even with the intensive research done on the intonation of “Peninsular” Spanish, the information presented in Table 3 is far from definite. It is quite likely that with even more empirical investigations, some additional contours not listed there might turn up, and it is certainly the case that the specification of the form-meaning correspondences given could bear considerable refinement (cf. Fliessbach 2023). These considerations notwithstanding, the discrepancy between what is theoretically possible and what is attested in terms of tonal combinations seems to be equally striking for this variety of Spanish as it is for American English or Bengali.

One explanation for this discrepancy is provided in the case of Bengali by Hayes & Lahiri (1991: 72–75) who point out that nearly all unattested contours are those that include a sequence of like tones (L-L or H-H) and therefore propose that the obligatory contour principle (OCP) is effective in Bengali “at the level of the entire tune”. The OCP is a prohibition against sequences of like tones based on observations first made in Leben (1973). It is discussed and named as such in Goldsmith (1976). There and in Odden (1986), it is established that even though evidence for a principle like the OCP seems abundant in most of the African tone languages studied there, it cannot be said to hold universally and that asymmetries exist in how the OCP is applied at different levels of linguistic structure, also in terms of its directedness. The OCP has since been extended to non-tonal aspects of linguistic sound structures (McCarthy 1986; Frisch et al. 2004), and is functionally motivated in a broader context as a tendency to avoid sequences of perceptually overly similar elements (cf. Boersma 1998, especially ch. 18; Flemming 2002, 2004). Boersma (1998: 416) also describes a tendency against the repetition of similar articulatory gestures, which he takes to be part of the articulatory functional motivation for the OCP. The language-specific aspect of the OCP is also acknowledged in Hayes & Lahiri (1991: 74), and it also shows in the examples we have discussed: while in Bengali, any sequence of like tones seems to be prohibited, the data from American English seem to suggest a robust but violable tendency to avoid sequences of high tones, yet sequences of low tones are quite frequent.

For the Spanish data from Table 3, the same argument cannot be as easily made: six of the prenuclear + nuclear contours contain sequences of like tones (b, c, d, g, i, and j), of those, four contain sequences of high tones (d, g, i, and j); of the nuclear contours themselves, three (b/c, g, and i) have sequences of like tones, two of those (g and i) consisting of high tones. Can we thus say that Peninsular Spanish entirely ignores the OCP, and if yes, what else might account for the discrepancy between possible and attested tone sequences? First of all, it is important to recognize that for all three languages we are discussing here, the facts we have seen suggest that at least at some level, it is the tonal sequence as a whole, without an identification of individual tones as belonging to pitch accents or boundary tones, that is important as a domain to which constraints or processes might apply. That assumption is inherent in calling upon the OCP to explain the (relative) absence of sequences of like tones, no matter whether they are pitch accents or boundary tones, both in American English and in Bengali, for which this assumption is explicitly acknowledged in Hayes & Lahiri (1991: 74). It is also inherent in the practice, followed in the descriptions of all three languages here as well as several others[40] of assigning intonational meanings to tonal tunes consisting of a combination of at least nuclear pitch accent plus boundary tone(s), the nuclear contour. Especially for Spanish, it has even been proposed that for the identity of the tune as a whole, it is somewhat irrelevant which of the tones it is made up of form the nuclear pitch accent, as long as they stay in the same order (Torreira & Grice 2018). The domain for such a holistic tune, linked to some pragmatic meaning, is sometimes taken to be the IP (Hayes & Lahiri 1991: 52). This finds its parallels on the meaning side by statements to the effect that an IP encodes “an informational unit or sense unit” (Heusinger 2007: 280, referring to concepts developed in Halliday 1967 and Selkirk 1984, respectively), that it represents one “informational chunk” which is processed as such in production and perception, similarly across languages (Himmelmann et al. 2018: 236 and the works cited therein), or that it conveys exactly “one new idea” (Chafe 1994: 108–119). Actually, Chafe (1994: 57–58) proposes the „one new idea constraint“ to hold for what he calls „intonation units“, which he states are more similar to the intermediate phrase than the intonational phrase in the description of Beckman & Pierrehumbert (1986). Since we here assume with Ladd (2008: 134) that it is the intermediate phrase, and not the intonational phrase, which defines

40 Cf. Grice et al. 2005: 72 for German, Prieto et al. 2015: 45–46 for Catalan, Gili-Fivela et al. 2015: 191–193 for several varieties of Italian, Frota et al. 2015: 278–279 for European and Brazilian Portuguese, Arvaniti & Baltazani 2005: 95 for Greek. The practice is not applied to Standard Dutch in Gussenhoven 2005: 137, who also claims that there are no restrictions prohibiting any of the possible combinations of nuclear pitch accents plus boundary tones (however, of the 24 possible resulting combinations, he only attests to 18, stating that the others are probably rare).

the nuclear accent and the nuclear contour, Chafe's idea is still compatible. Thus, the relevance of a holistic tune whose domain is the ip/IP seems relatively well grounded. If considered from the perspective of the (nuclear) tune, something like the OCP or rather its somewhat more abstract correlate of similarity avoidance in order to create perceptual contrasts (Flemming 2008) can probably still be made use of to explain the discrepancy between possible and attested tone combinations also in Spanish.

While we cannot provide a definitive answer here, the evidence seen so far suggests that only a relatively limited number of tunes at the level of the nuclear contour can be usefully associated with meanings in a language. Evidence that unfamiliar tunes can slow down processing (Braun et al. 2011) also suggests a way that tune frequency at this level can be brought into relation with notions like markedness. Potentially, it would make sense that a small number of acoustically least eventful contours are used most frequently, in a broad range of situations covering both unmarked meanings and more marked meanings when they are retrievable from context. This opens up a space for more eventful contours, acoustically and perceptually more prominent, to signal cases in which it is particularly necessary to convey a marked meaning even against expectations built up from a conversational context. They presumably need to be both acoustically salient and of relatively low frequency, so that the mere fact of their occurrence is sufficient to captivate attention to such an extent that a change of contextual expectations might be effected. At the same time, if they are to convey quite specific meanings, they need to be sufficiently recognizable, which means they must be spaced sufficiently far apart from each other in terms of their dynamic acoustic properties. This as well as their low frequency (by itself an obstacle to identification) presumably also contributes to setting an upper limit on the number of types that are practically differentiable, each with its separate meaning. We will revisit some of these issues in section 3.7.

3.5 Relation between tones and the segmental string: association and alignment

In the previous section, we considered aspects of intonation at the level of the (nuclear) contour. In this section, we will instead look at the level of individual tones, specifically the phonology and phonetics of how they relate to the segmental material, and how this differs both crosslinguistically and between types of tones.

Since Pierrehumbert (1980), intonational tones have been classed into different categories: the tones associated with metrically strong positions are pitch accents (T*) and those located at the edges of prosodic constituents are boundary tones (T%). In the rise-fall-rise contour L*H L- H%, that accounts for the starred L and the final H%, but leaves out the trailing H of the pitch accent and the L-. The latter was originally called a “phrase accent” in Pierrehumbert (1980) and was then reanalyzed as an intermediate boundary tone in Beckman & Pierrehumbert (1986: 288). The former, unstarred tone in a bitonal pitch accent is either called a leading or trailing tone, as already stated above. Even though we have just seen that there are aspects of tonal behaviour, especially in relation to their pragmatic function, that suggest it is useful to consider tonal contours in a somewhat holistic fashion, this does not mean that all of the tones in such a contour-forming tone sequence show the same properties with respect to how they relate to the text and the prosodic structure. In this regard, the “phrase accents” and unstarred tones in multitonal pitch accent have often been seen to behave differently from the starred tones and boundary tones. For the starred tone, metrical structure provides an association site (Pierrehumbert 1980: 32–33). For the IP-boundary tone, metrical structure is irrelevant but the edge of the IP provides a “straightforward” orientation (Pierrehumbert 1980: 32). Thus these two types of tones are seen as oriented by the prosodic and metrical structure, independent of each other and of their tonal environment. The phrase accent, on the other hand, “is found near the end of the word with the nuclear stress even when this is not a metrically strong position” (Pierrehumbert 1980: 32). Intonational descriptions of Spanish do not make use of the phrase accent at all, and use the ip boundary tone in a less restricted way, as we have already seen in 3.4.7. Regarding unstarred tones, Pierrehumbert concludes after investigation of a small corpus produced specifically for this purpose and in reference to the trailing H in a L*H pitch accent that it “is located at a given time interval after the L*, regardless of the stress pattern on the material following the accented syllable” (Pierrehumbert 1980: 77).

Even without saying anything about phonetic alignment (the timing of a tonal event relative to the segmental material), in these descriptions these two are therefore different from the starred tones and boundary tones because they aren’t orientated directly with reference to the metrical or prosodic structure but only indirectly by reference to other tonal events. From this description alone it would be hard to argue that either of the phrase accent or unstarred tone can be associated to a specified TBU, as the starred tone is. At the same time, they also show “a certain amount of variation” in their placement (Pierrehumbert 1980: 32), mirroring their less directly determined phonological status in phonetic terms. We have also already seen that tonal alignment can affect the (perception) of intonational meaning, in the discussion of the L*H L- H% vs. LH* L- H% contour of American English. Meaning differences have also been related to alignment for example in

German (Kohler 1987, 2005). Thus, a relation clearly seems to exist between alignment behaviour and phonological categories of tones (which may then also relate to meaning differences). This relation has since been further refined. In the following sections, we will first take a closer look at the phonetic alignment of pitch accents taken to be associated with prominent positions, and then at how the absence of such robust alignment can be seen as contributing evidence for the absence of such positions at the word level in some languages. The variation in degree of anchoring both across languages and types of tones is highly relevant for the present work because it also represents one of the dimensions of variation between Huari Spanish and Quechua.

3.5.1 Segmental anchoring in pitch accents

On the one hand, a number of findings on the relatively solid temporal alignment of the pitch turning points of pitch accents have led to the formulation of the "segmental anchoring hypothesis", according to which they are aligned "with specifiable points in the segmental string" (Ladd 2006: 20). For Greek prenuclear rising accents, Arvaniti et al. (1998, 2000) showed that the elbow of the low tonal target consistently aligned just before (5 ms on average) the consonantal onset of the stressed syllable, while the peak of the high tonal target aligned just after (mean of 17 ms) the onset of the vowel in the posttonic syllable. That is to say, the distance between the low and the high tonal target is dependent on the length of the stressed syllable plus the length of the onset consonant in the posttonic, but not upon that of the following vowel, and both tones can be said to be stably aligned in relation to the stressed syllable and independently of each other, unlike in the case of the English L*H. This leads Arvaniti et al. (2000) to propose that both tones in this bitonal configuration are associated with the stressed syllable, and starred, i.e. L*H*.

On the basis of similar findings, in particular because both the distance between valley and peak and the distance from stressed syllable onset to peak correlate with stressed syllable duration, O'Rourke (2005: 105–108) also proposes the association of both tones (symbolized also as L*H*) with the stressed syllable for nuclear and prenuclear pitch accents in both Cuzco and Lima Spanish. However, she also finds on the one hand that for her Lima data, no syllable boundary serves as an anchor for the nuclear peaks, but since the peak always occurs within the bounds of the stressed syllable, association with it is nonetheless assumed (O'Rourke 2005: 108). On the other hand, prenuclear peaks seem also quite strongly attracted by the right boundary of the word containing the stressed syllable, more so for Lima than for Cuzco, which concurs with more prenuclear peaks in Lima being realized after the stressed syllable than in Cuzco[41] (O'Rourke 2005: 79–84, 105–106). For the first prenuclear pitch accent in a phrase in Madrid Spanish, Prieto & Torreira (2007) find that the H peak position varies dependent on the syllable structure of the stressed syllable, but again because it is in general realized within the stressed syllable, they assume association of the H tone with it. For Cuzco Quechua, O'Rourke (2009) finds that peaks in final words are realized significantly earlier than in prefinal words, with both however still occurring mostly within the tonic syllable. She therefore proposes an association of the low tone with the stressed syllable (taken to be the word penult), symbolized as L*H, for prefinal words, and of the high tone, symbolized as LH*, for final words. Another possible analysis she considers is to assign LH* throughout and to attribute the differing alignment to the presence or absence of an incoming phrasal boundary tone (O'Rourke 2009: 309, note 16).

Mücke et al. (2009) investigate peak alignment in German rising contrastive pitch accents, comparing Northern (Düsseldorf) and Southern (Vienna) varieties, prenuclear and nuclear position, syllable structure (open vs. closed syllables) and acoustic vs. articulatory measurements. The picture that emerges from the acoustic measurements is that while both dialectal background and syllable structure have measurable effects (Southern varieties and closed syllables have later peaks), these are also (as in O'Rourke's findings on Lima and Cuzco Spanish) subject to considerable speaker variation (n=2 each for both varieties), resulting in syllable structure being a significant factor only for nuclear accents overall, and dialectal background for neither. Yet the alignment difference between nuclear and prenuclear position was found to be significant and substantial for all speakers, leading Mücke et al. (2009: 336–337) to call the dialectal differences gradient and those between accent status discrete and symbolical. Articulatorily, their findings show that latencies between peaks and articulatory anchors (opening and closing gestures) are smaller than those measured acoustically between peaks and segmental anchors, but do not differ much in variability (Mücke et al. 2009: 336).

41 Note that considerable individual differences between Cuzco Spanish speakers with regard to prenuclear peak placement lead O'Rourke to group them into 4 different patterns. The first pattern (A) is essentially the same as that exhibited by Lima speakers, with all prenuclear peaks realized post-tonically, while in the last (D) almost all prenuclear peaks are realized within the stressed syllable. Patterns B &C are in between, with prenuclear late peaks more frequent on subjects than on verbs in the SVO sentences O'Rourke elicited. Pattern D was produced by the largest group (A: n=3, B: n=4, C: n=3, D: n=5). Interestingly, the grouping of speakers according to these patterns could not be correlated to their status as either Spanish monolinguals, early Quechua-Spanish bilinguals (<5 years) or late Quechua-Spanish bilinguals (having learned Spanish only after entering school), but as a whole, Cuzco speakers thus behaved clearly differently from Lima speakers (for details, see O'Rourke 2005: 79–86).

In summary, it seems that languages, speakers and individual tones in pitch accents differ with respect to the degree that they have invariable segmental anchoring behaviour. It is perhaps worth emphasizing the relatively large role of individual variation found by the studies discussed in this paragraph especially when considering that with the exception of O'Rourke (2005), none had more than 5 experimental subjects. Analysis of data from 72 speakers in total of German (n= 35) and Neapolitan (n= 17) and Pisa Italian (n= 20) has shown that this alignment variation is so far-ranging that Niebuhr et al. (2011) suggest speakers might even follow different global strategies, with one group broadly seeking to align f0 turning points with segmental landmarks, and the other seeking to realize contour shapes specific to pitch accents. Another summary finding seems to be that there is both evidence for independence between the individual tones making up bitonal pitch accents (their individual segmental anchoring) and for considering their movement as a single unit (the fact that their alignments are never totally independent of each other, cf. Ladd 2006: 27–28). Different bitonal pitch accents in different languages tend more to one or the other side of this continuum: in the English L*H, with the "fixed temporal distance" at which the H follows the L*, the two tones are less independent of each other than in the Greek L*H*. This difference recalls one that Grice (1995) has proposed to differentiate *within* English bitonal pitch accents: those with trailing tones (T*T), with the trailing tone occurring at a given distance after the starred one, are "melodic units", in which the unstarred tone is not directly associated, while those with leading tones (TT*) are "tone sequences", in which both are somewhat independently associated (Grice 1995: 216–219). Investigations into the coordination of articulatory gestures (e.g. Katsika et al. 2014; Tilsen 2016, 2019) are likely to provide more fine-grained and adequate explanations also of the effects of larger prosodic domains on peak alignment (cf. also Ladd 2006: 33–35), but this cannot be dealt with here. In sum, tones belonging to pitch accents have shown themselves to be aligned relative to a TBU assigned to a prominent position in the metrical structure, indicating association with that TBU. Where such independent alignment cannot be found, independent association is also in doubt.

3.5.2 Secondary association and tonal spreading

After having introduced segmental anchoring in 3.5.1, we can now turn to two mechanisms by which phonological tones can relate to the segmental string without independently associating, secondary association and tonal spreading. They allow to treat contours as similar that share the same number and type of turning points but differ in the way high or low stretches are extended across material, and to differentiate between points of similar tonal height according to whether pitch is actively manipulated there due to a relevant prosodic position, or simply maintained due to a specification from elsewhere. An optimality-theoretic approach to these issues is also introduced. This is crucial for the later analysis in sections 5.3 and 6.3, because it will allow for a principled generalization across superficially different contours containing such stretches in both Huari Quechua and Spanish, and to demonstrate how the different intonational variants we will encounter are relatable to each other via stepwise changes in constraint rankings.

Grice et al. (2000) study the "Eastern European Question Contour" (EEQT), a polar question contour consisting of a final rise-fall (LHL) in several related (varieties of) Eastern European languages, Standard and Transylvanian Hungarian, Standard and Cypriot Greek, and Standard and Transylvanian Romanian. They argue that in all varieties, the phonological representation is L* H- L%, i.e. the H tone is a phrase accent not associated with the nuclear accent and thus not "prominence-lending" in either of the varieties, but the position of its peak seems to vary discretely, rather than gradually, between the varieties. In a configuration in which nuclear stress is on the prefinal word in an IP, producing a low tonal target on the nuclear stressed syllable followed by a low stretch throughout the nuclear word, the H peak is aligned on the penult of the final word in Standard Hungarian (unless that is also the initial, i.e. stressed syllable, in which case the peak moves to the final syllable) and on the penult or final syllable in Cypriot Greek, independent of which syllable bears stress in the final word, and on the postnuclear stressed syllable (in the final word) in Standard Greek and Romanian. The Transylvanian varieties show the same behaviour as their standard counterparts, except that instead of the low stretch following the L*, pitch directly rises to form a plateau extending to the position at which the peak occurs in the standard varieties. In order to analyze this behaviour, Grice et al. (2000) take recourse to a mechanism originally proposed in Pierrehumbert & Beckman (1988) for Japanese, *secondary association.* Pierrehumbert & Beckman take the AP-initial H- tone in Japanese to associate *primarily* to the left edge of the AP, but *secondarily* to the second mora of the AP. For Japanese, this (secondary) association is reflected in the findings by Ishihara (2006: 72) that the AP-initial peak in unaccented words is quite stably aligned just at the beginning of the third mora, although Venditti (2005: 181) states that its alignment can also vary considerably. Returning to the EEQT, Grice et al. (2000) propose to analyze the H- phrase accent as primarily being associated with the right phrase edge, but secondarily with the position at which the peak is found. In the Transylvanian varieties, it is secondarily associated both with the nuclear stressed syllable *and* that additional position, via tone copying, not spreading (i.e., the tone does not associate with each intervening syllable; see Table 4 for a summary of the secondary association sites).

Table 4: Secondary association of H- phrase accent in the Eastern European Question Tune (EEQT), adapted from Grice et al. (2000: 158, 161).

	nuclear accent in non-final word	***nuclear accent in final word***
Standard Hungarian	penult	penult
Standard Greek	postnuclear stress	final
Cypriot Greek	penult/final	final
Standard Romanian	postnuclear stress	final
Transylvanian Romanian	nuclear syllable *and* postnuclear stress	nuclear syllable
Transylvanian Hungarian	nuclear syllable *and* penult	penult

Gussenhoven (2000a, 2000b, 2002b, 2004) develops a slightly different intonational phonology based on similar considerations. He takes phrase accents to be boundary tones, and adopts the notion of phonological alignment from prosodic morphology (cf. McCarthy & Prince 1993, 2001[1993]). Prosodic constituents (among which tones are counted) can align with the edges of other prosodic constituents. The relative ranking of optimality-theoretic phonological alignment constraints for the tones in a given input determines both their sequence and proximity to each other in the output.[42] Because his model is couched in OT-terms, constraints are formulated together with their conflicting counterparts. For alignment, this leads easily to situations in which tones are aligned in opposing directions (*multiple* alignment), and without any higher-ranking constraints intervening, the results are long pitch stretches, such as the plateaux in the Transylvanian varieties of the EEQT, in which the intervening TBUs are not associated to the tone (which would be spreading). The endpoints of such stretches do not have to have the same phonological status. Gussenhoven only allows association with TBUs, but not with edges of prosodic constituents. Association acts in two directions via two families of constraints, those that associate classes of tones with TBUs (e.g. H → TBU) and those that associate classes of TBUs with tones (e.g. σ ← T). In this way, a tone available at a position can associate with that position because it is a tone that has to associate, or because it is a position that has to associate, or it does not associate at all (if the constraints that would cause association in this case are outranked by a constraint militating against association). That means that in principle, tones in his model can be only aligned but not associated, and still be realized, if no constraint deleting unassociated tones is high-ranked. Secondary association, in Grice et al's terms, can then result out of a combination of multiple alignment, with the tone associating with a relevant TBU at both edges, or only at one of them. Additionally, an extended pitch stretch could also not be associated at either edge in Gussenhoven's model. Importantly here,

42 For a more detailed introduction into Gussenhoven's OT model of intonation, see section 5.3.

the status of a tone as associated with a TBU or "only" aligned with some prosodic constituent is taken by Gussenhoven (2018: 406–407) to be reflected in its phonetic alignment behaviour: if a tone's peak or valley alignment can be shown to be indifferent to phonological structure, this is taken as indication for a lack of association.

With this additional specification, the Gussenhoven model is in principle capable of distinguishing between more alignment-association-scenarios than the Grice et al. (2000) model: in the study on German mentioned above, Mücke et al. (2009: 335–336) identify a second low tonal target in their data as evidenced by a low elbow following the nuclear (but not the prenuclear) pitch peak at a fairly invariable distance (some 150 ms later), independent of the segmental material. They take the presence of this additional low tonal target in the nuclear condition to be the reason for the significantly earlier peak alignment of the nuclear pitch accents compared to the prenuclear ones, comparable with the analysis for postnuclear deaccentuation in German by Féry (2017) and also with the analysis of nuclear phrasing for Spanish espoused here. However, they analyse it as a phrase accent secondarily associated with the nuclear stressed syllable. This gives this low tone phonologically the same status as e.g. the Japanese AP-initial H, or the EEQT high phrase accent tone. However, the latter Grice et al. (2000) have shown to be sensitive to discretely variable aspects of structure in its dialectal variation (see Table 4), and the former is also much more sensitive to structure as evident from its temporal alignment (cf. Ishihara 2006). This difference in sensitivity to structure between the German postnuclear low tone, on the one hand, and the EEQT high phrase accent and the Japanese AP-initial H, on the other, could be seen as an argument against assigning them the same status. Indeed, it should be at least possible to keep them apart in their representation: in the Gussenhoven-style model, the German low tone found by Mücke et al. (2009) would certainly be analyzed as as an additional leftward alignment of the upcoming low IP-boundary tone without association. This can be seen from analyses of comparable low phrase accents in Dutch by van de Ven & Gussenhoven (2011), in West Germanic varieties by Peters et al. (2015), and in English according to the analysis by Gussenhoven (2018) of the findings reported on in Barnes et al. (2010).[43] Because in principle it allows for a more fine-grained representation (which does not mean that it is necessarily the more

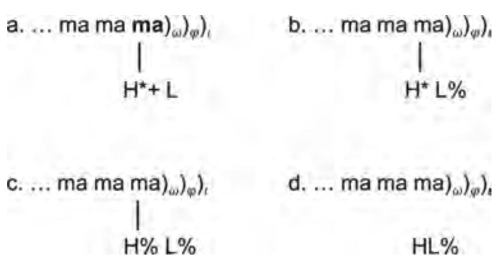
43 Barnes et al. (2010: 989) themselves take the turning point to be "epiphenomenal, the result of a constraint on the phonetic implementation of deaccenting in the postnuclear region: The fall from the level of the H* maximum to the level of the eventual L- at the phrase boundary (i.e., E2) must be accomplished sufficiently swiftly to avoid creating the percept of a pitch accent (L* or downstepped H*, for example) on a lexically stressed syllable within the postnuclear region." Arguably, the unblocked leftward alignment of the L%, resulting in a flat postnuclear stretch is precisely what would constitute deaccentuation here.

adequate representation, further studies will have to decide that), I am adopting the Gussenhoven-style model here together with its heuristic about the association status of a tone based on its alignment behaviour.

3.5.3 Representational possibilities for the association and alignment of tones in languages without word stress

We can now extend the discussion on variability in tonal anchoring to languages without word stress. Since we are pursuing the hypothesis that Huari Quechua has no word stress at all or that it is almost irrelevant in its phonological system, this section will provide criteria from alignment and association that can be brought to bear on the matter. Following the Gussenhoven-style approach, a phrase-final HL sequence can be represented in at least four different ways:

(13) Representational possibilities for a phrase-final HL sequence, adapted from Maskikit-Essed & Gussenhoven (2016: 355)



In (13)a, the starred H tone is associated with the syllable bearing word stress, and the L tone is trailing (which means that it is left-aligned with the right edge of the H). In this phrase-final configuration, it is unclear whether the model would have different phonetic alignment predictions for H*+L and H* L% here (L% being minimally right-aligned with the right edge of ϕ/ip or ι/IP). This would also depend on the constraints on tonal crowding. Rathcke (2016) demonstrates that falls in pitch accents and those due to boundary tones in German and Russian are treated differently in terms of compensation strategies under time pressure (truncation, compression, tonal re-alignment), but also that these strategies are language-specific to a certain degree.

3.5.3.1 French

(13)b, in which no word stress exists but the H tone forms a pitch accent associated with a postlexically determined position (the final syllable in the phrase), is exemplified by French, according to Maskikit-Essed & Gussenhoven (2016: 356). In (13)c, two boundary tones simply align rightmost and associate with the rightmost available TBUs. In (13)d, finally, tones are only aligned with the right edge of the phrase, but remain unassociated. Maskikit-Essed & Gussenhoven argue that (13)b is different from (13)c, the latter of which they say is exemplified by Korean AP boundary tones, because in French, whether a postlexically assigned phrase-final position is eligible for accentuation can be exceptionally determined by the morphophonological makeup of the phrase (see Figure 10 for the intonational structure of (Seoul) Korean and French): without citing sources, Maskikit-Essed & Gussenhoven (2016: 356) state that in French *que sais-je?* "what do I know?", the AP-final peak cannot be aligned on and the corresponding H tone not associate with the pronominal clitic *je*, but that this is possible for the pronominal clitic *le* in *prends-le* "take it".

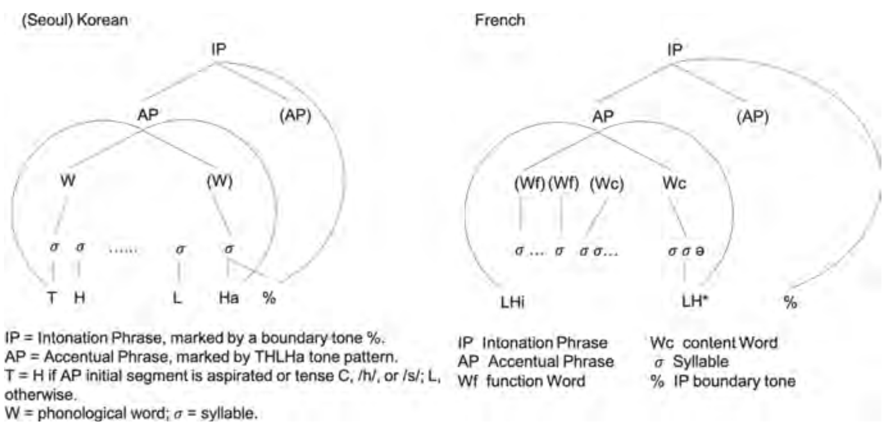


Figure 10: Intonational structure of (Seoul) Korean, left, and French, right, adapted from Jun (2007: 151), and Jun & Fougeron (2002: 152), respectively.

This is a somewhat controversial analysis. According to Welby (2006: 347), the general view is that the AP-final H will associate with the last full vowel, which excludes schwa, the vowel in both *le* and *je* (if one is realized at all). Welby & Loevenbruck (2006: 114) find that in contrast to this generalisation, the final peak can occasionally align on an AP-final schwa, but do not relate this to any lexical or morphosyntactic difference. Féry (2017: 182) distinguishes between two kinds of AP-final schwa in French: a postlexically inserted schwa, which cannot act as TBU for the final H tone, and an "underlying" schwa, which can. She explicitly gives examples including both *le* and *je* for the latter category. In spite of this, there is some agreement that the AP-final accent (LH*) in French is a pitch accent, indeed associated with what should perhaps be called the final licit TBU in the phrase, while the AP-initial accent (LHi or LH-) is a phrase accent and not associated with a

TBU (Jun & Fougeron 2000, 2002; Welby 2006; Welby & Loevenbruck 2006; D'Imperio et al. 2007). This is argued to be so because while the former is stably temporally aligned (allowing only for a kind of discrete variation in alignment with either the AP-final vowel or the penultimate one if the final one is a schwa) and the syllable on which it is realized is longer and louder than its neighbours, the latter is optional, varying in its alignment within the first three syllables of the AP and not accompanied by more loudness and longer duration (Jun & Fougeron 2002: 159).

Welby (2006) is able to further differentiate this picture based on measurements of phonetic alignment: in the initial rise, the L is stably aligned at the left edge of the first content word and thus taken to be associated both with that word's left edge and optionally secondarily with the left edge of the AP[44] (Welby 2006: 364–366). In Gussenhoven's model, association is only possible with TBUs, but not with constituents (Gussenhoven 2004: 155, 2018: 408–409). Yet the stable alignment of the initial elbow with the left edge of the first content word in Welby's data indicates association. It could be modeled with constraints that align it with both edges (with that aligning it with the content word edge ranked higher) and a constraint that stipulates that it must associate with whichever TBU is available in its location (L → TBU). The AP-final H is consistently aligned in the final syllable, accompanied by syllable lengthening and also interpreted as the starred tone of a pitch accent, albeit one that is not "prominence-lending" (Welby 2006: 364–365). Thus the two outside tones are stably aligned with segmental landmarks, even across different speech rates and taken to be associated. The two middle tones are different: the initial H has highly variable alignment not just within one syllable but extending over at least the initial two syllables of the first content word and the final L varies equally in its alignment. Both are taken not to be associated, yet "edge-seeking", i.e. showing alignment tendencies towards their respective AP-edges (in terms of the Gussenhoven model), but they are also not realized at a relatively constant distance to their initial L and the final H, respectively, varying also with speech rate (Welby 2006: 366–367). Thus, the French AP seems to consist of tones belonging to at least two of the configurations above. The position of all tones is postlexically determined, so (13)a is not appropriate. However, according to Welby's (2006) analysis, both the AP-final H* and the AP-initial L should be classed in category (13)b because they are clearly associated with a postlexically determined position whose eligibility is itself subject to (morpho-)phonological constraints. Both the middle tones, initial H and final L, should most appropriately be classed in category (13) d, since they are not associated (independent of whether they form part of bitonal

44 Because it forms a low stretch if the content word is preceded by function words that in most cases extends to the beginning of the AP.

accents with the external tones). Category c) is perhaps best occupied by the Japanese AP-initial H- in cases of unaccented words, because it is associated (according to the findings of stable alignment by Ishihara 2006), yet its position can simply be described as aligning as leftmost as possible and associating where the leftmost free TBU (mora) is available, resulting in association with the second mora of the AP, because the initial one is occupied by the L tone. For Korean, which Maskikit-Essed & Gussenhoven (2016: 355) take to exemplify this category, it has been found that at least the AP-initial peak is also sensitive in its alignment towards the morphosemantic makeup of the words phrased in it (Kim 2013: 79–84) and so the AP-initial H can therefore also be thought of as exemplifying category (13)b. Note that the difference between the French AP-final H*, which is analyzed as a pitch accent because it also affects duration and intensity, and the AP-initial L, which does not, is actually not expressible in the typology in (13). This shows that in a language arguably without word stress (cf. also Féry 2017: 180–184), accents can also differ along this dimension.

3.5.3.2 Tashlhiyt Berber

A further relevant case is that of Tashlhiyt Berber as investigated by Roettger (2017). Using a statistical assessment of durational and intensity differences between syllables as a proxy for prominence in a production study, Roettger (2017: 48–58) finds no consistent asymmetries that could be interpreted as evidence for word stress, and in particular no evidence that word stress is final as claimed before in the literature. His findings on peak alignment are highly interesting because they relate variability in peak placement with “phonetic enhancement” of tonal events, i.e. when syllables on which tones occur are also longer, louder and produced with more peripheral vowels (Roettger 2017: 47, 135, 144). In Tashlhiyt Berber, polar questions (both neutral and echo) and statements with a contrastive interpretation are realized with a very similar rise-fall (LHL) tonal movement. In questions, this is realized on the IP-final word, while in contrastive statements it is realized on the word which is contrasted against an alternative in the context. In cases where the contrasted word is phrase-final, these two positions obviously fall together. Yet the utterance modalities also differ in terms of their peak alignment in two additional ways:[45] on the one hand, quasi-discretely, such that peaks in statements are bimodally distributed with one mode towards the end of the penult and the other some

45 They also differ in terms of global pitch level over the utterance and local pitch excursion on the pitch peak, with scaling in both cases such that polar questions > echo-questions > (contrastive) statements, a difference that also proves robust in perception (Roettger 2017: 76–79, 93). This reflects a crosslinguistically frequent tendency (cf. the works cited in Roettger 2017: 65).

way into the final syllable of the word bearing the pitch peak, while in questions, peaks are unimodally distributed with a majority (but by no means all) located inside the final syllable (Roettger 2017: 80–81). On the other, they also differ gradually in that even only within the tokens that realize the peak on the final syllable, statements align the peak earlier than questions. In a perception experiment only the former, but not the latter, difference significantly affected ratings as either statement or question, and with a considerable amount of variability both between and within test subjects (Roettger 2017: 92–95). The syllable on which the peak was realized was phonetically enhanced in so far as it was louder and longer, and also sounded more “prominent”, although this was not systematically tested (Grice et al. 2015: 249, 254; Roettger 2017: 85, 135). In addition to utterance modality, peak placement is also affected by sonority differences between the syllables making up the word. Tashlhiyt has few vowels (/i a u/) but many and diverse consonants, all of which, including voiceless stops, are allowed as syllable nuclei, resulting in a large number of vowelless words in the lexicon and comparatively long vowelless strings in actual speech (Roettger 2017: 37–38). Grice et al. (2015) found in a study on bisyllabic words that peak placement was attracted to the syllable containing the more sonorous nucleus, showing a preference both for vocalic nuclei over consonantal ones as well as for more over less sonorant consonants, with liquids > nasals > voiced obstruents > voiceless obstruents (thus, e.g. /tu.gl̩/ “she hanged” and /tr̩.ks̩/ “she hid” were more likely to realize the peak on the penult, and /tn̩.za/ “it was sold” and /ʜf̩.dˤʁ̩/ “I learnt” more likely to realize it on the final syllable, cf. Grice et al. 2015: 248–251, 263). It also showed a preference for heavier (with a coda) over lighter syllables and a general preference for being rightmost; all these factors (including utterance modality) seem to be independent of each other and probabilistic rather than categorical: only when several of them came together in the same direction were counts of 100% reached for peak placement in one of the two syllables, but not by all speakers, and with within-speaker variability present even in the exact same condition on the same word (cf. Grice et al. 2015: 251; Roettger 2017: 86–88). The situation is exacerbated further when words are considered that contain no sonorants, with either voiced (e.g. /tb̩.dg̩/ “she was wet”) or unvoiced obstruents in the syllable nuclei (e.g. /tk̩.ʃf̩/ “it dried”): in those cases, the peak can either a) not be realized at all, or b) it is realized on the word preceding the non-sonorant target word, or c) it is realized on a schwa-like vowel inserted somewhere in the word (Grice et al. 2015: 254–255; Roettger 2017: 98–101). These three alternatives were found regardless of utterance modality, but the presence of an inserted schwa in the target word and the realization of the peak on the preceding word (b) never occurred together in a word consisting only of voiceless obstruents, i.e. when a schwa was inserted in such a word, it always carried the tonal peak (Roettger 2017: 100). Roettger proposes that there are two types of schwa in Tashlhiyt: one is purely the result of articulatory timing, when two constricting gestures are timed far enough apart that a transitional period occurs (“gestural underlap”) during which airflow can pass unhindered, resulting in a vowel-like sound (a “transitional vocoid”) as long as the vocal folds are vibrating (Roettger 2017: 106–108). Such an account is of course insufficient for words consisting only of voiceless segments, since no phonation will occur. For those, he proposes a true epenthetic vowel, a schwa that is postlexically inserted in voiceless environments precisely in order to act as a TBU (Roettger 2017: 125–128). Its distribution is not categorically determined, and seemingly affected by a variety of heterogeneous factors, including sociolinguistic ones, that do not correlate with peak alignment variability (Roettger 2017: 113–114, 124). The difference in status between the two schwas accounts for the fact that in voiceless words, a schwa never cooccurs with the peak realized on the preceding word: because this is the postlexical epenthetic vowel inserted to serve as TBU, whereas in words with voiced obstruents, the transitional vocoid can occur but not serve as TBU because it is invisible to the phonology (cf. Roettger 2017: 126–128).

Roettger (2017) does not offer a full phonological analysis of Tashlhiyt intonation, because only certain aspects of its system were considered. However, he takes the discrete variability in stable alignment to different syllables depending on the factors outlined and the “phonetic enhancement” of the syllables the peak is aligned with as evidence that the H tone is not just aligned, but associated with the syllable it occurs on (Roettger 2017: 144). Note that this is not the same as saying that these syllables bear word stress, which is clearly not the case (since there are realizations of the same lexical item by the same speaker under the same experimental conditions, in one of which the peak is located on the penult, while in the other it is on the final syllable, and that syllable is phonetically enhanced in both cases); thus the Tashlhiyt case is further evidence that the determination of a culminative position at the lexical level and the acoustic correlates of increased duration and intensity, usually associated with stress accent, are orthogonal to each other (we have already seen that Japanese is in a sense the inverted case). It also seems that this behaviour is conditioned by the type of word the tonal movement is realized on and/or the utterance type: in a study on wh-words (question words) in Tashlhiyt, Bruggeman et al. (2017: 20–21) found that essentially the same rise-fall (LHL) contour is realized on them as on the final word in polar questions and the contrasted word in contrastive statements, but that while the peak is always aligned within the question word, no evidence for systematic alignment with any syllable nor for phonetic enhancement of the syllable realizing the peak could be found, making an analysis in terms of association with a TBU unlikely. For both the question words in Bruggeman et al. (2017) and the tone-bearing words in Roettger’s (2017) study, secondary association to a higher prosodic constituent is also proposed: for the question words and the contrasted words in statements, this association is with the prosodic word, for polar questions, it is with the IP (Roettger 2017: 137–140). Tones are also aligned with the right edge of the domain they are associated with.

Despite the rather complex situation in Tashlhiyt and several unsolved puzzles,[46] it seems that parallels can actually be drawn to the French case. In both languages, an H tone (the AP-final one in French) seeks to occur rightmost in its domain and will associate with an available legitimate TBU. Assuming the account of schwa by Féry (2017), both languages have two different types of schwa, only one of which can serve as a TBU. While in French, that is essentially the only additional constraint able to prevent a rightmost placement of the H tone, in Tashlhiyt the segmental environment is much more adverse to tonal realization, resulting in cases in which the rightmost available TBU is actually placed before the target word. What a legitimate TBU is is thus affected by different factors in the two languages, but the generalization that the tone seems to almost occur as rightmost as possible, everything else being equal, makes a good case for the adequacy of the alignment constraints as proposed in the Gussenhoven model (cf. Grice et al. 2015: 260 for this argument, but Roettger 2017: 143–144 for a more sceptical view). The Tashlhiyt H tone in polar questions and contrastive statements would thus be classed, like the French AP-final H* pitch accent, under category (13)b, while its counterpart in wh-words would fall in category (13)d.

3.5.3.3 Ambonese Malay

Let us make a final comparison with another language without word stress, Ambonese Malay as discussed in Maskikit-Essed & Gussenhoven (2016). Using durational and peak alignment measurements on target words from a task eliciting small read dialogues, they make the case that Ambonese Malay also has no word stress. In a comparison with data from a similar task made with Dutch speakers, where stress is uncontroversial, they find that peak alignment in Ambonese Malay does not correlate strongly with any segmental position in the word, especially not one in the penult, which has been proposed to bear stress in previous studies based on impressionistic analyses; the peak effectively varies relatively freely in its placement on the last two syllables (Maskikit-Essed & Gussenhoven 2016: 364–366). This leads them to propose that the H tone is aligned rightmost, but remains unassociated, thus belonging to category (13)d. Note that even though we have already mentioned above that the Gussenhoven model does not allow for association with higher prosodic constituents, Maskikit-Essed & Gussenhoven (2016: 382) state that

46 The most theoretically challenging one of which is perhaps that the Tashlhiyt data do not in any way support a deterministic relation between form and function in intonation, an issue which is discussed in Roettger (2017: 145–148).

"words are referred to as domains within which the rising-falling pitch movement is placed", because the peak always occurs within the target word and because its distance from the word edge correlates with the duration of the final two syllables as well as of the word itself, more so than in the measurements on Dutch. If "reference to a domain" is not to be proposed as a third, as yet somewhat vaguely defined, means of tune-text linking (besides association and alignment), then this must be taken as a covert admission of association to higher-level prosodic constituents. In fact, this would not be a surprising finding if Grice et al. (2000: 148) turn out to be right in suggesting that (secondary) association to a TBU rather than a higher domain is typologically more frequent.

Intonation in Ambonese Malay seems also not to signal prominence in contrastive discourse contexts in terms of pitch scaling or any other means[47] (cf. Maskikit-Essed & Gussenhoven 2016: 372–374, this is also in line with results from prominence perception experiments on the related language Papuan Malay by Riesberg et al. 2018), and as far as their data allows them to say, the only functional differentiation intonation makes in that languages is one between declaratives, with a final low boundary tone, and non-declaratives (polar questions and continuation rises), with a final high boundary tone (Maskikit-Essed & Gussenhoven 2016: 377–380).

We have thus seen that also in languages without word stress, there is considerable variation in intonational systems in terms of association and alignment of different tones and their relation to various functions such as the signaling of prominence. It seems that the best measurable indicators for association are stable phonetic alignment with a consistently identifiable position, on the one hand, and "phonetic enhancement", on the other. However, phonetic enhancement is certainly not a necessary condition for association, as at least Japanese shows. In addition, neither is evidence for word stress: only when they indicate patterns that consistently point to a unique position in each lexical word can that connection be made. In absence of such a pattern, a postlexical pitch accent hypothesis is at least as likely, and more so when peak placement is discretely variable, as in Tashlhiyt Berber and French. Table 5 attempts to give an overview over the interrelation between association, phonological and phonetic alignment to lexically/morphologically specified positions in a word or postlexically determined ones in a larger prosodic unit, phonetic enhancement and word stress or lexical accent via tones from various languages discussed in this section. Two cells contain no data. For a type of tone to be entered there, it would need to exhibit continuously variable alignment that spans

47 The results in Maskikit-Essed & Gussenhoven (2016: 372–374) show some variation between the speakers, in that for 2 out of the 4 speakers, the difference between focus conditions was actually significant, but with very small effect sizes (0.23 semitones on total average). In general, their small study size certainly leaves room for doubt until further results corroborate their findings.

Table 5: Attested possibilities for tonal association and alignment from some intonational systems.

		phonetic enhancement	***no phonetic enhancement***
Association to a TBU – stable phonetic alignment with a segmental or syllabic landmark	lex./morph. specified position	starred tone(s) of pitch accent on a stressed syllable (Germanic and Romance, Cuzco Quechua (?))	starred tone of pitch accent on a lexical accent position (e.g. Japanese, Turkish (Levi 2005)), H phrase tone in EEQT in Standard Greek (sec. assoc. to final stressed syl according to Grice et al. 2000)
	postlex. determined position	Tashlhiyt H in rise-falls (Roettger 2017); AP-final H in French (Welby 2006)	Japanese AP-initial H (secondarily associated to a mora, according to Pierrehumbert & Beckman 1988, closely aligned according to Ishihara 2006), H phrase tone in EEQT in Standard Hungarian (sec. assoc. according to Grice et al. 2000); AP-initial L in French (Welby 2006); AP-initial H in Korean (Kim 2013); some phrase-final boundary tones (?)
Phonological alignment to a constituent edge (including another tone) – variable phonetic alignment across landmarks	lex./morph. specified position	–	some leading or trailing tones of a bitonal pitch accent on a stressed syllable (Germanic / Romance)
	postlex. determined position	–	some phrase-final boundary tones (?); H phrase tone in EEQT in Standard Hungarian according to Gussenhoven-style analysis; H tone in Ambonese Malay (Maskikit-Essed & Gussenhoven 2016), H phrase tone in EEQT in Cypriot Greek, AP-initial H in French, AP-final L in French (Jun & Fougeron 2002; Welby 2006), H tone in Tashlhiyt Berber wh-words (Bruggeman et al. 2017)

at least two syllables, and at the same time there would have to be evidence that whatever syllable the tone is realized on is phonetically enhanced. That is conceptually not impossible, but perhaps not overly likely. Spanish in general contributes to the data resumed in the table, but Quechua mostly represents a large gap in our knowledge in this respect. I have tentatively included Cuzco Quechua among those languages where phonetic enhancement occurs on the stressed and pitch accented syllable, but this has not been conclusively shown (O'Rourke 2009: 298–299).

For Huari Quechua Quechua and Spanish I will present an analysis in section 6.1.6 that would allow us to locate them on this table and weigh in on the question of word stress in Huari Quechua. Since vowel length is contrastive in this Quechua variety (unlike in Cuzco Quechua), it should be less likely that duration will also be exploited strongly to provide phonetic enhancement for pitch accented positions. This seems to be corroborated by the results in Hintz (2006: 507), but see the findings in section 6.1.6 for a qualification.

3.6 Recursive prosodic structure

In this section, we will revisit the hierarchy of prosodic constituents initially introduced in section 3.2. We will consider arguments and evidence from several languages that the prosodic units above the prosodic word can have a recursive structure. In section 5.2 on complex Huari Spanish utterances, I will argue that the data presented there is also best analyzed as representing a recursive prosodic structure. The discussion about recursive prosody and cues to it ties in both with that about tonal scaling and which role it plays in conveying phonological tones and prosodic constituents from section 3.4.2, as well as that about the role of prosody in the cueing of information structure, which we will come to in section 3.7.

3.6.1 The Strict Layer Hypothesis and prosody-syntax mapping

In the early days of the development of the prosodic hierarchy, the following principles were thought to universally hold for it, besides the hierarchy and its levels being themselves universal:

(14) Architectural principles of the prosodic hierarchy (Nespor & Vogel 2007: 7)
Principle 1. A given nonterminal unit of the prosodic hierarchy, X_i, is composed of one or more units of the immediately lower category, X_{i-1}.
Principle 2. A unit of a given level of the hierarchy is exhaustively contained in the superordinate unit of which it is a part.
Principle 3. The hierarchical structures of prosodic phonology are n-ary branching.
Principle 4. The relative prominence relation defined for sister nodes is such that one node is assigned the value strong (s) and all the other nodes are assigned the value weak (w).

Principles 1 and 2 have also been formulated by Selkirk (1984: 26–27) under the heading *Strict Layer Hypothesis* (SLH). It claims that the prosodic hierarchy is strictly ordered and non-recursive: constituents of level n_i exhaustively dominate all constituents of level n_{i-1} in their domain and are themselves exhaustively dominated by a constituent of level n_{i+1}, with the index on n being prohibited to skip a number, so that structure (15)a is well-formed, while structures (15)b-d are not (deviant elements in bold[48]):

(15) Possible and impossible prosodic structures according to the SLH
- a. $[_{\phi}[_{C}[_{\omega}[_{\Sigma}\sigma\ \sigma]_{\Sigma}\ [_{\Sigma}\sigma\ \sigma]_{\Sigma}]_{\omega}]_{C}]_{\phi}$
- b. $*[_{\omega}[_{\Sigma}\sigma\ \sigma]_{\Sigma}[_{\Sigma}\sigma\ \sigma]_{\Sigma}\ \boldsymbol{\sigma}]_{\omega}$
- c. $*\mathbf{[}_{\boldsymbol{\omega}}[_{\omega}[_{\Sigma}\sigma\ \sigma]_{\Sigma}[_{\Sigma}\sigma\ \sigma]_{\Sigma}]_{\omega}\mathbf{]}_{\boldsymbol{\omega}}$
- d. $*\mathbf{[}_{\boldsymbol{\phi}}[_{IP}[_{\phi}[_{C}[_{\omega}[_{\Sigma}\sigma\ \sigma]_{\Sigma}[_{\Sigma}\sigma\ \sigma]_{\Sigma}]_{\omega}]_{C}]_{\phi}]_{IP}\mathbf{]}_{\boldsymbol{\phi}}$

While structure (15)a conforms to all stipulations of the SLH, (15)b does not, since it contains a syllable (σ) that, while being contained inside a phonological word (ω), is not contained within its next-higher constituent level, the foot (Σ). The level of the foot has been skipped. (15)c, on the other hand, is disallowed because it is recursive: a phonological word is contained within another phonological word. (15) d, finally, is out because in it a constituent of the level of the intonational phrase (IP) is dominated by a lower-level constituent, that of a phonological phrase (ϕ). Claims for the attestation of structures like (15)b-d have all been variously made in the literature, but we will be mainly concerned with (non)-recursive structure here. Principle 3 has to be understood at least to some extent in the context of principles 1 and 2: clearly, if there was a restriction to binary branching of constituents (as there still is, for example, in Liberman & Prince 1977), it would be impossible to maintain principles 1 and 2 with the same set of prosodic levels and without questionable results, such as having to posit two different prosodic domains for *white rabbit* ((16)a) and *white fluffy rabbit* ((16)b), while the problem does not arise with n-ary branching ((16)c).

(16) Differences between only binary and n-ary branching
- a. $[_{C}[_{\omega}\text{white}]_{\omega}\ [_{\omega}\text{rabbit}]_{\omega}]_{C}$
- b. $[_{\phi}[_{C}[_{\omega}\text{white}]_{\omega}]_{C}\ [_{C}[_{\omega}\text{fluffy}]_{\omega}\ [_{\omega}\text{rabbit}]_{\omega}]_{C}]_{\phi}$
- c. $[_{\phi}[_{C}\ [_{\omega}\text{young}]_{\omega}\ [_{\omega}\text{white}]_{\omega}\ [_{\omega}\text{fluffy}]_{\omega}\ [_{\omega}\text{little}]_{\omega}\ [_{\omega}\text{rabbit}]_{\omega}]_{C}]_{\phi}$

There is another important consideration influencing the positing of principle 3, however: the flatter structures created in this way are to reflect the finding that while there is a certain degree of correspondence between morphosyntactic structure and prosodic phrasing, the constituents of the prosodic hierarchy are

48 In order to hopefully improve legibility I use double marking of brackets in this section, using the same symbols for the prosodic constituents as described in section 3.2 for subscription. Thus, $[_{\phi}XYZ]_{\phi}$ marks one phonological phrase.

decidedly not isomorphic with morphosyntactic ones (Nespor & Vogel 2007: 1–2). This is even held to be true for the phonological phrase, which is at the same time considered to be the most important interface category with the morphosyntactic derivation (Nespor & Vogel 2007: xx–xxi). In some subsequent work, this stipulation of non-isomorphism between syntax and prosody has been moved away from to a certain degree and instead, a close correspondence between the two levels has been espoused as the norm that can occasionally be deviated from. This is reflected for example in how the following optimality-theoretic constraints are formulated:

(17) MATCH-constraints (Selkirk 2011: 439)

i. Match clause

A clause in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ι, in phonological representation.

ii. Match phrase

A phrase in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ϕ, in phonological representation.

iii. Match word

A word in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ω, in phonological representation.

Here, "matching" means that both the left and right edges of the prosodic constituent coincide with those of the morphosyntactic one. The idea of close mapping has taken such hold that for instance in Féry (2017: 36), the prosodic levels above the foot are introduced with the 'corresponding' syntactic units as an indicator of their size; the prosodic word "corresponds roughly to a grammatical word", the prosodic phrase to a syntactic phrase (NP, VP, PP, AP etc.), the intonational phrase to a clause and the utterance to a paragraph. It is somewhat unfortunate for those wishing to assess this correspondence claim that 'paragraph' is a notion not regularly discussed in syntax, and that even 'clause', which is much more ubiquitous, lacks a definition relevant for prosody-syntax interactions in "current syntactic theory", as Myrberg (2013: 94) admits. Furthermore, it is clear that 'corresponds roughly' must here in fact often mean 'does not correspond at all' when considering that utterances in actual conversation often consist of syntactic fragments only a few words long; nevertheless, they are utterances and intonational phrases: it is well-known in prosody research that recordings of words pronounced in isolation should not be used to draw conclusions about word-level prominences precisely because they are then always pronounced within their own intonational phrase, cf. Jun & Fletcher (2014: 495); Féry (2017: 179). Evidence from ellipsis phenomena also underscores the point that a close mapping to morphosyntax can be only one of several competing constraints affecting prosodic phrasing.[49] This insight is captured in the nature of these constraints as violable, as usual for OT. More recent studies have also reformulated the SLH in a family of OT-constraints. Selkirk (1996: 189–190) proposes the following four constraints, where C_i is a prosodic constituent of level *i*, PWd is a prosodic (phonological) word, Ft is a foot:

(18) SLH as violable OT-constraints (Selkirk 1996: 189–190)
 (i) *LAYEREDNESS* No C_i dominates a C_j, j > i, e.g. "No σ dominates a Ft."
 (ii) *HEADEDNESS* Any C_i must dominate a C_{i-1} (except if C_i = σ), e.g. "A PWd must dominate a Ft."
 (iii) *EXHAUSTIVITY* No C_i immediately dominates a constituent C_j , j < i-1, e.g. "No PWd immediately dominates a σ."

49 Consider the following short dialogues. In i), A's monosyllabic utterance would be uttered with question intonation, i.e. IP-level boundary tones. B's answers are equally full IPs but of varying syntactic status. That B can felicitously answer with a-d in i) but only a in ii) makes an account of A's question as an elliptical version of a full question sentence doubtful because of question-answer congruence (but cf. Gretsch 2003 for the same observation and a syntactic account). The "elliptical" question by A encoded by the question intonation is clearly underspecified in a way which a "full" syntactic version can never be, either as a wh-question (Where are the keys? – Here / # Yup) or a polar question (Do you have the keys? – Yup / Here / # My pockets). In iii), it is clear that syntactic accounts can hardly explain the unfelicitousness of B's c. compared to a. A's question is here specified by the context (with a strong bias for expecting the answer to be epistemically accessible to the interlocutors but not the speaker) in a different way to how a "full" syntactic question would be specified (Who has the keys? - You [declarative intonation]). The constraints for prosodic well-formedness at the level of the IP are therefore primarily context-dependent and pragmatic, such that information which is sufficiently encoded to provide a context update given the discourse context can form an IP, but not less information (or more, according to the "exactly one idea"-proposal by Chafe 1994).

i. Context: couple leaving their flat together
 A: Keys? [question intonation]
 B: a. Here. b. Yup. c. Got them. d. My pockets [declarative intonation]
ii. Context: the couple have just closed the door behind them after leaving the flat and often forget to turn off the lights when leaving
 A: Lights? [question intonation]
 B: a. Yup. b. # Here. [declarative intonation]
iii. Context: three flatmates sharing one pair of keys leaving their flat together
 A: Keys? [question intonation]
 B. a. Me. b. You got them. c. # You. [declarative intonation] d. You [insistent intonation]

Here and elsewhere, the hash (#) before examples is used to indicate that the example is well-formed (grammatical) in the language, but infelicitous/pragmatically odd in the given context. The asterisk (*) in the same position indicates that the example is not well-formed (ungrammatical), independent of context, according to what is assumed about the grammar of the language in question.

(iv) *NONRECURSIVITY* No C_i dominates C_j, j = i, e.g. “No Ft dominates a Ft.”

For Selkirk (1996: 191), *EXHAUSTIVITY* and *NONRECURSIVITY* are violable constraints, so that level skipping as in (15)b and recursive structures as in (15)c can be allowed in some languages if other constraints are ranked higher. The first two constraints however, *LAYEREDNESS* and *HEADEDNESS*, are inviolable universally (i.e. in all natural languages) in her conception.[50] This is also maintained in Myrberg (2013), but she proposes to further complement *EXHAUSTIVITY* and *NONRECURSIVITY* with a slightly different constraint, *EQUALSISTERS*:

(19) *EQUALSISTERS* (Myrberg 2013: 75)
Sister nodes in prosodic structure are instantiations of the same prosodic category.

EQUALSISTERS is argued by Myrberg (2013: 78) to be able to explain preference patterns for prosodic structures in Stockholm Swedish which *EXHAUSTIVITY* and *NONRECURSIVITY* cannot differentiate between (the latter but not the former being violated uniformly in all of them). She argues that these structures are recursive (with an intonational phrase containing one or two intonational phrases) based on the distribution of IP-initial and IP-final as well as lexical and focal pitch accents and their downtrend behaviour (Myrberg 2013: 98–100, 108–110), i.e. on observable empirical grounds.

3.6.2 Empirical evidence for recursive prosodic structure from systematic pitch scaling

While some of the arguments for recursive prosodic structure in the literature are clearly motivated by maintaining a tight mapping between syntax and prosody and are principally theoretical, there is another line of arguments for it that is based on empirical findings and can be evaluated independently of assuming such a tight mapping. They mostly come from findings about pitch scaling, which is an area that AM is still

50 Féry (2015) discusses examples from non-extraposed German relative clauses that she calls “prosodic monsters” and which she argues to also violate *Layeredness* (i.e. structures like (15)d) by containing an IP within a phonological phrase. An important point she wishes to make is that syntax and prosody interact as equals: syntactic and prosodic constraints must be in the same ranking hierarchy in order to account for the acceptability differences between the examples she gives. However, her analysis of the prosodic phrasing is based only on a close mapping between syntactic and prosodic categories and she does not provide any measurable intonational evidence against alternative analyses with less complex prosodic structure and a looser mapping.

struggling to gain a solid theoretical grasp on (cf. section 3.4.2). We will review some of those that argue for a recursive category at the top of the prosodic hierarchy (IP) in the following. Terminologically, "downstep" will be used to label a process analyzed as phonological and resulting in a lowering of pitch level between two successive units; "declination" is the phonetically (physiologically) determined lowering of pitch over longer stretches, and "downtrend" is used to describe either of these without committing to an underlying cause. "Upstep" is the counterpart of downstep in the other direction.

3.6.2.1 Stockholm Swedish

Myrberg (2013) argues for prosodic recursion based on Stockholm Swedish data such as the utterance shown in Figure 11. Swedish has a lexical pitch accent so that content words belong to one of two accent classes, called accent I and accent II. IPs are delimited at their left edge by a so-called "initiality accent" located on the first prosodic word, L*H on an accent I word, and H*LH on an accent II word. At their right edge, they are delimited by a high or low boundary tone (H% or L%). The most prominent phonological phrase (PP here) in an IP gets assigned a "focal accent" on the rightmost prosodic word which is almost of the same form as the "initiality accent". The "initiality accent" is argued not to be the same as the "focal accent" because unlike the latter, its position is only determined by the position in the IP and not at all by information structural criteria, appearing also on backgrounded and given words. Its final H is also aligned later and more variably than that of the "focal accent" (Myrberg 2013: 85–87, cf. also Horne et al. 2001; Roll et al. 2009). In utterances like the one shown in Figure 11 from a corpus of read utterances by three speakers (n=27, cf. Myrberg 2013: 93–94), the tonal movement realizing the initiality accent is produced twice (on *andra* and *Anna*), and that of the focal accent as well (on *utklädda*, which has a secondary stress on the second syllable leading to the association of the L tone as L*, and on *med*). The focal accent pitch movements are both followed by clear low valleys indicating an L% tone at the right boundary of the IP. Thus judging from the presence of these initial and final pitch movements, it is clear that these examples consist of two IPs, with each extending over one of the main clauses making up the coordinated syntactic structure (see the lower part of Figure 11). These utterances however all (Myrberg 2013: 98) also display a steady downtrend in the relative scaling of their local pitch maxima and minima, as shown by the dashed lines in the pitch track in the upper part of Figure 11. The fact that this downtrend continues across the utterance and that no or only partial pitch reset takes place at the boundary between the two IPs is taken to be evidence for the presence of a third, larger, IP which fully contains the two smaller ones.

Pitch reset normally occurs at the boundary between IPs in Swedish (Myrberg 2013: 108–109), and is also taken as a universal cue for IPs by Himmelmann et al.

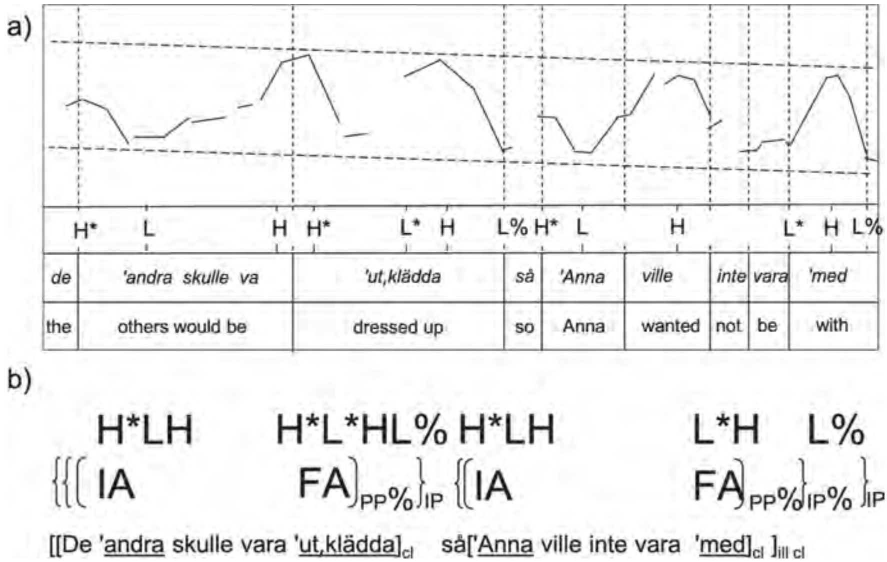


Figure 11: Pitch track of a Stockholm Swedish utterance, *de andra skulle vara utklädda så Anna ville inte vara med* "the others were getting dressed up, so Anna didn't want to join", with dashed lines indicating downtrend of the pitch maxima and minima over the time course (a), and its proposed phrasing as two coordinated minimal IPs in a maximal IP (b), adapted from Myrberg (2013: 97, 99).

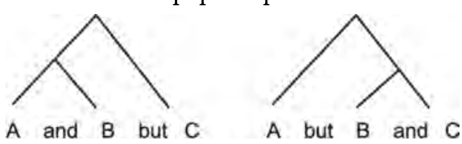
(2018). That it does not occur in the Swedish examples is thus evidence that as a whole they consist of a constituent in whose domain reset is suspended, i.e. an IP.

3.6.2.2 British English

A very similar argumentation has been brought forward by Ladd (1986, 1988, 1990, 1993, 2008: 294–297) based on evidence from somewhat more complex English data. Ladd (1988) presents read speech data from speakers of British English in two conditions, the "A and B but C" condition and the "A but B and C" condition, exemplified by (20)a and (20)b, respectively. Their hierarchical structure[51] is given in (20)c, intended to show that the clauses connected by *and* are sister nodes under a higher constituent, while the clause preceded by *but* is a sister node at the level of

51 Note that Ladd (1988: 531) argues for the hierarchical asymetry in clause status and boundary strength based on semantic and pragmatic considerations, not on syntactic ones: "The most natural interpretation of these sentences appears to be one in which the but opposes one proposition to the two conjoined by and [. . .]". The experimental materials from all the studies in this section can be interpreted in this way, suggesting that hierarchical pitch scaling at least partially serves the function of cueing discourse cohesion and coherence.

that constituent. This leads to the intuitive understanding of the boundary before *but* as stronger than the one before *and* in these constructions, which Ladd (1988: 531) hypothesizes is reflected in pitch scaling.

(20) "A but B and C" and "A and B but C" sentences from Ladd (1988: 532)
   a. Allen is a stronger campaigner, and Ryan has more popular policies, but Warren has a lot more money
   b. Ryan has a lot more money, but Warren is a stronger campaigner, and Allen has more popular policies
   c. 

When read by Ladd's test subjects, the sentences were consistently produced with several pitch accents on the prominent words and a final falling boundary movement at the end of each clause, indicating that each clause is produced as an IP (Ladd 1988: 532). In two experiments, the first one involving examples with three accented words per clause (4 speakers x 18 sentences x 2 conditions = 144 utterances), the second with four accented words per clause (same number of utterances), peak measures taken for the accents showed a downtrend across each IP corresponding to an individual clause, but with partial, not full, reset between the IPs, indicating them all to be part of a larger constituent in which full reset is suspended (Ladd 1988: 535, 539). Average peak measurements on the first accent following a *but*-boundary were higher than those on the corresponding accent in the other condition, when it followed an *and*-boundary, and this difference in boundary strength was also mostly supported by durational measurements at the boundaries (Ladd 1988: 535–536, 539). These differences were shown to be statistically significant in the second experiment for all speakers (significant interaction in ANOVA of sentence type x clause type x accent and/or sentence type x clause type), and for three in the first (Ladd 1988: 535, 539). A third experiment served as a control to ensure that it is not a local effect of *but* which is responsible for the observed scaling differences, but indeed the hierarchical structure.

3.6.2.3 Northern German

Truckenbrodt & Féry (2015) (see also Féry & Truckenbrodt 2005) essentially replicate Ladd's (1988) study with German data, using similar experimental conditions in a reading task with 5 northern German speakers as subjects.

(21) Example sentences for the three experimental conditions from Truckenbrodt & Féry (2015: 25–27), underlined syllables are expected to be accented

a. AX condition (Ladd's A but B and C condition): *A während [B und C]*$_X$
Context: Warum meint Anna, dass Handwerker teurere Autos haben als Musiker?
'Why does Anna think that craftsmen have more expensive cars than musicians?'
Sentence: Weil der Maler einen Jaguar hat, während [die Sängerin einen Lada besitzt, und der Geiger einen Wartburg fährt]
'Because the painter has a Jaguar, while [the singer owns a Lada and the violinist drives a Wartburg]'

b. XC condition (Ladd's A and B but C condition): *[A und B]*$_X$ *während C*
Context: Warum meint Anna, dass Musiker teurere Autos haben als Sportler?
'Why does Anna think that musicians have more expensive cars than sportsmen?'
Sentence: [Weil die Sängerin einen Jaguar hat, und der Geiger einen Daimler besitzt], während der Ringer einen Lada fährt
'[Because the singer has a Jaguar and the violinist owns a Daimler], while the wrestler drives a Lada'

c. No-X condition (control condition)
Context: Warum meint Anna, dass ihre Nachbarn teure Autos haben?
'Why does Anna think that her neighbours have expensive cars?'
Sentence: Weil Möller und Hummel einen Jaguar haben, Meyer und Lerner einen Daimler besitzen, und Wollmann und Lehmann einen BMW fahren
'Because Möller and Hummel have a Jaguar, Meyer and Lerner own a Daimler, and Wollmann and Lehmann drive a BMW'

As the examples in (21) show, their experimental sentences all create an argumentative contrast between two propositions one of which, X, is itself a coordination of two propositions. Each proposition is expressed through an individual clause, the two coordinated ones separated by *und* "and", and the contrasting one separated by *während* "while". In the control condition, no argumentative contrast exists. As in Ladd (1988), each clause is realized with several pitch accents (here, one prenuclear one and a nuclear one on the final accented word in the IP, both L*H) and a separate final boundary tone (H% at the end of the prefinal and L% at the end of the final IP) and thus analyzed as forming IPs (Truckenbrodt & Féry 2015: 27–29). One important difference to Ladd's experiment is that because in German, the nuclear accent in an

IP is upstepped (Féry & Kügler 2008), the scaling of this upstep[52] is also hypothesized to be affected by the proposed hierarchical prosodic structure (Truckenbrodt & Féry 2015: 24), in addition to the scaling on the IP-initial accents. Their results confirm Ladd's in that the difference in boundary strength is reflected in the scaling of the initial accent between the two conditions (see A in Figure 12), and no such effect is found in the control condition, which shows downtrend with partial reset across the three IPs, as in Swedish (Truckenbrodt & Féry 2015: 30). In order to model this effect, they use the concept of phrasal reference lines, first proposed in van den Berg et al. (1992) for Dutch. There it was observed that peak values in downstep contexts seem to orient themselves towards an abstract reference line extending across a phrasal constituent. Truckenbrodt & Féry (2015: 31–32) analyze their results as showing that a phonological process of downstep is responsible for the observed effects. Downstep is implemented as a lowering of the respective phrasal reference lines. It only applies between sister nodes in the prosodic structure (as also proposed by Ladd 1988: 541–542), lowering the second sister, and its effects manifest as less strong between higher constituents than between lower ones.[53] This is schematically represented[54] in B in Figure 12. In the AX condition, lowering takes place between the IPs A and X (=B + C) and between B and C. In the XC condition, it takes place within X, i.e. between A and B, and between X and C, but not between B and C, as they are not sisters. The actual pooled values reflect this: the initial peaks in B and C in the AX condition are both lowered with respect to the preceding initial peak, whereas in the XC condition, they are at the same level, and B is lower here than in the AX condition.

The results on the upstepped peaks also confirm the general predictions of this model but show some additional variation. In the AX condition, upstepped IP-final nuclear peaks are roughly at the same height as the preceding initial peaks (no significant differences in paired t-tests), but in the XC condition, the height of the nuclear accent in the second clause (B), labeled H4 in Figure 13, varies. For two speakers, it is significantly higher than the preceding IP-initial peak H3 and not significantly different from H1, the initial peak in the first clause, while for two others, it is the other way around, with H4 being at broadly the same height as H3

52 The nuclear accent in the experiment sentences always fell on a lexical item signifying a car name in each IP, which was contrasted with a different item of the same type in the other clauses.
53 Correlating boundary strength with amount of lowering is the proposal by Ladd (1988), reformulated via depth of embedding in Féry & Truckenbrodt (2005). Truckenbrodt & Féry (2015: 32–33) relate the difference between the initial accents in B in the two conditions to an additional stipulation of final lowering (Liberman & Pierrehumbert 1984) applying to the phrasal reference line of the last sister under a node. The different proposals do not make any differing predictions with regards to the data discussed in Truckenbrodt & Féry (2015).
54 In reality, the reference lines should presumably not be thought of as horizontal, but as always including some slowly increasing amount of decay, the effect of phonetic declination over any utterance.

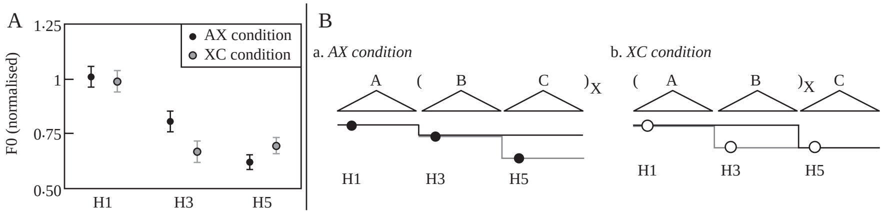


Figure 12: A: Mean values and 95% confidence intervals for the IP-initial peaks, normalized and pooled values of speakers S1–S5 for the two conditions AX (n= 67) and XC (n= 66). *H1, H3,* and *H5* are the respective initial peaks of the clauses A, B and C, from Truckenbrodt & Féry (2015: 31).
B: Schematized phrasal reference lines and IP-initial peaks for the two experimental conditions, from Truckenbrodt & Féry (2015: 32).

and lower than H1 (Truckenbrodt & Féry 2015: 35). This is reflected in the average value for H4 in the XC condition seen in B in Figure 13. The interpretation is that the height of H4 is variably orientated either towards the higher reference line of X, or the downstepped one of B in the XC condition, as shown in A of Figure 13. The fifth speaker always upstepped H4 almost to the height of H1 in both conditions, pointing once again to the need for further studies with more speakers to illuminate the role of inter-speaker variation in these phenomena.

3.6.2.4 European Portuguese

European Portuguese is a Romance language for which comparable empirical findings about prosodic boundary strength have been connected to a hypothesis about prosodic recursion (Frota 2012, 2014). This argument is based on gradual differences corresponding to boundary strengths not only with respect to pitch scaling, but also phrase-final lengthening and even the likelihood of the application of segmental prosodic processes. The IP in European Portuguese, according to Frota (2014: 11–13), is the domain of a number of processes, including the following: a) it is the domain at which at least one pitch accent is obligatory, on the strongest word in the final PhP in the IP; b) it has an obligatory boundary tone at its right edge and an optional one at its left edge; c) word-final fricatives are voiced when followed by a word-initial vowel within the IP, but voiceless at its edges; d) it is the domain of final lengthening; and e) clitic elements have a tendency to be realized in their stronger forms when positioned at the left edge of the IP. The distribution of pitch accents and boundary tones (a+b) is shown to agree with the phrasing cued by fricative voicing (c) and final lengthening (d) in (22)a and b and the corresponding pitch tracks given in Figure 14. The tonic and posttonic syllables of the words preceding an IP-boundary as marked in (22)a and b are lengthened (underlined in the examples), a pitch accent is realized on the final stressed syllable followed by a boundary

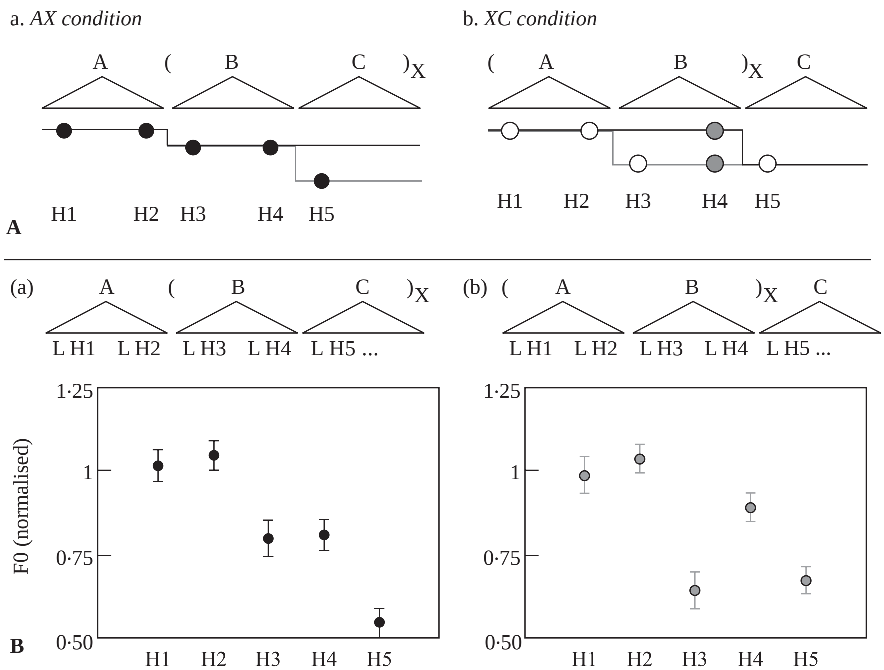


Figure 13: A: Schematized phrasal reference lines and upstepped nuclear peak values for the two experimental conditions, from Truckenbrodt & Féry (2015: 34). **B:** Mean values and *95%* confidence intervals for upstepped nuclear peaks *H2* (clause A) and *H4* (clause B) in relation to the three IP-initial peak values *H1, H3*, and *H5* in the normalized pooled data of speakers S1–S4. (a) AX condition (n=51); (b) XC condition (n=50), from Truckenbrodt & Féry (2015: 35).

tone, as evidenced by the pitch movements in Figure 14, and IP-final fricatives are realized as voiceless [ʃ], while IP-internally, they are realized as voiced [z].

The difference between (22)a and (22)b is proposed to be that while in (22)a, all IPs are of equal status, in (22)b the “short IP” (Frota (2014: 11)) *as alunas* “the students” is embedded within the larger IP extending until *até onde sabemos*, forming a recursive “compound prosodic domain” (Ladd 2008: 297–309; Frota 2012, 2014). The short IP is argued not to be an entirely different category, e.g. an intermediate phrase, because the differences to a longer IP are mostly only gradual: while fricative voicing extends throughout the larger IP domain in (22)b (visible also in the clearly different spectral quality of the segment corresponding to the fricative in B compared to A in the original pitch tracks[55] in Frota 2014: 13), final lengthening

55 Voice onset time (VOT), the phonetic exponent of voicing, is of course also a continuous parameter. It might turn out that both VOT and the place difference between [z] and [ʃ] as realized in terms

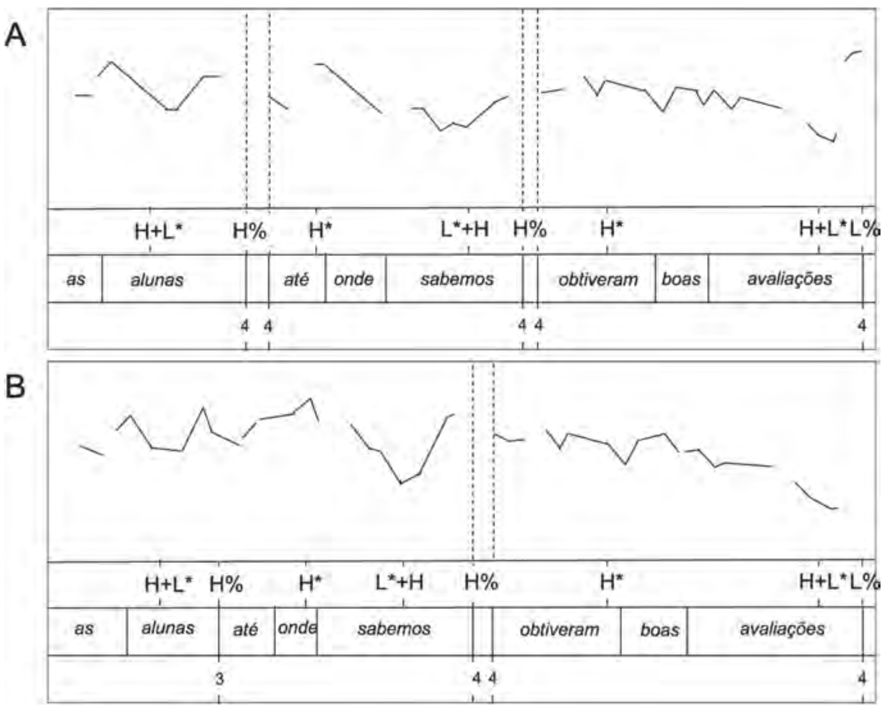


Figure 14: Two versions of an elicited European Portuguese utterance, corresponding to the proposed phrasing structures in (22)a (A) and (22)b (B). Adapted from Frota (2014: 13).

and pitch scaling are decreased (but still present) at the right edge of the short IP in comparison to the right edges of larger IPs, as exemplified in Figure 14.

(22) European Portuguese example sentences with different proposed phrasings (Frota 2014: 14)

a. { a[z] a<u>luna</u>[ʃ] }$_{IP}$ { até onde sa<u>bemo</u>[ʃ] }$_{IP}$ { obtiveram boa[z] avalia<u>çõe</u>[ʃ] }$_{IP}$

b. {{ a[z] a<u>luna</u>[z] }$_{IP}$ { até onde sa<u>bemo</u>[ʃ] }$_{IP}$ }$_{IP}$ {obtiveram boa[z] avaliaçõe[ʃ] }$_{IP}$

“The students, as far as we know, have got good grades”

For the fifth boundary cue given above, the realization of clitics in their strong or reduced form (e), the data presented in Frota (2014) does not directly bear on the question of whether the IP is recursive or not in European Portuguese. However,

of spectral quality are actually also subject to gradual effects of boundary strength in European Portuguese utterances like these.

the percentages given with the examples in (23) indicate that this is also not a discrete but gradual difference, which only comes out when quantified: in (23)a, the clitic element in question *a os* "to the" is placed IP-medially, and realized in the strong, i.e. diphthongized, form [awʃ] in 20% of the elicitations, and in the reduced monophtongized one [ɔʃ] in 80% of them. When the clause *a os jornalistas* "to the journalists" receives its own phrase, either fronted as in (23)b or in the same position ((23)c), which in both cases effects a topical interpretation according to Frota (2014: 14), so that the clitic elements are IP-initial, this number goes up to 88% and 92%, respectively. Although cases testing this are not reported on in Frota (2014), the difference between 20% and around 90% is large enough to hypothetically allow for intermediate prevalences (say, around 55%) that might characterize the behaviour of cases that are in-between: at the left edge of the second element in a compound IP (comparable to the beginning of *até onde sabemos* in (22)b), for example. In such a position, a clause like *a os jornalistas* would be adjacent to an initial boundary, unlike in (23)a, but this boundary would be weaker than the boundary the clause is adjacent to in (23)b and (23)c following Frota's analysis. Should this turn out to be the case, it would indicate that boundary strength of recursive domains is also expressed by the quantitative prevalence of an otherwise discrete phenomenon.[56]

(23) Three European Portuguese test sentences with reported prevalence in realization as its strong form [awʃ] instead of the reduced form [ɔʃ] of the clitics *a os* "to the", from Frota (2014: 14)

a. [a[z] angolana[z] ofereceram especiaria[z] [ɔʃ] jornalista[ʃ] $]_{IP}$
'The Angolan women offered spices to the journalists' (20% occurrences diphthongized)

b. [[awʃ] jornalista[ʃ] $]_{IP}$ [a[z] angolana[z] ofereceram especiaria[ʃ] $]_{IP}$
'To the journalists, the Angolan women offered spices' (88% occurrences diphthongized)

c. [a[z] angolana[z] ofereceram especiaria[ʃ] $]_{IP}$ [[awʃ] jornalista[ʃ] $]_{IP}$
'The Angolan women offeres spices, to the journalists' (92% occurrences diphthongized)

56 Although this is presumably a simplification. Just as VOT and fricative spectral energy are located on a continuum, so are the differences in formant frequency change between [ɔ] and [aw]. If it were the case that non-maximal IP boundaries really resulted in an intermediate rate of monophtongization, it would be interesting to see whether measurements of these continuous parameters also revealed intermediate realizations. This would then presumably be reflected quantitatively as differing rates of realization of a discrete variable because of our categorical perception.

3.6.2.5 (Tokyo) Japanese

For (Tokyo) Japanese, Kubozono (1989) reports on findings that again also involve tonal scaling. It is usually assumed that the intermediate phrase in Japanese is the domain of downstep (Beckman & Pierrehumbert 1986; Venditti 2005), such that the pitch accents and initial rise of a second accentual phrase within the same intermediate phrase are realized at a lower pitch than those of the preceding one. However, when they are grouped into two separate intermediate phrases, the second AP's tones are realized at the same height as that of the preceding one, and reset takes place. In addition, the perceived disjuncture between words belonging to two separate intermediate phrases is larger than between two APs, e.g. pauses are more likely to occur or to be longer[57] (Venditti 2005: 176, 185–186). Kubozono's (1989) findings show that phrases composed of right-branching complex noun phrases containing only accented words (i.e. made up of as many APs, since it is usually taken to be definitional of APs to contain a single lexical pitch accent, cf. Ito & Mester 2012: 284) such as (24)a, realize a "pitch boost", an effect similar or equivalent to upstep, on the second accent, i.e. the one following the stronger boundary, while those composed of left-branching noun phrases, such as (24) b, do not. This upstep also takes place on the third accent in phrases realizing balanced structures, such as (24)c, again after the stronger boundary.

(24) Japanese test phrases with prosodic bracketing, adapted from Kubozono (1989)

a. right-branching complex noun phrase[58]

${}_{\text{iP}}[{}_{\text{AP}}[\text{kowa'i}]_{\text{AP}}]_{\text{iP}}$ ${}_{\text{iP}}[{}_{\text{AP}}[\text{me'-no}]_{\text{AP}}$ ${}_{\text{AP}}[\text{ya'mai}]_{\text{AP}}]_{\text{iP}}$

terrible eye-GEN disease

"terrible eye disease" (Kubozono 1989: 42)

57 Note however that Venditti (2005: 186) calls pauses neither a necessary nor a sufficient condition for the perception of a disjuncture of the kind that occurs between intermediate phrases. She also points out that there are cases where the perceived disjuncture does not match with the tonal cues assumed for a given category.

58 Kubozono (1989) locates the difference between these three types of sentences in the syntax, and assumes that this is then what is reflected in differing prosodic structures. The argument for syntactic difference however is only built on the lexical semantics and world knowledge: in *kowai me-no yamai* "terrible eye disease", *kowai* "terrible" is only understood to modify *me-no yamai* because that is the most likely interpretation. A marginal interpretation of "the disease of terrible eyes" is presumably possible. If *kowai* were replaced by *akai* "red", the noun phrase would most likely be understood as "the disease of red eyes" and correspondingly parsed and phrased as [akai me-no][yamai], because redness is something more likely to be said of eyes than of an eye disease. Similarly, in the other two examples, replacement of one of the lexical items by another of the same class, without changing anything in the observable syntactic structure, would result in an interpretation with different branching. It would be interesting to see if different kinds of prosodic phrasing were available in ambiguous cases, such as a variant on (24)b, *kanbashii ayu-no nioi*, where *kanbashii* "aromatic" could modify either *ayu* or *ayu-no nioi*.

b. left-branching complex noun phrase
$_{iP}[_{AP}$[na'mano]$_{AP}$ $_{AP}$[a'yu-no]$_{AP}]_{iP}$ $_{iP}[_{AP}$[nio'i]$_{AP}]_{iP}$
raw type.of.fish-GEN smell
"smell of raw *Ayu*" (Kubozono 1989: 44)

c. balanced complex noun phrase
$_{iP}[_{AP}$[na'oko-no]$_{AP}$ $_{AP}$[a'ni-no]$_{AP}]_{iP}$ $_{iP}[_{AP}$[ao'i]$_{AP}$ $_{AP}$[eri'maki]$_{AP}]_{iP}$
Naoko-GEN older.brother-GEN blue muffler
"Naoko's older brother's blue muffler" (Kubozono 1989: 51)

Parallel effects of prosodic boundary strength on the scaling of the AP-initial rise were found by Selkirk et al. (2003) in a reading task with differing syntactic structures. They found that scaling depended on whether the AP was initial or non-initial in an iP. Based on these findings by Kubozono (1989) and Selkirk et al. (2003), Ito & Mester (2012) propose to reduce the two levels of the AP and iP in Japanese into one, as recursive instantiations of ϕ, the phonological phrase. In their interpretation, downstep applies at that level. Because it is definitional for the minimal ϕ (=AP) to contain only one lexical pitch accent (and one phrase-initial accent) and downstep only applies between two pitch accents, they argue that the effect of downstep is only visible at a recursive instance of a non-minimal ϕ, i.e. an ϕ that contains at least another ϕ, and the scaling of the AP-initial rise is increased with increasing levels of recursion of the ϕ (Ito & Mester 2012: 286). The same boundary strengthening effect is what they invoke to explain the upstep in Kubozono's data (Ito & Mester 2012: 294).

Independent of prosodic recursion, all of these studies provide evidence that initial prosodic strengthening (Fougeron & Keating 1997, cf. also 3.2.2) also affects pitch scaling and is sensitive to differing boundary strengths, including differences that are not predicted by the traditional prosodic hierarchy. They also provide evidence that pitch scaling, especially in the case of regular downtrend, cannot be modeled purely locally, in contrast to what Pierrehumbert (1980); Liberman & Pierrehumbert (1984) had suggested, and that it instead seems to make reference to a hierarchical organization of larger prosodic units. From there, assuming prosodic recursion proceeds by two main arguments: for one, the fact that the prosodic unit of the smaller subparts and that of the whole utterance do not differ in their discrete cues (boundary tones or constraints on accent placement within them) but only gradually in their boundary strength as cued by pitch scaling of the boundary-adjacent movements, duration and degrees of pitch reset, is seen as most compatible with an account taking them all as instantiations of the same recursive phrasal category (Ladd 1986: 327–328; Frota 2012: 260–261; Ito & Mester 2012: 294; Myrberg 2013: 108–110; Truckenbrodt & Féry 2015: 37). Secondly, if the component units were taken to be categorically different from the overarching unit based on these gradual differences, then sufficiently long utterances with sufficiently complex structure would require postulating a large number of additional categories ad hoc, with no principled way of accounting for their similarities (the fact that they all share the same boundary phenomena with only gradual differences, cf. Ladd 2008: 294–295).

3.6.3 Separating arguments for prosodic recursion from assumptions of universality and a close prosody-syntax mapping

I would like to restrict arguments for recursive prosodic phrasing to those based on the kind of evidence described in the preceding sections, because they require the fewest theoretical presuppositions. There are other works invoking recursive prosodic structure at a level above the prosodic word for the analysis of their observations, but they are based on additional theoretical assumptions. For example, Elfner (2015) proposes recursive phonological phrases to account for the distribution of phrase-initial LH and phrase-final HL pitch accents in Connemara Irish: the HL accent appears on the stressed syllable of the final word in the phonological phrase, but the LH accent only appears on the stressed syllable of the leftmost word in a ϕ that is not minimal, i.e. one that dominates another ϕ (Elfner 2015: 1180, 1182), similar to the argument in Ito & Mester (2012). While her analysis seems well-suited to the data she presents, this sort of evidence is of a somewhat different nature: instead of gradual differences between instantiations of what is the same prosodic category by all other cues, as we have seen in the other cases discussed in this section, in the Irish case under Elfner's proposal, the different recursive levels of ϕ actually differ categorically, in their tonal makeup (LH initial in all ϕs, LH initial and HL final in all non-minimal ϕs). Bickel et al. (2009); Schiering et al. (2010) argue for up to 4 prosodic units close to the prosodic word in Limbu (Sino-Tibetan), based on the observation that they all serve as the application domain of quite distinct processes. They discuss and explicitly discard the option to solve this puzzle via a recursive prosodic word domain, precisely because the domains in question do not have the same phonological properties (Bickel et al. 2009: 50). However, the necessity for a recursive solution here only even arises when there is some incentive to apply the same set of prosodic categories universally for all languages. That is clearly the case for Ito & Mester (2012: 287–289), who posit recursive versions of the prosodic word (ω), phonological phrase (ϕ) and intonational phrase (ι) as universal interface categories with the syntax and who explicitly reject the postulation of prosodic categories based only on observable phonetic effects. Elfner (2015: 1203) equally voices some optimism that such a universal reduction would facilitate typological comparison because it is "uncontroversial to assume that all languages distinguish an intermediate category" between ω and ι.[59]

While this assertion is perhaps premature in the face of reliable empirical studies on the prosody of probably less than 5% of all living languages, it shows that some of the driving motivation for the proposal of prosodic recursion lies in this potential for universalization as well as the close correspondence with syntax. If these are not presupposed axiomatically, it is not problematic to assume 4 distinct categories resembling the prosodic word in Limbu, for example, if they can indeed all be shown to be the domains of distinct processes. The evidence on prosodic boundary strengthening we have discussed is of a different kind, because if recursion is not assumed in those cases then the number of categories of the prosodic hierarchy will potentially rise indefinitely, assuming that gradual boundary differences can still be found in sufficiently long and complex utterances (which hasn't been tested so far).[60] This is conceptually problematic even if no assumptions about the universality of prosodic categories or a tight mapping between syntax and prosody are made. It is clear in any case that prosodic recursion must have its limits, as argued for by both van der Hulst (2010) and Ladd (2008: 297–299): no one has ever proposed an intonational phrase inside a syllable, for instance, and the relative "flatness" of prosodic structure as compared to the syntactic structure seems to be still agreed upon by everyone. In addition to clear evidence of abundant mismatch between syntax and prosody, there is also evidence that the same kind of syntactic embedding does not correlate with similarly recursive prosodic structures in different languages: Féry & Schubö (2010) found evidence from downstep indicating recursive prosodic phrases in German utterances with syntactic center-embedding, but not for similar Hindi utterances.

In the interest of adequate descriptions for each language under investigation, universal claims should be made very cautiously, while the general idea of different prosodic levels that stand in a hierarchical relation to each other and serve as

59 This is somewhat ironic because for the analysis of her Irish examples, all full utterances, she makes no recourse to a second phrasal category (namely ι) at all, having no need for anything above a maximal ϕ in terms of unexplained cues or structural analysis.

60 For Korean, Jun (2007) takes this path: she proposes the intermediate phrase based on findings about cues that resemble those for the AP, but in a stronger version (increase in tonal scaling, but depending on the AP-initial tones, which are in turn dependent on the AP-initial segments, cf. Jun 2007: 160–161). Above the iP there is also an IP, which is clearly separate because it has different cues (additional boundary tones, cf. Jun 2007: 155). Whether it is more parsimonious to take the Korean intermediate level also as a recursive version of the AP, as Ito & Mester (2012) propose for Japanese, or as its own category, as Jun (2007) does, is so long a question of theoretical choice as it is unclear whether the cue strength for it differentiates only between two versions (AP and iP) or indefinitely and correlating with the number of proposed levels of recursive AP-embedding.

the domain of phonological rule application should always be kept in mind. Ladd (2008: 298–299) proposes a weakened version of the SLH, in which the different levels are still ranked, but where compounding of levels is allowed in order to reduce the number of categories necessary for the adequate description of each language, explicitly with the motivation that this would "allow us to identify any given boundary as being of one category or another on purely phonetic and phonological grounds, without as it were looking over our shoulder at the theoretical consequences for prosodic structure or for syntax-prosody mapping".

3.6.4 What is (not) known about prosodic recursion in Spanish and Quechua

As far as I have been able to determine, similar empirical studies have not been done on Spanish (or on any variety of Quechua). I will review some results from related studies on Spanish, but the issue of prosodic recursion based on empirical evidence itself so fas has not been tackled. Garrido et al. (1995) attempt to correlate degree of final lengthening, pitch reset and pre-boundary pitch contours in read Spanish utterances with differing syntactic boundaries[61] but only find non-significant tendencies for final lengthening. Because they treat pitch reset and pre-boundary contours only as categorical variables (presence or not of reset and a pre-boundary peak), their findings in this respect are not interpretable for the issue at hand. Rao (2008) investigates phrasing patterns in Barcelona Spanish, via SVO sentences read by 18 speakers that systematically differ in terms of the length (in prosodic words) and complexity (depth of syntactic branching) of both subject and object. He assumes that under a tight mapping between syntax and prosody and following a proposal originally from Ladd (1986), postnominal appositives should form a recursive phrase together with their head NP (Rao 2008: 100–101). He also provides the pitch track of an individual example where tonal scaling might reflect a difference in boundary strength:

In Figure 15, pitch is scaled clearly higher at the end of the postnominal adjectival modifier *inteligente y gordo de Barcelona* than after *el Javier*, with which it forms an NP together. However, Rao (2008: 101) does not claim that this is recursive prosodic phrasing, unlike in the case of an appositive. He also does not make any statements about whether this scaling difference is an individual occurrence or a robust finding in this or similar conditions. Instead, he simply assumes that whenever a phrasing such as [head noun]$_{\text{PhP}}$ [appositive]$_{\text{PhP}}$ occurs, this constitutes recur-

61 They do not refer to categories of the prosodic hierarchy, but under a close syntax-prosody mapping their different syntactic boundaries would likely result in recursive phrasing.

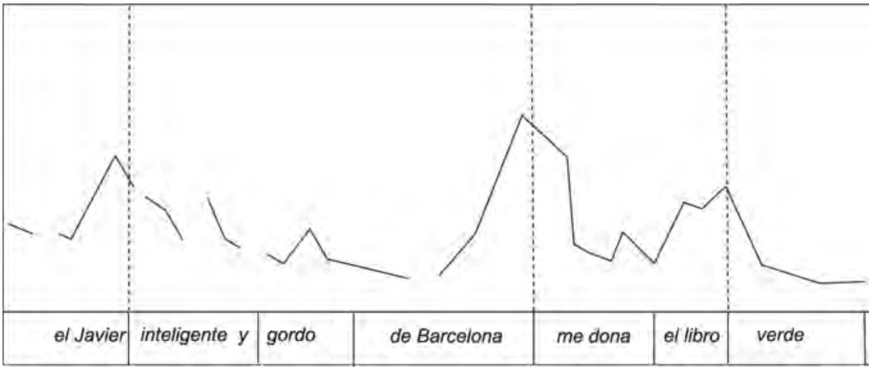


Figure 15: Pitch track of a read Barcelona Spanish utterance from Rao (2008: 97), *el Javier inteligente y gordo de Barcelóna me dona el libro verde* "The intelligent and fat Javier from Barcelona donates the green book to me". The phrasing according to Rao 2008 is (El Javier)ϕ(inteligente y gordo de Barcelona)ϕ(me dona el libro)ϕ(verde)ϕ.

sive prosodic PhP phrasing. That cannot be counted as empirical evidence for it. His criteria for identifying a PhP boundary (equivalent to an iP boundary) include the presence of continuation rises, decreases in pitch range, phrase-final lengthening and pauses (Rao 2008: 93), but when presenting the results, only the presence or absence of a boundary is marked, and no information on the cues given. Interestingly, his results show that even in the case of appositives, the "recursive" phrasing is never the only, and sometimes not even the most frequent, pattern (cf. e.g. Rao 2008: 107). At a broader level, they show that syntactically motivated phrasing constraints aiming to map relevant syntactic boundaries to PhP boundaries can frequently be overridden by prosodic constraints aiming to create balanced structures or those conforming to minimality or maximality constraints on the PhP, and that "as syntactic branching of utterances increases, prosodic concepts seem to play a more crucial role than syntactic conditions in determining the parsing of phrases" (Rao 2008: 120). In this regard, these results reported for Barcelona Spanish are similar to those made for Castilian Spanish in Prieto (2006) and for Lima Spanish in Rao (2007) based on similar methodology[62] (read speech, 3 speakers from Madrid and one from Burgos in Prieto 2006, 3 Lima Spanish speakers in Rao 2007). In both of those, phrasing patterns are explored via OT constraints, and

62 Prieto (2006: 42) uses slightly different cues for the boundaries of phonological phrases than Rao (2007, 2008): her criteria are "the perception of a prominent stress (together with a level 2 phrase break in the ToBI framework)" as well as optionally a boundary rise. In general, the status of the phonological phrase / intermediate phrase in Spanish is not settled, and no phrasing studies based on spontaneous speech have been undertaken (Hualde & Prieto 2015: 359–360).

their results also suggest that the purely prosodic ones often outrank those aiming for a close mapping between prosody and syntax. It has also to be noted that only in the study on Lima Spanish do any of the experimental conditions reach something approaching categorical preference for a given phrasing pattern, and then always the one most clearly following prosodic constraints instead of ones promoting a close mapping (Rao 2007: 96–97, 100–102). Overall, this is less strongly the case for the Castilian Spanish study, where considerable variation in pattern preference between the speakers is reported (Prieto 2006: 46).

In sum, the question of recursive prosody in Spanish (and Quechua) is still open to be answered via further research. These studies however suggest that the assumption of a tight mapping between syntax and prosody might not be the best point of departure to go looking for prosodic recursion, since they show prosodic phrasing in Spanish quite clearly to be subject to a number of constraints caring little about such a mapping.[63] It must also be noted that all the studies discussed in this section (not only on Spanish) are based on read speech by less than ten speakers each (except Rao 2008), and yet they also report on some considerable variability in their results. No comparable studies addressing prosodic recursion have been done with more naturalistic data to my knowledge. In section 5.2, I will argue for prosodic recursion at the level of the IP based on data from Huari Spanish uttterances that are both semi-spontaneous and quite long and complex, also at the level of information structure.

3.7 Information structure and prosody

In the preceding sections, we discussed relevant aspects of prosodic typology mostly without referring to the dimension of meaning. In this section, I will first give a brief overview over the types of meaning intonation can convey and then concentrate on information structure. I will give an introduction to the model of context and information structure that I follow in this work (section 3.7.1), lay out how it can be applied to the study of understudied languages without having to make prior assumptions about a relation between information structural categories and formal means of expression (3.7.2), and then conclude with (what is known about) the relation between information structure and prosodic cues in Spanish and Quechua (section 3.7.3).

63 The equal status of prosodic and syntactic constraints for prosodic structure is also emphasized in Féry (2015).

Prosody and intonation aid in the signaling of a variety of propositional and non-propositional aspects of meaning. Prosodic phrasing helps to disambiguate semantic and syntactic structures as well as word recognition both in spoken language (cf. e.g. Nibert 1999, 2000; Serrano 2010 on Spanish, cf. also Brown et al. 2015) and probably even in silent reading (the "implicit prosody hypothesis", cf. Fodor 2002, the works contained in Frazier & Gibson 2015). Intonation, especially contour choice, has been shown not only to play a decisive role in signaling utterance type in many languages, but also to convey a number of propositional attitudes and other non-at-issue content in several languages (among others, cf. Ward & Hirschberg 1985; Hirschberg & Ward 1992 on pragmatic meanings classified as epistemic uncertainty and incredulity in English; Vanrell et al. 2013; Vanrell et al. 2017 on uncertainty and evidential meanings in Catalan polar questions; Grice & Savino 1997; Grice & Savino 2004; Savino 2012 on uncertainty and the difference between information-seeking and confirmation-seeking polar questions in Bari Italian; Venditti et al. 1998 on meanings including incredulity and insistence in Japanese; Wollermann et al. 2010; Wollermann et al. 2014 on uncertainty in German declaratives; Escandell-Vidal 2017 on evidential meanings in polar questions and Fliessbach 2023 on mirativity and obviousness in Spanish). Contour choice can also have an effect on the salience of conversational implicatures (Kurumada et al. 2014, Prieto 2015, Marneffe & Tonhauser 2019). Not least, prosodic cues are essential for the smooth working of the turn-taking and repair systems, two strong candidates for perhaps truly universal systematic features of human language (Stivers et al. 2009; Bögels & Torreira 2015; Dingemanse et al. 2015; Levinson 2016).

3.7.1 Context and the question under discussion

Here we will only treat one particular aspect of how prosody and intonation help to relate individual utterances to their linguistic and non-linguistic context, the signaling of information structure. This is because it will be the most relevant aspect in the analysis of the Huari Spanish and Quechua data. Other aspects of intonational meaning will also feature occasionally, but not be treated systematically. The present work is mainly on prosody and not on formal pragmatics, but because these aspects of linguistic context are crucial factors for understanding prosody in spontaneous, ongoing discourses, some background to them will be provided here. Information structure relates to how the information contained in utterances is packaged with regards to the momentary knowledge states of participants, the discourse progression up to the point of utterance, and the intentions of participants for the discourse progression in the future (Chafe 1976; Krifka 2007). I assume a context model similar to the ones of Farkas & Bruce

(2010) and Roberts (2012b). One of the main components of these models is the *common ground* (CG, Stalnaker 1974, 2002). The common ground can be defined as the proposition specifying the set of propositions on which the participants have reached an agreement with regards to its truth value. More accurately, it is the proposition which speaker A believes that speaker B believes that speaker A believes that speaker B believes . . . specifies the set of propositions which the speakers have agreed upon. This formulation preserves the fact that the common ground is always a doxastic projection by each individual participant about what they believe the common ground to be. Crucially, it is also necessary to include a performative aspect here in that the common ground is what speakers act like it were the common ground: this can be out of pretense or forgetfulness or for another reason (Stalnaker 2002: 704–705). However, it is also assumed that speakers are basically rational in the pursuit of their intentions and communicating cooperatively in the sense of Grice (1989); this assumption effectively allows speakers to ascribe reasons (intentions) to their interlocutor's behaviour and to keep calculating it rationally even when it deviates from what they themselves would expect their interlocutor to behave like given the state of CG, instead of having to assume that they behave randomly.[64]

What is not in CG can be seen as the set of propositions for which no truth value has yet been agreed upon. That is to say, these propositions define the set of all possible worlds that are compatible with CG (the context set in Roberts 2012b: 4). The assertion by one speaker of a proposition, publicly committing to a truth value for it, and the acceptance of this assertion by the interlocutor(s), moves the proposition to CG, reducing the context set. This movement of propositions from context set to CG is the basic drive of the discourse progression.[65] In order to model its directedness and coherence in a hierarchical structure, Roberts (2012b) employs

64 Recasting this point somewhat, one could say that the social act of committing to an action is more relevant here than whatever mental states are responsible for it: acting like a certain proposition is part or not part of CG is a commitment by a speaker to another to be consistent with respect to this position, and breaking such commitments has the social consequence of not being treated as a reliable communication partner anymore (cf. Geurts 2019).

65 This is necessarily an extremely simplified picture. Not only are there speech acts other than assertions and questions, such as directives, that have the aim of getting another speaker to change the way things in the world are instead of increasing shared knowledge about how they are, but there are also assertions with additional presupposed or non-at-issue modal meanings such as surprise and obviousness that force interlocutors to revise some of the assumptions already shared by them and contained in CG (surprise / mirativity) or that make statements about the fact that the content of an assertion is already contained in or entailed by CG and therefore not newsworthy (obviousness). See Reich (2018); Fliessbach (2023) for a discussion of these meanings.

the device of the *question under discussion* (QUD, also used in Farkas & Bruce 2010; Ginzburg 2012, and several others). In her QUD model, discourse progresses via the posing and answering of (mostly implicit) questions. By accepting a QUD as the current one, discourse participants agree to keep making attempts at answering it until they either agree that it has been fully answered or that they must leave it unanswerable for the moment (Roberts 2012b: 6–7). Questions are only felicitous when their answers are not entailed by CG[66] and when they can be construed to be *relevant*, that is when answers to them contribute to providing (partial) answers to the current QUD in the discourse[67] (Simons et al. 2010: 316–317; Roberts 2012b: 21), with a partial answer defined as providing an evaluation of at least one of the propositions comprising the *alternative set of the question*, i.e. all possible answers to it (Roberts 2012b: 11–12, based on the semantics of questions first proposed in Hamblin 1958, 1973). Assertions are also only felicitous when they are not entailed by CG and when they are relevant, providing a (partial) answer to the current QUD. The entailment and relevance conditions define permissible *question strategies*, i.e. those that aim to completely answer sub(-ordinate) questions in order to partially answer super(-ordinate) ones; once a sub-QUD has been answered, the super-QUD it is entailed by becomes the current QUD again, giving discourse overall a hierarchical QUD structure. The question alternative set is pragmatically restricted by the discourse context, as in (25), where the context reduces the question alternative set to three members from the potentially infinite number of propositions about Alice winning some prize.

(25) *Context*: Alice, Claire, and David each won a prize at a sheep shearing competition. They bring home a pair of golden scissors, a silver tuft of sheep's wool, and a bar of lanolin soap.
Question: What did Alice win?
Pragmatically restricted question alternative set: {Alice won the golden scissors, Alice won the silver tuft of wool, Alice won the lanolin soap}

66 This is again a simplification. In practice, questions whose answers are entailed by CG do occur and are likely to provoke an assertion that is marked for obviousness (cf. Fliessbach 2023). Effectively they and their responses represent highly marked discourse moves.

67 Based on a discussion of utterances containing epistemic modals and evidentials in discourse, Roberts (2015: 52–53, 2017: 30–31) introduces a revised notion of relevance which allows discourse moves (questions and assertions) to be indirectly relevant, i.e. felicitous, when they provide an assessment in terms of the *likelihood* that a relevant proposition is true or false (see also Simons et al. 2010: 316–317, note 3).

Propositions forming the question alternative set of a wh-question all share a similar structure in that they only differ with regards to which element replaces the wh-element in the question, with the appropriate type of entities specified by the choice of the wh-element (cf. Roberts 2012b: 10). Polar questions have a question alternative set comprising only two members that only differ with respect to their absolute polarity (cf. especially Farkas & Bruce 2010 for a treatment of polar questions in discourse). Alternative questions have effectively the same kind of question alternative set as wh-questions, with a specified constituent in the question to be replaced by one in the answer, but the answers treated as likely by the speaker are usually fully listed in the question, and the restriction on appropriate types of entities is provided only by context, not the choice of wh-element (e.g. *Do you prefer Canada, going home, or Chadwick Boseman?*; cf. on alternative questions Sadock & Zwicky 1985: 178–179; Riester & Shiohara 2018: 292). A felicitous answer to a question must be the assertion of a proposition that is a member of the question alternative set; this *question-answer coherence* is pragmatically ensured via the Gricean maxim of relevance/relation, according to Roberts (2012b: 21) (cf. also Grice 1989; Krifka 2007: 22–23).

While this conception is probably not fully capable of providing a realistic model of discourse (cf. Riester et al. 2018: 405 for this view as well as e.g. Roberts 2012a: 8–14 for a number of outstanding issues), it couples two major mechanisms: discourse coherence (via question-answer coherence and the hierarchical QUD structure) and the internal division of utterances via their information structural status. The latter has two aspects: the division between what is and what isn't *at-issue* and the one between *focus* and *background*. The former of these describes a meaning distinction at the level of propositions: in an utterance, only those propositions are *at-issue* (AI) that are relevant with regards to the current QUD (Simons et al. 2010: 317; Beaver et al. 2017: 280; Roberts 2017: 9), those that aren't are *non at-issue* (NAI). The AI/NAI-division is a broad label encompassing aspects of a range of phenomena such as pragmatic presuppositions, appositives, non-restrictive relative clauses, evidential meanings, conventional implicatures, and the projection behaviour of factives (cf. e.g. Tonhauser et al. 2013; Faller 2014; Bianchi et al. 2016; Tonhauser et al. 2018; cf. Potts 2015 for an overview). One diagnostic for at-issueness is whether content can be directly targeted by polarity particles in responses, according to Roberts (2015: 44–46, 2017: 9). According to this diagnostic, some of the meanings conveyed by intonation are also non-at-issue (cf. for this view also Potts 2005: 26, 36–37; Prieto 2015):

(26) *Context*: Alice, Bob, and Doreen are living together. They are going to have friends over for dinner.
Alice: Who's coming for dinner?
a) *Bob*: Jonathan (L*H L-H%)
Doreen: a. No [he's not coming]
b. # No [you're not uncertain]
c. Why so unsure about it? You told me he was coming five minutes ago!
b) *Bob*: I'm not sure about Jonathan
Doreen: Yes [you are – you told me he was coming five minutes ago]

In (26)a), Bob answers Alice's question with the uncertainty contour on *Jonathan*. Doreen can deny the truth of the at-issue proposition that John is coming (answering the current explicit QUD, (26)a)a), but she cannot deny the non-at-issue content that Bob is uncertain conveyed by the intonational contour directly by using only the response particle *no* ((26)a)b). In order to do that, she has to perform a marked[68] move requiring much more explicit elaboration, like (26)a)c. That it is not the fact that she cannot target Bob's uncertainty for reason of it constituting a subjective mental state accessible only to him is shown in (26)b), where the uncertainty meaning is lexically conveyed in the main clause and at-issue (in the sense that Bob's public commitment about being unsure has a bearing on the probability that the answer to the relevant question *Is Jonathan coming?* can be taken to be true, cf. Roberts 2015, 2017). For Roberts (2017: 7), the focus-background division is part of the mechanism that can separate AI from NAI content, as in (27):

(27) (from Roberts 2017: 7)
Context: A, B, and C are discussing the race that B and C just watched, in which their mutual friend Mary was supposed to be a participant.
A: How did Mary run?
B: She ran [QUICKLY]$_F$—one of her best times ever!

68 Farkas & Bruce 2010 provide reasoning about what kind of conversational moves are more or less marked depending on context. They model the relatively privileged status of some *responses* to certain *provocations* using the device of the *projected set*, which contains the immediate future state(s) a context will assume upon an unmarked response to a given provocation. For example, they argue that the unmarked response to an assertion is to accept it (and thus for the asserted proposition to enter CG), and that to a neutral polar question is to commit to either of the propositional alternatives provided by the polarity, but that biased polar question types differ in precisely this property, i.e. that they only contain one of the polar alternative propositions in the projected set (Farkas & Bruce 2010: 93, 96–97).

C: a. No she didn't!
b. Hey, wait a minute! Mary didn't run at all! / But Mary didn't run at all!

In (27), the explicit QUD posed by A already presupposes that Mary ran, it asks only for how she ran, which is answered in B's response by the focused adverb *quickly*. It is only this assertion about Mary's quickness which C can target in their contradiction by saying *no she didn't*; if they wanted to challenge the backgrounded content that she ran, they would again have to resort to a more marked response like the ones in (27)b. In this conception, the focus-background division is a special case of the AI-NAI division as applied to a particular type of content, the *proffered* content: "the compositionally calculated truth conditional content of the expression; what it contributes to what is asserted, asked or directed by an utterance in which it occurs" (Roberts 2017: 6). Only (parts of) the proffered content can then be at-issue at all. Other types of content, the presupposed and the auxiliary content, which include conventionalised implicatures and the type of meaning conveyed by intonation in (26), for example, are always non-at-issue. Under this view, focus and background can essentially be subsumed via cross-classifying types of contents and at-issueness: focus would be [+at-issue, +proffered] and background [-at-issue, +proffered]. Others, although using the same terminology, draw the distinction elsewhere: Riester et al. (2018); Riester & Shiohara (2018); Riester (2019) develop a model for information structural annotation of actual conversations based on QUD structure, but they reserve the label of "non-at-issue" for those types of content Roberts would class as presupposed and auxiliary, and explicitly exclude backgrounded but proffered material from it (cf. Riester et al. 2018: 428). Although Roberts' proposal is theoretically neat because it achieves a comprehensive unification, I will here adopt the convention followed by Riester and colleagues, simply because their goals align more immediately with my own in this work: to operationalize the QUD model for the categorization of segments of actual spoken conversation according to information structure, independently of linguistic form, so that prosodic exponents can then be related to them without entering a circular argumentation (cf. Riester et al. 2018: 405–406). In the following, I will use their definitions to discuss the information structural notions of *focus*, *background,* and *topic*.

3.7.2 QUD-based annotation of information structure independent of form

Focus is very heterogeneously defined in the linguistic literature (cf. Krifka & Musan 2012: 6–7, 17–18; Riester & Shiohara 2018: 290–291 for overviews). The definition adopted by Riester et al. (2018: 417); Riester & Shiohara (2018: 291–292); Riester (2019: 166) is to say that the *focus* in an utterance is exactly that part which answers the current QUD, while its counterpart, the *background*, is material that is proffered but already given in the question itself. I will also adopt this definition of focus here. It has the crucial advantage of being entirely based on context (as described in QUD terms).

(28) QUD-focus relation in wh-questions
  a. Q: What happened?
     A: [Sharon sheared her sheep on the first Saturday in September]$_F$
  b. Q: What did Sharon do?
     A: Sharon [sheared her sheep on the first Saturday in September]$_F$
  c. Q: When did Sharon shear her sheep?
     A: Sharon sheared her sheep [on the first Saturday in September]$_F$
  d. Q: Who sheared their sheep on the first Saturday in September?
     A: [Sharon]$_F$ sheared her sheep on the first Saturday in September
  e. Q: When in September did Sharon shear her sheep?
     A: Sharon sheared her sheep [on the first Saturday]$_F$ in September

(29) QUD-focus relation in polar and alternative questions
  a. Q: Did Sharon shear her sheep on the first Saturday in September?
     A: [Yes]$_F$ / [Yes, she did]$_F$ / [Yes, Sharon sheared her sheep on the first Saturday in September]$_F$
  b. Q: Did Sharon shear her sheep or her goats on the first Saturday in September?
  c. A: Sharon sheared [her sheep]$_F$ on the first Saturday in September

In (28), the same utterance, *Sharon sheared her sheep on the first Saturday in September* has different focal domains depending on the preceding question context. The focused element(s) correspond exactly to the open variable in the wh-question and are marked by square brackets with a subscripted F. All material that is already given in the question is not part of the focus in the answer and thus backgrounded. As mentioned above, in the conception by Roberts (2017), it would also be non-at-issue. In actual conversation, the full answer utterances would most likely not be produced. Instead, most or all of the material in the background would probably be left out, but note that the focal material could not be further reduced given the QUD. This is true also in (29)a, where if the two first variants aren't chosen, then the full utterance has to be produced to be felicitous. The bracketed notation for the focal material is intended merely as a notational aid: the property of being in focus is bestowed entirely by the context in this conception, and no specific focus marking is necessary, although various expressive means, a number of them prosodic, are employed to various degrees in different languages to aid in the interpretation of which material is focal and which is backgrounded (see below).

In that, this focus definition differs from another very influential one, originally proposed in Rooth (1985, 1992) and defended in Krifka (2007); Krifka & Musan (2012), namely that focus indicates the presence of alternatives relevant for the interpretation. Because of the way the context is modeled, with felicitous answers effectively coming from the question alternative set of the QUD (Roberts 2012b), the two definitions cover a lot of common ground. However, while in the definition adopted here, focus is essentially just a label for that part of an utterance which directly corresponds to the QUD – and this means that alternatives will be relevant for its interpretation by definition –, in the other it is something (a feature) which actively acts to indicate that alternatives should be taken into consideration for interpretation. This is made explicit in the revised definition given in Krifka (2007: 19): "A property F of an expression α is a Focus property iff F signals (a) that alternatives of (parts of) the expression α or (b) alternatives of the denotation of (parts of) α are relevant for the interpretation of α." Here, focus is a *property of an expression* which *signals* a meaning, just like e.g. a morphological tempus marker. After some discussion that alternatives are also relevant in the interpretation of other expressions lacking focus, Krifka & Musan (2012: 7) refine the definition given there by saying that it "means to say that focus especially stresses and points out the existence of particular alternatives", which again implies the presence of some kind of focus signifier. From this perspective, it also makes sense to state, as they do, that it seems that some African languages do not make "use of focus to mark question-answer coherence" (Krifka & Musan 2012: 10, note 4) given findings that these languages do not employ any expressional means to mark the difference between focus and background in question-answer pairs. From the perspective of the definition adopted here, the argument is subtly different: the answer to the QUD, as long as it can be clearly established, is the focus by definition – but it is clearly an open question typologically whether and how individual languages will employ any means that aid to reconstruct such a context in interpretation.

It has repeatedly been observed that languages differ with respect to the expressional dimension of focus, and plausible hypotheses have been made about what makes focus marking more likely, e.g. based on an asserted proposition's status with respect to expectations built up in the discourse (Zimmermann 2008), or on a notion of scale of strength applied to different proposed types of foci (Féry 2013: 689–690). However, if focus is fundamentally conceived of as some kind of property of expressions, then making any kind of comparison between what can count as a focus marker becomes a circular enterprise because context cannot serve as a true third of comparison, but only as an insufficient indicator. This circularity and its gloomy consequences for the investigation of information structure across languages is the explicit reason why Riester et al. (2018: 406); Riester & Shiohara (2018: 290) define focus and all other information structural notions strictly based on context instead of on any linguistic form. Doing so means to assume that the underlying mechanism of information transmission and the essential pragmatics concerned with its packaging are effectively universal or at least sufficiently similar across languages, to the extent that a model can accurately describe them (Riester et al. 2018: 405–406). Effectively, the assumption is that this mechanism is shaped directly by universal aspects of human communication, such as the Gricean cooperative principle and the maxims resulting from it.[69] In this regard, the position adopted here from Riester and colleagues, locating the universality in the pragmatics, is different from both the Rooth/Krifka position, which locates it in some part of the expressional machinery, presumably in the syntax, and that taken by Matić & Wedgwood (2013), who deny any universality to focus as an information structural category. Whichever of these positions might eventually turn out to be true, the approach by Riester and colleagues has the clear methodological advantage that focus and other IS-notions can be identified across languages and from context alone, and thus allows to correlate definable context conditions with whatever means of expression a language might or might not employ. In order to do that in the analysis of naturally occurring conversation, criteria for the identification of implicit QUDs must be set down. This is highly important because in many types of conversations, explicit questions are quite rare, and instead, sequences of assertions occur, that nonetheless are taken to follow a hierarchical QUD structure (Riester & Shiohara 2018: 292). For them, the implicit QUDs have to be reconstructed, as it were, in order to be able to make statements about the information structural division of the assertion at hand.

The first and broadest of these criteria is question-answer congruence, i.e. that the observed assertion must be a felicitous answer to the implicit QUD (Riester et al. 2018: 411–412). Question-answer congruence allows any of (30)a-c (and some others) to be identified as implicit QUDs for the assertion in (30), but prohibits questions such as (30)d-e because they are not answered by the assertion.

69 Presumably, this can also be brought into relation with Chafe (1994: 109–110)'s „one new idea", and also with how information transmission is described by Shannon (1948), namely that it eliminates possible states a system could assume. If no indication in an utterance were available from context-based expectation which the relevant location for the elimination of such possibilites is, all transmitted material would be equally informative and no inference about the intended direction of the discourse were possible, which would arguably divest humans of a large part of their intention-reading skills, effectively grounding communication to a halt.

(30) Question-answer congruence
Assertion:
Alice went down the rabbithole
Question:
a. What happened?
b. Where did Alice go?
c. Who went down the rabbithole?
d. # What is a rabbithole?
e. # When did Alice go down the rabbithole?

Possible implicit QUDs are further constrained by two principles that capture a crucial difference between implicit QUDs and explicit ones: implicit QUDs, not being actually performed discourse moves, cannot possibly introduce any new material, they can consist only of material that is already given[70] in the context (and combine it with operators, such as a wh-element). This principle is called *Q-Givenness* in Riester (2019: 174). The second principle is called *Maximize-Q-Anaphoricity* (Riester et al. 2018: 412; Riester 2019: 175) and it states that implicit QUDs should not only consist only of given material, but that they "should contain as much given (or salient) material as possible". "As possible" is intended to be constrained by the preceding context and the observed assertion: this means that all material that can be determined as given in the answer (from the preceding discourse context) will be contained in the reconstructed implicit QUD. Effectively, *Maximize-Q-Anaphoricity* thus works to narrow the focus as much as possible both in the implicit QUD and in turn, in the assertion (cf. Riester 2019: 175).[71] We can see how the three principles act together in identifying implicit questions in the analysis of a short example excerpt from the Quechua corpus:

70 With a definition of givenness adopted from Baumann & Riester (2012), where it is broadly defined as an expression or a referent having been made available in the preceding discourse context, with a separation between referential and lexical givenness and a temporal decay of cognitive activation included in the model. See also Riester (2019: 174, note 8); cf. section 6.2.3.3 for a more detailed discussion.
71 Thus, while implicit QUDs are considerably constrained by the assertions they are reconstructed from, explicit questions, with only general constraints of coherence placed on them, can truly change the direction of discourse (cf. Riester 2019: 174–175).

(31) TP03_KP04_MT_Q_2640–2704[72]

time	*KP04 (with the path on the map)*	*TP03 (without the path on the map)*
264.0	alli-m good-ASS *alright*	
265.0	naa tsawra tillaku-pita-qa PSSP then lightning-ABL-TOP *well then from the lightning*	
266.8		ya yes *yes*
267.5	pasa-rku-y manka-yaq pass-DIR-INF pot-TERM *go up to the pot*	
268.7	manka-pa hana-n-pa-m pasa-n pot-GEN above-3-GEN-ASS pass-3 *it goes above the pot*	

(31) is a short excerpt from a *maptask* experiment forming part of the corpora analyzed in this work. As Riester et al. (2018: 408) emphasize, it is very important that a corpus analyzed in this way should be well understood by the analyst. This also includes knowledge about the type of conversation (in this case, the kind of experimental communication game), its overall context, and the speakers involved. General information on these matters is provided in sections 2.3 and 2.4. Here, it is relevant to know that the speakers are in the process of tracing the path for a second time already, after the first attempt led to confusion (because two landmarks are intentionally different on the two speakers' maps, cf. Figure 188 and Figure 189 in Appendix B). Their discourse proceeds from landmark to landmark, having begun with one where they were sure to be still in agreement. At 264.0, KP04 signals that the discussion about the preceding landmark is complete and that they can move on to the next issue; in terms of Farkas & Bruce (2010), the Table of current issues is empty (but in terms of QUD structure following Roberts 2012b, very superordinate QUDs such as WHAT IS THE OVERALL TRAJECTORY OF THE PATH? are still active). We will proceed backwards through the example, for reasons that become clear soon. At 268.7, the only new element in the assertion is *hana-n-pa-m* "above"; all the other elements are given already in the preceding turn. Thus by the principle of *Q-Given-*

72 https://osf.io/d7tzc

ness, the meanings encoded by *manka-pa* "of the pot" and *pasa-n* "s/he/it passes" are included in the implicit QUD, a good candidate for which would be something like WHERE IN RELATION TO THE POT DOES THE PATH GO?, so that only *hana-n-pa-m* is in focus in the assertion. Note that also *Maximize-Q-Anaphoricity* prevents WHERE DOES THE PATH GO NOW?, with a correspondingly broader focus [manka-pa hana-n-pa-m]$_F$. Moving one turn backwards to 267.5, although this is a directive, we can apply basically the same kind of reasoning: *manka-yaq* "to the pot" is not given in the context, so the implicit QUD must at least ask for it and it is in focus. Regarding *pasa-rku-y* "move in a direction", strictly speaking it is not given in the preceding context and thus the most apt implicit QUD could be something like WHAT NEXT?, effecting the broadest possible focus. On the other hand, it could be argued that the action of passing from one place to the other in this conversational game is so commonplace that it is always salient and thus given. Then, by *Q-Givenness* and *Maximize-Q-Anaphoricity*, the implicit QUD would be something like WHERE TO GO?, with a narrower focus in the assertion just on *manka-yaq*. Deciding this issue can be facilitated by considering the entire corpus, but in the end it is probably more important that the analytical decision of how to treat such verbs and the actions denoted by them should be consistent. Moving backwards by yet another step to the peninitial turn (in the excerpt) by the same speaker, a less simple decision awaits us. At a first glance, we could reconstruct an implicit QUD like FROM WHERE THEN? for *naa tsawra tillakupitaqa* "well then from the lightning" at 265.0, with *Maximize-Q-Anaphoricity* working to include *tsawra* "then", and perhaps even the meaning encoded by the ablative suffix *–pita*, into it. However, this turn is not an assertion by itself. The only argument for that could be that it is directly followed by backchanneling by the other speaker, but if the turns by KP04 at 265.0 and 267.5 were produced in a single utterance, it would be clear that *tillakupitaqa* does not constitute an independent speech act. Instead, it seems to fulfill a topical function here:[73] the directive at 267.5 is referentially disambiguated only by its presence. I want to argue further that it is a contrastive topic, albeit a type called *implicit contrastive topics* by Riester & Shiohara (2018: 300–301).

73 The suffix *–qa* or its variants is usually identified as a „topic marker" in Quechua grammars (cf. amongst others Parker 1976; Weber 1989; Cusihuamán 2001; Adelaar & Muysken 2004), and the gloss used here also indicates this. However, topic definitions are almost as numerous as focus ones (see Roberts 2011), and Weber (1989) describes uses of it for Huallaga Quechua that do not square well with all of them. In addition, in keeping with the approach laid out by Riester and colleagues, I try to refrain as much as possible from building arguments for information structural analysis from evidence based on formal linguistic features, instead of context-content relations. Section 6.4 will be particularly concerned with disentangling the cues for IS from different domains from context-based interpretation in Huari Quechua.

Contrastive topics, according to Büring (2003); Roberts (2012b), are the result of a complex strategy of inquiry that involves parallel subquestions. That is to say, when a complex QUD containing two wh-variables, such as WHO SHEARED WHICH SHEEP? is to be answered, then this is often done by asking the subquestions that result from filling one of the variables with the individual members of its answer set, such as SEAN SHEARED WHICH SHEEP?, SIAN SHEARED WHICH SHEEP?, SHANE SHEARED WHICH SHEEP?. This implies providing (partial) answers to the first variable. Thus the superquestion WHO SHEARED WHICH SHEEP? is answered by first answering WHO SHEARED SHEEP?, and then substituting the answers into the superquestion, yielding subquestions that are parallel. This is a viable strategy because each subquestion is entailed by the superquestion, each answer to a subquestion is a partial answer to the superquestion and thus relevant. In the answers to the subquestions, yielding parallel structures such as *Sean sheared Bessy, Berta, and Balu; Sian sheared Billy, Bartholomew, and Boris*; *Shane sheared Bob, Bonita, and Brenda*, the sets of sheep {*Bessy, Berta, Balu*}, {*Barbara, Bartholomew, Boris*}, and {*Bob, Bonita, Brenda*} are at-issue with regards to the current QUD, and the shearers *Sean, Sian*, and *Shane*, as contrastive topics, are at-issue with regards to the preceding QUD. According to Riester & Shiohara (2018: 295); Riester et al. (2018: 422–423), the parallel structure of such assertions indicates the complex relationship they are in, namely that they aren't just relevant for the current QUD, but answer such a superquestion in concert.[74] However, it is precisely this overt parallelism which at first sight seems to be missing in our example (31). Riester & Shiohara (2018: 300) remark that in their Sumbawa corpus, such overtly parallel structures do not occur, and it seems likely that they are in general quite rare in naturally occurring discourse (they do come up on occasion in our own corpora). They nevertheless argue that turns similar to our example should be analyzed as containing *implicit* contrastive topics because, while only realizing one of the parallel answers overtly, their interpretation in context implies the existence of such relevant (with regards to a super-QUD in the discourse) parallel predications made about other entities that would also be contrastive topics. They cite considerations that continuing topics, which are not contrasted, usually do not need to be overtly expressed at all (of course depending on the possibility for the language in question to not realize given arguments), and that when such elements are then overtly realized, they often indicate a topical change, i.e. a contrast, and such an implicit parallel structure if this is corroborated by the broader context (Riester & Shiohara 2018: 301). In our example, the

74 Cf. also the definition for so-called *delimitators*, comprising contrastive topics and frame-setters together, in Krifka & Musan (2012: 32–34), which also emphasizes that such structures cannot be interpreted purely locally.

case can be argued quite well that *tillakupitaqa* is an implicit contrastive topic. In the preceding discourse, speaker KP04 (the instructor in the *maptask*) had already tried to move the conversation to where the path leads from the lightning (as part of the repeated instruction going stepwise from landmark to landmark), this was then followed by the other speaker (TP03) backtracking because he could not follow the instructions KP04 was giving subsequent to moving from the lightning on his map. This is the issue that KP04 signals as having been settled by uttering *allim* at 264.0, and with the turn at 265.0, he thus indicates that the following path instructions are to be understood with reference to the lightning (again), as opposed to the subsequent landmark that left TP03 stranded. Because of this, and the stepwise progression from landmark to landmark, it seems reasonable to assume that a superquestion something like FROM WHICH LANDMARK, WHERE DOES THE PATH GO? plays an important role in the structure of their discourse, and that the assertion at 265.0 and 267.5 is to be understood as a partial answer to that, with *tillakupitaqa* an implicit contrastive topic in the sense laid out above. In summary, we can see that the approach pioneered by Riester and colleagues can be used to make quite a bit of headway[75] into an information-structural analysis of spontaneous speech corpora without making reference to formal means of encoding them, and it will be used in this way in parts of this work to separate these meaning-based categories from potential prosodic cues.

3.7.3 Prosodic cues and information structure

A strictly context-based definition of information structural categories still leaves the question open whether formal means of expression signal information structure directly. However, a view often adopted in the literature on prosody and intonation is that this signaling is only indirect (Ladd 2008; Calhoun 2010b; Kügler & Calhoun 2020). According to this view, a constituent e.g. being in focus is not automatically linked to some kind of prosodic exponent; instead, information structural division of utterances is signaled indirectly via metrical and prosodic structure, which is what is cued by phonetic exponents in turn. In the following two sections, I will review evidence for this hypothesis and refine the relation between information structure and prosodic cues first for Spanish, and then for what is known in this regard about (Cuzco) Quechua.

75 Although the analysis especially of implicit contrastive topics is not yet fully satisfying, as Riester & Shiohara (2018: 301) admit. My own analysis will often be based on reasoning drawing on a consideration of larger parts of the nonlocal discourse context, as exemplified above.

3.7.3.1 Spanish

In fact, much of the literature on prosodic focus marking in Spanish can be read as evidence for precisely this indirect relation. Cases in point are Gabriel (2007) and Vanrell & Fernández Soriano (2018). Both investigate prosodic and syntactic strategies of signaling information structure in assertions via a question-based elicitation paradigm involving speakers from several varieties of Spanish.[76] They find that the prosodic cues involved in signaling the information structural division between background and focus in declaratives include a high iP-boundary tone (H-) occurring at the right edge of the prefocal background material,[77] an LH* pitch accent on the final prosodic word of the focal material, a low iP-boundary tone (L-) at its right edge, and pitch accent compression or deaccentuation on the postfocal background material. While all of these cues were attested in productions from speakers of all varieties in both studies, they were also all optional to some extent. H- at the right edge of the prefocal background occurred in 81.3% of cases (113 of 139 possible utterances), with individual occurrence rates ranging from about 55% (3 speakers) to 100% (4 speakers) and the other speakers distributed evenly among intermediate rates in Gabriel (2007: 276–277). Across speakers, occurrence was more frequent when the focus domain was narrower (i.e. the focal material consisting of fewer elements in correspondence to the QUD) than when it was broader. Postfocal deaccentuation/pitch accent compression (cf. section 3.4.6) has slightly lower occurrence rates, ranging from 40% to 83% across the relevant elicitation contexts and also showing strong individual differences between speakers. Again, the width of the focus domain seemed to play a role, with a group of utterances where the focal material consisted of more than a single prosodic word never occurring with deaccentuation on the available postfocal backgrounded material (Gabriel 2007: 282). Vanrell & Fernández Soriano (2018) report on similar tendencies with regards to these two cues, but do not offer precise numbers. With regards to the realization of

76 In the case of Gabriel (2007): 18 speakers, of which 14 from various Spanish regions, all judged to be speakers of “standard Castilian Spanish”, and one each from El Salvador, Colombia, Mexico, Argentina, of which only the last two are judged to not be speaking “standard” Spanish by Gabriel (2007: 269). In contrast, Vanrell & Fernández Soriano (2018: 38) present data from 9 speakers, 2 from the Canarian Islands, 2 from the Basque country with Basque as L1, another two from the Basque country with Spanish as L1, three from Madrid; they take them to speak at least four different varieties of European Spanish.

77 Gabriel (2007: 280–281) also finds evidence that the segmental prosodic process of *sinalepha*, the assimilation of adjacent vowels across word boundaries, is blocked between the right edge of the final phrase containing backgrounded material and the left edge of the phrase containing focal material. For a study on variant phonetic realizations of the H- boundary tone, see Gabriel et al. (2011).

the focal material itself, both studies find that one alternative to the LH* L- contour is L* L%, which occurs only when the focal material is final in the utterance (but then not always, cf. also Hualde & Prieto 2015: 364). In the data reported on by Vanrell & Fernández Soriano (2018: 54), it is also not restricted to cases of broad focus. Vanrell & Fernández Soriano (2018: 43) attest to a third option, L<H* H-, which they describe as a "rise throughout the accented syllable which continues to the end of the intermediate phrase". This contour they only describe in cases where the focal constituent is a subject and occupies an initial position due to it-clefting (e.g. *Fue [Juanita]*$_{F}$ *la que vio siete armadillos*);[78] it is not followed by postfocal deaccentuation or compression (Vanrell & Fernández Soriano 2018: 43–54). The latter fact suggests a possible connection between the presence of the L- and subsequent deaccentuation or compression.

Both studies interpret their results in such a way that the right edge of the focal material seeks to align with the right edge of an iP, and with the most prominent position in that iP. In view of the definition of nuclearity adopted here from Ladd (2008) (cf. section 3.4.4), where the nuclear accent is the only obligatory and final one in an iP, we can simply say that focus is preferentially aligned with a nuclear accent. What the results of these studies (and many others) further show is that it is misleading to speak of prosodic cues for focus. The cues discussed above are not only used in the contexts described: H- and L- are simply cues for iP-phrasing, which is certainly not wholly determined by information structure. Deaccentuation and locally reduced pitch span also occur in contexts that are not postfocal (cf. Ortega-Llebaria & Prieto 2007; Torreira et al. 2014). Summing up, the division between focus and background is clearly one important factor influencing iP-level phrasing and nuclear accent placement in Spanish, but its signaling does not make use of a unique phrasing strategy, or one that is reserved exclusively for this function.

---

78 Note that by and large, syntactic strategies are found to be at least as variably associated with information-structural configurations as prosodic ones in both of the studies: preferences for different types of clefts, fronting, *in-situ* word order (with prosodic marking as discussed in the text) and the postposing of the focal material (*p-movement* in Zubizarreta 1998) vary somewhat across focus type and variety, but mostly across speakers, and are thus only ever tendencies, with the gist being that clefts are preferentially used in contrasting or correcting contexts, and that *in-situ* word order is overall far more preferred than p-movement, which is rather marginal (81/182 for in-situ vs 26/182 for p-movement in "information focus" contexts, 114/300 for in-situ vs 34/300 for p-movement in "contrastive focus" contexts in Vanrell & Fernández Soriano 2018: 48,54; cf. Gabriel 2007: 283, 289–290). However, some type of syntactic strategy resulting in a deviation from unmarked word order seems to be employed in roughly half of the cases.

While results on phrasing in Spanish indicate that focus is associated with the highest prominence at the iP-level, variably expressed by a range of phonetic cues, results on the paradigmatic choice between nuclear configurations also indicate that focus is not the determining factor for the choice of pitch accents and boundary tones. Nuclear configurations occur in focus position, but the choice of configuration can encode additional pragmatic meaning such as illocutionary force and modal evaluative meaning.

Returning to the three focal nuclear configurations just discussed, both the paradigmatic choice between LH* L-, L* L%, and L<H* H-, and indeed of these three variants against other nuclear configurations, does not differentiate between what is focal and what isn't.[79] Between those three, on the one hand, the choice is largely determined by position, with L* L% occurring IP-finally, LH* L- IP-medially, and L<H* H- seemingly optionally when important material is phrased separately IP-initially.[80] With regards to the contour conventionally transcribed as L* L%, its most

79 Or between „informational“ focus and „contrastive“ focus (a focal choice between a finite set of salient candidates), as is sometimes claimed: both Gabriel (2007) and Vanrell & Fernández Soriano (2018) explicitly test for this difference and come to the conclusion that it does not affect prosodic realization in any of the dimensions discussed here (“contrastive” focus is also sometimes related to a notion of “emphasis” that is supposedly expressed in increased pitch scaling).

80 Gabriel (2007: 285–287) also finds this contour in part of his data elicited through a reading task, where the context together with the word order of the elicitation sentence forces a reading in which the focal constituent is fronted. He dismisses instances of L<H* H- on such fronted constituents that denote the answer to the (explicitly given) QUD as an inappropriate reinterpretation of them as (left-dislocated) topics by the speakers. Given that fronting as a focusing strategy is never attested in his more spontaneous data and that in clefts, the occurring tonal movement is analyzed as LH* L-H% (Gabriel 2007: 283), this characterization is certainly not unreasonable. However, fronting (without claims about its intonational specification) has been suggested to convey a mirative or counterexpectational import in Spanish (Leonetti & Escandell-Vidal 2009; Reich 2018; Cruschina 2019). The lack of such a pragmatic specification in the elicitation contexts might go some way in explaining its scarcity in the data reported on by Gabriel (2007), and in the light of the findings by Vanrell & Fernández Soriano (2018), it seems plausible that L<H* H- is not a misinterpretation (under the purely contextual definition of focus adopted here, that is strictly speaking impossible if the experimental task was not itself misunderstood), but simply an option available for phrasing important IP- or utterance-initial material in its own iP. That could occur both in it-clefts and constructions with initial topics, so that “important material” would have to remain as placeholder until further research provides more precise insights. Under the view adopted here, prosodic configuration, syntax, and discourse context together would then act together to yield an interpretation disambiguating the information structural role of the fronted constituent.

An additional point could be made regarding the interpretation of L<H* H- based on its phonetic realization: Gabriel et al. (2011) show that the phonetic realization of H- has a range of variants, in most of which the pitch is affected not only in the final syllable of the phrase, but already in the preceding ones including the tonic and posttonic. That makes it difficult to tell whether the

conspicuous feature is the absence of any pitch excursion on the nuclear accented syllable, as noted by Hualde & Prieto (2015: 364), who also call this an “accent without tonal correlates”. That such a featureless contour is regularly interpreted to cue rightmost prominence in the IP can best be explained by the strong expectational bias for this metrical configuration, an observation already partly encapsulated in the formulation of the “nuclear stress rule” (Chomsky & Halle 1991 [1968]: 17–25). Prosody corresponding to this unmarked rightmost default prominence does not seem to need to be cued overtly most of the time, only marked deviations from it do (cf. Ladd 2008: 223, 257–259). This explains why focus in prefinal positions is associated with the acoustically more prominent nuclear configurations LH*L- and L<H* H-.

The paradigmatic choice between these three and other nuclear configurations, on the other hand, seems to encode pragmatic meaning at the illocutionary level and that of additional non-at-issue meanings: Fliessbach (2023) shows conclusively that non-at-issue meanings conventionally described as mirativity and obviousness, as well as whether an assertion is an agreeing or disagreeing response or a neutral provocation, have a clear effect on the nuclear configuration in declaratives even when focus is kept totally constant. A similar intuition also informs the data in Table 3 from Hualde & Prieto (2015) and much of the literature contributing to it. This means that LH* L-/L<H* H-/L*L% should be seen to signal something like neutral or unmarked declarative, to which focus position is orthogonal.

Based on this discussion, a more appropriate conceptualization for the relation between information structure and prosody is the following: the suprasegmental realization across the entire utterance cues a prosodic structure and a prominence profile (i.e., a metrical structure) that can be brought into relation with information structurally relevant divisions of the utterance. This cueing is asymmetrical because it is based on default expectations: as seen above, the expected = unmarked case of rightmost prominence at the highest level of metrical structure is often not given acoustically prominent pitch cues at all (cf. Calhoun 2010b: 11–13). The argument can further be made that even the nuclear accent is truly a metrical-prosodic category whose association to focus is only preferential, but not categorical (cf. Ladd 2008: 263–273). Evidence that this is true for Spanish comes from Calhoun et al. (2018), with similar arguments made for English in Calhoun (2010b). Calhoun et al. (2018) investigate the prosodic and syntactic realization in Venezuelan Spanish utterances (n=651 from 9 speakers from Valera) consisting of intransitive sentences under different information structural conditions and accounting

pitch accent on the iP-final word is LH* or L<H* based on peak alignment. While Vanrell & Fernández Soriano (2018) opt for the delayed version (L<H*), Gabriel et al. (2011) analyse it as LH*.

for the difference between unaccusative and unergative verbs. They performed a picture-based elicitation task that allowed speakers to produce utterances freely in terms of syntax and prosodic realization. See (32) for examples.

(32) Unergative and unaccusative Spanish example sentences, after Calhoun et al. (2018)

a. La chica estornudó — *unergative*[81]
"The girl sneezed"

b. La chica apareció / Apareció la chica — *unaccusative*
"The girl left"

Utterances were coded by the three authors for the position of nuclear stress (i.e. highest prominence in the utterance, on either the initial or final of the two lexical words) independently of the elicitation context. Acoustic analysis shows that this coding corresponds to a clear mean F0 difference (final stressed syllable 23 Hz lower) in favour of the initial accent in the case of initial stress and roughly equal pitch height (final stressed syllable 1 Hz lower) in the case of final stress, but only a relatively smaller difference in syllable length (final stressed syllable 50 ms longer) for initial stress, as opposed to final stress (final stressed syllable 67 ms longer, Calhoun et al. 2018: 16). That the coding thus corresponds to such a complex acoustic correlate is in itself evidence that strength relations as modeled via metrical structure are above all based on complex, context-dependent expectations (see also section 3.3.2).

In the analysis of their data, Calhoun et al. (2018: 18–20) find that in the overall majority (56%), their intransitive utterances have the word order subject-verb ((32) a and the first alternative of (32)b) with rightmost nuclear stress. All in all, 16% of utterances have the word order verb-subject (the second alternative in (32)b), mostly also with final nuclear stress. The remaining 28% exhibit subject-verb word order with initial nuclear stress. These ratios vary considerably across conditions, with contexts eliciting a correction (corrective focus) on the subject, those in which the subject is focused as answer to the QUD but not in correction (information focus), and those asking the question "what happened?" (broad focus) having significantly different rates of nuclear stress on the subject (53.3%, 32.5%, and 12.8%, respectively). Note that even in the corrective focus condition, 20.7% of utterances

81 The classification here is taken from Calhoun et al. (2018: 7–10) and the works their discussion is based on. In generative syntax, unergative verbs are those whose single argument is VP-external, while the single argument of unaccusative verbs is VP-internal. This has been related to semantic verb types, such as ones denoting an uncontrolled process (unergative), or changes of location (unaccusative).

still have subject-verb order with final nuclear stress, i.e. the focal position is not marked. For information focus, this number increases to 49%, so that nearly half of all occurrences do not mark the focal position (the remaining 26% and 18.5%, respectively, have verb-subject order). In contrast, in the broad focus contexts, less than 30% (28.8%) of occurrences do not have subject-verb order with rightmost stress. Only here, the verb type also accounts for a large and significant difference: with unergative verbs in the broad focus condition, 88% are subject-verb with rightmost stress, whereas with unaccusative verbs, only 54% are, while 19% have the same word order with initial stress, and 28% have verb-subject word order. In the condition of information focus, the difference due to verb types is much smaller and just about reaches significance (p= 0.04), while it entirely disappears in the corrective focus contexts. This means that even though both information structure as induced through discourse context and verb type (to a lesser degree) have demonstrable effects on nuclear stress placement, there is a clear default to place it rightmost, even when this means providing no overt cues to focus position. The relationship between prosodic/metrical structure, semantics and syntax in the signaling of information structure is clearly not categorical, but can only be conceived of as probabilistic and heavily shaped by contextual expectations (cf. Calhoun 2010b; the assumption of a distributional relation between information structure, prosody, and syntax is also evident in the use of stochastic OT for their modeling by Gabriel 2007). The role of expectations is also implicitly invoked by Calhoun et al. (2018: 22) when they say that the corrective focus condition is more often overtly cued because it is more informative. Such degrees of informativity can be modeled in a discourse model such as the one by Farkas & Bruce (2010), where the assertion p of a correction against a proposition q which is already in the commitment set of the interlocutor would entail an empty projected set and would therefore be a marked move. Since the acceptance of such a move necessitates the removal of q from the commitment set of the interlocutor in addition to entering p into CG, it is more informative than an assertion that is not a correction. A plausible hypothesis is therefore that the more unexpected (i.e. informative) and therefore pragmatically marked a move is, the more likely it is to also receive a prosodic form that is also marked in the sense that it cues a metrical structure that is not the default.

Interesting further implications arise from the conception of the nuclear accent as belonging minimally at the level of the iP, and the distributive relation shaped by expectations between metrical/prosodic structure and information structure: it makes the prediction that several nuclear accents can stand in a metrical relationship with each other, and that utterances produced with several of them can still be assigned e.g. an unmarked, i.e. rightmost relationship overall, and thus receive essentially the same interpretation in terms of information-structural division as smaller utterances consisting of only a single iP/IP (see also Ladd 2008: 271).

Calhoun (2010b: 6) claims that this is the case for English and uses the constructed example given in Figure 16 to demonstrate that what she calls an "emphatic rendition"[82] of *Arun bought a Porsche*, with each content word produced as a separate iP/IP (she assumes only a single phrasal category), can still be an answer to "What happened?", in that *bought* and *a Porsche*, each a phrase with a nuclear accent, are assigned the relation w-s at a higher node, which then receives the highest prominence at the highest node, assigning *Arun* w only at this level. While the idea in itself is compelling, it somewhat suffers from being not much more than a thought experiment in this fashion. From an intonational perspective, investigating such utterances consisting of several iP/IPs is also compelling in terms of questions about how prosodic structure might deal with this and what it might mean for the discussion about recursive prosodic structure (cf. section 3.6). As a first approximation to how such multi-iP/IP utterances might actually play out, we can consider the two utterances in Figure 17 and Figure 18, which are basically attempts at recreating something like Calhoun's invented example for (Lima) Spanish.[83]

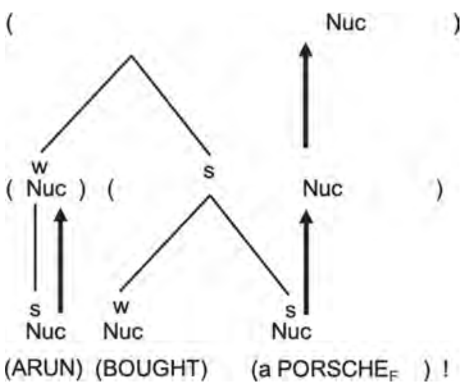


Figure 16: Metrical structure of an „emphatic rendition" of the sentence Arun bought a Porsche, adapted from Calhoun (2010b: 6).

82 The context Calhoun (2010b: 5) evokes for the constructed example is that „the speakers are so surprised they produce every word in a separate phrase". This is perhaps not fully convincing, especially without further intonational specification, but intuitively, I would agree that such utterances where each or nearly each word is phrased separately, do exist.

83 The speaker is the same speaker from Lima who also produced the utterances shown in Figure 7, Raúl Bendezú Araujo, who was kind and patient enough to record himself speaking them according to my instructions, for which I owe him thanks. They are of course totally artificial, but only serve to illustrate the problem here.

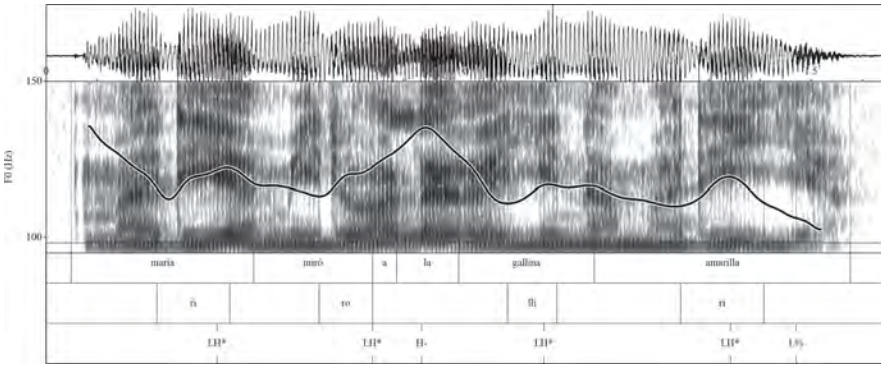


Figure 17: *María miró a la gallina amarilla* ‘Maria looked at the yellow chicken’, intended as an answer to “What happened?”, spoken by a Spanish speaker from Lima, normal rendition.

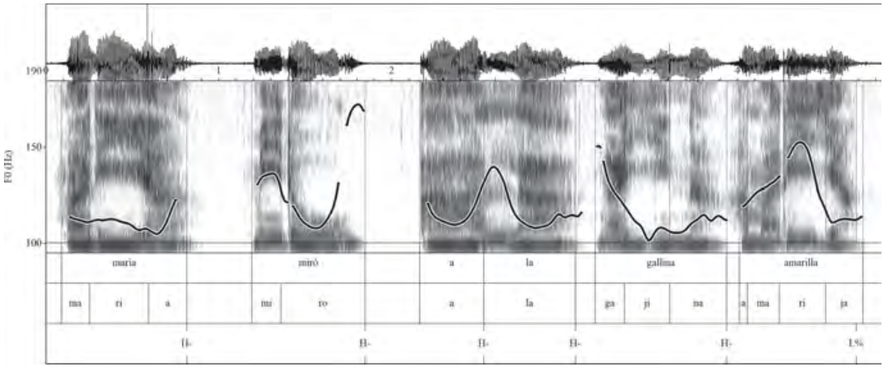


Figure 18: *María miró a la gallina amarilla* ‘Maria looked at the yellow chicken’, intended as an answer to “What happened?”, spoken by a Spanish speaker from Lima, “insistent” rendition.

Both Figure 17 and Figure 18 are productions of the utterance *María miró la gallina amarilla* “Maria looked at the yellow chicken” as answer to the question “What happened?”. While in the case of Figure 17, the speaker was told to produce the utterance as naturally as possible in this context, Figure 18 is the result of being asked to produce the utterance to the same QUD, but explicitly with added insistence, speaking slowly and hyperenunciating, as if talking to someone hard of hearing or slow on the uptake. I have refrained from transcribing pitch accents in Figure 18, both because Lima Spanish has not really been given an intonational account in the AM framework to my knowledge, apart from the inclusion of equally unanalyzed individual examples in the *Atlas interactivo de la entonación del español* (Prieto &

Roseano 2009–2013), and because the example is so obviously artificial.[84] It is still instructive to make a comparison with Figure 17. There, each content word realizes an LH* pitch accent, and a high rising pitch movement suggests the presence of a H- on or after *a la*, indicating that an iP-level boundary occurs there, while finally, a L% can be safely assumed. In comparison, Figure 18 seems to consist of a far greater number of phrases, visible not only from the pauses but also from the accompanying pitch movements indicating boundary tones, which I have tentatively transcribed. What is interesting is on the one hand that seemingly each single prosodic word (including the clitics *a la*, which seem to have been promoted to prosodic word status) is indeed realized as an individual iP and thus with its own nuclear accent, and that on the other hand, prominence relations between them can still be made out. The boundary rise delimiting *miró* is scaled far higher than any of the others, suggesting perhaps a stronger boundary there (roughly corresponding to the position of the H- in the "natural" Figure 17). The pitch accent on the final word, *amarilla*, is also quite certainly different from the preceding ones, further suggesting a structure above the level of the individual iPs, presumably an IP, with respect to which this pitch accent is final. Assuming that Figure 18 is indeed an appropriate answer to the question "what happened?", then this incidental comparison does seem to bear out the prediction by Calhoun (2010b) regarding the assignment of prominence relations at successive levels above that of the one at which the nuclear accent is assigned. In section 5.2 I will introduce a set of less artificial elicited Huari Spanish utterances that share the property of apparently assigning prominence relations at multiple levels of "nuclear accents". I will provide an analysis there for them that connects the indirect relation between phrasing, prominence, and information structural categories with the question of prosodic recursion in Spanish.

Broadening the focus somewhat, we can finally turn to Kügler & Calhoun (2020) to see how the prosodic cues to information structure found for Spanish relate to a wider typological context. They describe three main strategies in which information-structural categories (mainly focus) can be signaled in this indirect fashion via prosodic structure crosslinguistically. In our discussion of Spanish we have seen all three of them employed, namely prominence (focus seeks to align with the highest stress in the phrase which is often realized with the most acoustically prominent pitch movement), phrasing (focus seeks to align with prosodic edges) and pitch register (the focus-background division of the utterance corresponds at least partially

84 We might note that the majority of identifiable pitch accents seem to be falling, perhaps instances of HL*, which occurs as part of HL* L% in Table 3 with the label "insistent explanation/ insistent request".

to an asymmetrical difference in pitch register and span, in that postfocally, pitch movements are reduced and occurring at a lower register). Languages seem to differ in the degree to which they prefer either of these strategies and implement them.

3.7.3.2 (Cuzco) Quechua

In the case of Quechua, very little research has been done regarding prosodic cues for information structure. Cole (1982: 210–211) describes for Imbabura Quechua (Quechua IIB, Ecuador) that there is a single main intonational peak in an utterance and that it is normally located on the final word, followed by a final fall independent of whether the utterance is a declarative, polar interrogative, or wh-interrogative. The main peak can move to a non-final word if that word is "emphasized or contrasted". If an utterance consists of several breath groups (i.e. smaller phrases), they can each bear a main peak. For Cuzco Quechua (Quechua IIC, southern Peru), Cusihuamán (2001: 79–81) does not mention a similar shift in the position of a main pitch peak according to the position of what is felt to be the "contrasted" word; he only describes pitch movement on the last two syllables of the utterance (using a four-level system), indicating that while declaratives, interrogatives, and imperatives have a final fall (with different pitch spans), exclamative utterances have elevated pitch only on the final syllable and end high. These affirmations are based on impressionistic data; the only works that I am aware of that deal with prosodic cues for information structure on an instrumental basis are again concerned only with Cuzco Quechua. O'Rourke (2005: 61–68) tentatively proposes that there are no prosodic cues to information focus[85] in declaratives, neither in terms of peak

85 Her criterion for whether a constituent is in focus is that it is non-initial and marked by the evidential suffix *–m/mi*, after a proposal by Muysken (1995) that the evidentials *–mi/-chi/-shi* mark focus on the constituent they attach to (in initial position he states that they can also have scope over the entire clause). It seems however, that the relationship between the position of focus and the presence of the evidentials can at most be characterized in such a way that if there is an evidential in a sentence, the constituent it attaches to is focal, but not the other way round: this much is clear from the frequent attestation of sentences without any evidentials. The characterization in Weber (1989: 427–429) for Huallaga Huanuco Quechua (Quechua I, central Peru) is more cautious: he asserts that broadly, what parts of a sentence are thematic (i.e. topical), and which are rhematic (i.e. focal), can be determined from the interplay between the distribution of the "topic marker" *–qa*, the evidentials, and the position of the verb. Optional *–qa*-marked initial constituents are thematic, followed by a rhematic part which may contain constituents marked with the evidentials, and then the verb, followed optionally by further *–qa*-marked thematic constituents. He explicitly warns against simply identifying the evidential-marked constituent as the last or first rhematic one and also attests sentences with more than one evidential, which is ungrammatical in Muysken (1995: 381–382). This suggests a relationship (perhaps dependent on the language variety) between focus position and evidential marking that is quite similar to the distributional relationship be-alignment nor of postfocal downstep or deaccentuation, but stresses that further research is needed because her analysis is based on only a small number of individual examples. O'Rourke (2009) somewhat qualifies this assessment. She proposes a regular LH* pitch accent[86] on stressed syllables (taken by her to be the penult in Cuzco Quechua) in declaratives, combined with iP-initial L- and optional iP-final H- in non-final, and iP-final L-, in utterance-final iPs (O'Rourke 2009: 308–309). Overall, peaks are aligned within the tonic syllable and on average show downstep across the utterance, but based again on the observation of individual utterances, this downstep pattern is said to be overruled by the focal constituent (again identified via the presence of *–n/mi* marking) having the highest peak in the utterance, suggesting highest prominence (O'Rourke 2009: 302–304, 307–308). Again, no evidence for postfocal deaccentuation or compression is found.

I want to conclude the section on prosodic cues and information structure by discussing in some more depth Muntendam & Torreira (2016), a study that to date is unique not only in that it experimentally investigates Quechua prosody under different information structural conditions, but also in providing comparative data from two Spanish varieties. Because their findings are so relevant for the present work, I will also address some important methodological shortcomings, but it should be clear that their pioneering contribution is extremely valuable to the aims of my own study because in many aspects it represents the only comparable work on at least related varieties. Muntendam & Torreira (2016) investigate the effect of information structure on prosody in Cuzco Quechua and Cuzco Spanish by the same bilingual speakers (16 speakers), and "Peninsular" Spanish (7 from Castile and Leon, 1 from Murcia), using a question-based task to elicit short utterances mainly of a noun phrase made up of an adjective and a noun in contexts of broad focus, contrastive focus on the noun, and contrastive focus on the adjective. Speakers participated in pairs and were given a stack of cards containing preformulated questions and coloured objects from which answers were to be built. As can be seen from (33)b) and c), "contrastive focus" on the adjective and the noun were elicited by asking a polar question in which the other element in the noun phrase was given

tween focus and prosodic cues argued for here. Inarguably, the evidentials and other morphosyntactic devices in Quechua interacting with focus all convey an additional (paradigmatic) meaning that is orthogonal to their use as focus markers, see e.g. Faller (2002, 2003, 2014); Behrens (2012); Bendezú Araujo (2021) for accounts.

86 Actually she proposes an L*H for pitch accents in prefinal, and an LH* for those in final position in the utterance, based on peak alignment data. However, she also admits to the possibility that since peaks are nearly always realized within the stressed syllable in all positions, the pitch accent could also be taken to be LH* in general, and the alignment difference as phonetic (O'Rourke 2009: 309, note 16). The classification is clearly tentative and awaits further research.

and correct, while the adjective or the noun, respectively, did not correspond to the object shown on the card for the answering speaker and were thus intended to be corrected[87] in the elicited assertions. Broad focus was elicited by asking a wh-question about what object the answering speaker had on their card ((33)a)).

(33) Example elicitation dialogues in Spanish,[88] from Muntendam & Torreira (2016: 75)

a. Q: ¿Qué tienes?
"What have you got?"
A: Tengo una luna morada
"I have a purple moon"

b. Q: ¿Tienes una flor morada?
"Have you got a purple flower?"
A: No, tengo una luna morada
"No, I have a purple moon"

c. Q: ¿Tienes una luna negra?
"Have you got a black moon?"
A: No, tengo una luna morada
"No, I have a purple moon"

For each language variety, Muntendam & Torreira (2016) identify several attested intonational contours in the responses and their frequency of occurrence[89] across

87 Note that according to Farkas & Bruce (2010: 96), polar questions such as those asked in (33) b) and c) are unbiased with regards to their response (whether the proposition asked for is true or not). The answers elicited here thus constitute reversals, which are different from denials in the sense that no commitment has been made in the provocation. It could thus be argued that the difference between the "corrected" and the "uncorrected" element in the responses is simply one of relative givenness, which is very variably cued prosodically across languages (Cruttenden 2006; Calhoun 2010b). However, it could also be considered that these responses form a particular subclass of reversals that might be called partial reversals, in analogy to the partial denials that are described as possible responses to assertions (Farkas & Bruce 2010: 99–100).

88 Example elicitation dialogues for Cuzco Quechua are not provided in Muntendam & Torreira (2016). They would have been interesting with regards to possible differences in the placement of the polar question marker *–chu* (*-ku* in Conchucos Quechua) which can attach to different constituents and possibly be used to mark the question focus (specify the QUD with respect to which constituent it asks about, cf. O'Rourke 2005: 183). As far as I can tell, the intonational form of the elicitation question, produced by the participants themselves, was not controlled for. This applies also to the Spanish part of the experiment, for which it has been shown that the intonational form of the question can have a significant influence on the form of the response (Fliessbach 2023).

89 Originally, speakers produced 20 target utterances per condition, yielding 960 utterances in bilingual Cuzco Quechua and Spanish each, and 480 utterances in Peninsular Spanish (Muntendthe experimental conditions, reproduced together in Figure 19. For Peninsular Spanish, the results fall well in line with previous findings as well as with the theory of an indirect relationship between information structure and prosodic cues.

As can be seen from Figure 19, all three attested contours[90] for Peninsular Spanish occur in all three experimental conditions. Only tendencies can be made out in both directions of association: neither is an experimental condition exclusively linked to a single contour, nor an observed contour exclusively occurring in only one condition (with the (c) contour in the ContrN condition however coming closest). The same observation can be made correspondingly for the other two language varieties, making a direct encoding of information structure via prosody unlikely (Muntendam & Torreira 2016: 78, 84–85). The three contours identified

am & Torreira 2016: 75). For all languages, utterances including hesitations or longer pauses were excluded. Only for the Quechua data, all utterances containing case markers and all utterances in which the target NP containing the adjective and the noun were not utterance-final were also excluded, because "this is the natural position for the corresponding NPs in Spanish" (Muntendam & Torreira 2016: 76–77). For Peninsular and Cuzco Spanish, the elimination resulted in 396 and 600 remaining utterances for analysis respectively, but in the case of Quechua, only 227 utterances were included in the final analysis, a reduction by more than 75%. Assuming that a similar amount of elimination due to hesitation took place in the Quechua data and the Cuzco Spanish data, leaving some 600 utterances, then still more than 62% of the remaining data must have been eliminated only due to the presence of case markers or non-NP-final word order. It is difficult to assess how likely the utterances were to include case markers, because no Quechua example sentences are given in the paper and several constructions are possible in Quechua to convey the content of the examples in (33), some of which would not regularly have to contain any case markers at all. On the other hand, it can be assumed that the target utterances in Quechua were as simple as the Spanish ones, effectively consisting of only a verb and the target NP. That means that whether the target NP is final comes down to a binary option, and potentially the excluded occurrences represent the unmarked majority option, in which the verb is final. This issue is not addressed at all in the paper, but eliminating the word order option that represents the majority of cases should be a cause for concern. As it stands, there are already a number of studies about word order in Quechua, including the Cuzco variety, that all point to there existing a relation between word order and information structure, and a trend for verb-finality (Wölck 1972; Weber 1989; Muntendam 2010; Sánchez 2010). In the light of these assessments, it seems likely that the order V-NP is itself a marked word order and/or a relevant cue to information structure. The decision to eliminate it from the analysis perhaps thus means that data from a marked word order in Quechua is compared to that from an unmarked word order in Spanish. If this is the case, it could cast doubts on both the comparability of the findings on Quechua to those on the Spanish varieties (internal validity) and the possibility for generalization of the Quechua analysis beyond this sample of data (external validity).

90 Muntendam & Torreira (2016: 76) do not provide any details on how the contours were identified, and how ambiguous cases and disagreements in annotation were handled. They do assert, however, that the ToBI-style pitch accent and boundary tone labels "only serve [. . .] the practical purpose of distinguishing the contours" (Muntendam & Torreira 2016: 77, note 5) in their data, presumably as opposed to representing a fragment of grammatical analysis.

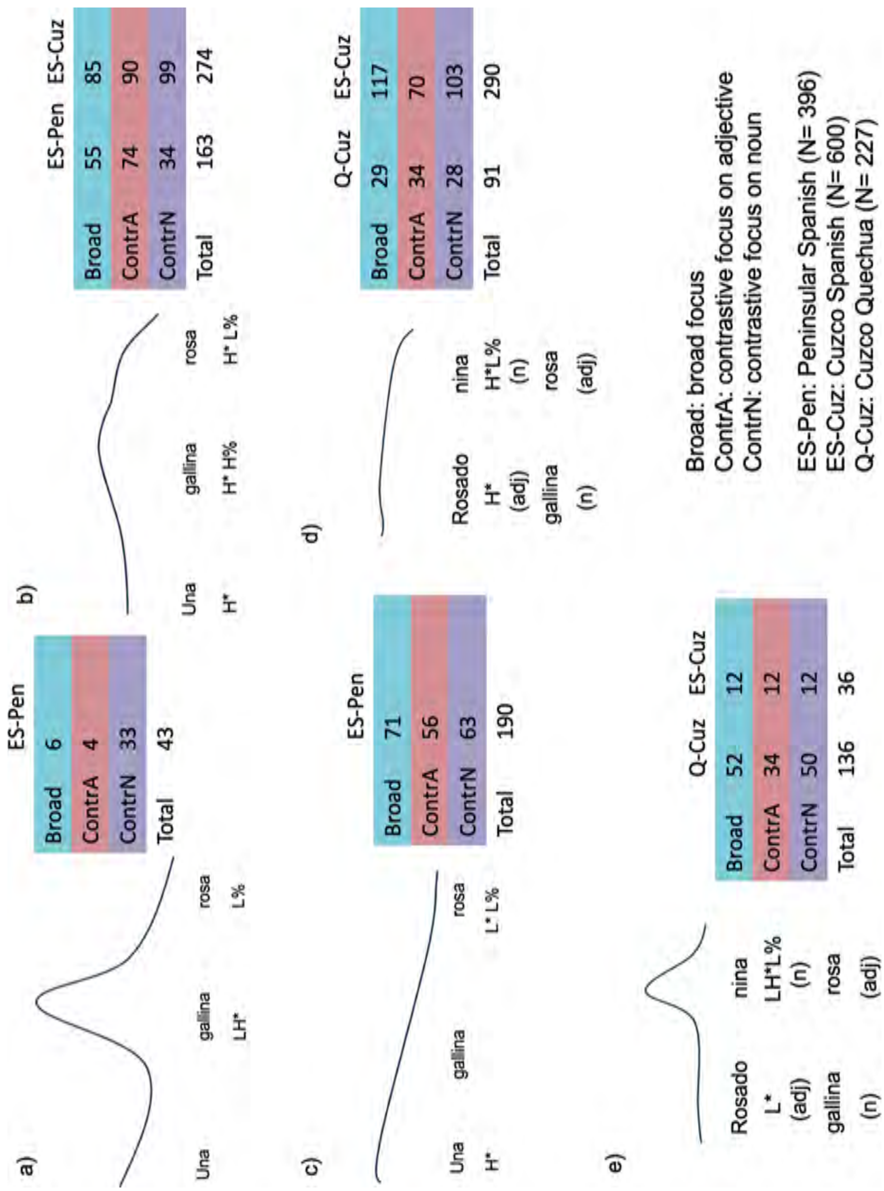


Figure 19: Schematized versions of attested intonational contours for noun phrases consisting of a noun and an adjective, *una gallina rosa* "a pink chicken"/ *rosado nina* "pink fire", from Peninsular Spanish (a, b, c), Cuzco bilingual Quechua (d, e), and Cuzco bilingual Spanish (b, d, e). Note the reversed word order between the two languages. Tables give number of occurrences of the identified intonational contours, per language and experimental condition, all adapted from Muntendam & Torreira (2016: 77–78, 81, 83).

for Peninsular Spanish are familiar and in broad agreement with the literature already discussed above. The (c) contour in Figure 19 is easily identified as the typical Spanish declarative nuclear contour in which the last word does not form an observable pitch peak at all, leading to an interpretation of final prominence by virtue of the expectational bias in that direction. Its status as an unmarked default is here confirmed by the fact that it is not only the most frequently used contour in the broad focus condition, but also in the condition with contrastive focus on the (prefinal) noun, and the most frequent overall (48% of all cases). Both other contours are arguably prosodically more complex because, if we equate the (a)-contour with the LH* L- contour from Gabriel (2007); Vanrell & Fernández Soriano (2018) and others, they involve an additional division of the IP into two iPs. The (b)-contour with the high prefocal iP-boundary tone is significantly more frequent in the condition with contrastive focus on the final adjective than the (c)-contour (Muntendam & Torreira 2016: 78), and the (a)-contour is clearly used most frequently (but not exclusively) in the condition with contrastive focus on the prefinal noun. Crucially, the relatively fewer occurrences of the (a)-contour in the condition of contrastive focus on the prefinal noun (33/130 occurrences, or 25%, in the ContrN condition) compared to the (b)-contour in the condition of contrastive focus on the final adjective (74/134 occurrences, or 55%, in the ContrA condition) cannot be explained via pragmatic or information structural markedness or degrees of informativity. In both conditions, focus is "contrastive" in the same way, but the (a)-contour is additionally marked in that the nuclear accent is not rightmost in the IP, i.e. the differential in the rate of occurrence can be explained when we make reference to the prosodic and metrical structure independently of the experimental conditions standing in for information structure. The expectation-based default of rightmost highest prominence seems strong enough to prevent a contour from being realized that would align focus position with a prefinal highest prominence in the majority of cases. The flipside is that this naturally makes an occurrence of such a contour that much more markedly informative, as already suggested in the discussion of Calhoun et al. (2018). These results thus provide further evidence for the independent relevance of prosodic and metrical structure, and the proposal about the nature of its relationship to both phonetic cues and information structure, which should be characterized as indirect: probabilistic, which is to say that from the perspective of an individual event, it is only possible to say that a certain information structural configuration will result in one of several realizations with a certain likelihood; and distributional, which is to say that only when observing many events can an association between an information structural configuration and a certain realization be made out, and intervening factors identified, in the form of trends in the distribution (cf. Calhoun 2010b).

Broadly, the same generalizations can be drawn from the Cuzco Quechua results in Figure 19: both contours are used in all three conditions, but there is a preference for the contour with clearer rightmost prominence (e) in the conditions in which focus is broad or rightmost (on the noun). Effectively, there does seem to be a tendency to differentiate between prefinal and final prominence, but no additional differentiation corresponding to the information structural division of broad vs. narrow final focus, as achieved in Spanish through the preferential use of contour (c), without iP-phrasing within the IP, vs (b), where the given material is phrased off with a H-. An open question is whether this is due to Cuzco Quechua not differentiating between broad and final narrow focus, or because this phrasing option is not available, or not used to separate given from new material. The Quechua results are also interesting in further ways. Both contours, (d), and (e), are at odds with the analysis of Cuzco Quechua declarative intonation in O'Rourke (2009). There, only a bitonal rising pitch accent LH* (see note 85) is proposed, but here the contours include the two monotonal pitch accents L* and H*. In addition, the high beginning in the (d) contour is also incompatible with the analysis in O'Rourke (2009: 304–305, 308–309), where a low boundary tone L- is proposed to be initial in every phrase. This suggests either that one of the analyses is incorrect, or that the contours in Muntendam & Torreira (2016) are not full contours in that they only characterize partial phrases. The discrepancy cannot be entirely explained away by saying that O'Rourke (2005, 2009) does not cover cases of contrastive focus; the L- LH* LH* L% contours we would expect from her analysis also do not occur here in the broad focus condition. It should be noted that since low pitch accents are notoriously difficult to identify and not much is known about how obligatory pitch accentuation is in prefinal position in Cuzco Quechua, the analysis of the L* in the (e)-contour is especially worthy of future investigation.

Cuzco Spanish, finally, makes use of contours (b), (d), and (e), also in general supporting the hypothesis about an indirect, distributional relationship between information structure and prosody because all contours occur in all conditions, with the difference in preference for (d) in the broad focus condition vs the preference for (b) in the condition with contrastive focus on the final adjective found to be statistically significant, but not the difference between the two contrastive conditions (Muntendam & Torreira 2016: 82–83). That is to say, Cuzco Spanish seems to mainly make a difference between a simple IP and one in which given material is phrased off in a separate iP. Based on the presence of contours (d) and (e) in both Cuzco Quechua and Spanish, Muntendam & Torreira (2016: 86), claim that this is evidence for cross-linguistic influence from Quechua to Spanish and even that this influence is "unidirectional: Spanish adopts prosodic features from Quechua but not the other way around". I would argue that their own evidence does not support this assertion, and that more research on the intonational phonology of Cuzco Quechua, but also on neighbouring varieties of Quechua or Spanish, is needed before any conclusions about directionality of influence can be drawn. Firstly, given the lack of a full intonational analysis of the contours in Quechua, with the proclaimed use of the ToBI-like labels solely for "the practical purpose of distinguishing the contours" (Muntendam & Torreira 2016: 77, note 5) but no information provided on the criteria for contour identification, comparing Quechua contours with Spanish ones cannot go beyond establishing a superficial phonetic similarity. Secondly, even if that issue were settled, there are no grounds for claiming that the (d)- and (e)-contours are in any sense "original" in Cuzco Quechua, and only "adopted" in Cuzco Spanish. The only statement supported by the facts is that the Cuzco speakers seem to use two of the attested contours in both of the languages spoken by them, and one in only one of them. The two "shared" contours might just as well "originate" from their use of Spanish and have made its way into Quechua, or be a shared innovation of the speaker community transgressing language boundaries, in the way of "diasystematic constructions" (Höder 2014a, 2018) in prosody, since virtually nothing else is known about the prosody of neighbouring varieties, of Quechua or Spanish.

In sum, even though considerable issues remain, Muntendam & Torreira (2016) make important headway into the study of prosody and IS in Cuzco Quechua. It seems that some of the same conclusions on the relation between information structure and prosody can be drawn as for Spanish: it appears to be distributional, insofar as there is not a categorical mapping between a contour type and a focus type.[91] Whether it is also indirectly mediated via metrical structure remains yet to be seen. For Conchucos Quechua, no comparable study exists, and not much more than basic facts of its prosody are known. In section 6.1, I will develop an account of Huari Quechua prosodic and intonational structure based on quantitative and qualitative data that takes some elements from O'Rourke's (2009) analysis, especially her use of initial and final iP-boundary tones, but greatly reduces the role of word stress and pitch accents. I will furthermore lay out how this proposal relates to information structure in sections 6.2 and 6.4, and develop an OT-model of intonation that describes a prosodic variation space between the attested forms of Huari Spanish and Quechua in sections 5.3 and 6.3. The insights gained there might even help shed some light on the outstanding issues in Cuzco Quechua and Spanish.

91 Roessig (2021: 81–87) also comes to the conclusion that no one-to-one mapping between focus types and pitch accent types seems to exist in West Germanic languages. Evidence from other studies considered there indicates that the distribution of continuous parametres like relative peak alignment and tonal onglide interact with the distribution of categorical pitch accent types across focus types to ensure that they can still be recognized correctly in perception.

4 Refined research questions

Before moving on to the actual analyses of Huari Spanish and Quechua intonation, this interim chapter serves to refine and expand the research questions based on the background provided in the two previous chapters. In the introduction, three broad leading questions (1), (2) and (3) were given, here repeated as (34)-(36) with their subquestions.

In large parts, this is an exploratory study on the prosody of two undescribed varieties of Spanish and Quechua, so even broad answers to these two questions should fill a research gap. However, based on the preceding theoretical discussion, all three of them can be further refined. Specifically, we can ask the expanded version of (1), (34), of both Huari Spanish and Quechua.

(34) What are the relevant properties of the intonational systems of Huari Spanish and Huari Quechua?
   a. What is their tonal inventory? Are tones only edge-seeking or can some be identified as pitch accents associated with a metrically strong position as well? How many options are available at each position?
   b. What evidence is there for which levels of prosodic structure?
   c. How are tones distributed across the units of the prosodic structure?

In particular for Quechua, one very important question regarding word prosody needs answering, without which the other questions cannot fully be answered. In the previous chapter, we saw that the evidence for word stress in Ancash Quechua varieties is mixed at best (section 3.3.3), especially when adopting a definition of word stress as culminative and obligatory, following Hyman (2014). Based on this discussion, the following question can be asked:
   d. What evidence is there that pitch events in Huari Quechua are affected by a metrically strong position at the word level that falls under Hyman's definition of word stress? In particular, can stress-sensitive pitch events be disentangled from edge-seeking ones based on evidence from tonal alignment?

This is a question tailored to Huari Quechua, but it will be useful to compare Huari Quechua here to Huari Spanish, for which the assumption of word stress is much less controversial.

Regarding (2), it is in a sense itself a subquestion to (1), so its expanded questions (35) also can be brought to bear on answering the overall question which

https://doi.org/10.1515/9783111304595-004

prosodic properties are relevant. In section 3.7 the scope of this question was restricted in that mostly, only those kinds of discourse meanings that are collected under the header of information structure are in focus here. However, other kinds of discourse meanings and their cues also have to be taken into consideration to some extent in order to disentangle the cues for information structural meanings from them.

(35) How and what kinds of interactional/discourse meaning do they encode?
  a. To what extent are paradigmatic tone contrasts (different boundary tones and/or pitch accents, different nuclear configurations) used to encode discourse meanings?
  b. To what extent are syntagmatic tonal devices (phrasing via edge tones, deaccentuation/dephrasing etc.) used to encode discourse meanings?
  c. Are prosodic cues used on their own, or in conjunction with, or vicariously for, other cues to discourse meanings (word order, morphology, particles)?
  d. Can the role of tonal scaling be shown to be restricted to a “phonetic” scalar encoding of “emphasis”? Does it contribute to the paradigmatic tonal inventory, or is it used nonlocally to cue prosodic structure and discourse meaning, as described for a number of languages (cf. section 3.6)? What does this mean for whether the prosodic structure itself is recursive or not?

Along more theoretical lines, answering these questions will also shed light on a question that emerged in section 3.7.3, but whose complete answer this thesis cannot provide.
  e. Do the identified cues exhibit a direct or even biunivocal relation to the proposed categories of information structure and other discourse meanings, or is it more intermediate and distributional, like discussed in section 3.7.3? Does this have effects on the conception of these categories themselves, and what role does context play?

Finally, (3) can also be expanded, as (36).

(36) Which of these properties are specific to one language, and which are perhaps common to both?
  a. Based on the answers to (1)/(34), do the differences between Huari Spanish and Quechua occupy neatly definable positions along the typological dimensions laid out in section 3.4?
  b. Do their variation spaces overlap and how can this be shown?

c. Especially assuming that a considerable amount of variation will be found, what does this mean for prosodic typologies and their proposed objects, languages?

Not all of these questions will be answered to equal degrees of depth across both languages. (34)d is clearly more aimed at Huari Quechua. In contrast, (35)d will be more thoroughly explored for Huari Spanish than for Quechua. As the whole study employs various methods, the type of evidence brought to bear on each individual question also varies. Broadly speaking, sections 5.1.1 and 5.1.2, and 6.1 aim at answering the first group of questions (34) for Huari Spanish and Quechua, respectively, with 6.1.6 particularly concerned with answering (34)d via a quantified comparison of peak alignment in Huari Quechua and Spanish. Sections 5.1.3 and 5.2, and 6.2 and 6.4, respectively, seek to provide answers to the second group of questions (35) for Huari Spanish and Quechua, with 5.2 and 6.4 both also particularly concerned with (35)d, but using different methods. I'm using the Spanish and Quechua OT-analyses in sections 5.3 and 6.3, respectively, as well as the concluding section 7.4, to tackle the third group of questions interested in the interrelationship of the prosodic systems (36). However, since the overarching goal of the study is to provide descriptions of the prosodic systems of the two languages, not each question has its own section where it is answered. Rather, the answers to the individual questions will result from the analyses as a whole.

5 Huari Spanish

This chapter is concerned with the description and analysis of the main intonational phenomena of Huari Spanish. It gives an overview over how intonation in Huari Spanish varies with pragmatic factors like utterance type and information structure. In the first section (5.1), the description is mostly based on relatively simple utterances. It will be established that in the majority of Huari Spanish declarative utterances, each accentable word forms an LH* pitch accent, also in phrase-final position. The pitch peak that is formed occurs within the stressed syllable, also in prenuclear accents. Utterances conforming to this description will be said to make up the "main" intonational variant of Huari Spanish. Besides describing the nuclear configuration also for interrogatives and documenting some more marginal phenomena, an important contribution of this chapter is the description of a variant accentuation behaviour for declaratives, encountered in a number of utterances, in which a single right-aligned rising (LH) or rising-falling pitch event (LHL) takes place over a number of accentable words. This intonational variant will be called "phrase accentuation" in order to separate it from the "main" variant. Possible factors explaining its occurrence will be explored.

In the second part of the chapter (5.2), a particular set of complex utterances from *Elqud* consisting of two topics and two comments will be described. They will be argued to present evidence for pitch scaling used to cue a hierarchical prosodic structure nonlocally and for a recursive prosodic structure, providing an answer to research question (35)d regarding Huari Spanish. In addition, a sizeable number of utterances of this type displaying the "phrase accentuation" will be analysed separately and the resulting insights then lead to the final analysis in the third part of the chapter.

There, some of the observations from the preceding two parts are formalized to produce an OT-analysis that allows for an understanding of the different intonational variants as the instantiation of a cluster of values that variable prosodic properties can assume. This will lead to a conception of how they can be related to each other as well as to the variants described for Huari Quechua.

5.1 Simple utterances

In this first section on Huari Spanish, a number of intonational phenomena will be described based on relatively simple utterances from *Conc, Maptask, Cuento,* and *Elqud*. I begin with declaratives (5.1.1), move on to interrogatives (5.1.2), and then discuss meaning-related aspected of prosodic variation (5.1.3).

https://doi.org/10.1515/9783111304595-005

5.1.1 "Main variant" neutral declaratives

This section describes the most frequent intonational behaviour for declaratives, what I will call the "main" variant of declarative accentuation. I will first give a qualitative description with examples and then present quantification results over a subset of the data.

5.1.1.1 Introductory description

In the majority of declarative utterances in Huari Spanish, each or nearly each content word is accented on its stressed syllable (see (37)–(40) and the corresponding Figures 20–23). Peaks are formed on and within each accented syllable, both if they are final and if they are prefinal in the phrase or the utterance, independent of whether a high or a low boundary tone follows. In continuation rises (with a high boundary tone, annotated as H- or H% in ToBI), when the last accented syllable before the phrase boundary is not also the phrase-final syllable, the pitch movements of the pitch accent and the phrase-final rise can very often be clearly distinguished (cf. *grande* and the two first instances of *gigante* in Figure 22 and *olla* in Figure 23). Again in the overwhelming majority (see also next section), the peaks that are observed on accented syllables are all aligned well within the stressed syllable, also on pre-final and proparoxytone final words, unlike in peninsular varieties of Spanish, but similar to what has been reported for Cuzco and Lima Spanish by O'Rourke (2005). Final low boundary tones (L%) realize their targets immediately following the peak on the final accented syllable, even if the final word is a proparoxytone and thus in principle providing some space for the realization of the boundary tone (cf. Figure 23). The final boundary tone is mostly not deleted even in final oxytones, realizing a full low target (cf. Figure 21), although it might be truncated insofar as it does not reach the same level as other low targets after a very high final peak (cf. Figure 20). Peaks are preceded by clear troughs or valleys that often extend from directly after the last accented syllable and which are only eliminated under strong time pressure conditions. These low stretches are taken to be evidence of low tonal targets realizing an L tone, just as the peaks are high targets realizing a H tone. It seems that in the majority of utterances, the pitch peak in the stressed syllable is reached roughly in the middle of the vowel, so that at the end of the stressed syllable, pitch has already slightly fallen from the local maximum. The elbow marking the end of the trough preceding the peak usually occurs in the syllable before the stressed syllable, or at the latest right at the beginning of it. These observations hold if no other tonal event is encroaching too closely (less than two syllables away) upon the peak, and if no particular pragmatic conditions obtain that seem to effect a divergent target placement (see section 5.1.3.3 below) and if not another, more edge-oriented mode of accentuation is employed, either also for pragmatic reasons or due to interspeaker variation (see section 5.1.3.1). The general analysis for Spanish declaratives is therefore here that each tonal movement related to an accent, also in final position, consists of an LH tone sequence, associated as a pitch accent annotated in ToBI as LH*, with the high target aligned within the stressed syllable and the low target directly before it, creating low stretches to the left. This is exemplified on typical examples like (37)–(40) below:[92]

(37) ZR29_Cuento_ES_0083
des**pués** de *un* **ra**to **lle**ga **un** coli**brí** que le **di**ce al o**í**do que su **nie**ta le nece**si**ta
LH* LH* LH* LH* LH*H- LH* LH* LH* LH* L%
"after a while a hummingbird comes that tells him in his ear that his granddaughter needs him"

(38) TP03_ Cuent_ES_0513
tan **gran**de *co*mo el gi**gan**te que le di- di**ó** **mie**do al gi**gan**te y el gi**gan**te hu**yó**
LH* H- LH* H- LH* LH* LH* H- LH* LH* L%
"as big as the giant so that it gave the giant a fright and the giant fled"

(39) KP04_MT_ES_0799
por **e**so p(u)e(s) le **ha**go *un* **cír**culo *ha*cia a**ba**jo
LH* LH* LH* LH* L%
"right because of that I'm making a circle around it downwards"

(40) ZZ24_MT_ES_1269
de la **o**lla de**ba**jo del mur**cié**lago
LH* H- LH* LH* L%
"from the pot, under the bat"

92 Stressed and pitch accented syllables are given in bold, with the pitch accents aligned under them. Stressed but not accented syllables are given in italics. Indefinite articles (*un, un/a/o*) are treated as regularly accentable, i.e. stressed even if not pitch accented because in the data here they often do realize an identifiable pitch accent, unlike the definite articles (*el, la, los, las*), which are never accented or stressed. Cf. Quilis (1993: 390–395); Hualde (2009), where the same distinction is made.

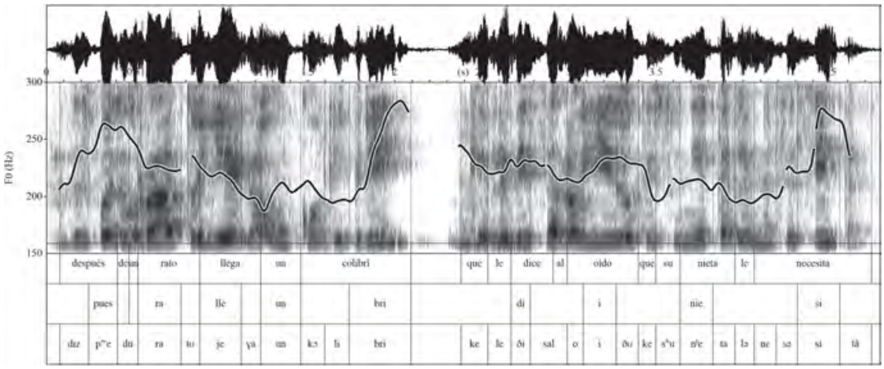


Figure 20[93]: ZR29_Cuento_ES_0083[94,95] (main accentuation declarative) Cf. (37).

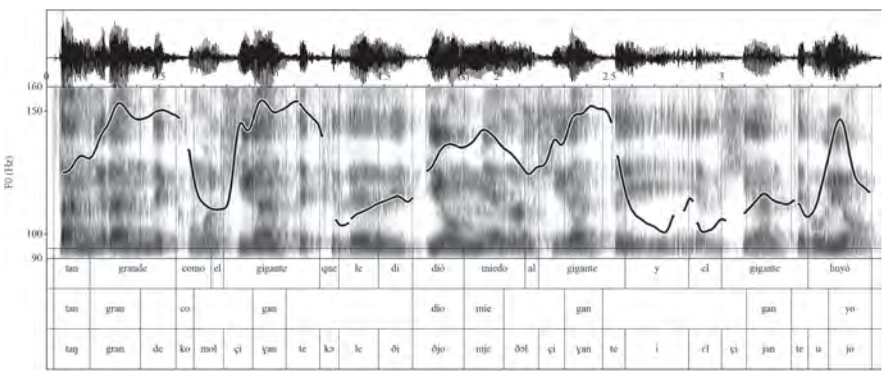


Figure 21: TP03_Cuent_ES_0513[96] (main accentuation declarative). Cf. (38).

The placement of the low tonal targets allows us to surmise both about the prosodic structure of accented and unaccented words and the spreading behaviour of the tones involved. In all of the examples discussed here, we see that the transition from a high target realizing the H tone in an LH* or a H- continuation rise to the upcoming low target realizing the leading tone of the following LH* suggests left-

93 Visualizations of Huari Spanish and Quechua examples in the figures consist minimally of a pitch track and spectrogramme, with a time-aligned transcription. For the Huari Spanish examples, the transcription is orthographic and in at least two tiers. The first tier is segmented according to word boundaries, the second gives the boundaries of stressed syllables.

94 The pitch object for this picture was created with a voicing threshold of 0.45 and some disturbances were manually removed.

95 https://osf.io/vyxtp/

96 https://osf.io/ypmb4/

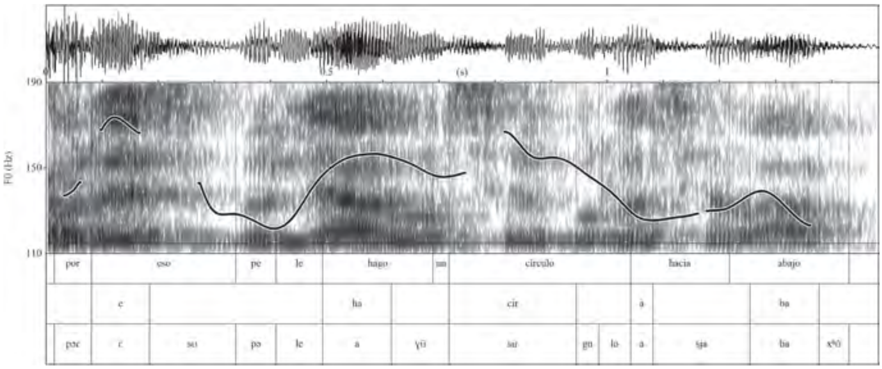


Figure 22: KP04_MT_ES_0799[97] (main accentuation declarative). Cf. (39).

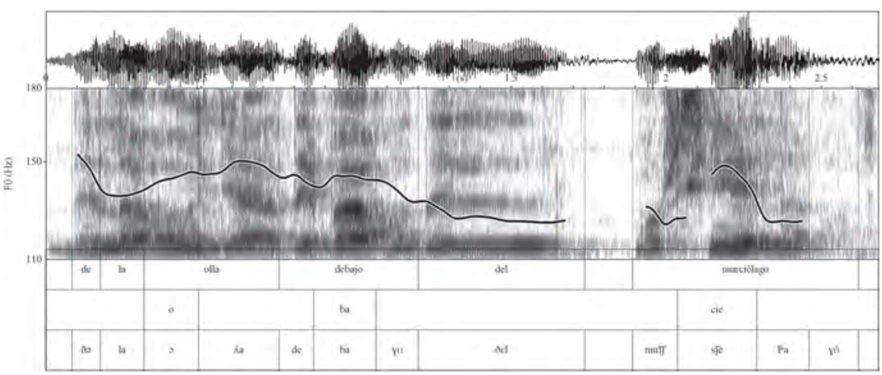


Figure 23: ZZ24_MT_ES_1269[98] (main accentuation declarative). Cf. (40).

ward spreading or multiple alignment of this low target. This is especially evident after the continuation rises in Figure 21, which are followed by an abrupt reset to low, even though several syllables intervene between the next accentuation event, in principle leaving time to reach the next low target more gradually. The same leftward encroachment of the low stretch is also seen in Figure 23 at the end, after the stressed syllable of the final proparoxytone *murciélago*. Since this final L can be taken to belong to the end of the prosodic constituent, the IP, of which it is the boundary tone, its relatively early realization should be seen as multiple alignment or leftward spreading of the low target. This holds also for the behaviour of the L tone in the LH* pitch accents: as leading tones they belong to their H*s, but the

97 https://osf.io/87j56/
98 https://osf.io/hv4ab/

low stretch formed by them is evidence that they seek to align in both directions. They often form an elbow before the rise to the high target immediately before the stressed syllable. The question remains how far the low target extends leftwards. The difference between Figures 22 and 23 is helpful here. While in Figure 23, the target of the final L% is realized immediately following the stressed syllable of the final *murciélago*, in Figure 22 there is a much gentler drop after the stressed syllable on the equally proparoxytone *círculo*, reaching the low target only at the beginning of the next word, the unaccented *hacia*. These and similar cases in the data suggest that the low stretch preceding a pitch accent is blocked both by a preceding high tone and the boundary of a prosodic constituent of the same level or higher than the prosodic word. For this proposal to work, it has to be assumed that unaccented words form a prosodic word together with the next accented word to their right, but not to their left.

I call the accentuation behaviour whose properties were just described the "main" variant of Huari Spanish, both because of its frequency (see next section) and its similarity to other varieties of Spanish. In regularly realizing a rising accent also on the final accented word in an utterance, the data here not only resemble those described for other varieties of Peruvian Spanish, but also for central Mexican Spanish (De-la-Mota et al. 2010), where a so-called "cirumflex contour", similar to what is described here, also most frequently forms the nuclear pitch accent in contexts of "broad focus". This is in contrast to findings on Madrid Spanish, where "broad focus" contexts have been found to often correlate with a low nuclear configuration (L*L%; cf. Estebas Vilaplana & Prieto 2010; Hualde & Prieto 2015). For Mexican Spanish, the L*L% configuration is reported to be only a minor variant, and in our data, as the next section will show, it is virtually nonexistent in such contexts. Another difference concerns peak placement. According to De-la-Mota et al. (2010: 324), the peak in the final accented word is aligned at the end of the stressed syllable in Mexican Spanish, whereas here it is aligned in the middle of the stressed vowel. Mexican Spanish also regularly has the delayed prenuclear pitch accents transcribed as L+<H*, which are exceedingly rare here. Thus while the "main" variant is similar to other intonational varieties of Spanish in its very regular pitch accentuation of accentable words, it also differs in several aspects from them.

5.1.1.2 Quantitative results on accentuation in data from Conc

This section provides a quantified perspective on the frequency of the central components of the "main" variant, i.e. that each accentable syllable realizes an LH* pitch accent whose peak is aligned within the stressed syllable. It is based on a subset of the Huari Spanish data. Annotations and measurements were made on all content and all polysyllabic function words from seven corpora from the *Conc* game, by the speakers TP03 & KP04, QZ13 & OZ14, SG15 & QF16, AZ23 & ZZ24, ZR29 & HA30, XU31 & OA32, and XQ33 & LC34, to provide quantifiable results about how often words are pitch accented and the frequency of the LH* pitch accent. These *Conc* corpora consist almost exclusively of declaratives. Content words were taken to be nouns, adjectives, adverbs, and verbs (except the two copulas *ser* and *estar*); function words were all others (articles, prepositions, demonstratives), with the token majority of polysyllabic function words in *Conc* consisting of deictic expressions such as *ahí*, *allá*, *acá* etc. The two copular verbs s*er* and *estar* were taken to be function words, so only polysyllabic forms of them were considered. Apart from monosyllabic function words, words were also excluded due to noise or when their pitch track was otherwise very fragmented. Words were counted as being accented with a pitch peak when the highest measured pitch in the sonorant part of the stressed (tonic) syllable was at least 7 Hz higher[99] than the lowest measured pitch in the sonorant part of the rhyme of the pretonic, or in the sonorant part of the posttonic. Only if the pitch range difference to the pretonic obtained was the pitch accent identified as LH*. If no pitch difference >7 Hz was found, the word was counted as unaccented. Table 6 gives the counts for accented vs. unaccented words as just described for all speakers, sorted according to word type (content/function). The two right columns give accentuation ratios, as words per pitch accent in the penultimate column, and as percentage of accented words among all considered words in the final column.

Table 6: Pitch accentuation counts in seven Spanish *Conc* corpora, sorted according to word type.

Word type	Accented	Unaccented	All	ratio words/accent	% accented words
Content	**391**	**46**	**437**	**1.12**	**89.5**
Function	**66**	**38**	**104**	**1.58**	**63.5**
All	**457**	**84**	**541**	**1.18**	**84.5**

The first result presented in this table is that on average, nearly 90% of all content words are pitch accented, or that an accent occurs once every 1.12 content words. This is in broad agreement with previous findings about Spanish accentuation which state

99 This threshold was chosen for comparability to the results in O'Rourke (2005: 62, note 10; 76, note 6,7) where it is used as a compromise between the results regarding psychoacoustic measurements of *just noticeable differences* in f0 in Klatt (1973) and Pierrehumbert (1979). The discussion in Fastl & Zwicker (2007: 182–188) suggests that it might be a slightly high threshold for the range of 75–700 Hz, the f0 range of the majority of human speech. Their data is however based on perception experiments with artificial sounds, not actual conversations with normal levels of background noise. The same 7 Hz threshold is also chosen in Rao (2009).

that nearly every content word is pitch accented.[100] The overall accentuation percentage of nearly 85% of all words considered is comparable to the 77% given in Rao (2009: 15) for spontaneous speech in Barcelona Spanish. The table also attests to a pronounced difference in accentuation between content and function words, with function words more than 25% less frequently accented than content words. A X^2-test was done on the cells counting accentuation for all speakers (shaded in grey in the table), with its result suggesting that the difference in word type is indeed highly significantly associated with the observed difference in accentuation (Pearson's X^2 (1) = 43.338, $p < 0.001$). That function words are less frequently accented than content words is also a result broadly in keeping with the literature on Spanish prosody (cf. Hualde 2009; Rao 2009).

Table 7: Pitch accentuation counts in seven Spanish *Conc* corpora, sorted according to whether words occurred in isolation or as part of phrases containing several words and word type.

Multi-word phrases					
words per phrase	**Accented words**	**Unacc. (of which function words)**	**All**	**ratio words/ accent**	**% accented words**
4-word-phrase	18	10 (5)	28	1.55	64.3
3-word-phrase	68	13 (6)	81	1.19	84
2-word-phrase	196	32 (16)	228	1.16	86
All words in multi-word phrases together					
Word type	**Accented**	**Unaccented**	**All**	**ratio words/ accent**	**% accented words**
Content	236	28	264	1.12	89.4
Function	46	27	73	1.59	63.0
All	282	55	337	1.20	83.7
Single-word phrases					
Word type	**Accented**	**Unaccented**	**All**	**ratio words/ accent**	**% accented words**
Content	155	18	173	1.12	89.6
Function	20	11	31	1.55	64.5
All	175	29	204	1.17	85.8

Table 7 explores a further possible factor influencing accentuation, namely whether words are more likely to be accented depending on their realization in isolation or as

100 The rate of 1.12 content words per pitch accent can be compared to the ratio of 1.27 content words per phonological phrase (defined as having a single identifiable pitch contour) found for Quechua in section 6.2.3.2 and based on a sample of similar size and type, consisting of all nominal sequences from the seven *Conc* games in Quechua by the same speakers considered here, plus one *Maptask* and one *Cuento* corpus.

part of a phrase containing several (>1) words. Torreira et al. (2014) state that words in phrase-medial position in Madrid Spanish are more likely to lack pitch accentuation than phrase-peripheral words. This would mean that on average, words in multi-word phrases should be less frequently accented than words realized alone, which are always peripheral. Words were counted as being together in a phrase if no discernible disjuncture was perceived amongst them, e.g. a hesitation, short break, or intonational boundary movement, corresponding roughly to break index level 2 or 3 of Sp_ToBI (cf. Aguilar et al. 2009). Monosyllabic functional elements such as articles or prepositional clitics were again disregarded, i.e. a multi-word phrase contains at least two words that are either content words or polysyllabic function words. The results in Table 7 indicate that the ratio of accentuation is not different overall between words in multi-word phrases as opposed to words produced in isolation or only together with clitics. Only for 4-word phrases does there seem to be a stronger tendency for words to be unaccented, but this result is statistically not fully conclusive and would need more data from words in 4-word phrases to be corroborated.[101] The results of a X^2-test on whether being in a multi-word phrase (without internal differentiation) vs being in a single-word phrase is associated with a difference in accentuation turned out not to be significant (Pearson's $X^2(1)= 0.43$, $p=0.51$), as could be expected from the very similar accentuation ratios in the table. Overall, this suggests that if an effect of being phrase-medial exists for the likelihood of a word to be accented, it seems too weak to emerge outside of four-word phrases or at all in this sample.

In answer to the question about the frequency of LH* as pitch accent and what other variants were encountered (see Table 8), 126 or 23.3% of all words were identified as not having an LH* pitch accent. Those include all words counted as unaccented. Of the 457 words counted as accented (with an identifiable peak), 42 (9.2% of accented words) were identified as having a different pitch accent than LH*. Of those, 26 are cases where the word in question occurred directly in phrase-initial position with the initial syllablle also being the tonic and where no rise leading up to the peak was found, so that the pitch accent was classed as H* instead of LH* (cf. OA32_Conc_ES_0298, the right image in Figure 24). With these cases it is probably the prosodic context that is responsible for this realization, i.e. the lack of pretonic material on which the rise can be realized leads to an increased likelihood of rise truncation. The context did not categorically lead to this realization, as OA32_Conc_ES_1549 (left

101 A X^2-test was done to investigate association between accentedness in phrases of the four different lengths, yielding a significant result (Pearson's $X^2(3)= 9.37$, $p = 0.02$), but expected counts in the cell of unaccented words in 4-word phrases were below 5, rendering the test somewhat unreliable (cf. Field et al. 2012: 818), especially since this was also the only cell in which the standardized residual was greater than $|1.96|$, suggesting it mainly contributed to the test being significant. A Fisher's exact test (two-sided, p=0.04) was just about significant.

Table 8: Words not counted as having an LH* pitch accent in seven Spanish *Conc* corpora.

	Accented (percent of all words)	Unaccented (percent of all words)	All (percent of all words)
Words not counted as having an LH* pitch accent	42 (7.8)	84 (15.5)	126 (23.3)
– *Of which* H* without leading rise on words in phrase-initial position with initial tonic	26 (4.8)	0 (0)	26 (4.8)
– *Of which* H tone part of preceding plateau (H*L/HL*)	14 (2.6)	24 (4.4)	38 (7)
– *Of which* only H% phrasal boundary	1 (0.2)	11 (2)	12 (2.2)
– *Of which* L*L%	0 (0)	2 (0.4)	2 (0.4)
– *Of which* others	1 (0.2)	47 (8.7)	48 (8.9)

image in Figure 24) with a more fully pronounced rise shows, but conversely, the truncated realization was only found in this context. Since the discourse context for the more LH*-like realizations and the H*-like realizations is the same, I assume that the realization without a rise is here only a prosodically conditioned truncated variant of the LH* pitch accent. Consequently, counting them together, a remaining 16 of 457 accented words (3.5%) were identified as having a different pitch accent than LH*. In particular, words in phrase-final position also realized the LH* pitch accent most frequently, and what could be identified as the nuclear configuration L* L%, familiar from the literature on many other Spanish varieties, was only found twice. Equally, the delayed rising accent L+<H*, frequently attested on prenuclear words in peninsular Spanish varieties, was only identified once here, and only tentatively. In the vast majority here, peaks of rising pitch accents (those identified as LH*) were thus found to be realized within the tonic syllable (a result also confirmed on another dataset in section 6.1.6), in agreement with what O'Rourke (2005) reports for Cuzco Spanish.

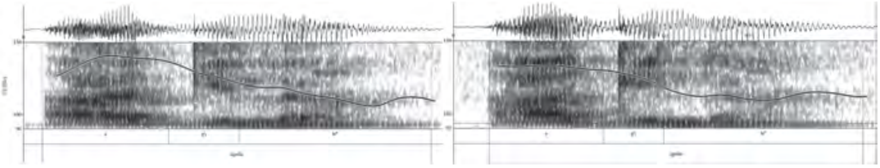


Figure 24: OA32_Conc_ES_1549[102] (left) and 0298[103] (right) (*águila* 'eagle').

102 https://osf.io/z8vkq/
103 https://osf.io/fxujy/

In the other cases where tonal movement on the tonic could not be identified as realizing an LH* pitch accent (in words both with and without peak), apart from simply flat realizations that make up a majority of the words counted as unaccented, two particular phenomena were observed. On the one hand, the rise expected on the tonic sometimes occurred only posttonically, in phrase-final words (12 cases). Instead of classing them as pitch accents with a delayed peak (L+<H* or L*H), I suggest that these cases look most likely to exhibit only boundary movement, i.e. H- or H%, but not a pitch accent. This is because they only occurred phrase-finally and because in proparoxytones, not only the tonic, but also the subsequent posttonic syllable was found to be low, with the rise taking place only on the (phrase-)final syllable, as in XQ33_Conc_ES_0966 (Figure 25). Because the tonic was low in most of these, they were counted as unaccented (11 of 12). These cases are similar to what is described as the "phrase accentuation" variant in section 5.1.3.1.

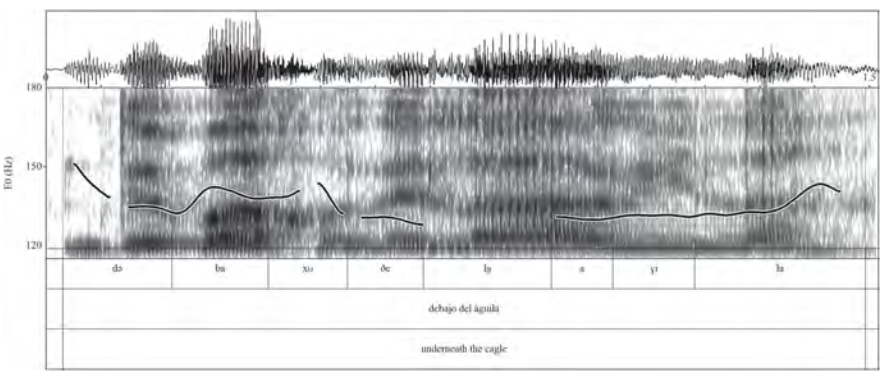


Figure 25: XQ33_Conc_ES_0966[104] (*debajo del águila* 'underneath the eagle').

On the other hand, in some cases the elevated pitch expected on the tonic was found already on preceding syllables. The tonic was then either also realized with high pitch, followed by a fall on or into the posttonic, or the fall already took place on the tonic itself, or it was even realized largely low after the fall from the pretonic.

I'm providing several examples (in Figures 26–30) to demonstrate the variability of the observed phenomenon. Perceptually, some of these sound like stress has shifted towards the pretonic, or even towards the preceding indefinite article in some cases. What they have in common is the lack of low targets on the pretonic, substituted instead with a high plateau-like realization, and that the tonic, instead of being a location for a pitch peak, seems to instead serve as a landmark on

104 https://osf.io/4685c/

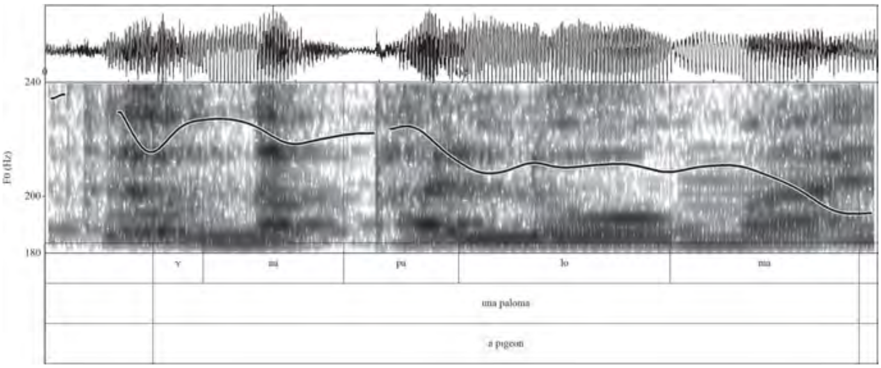


Figure 26: ZR29_Conc_ES_1065[105] (*una paloma* ‘a pigeon’).

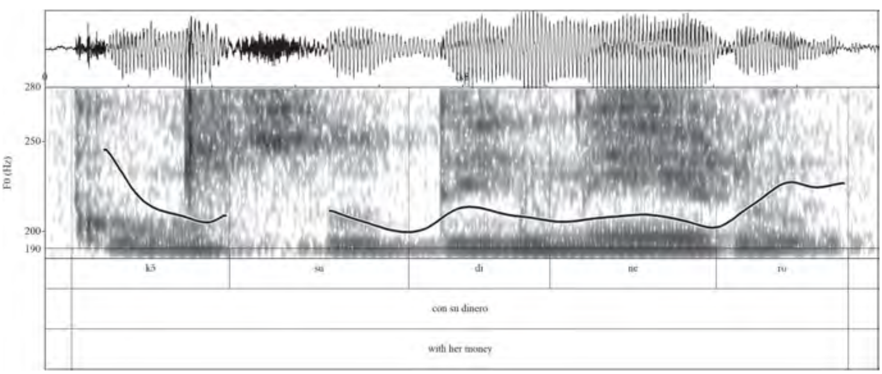


Figure 27: HA30_Conc_ES_0336[106] (*con su dinero* ‘with her money’).

which this high plateau ends. In those examples that include an indefinite article (Figures 26, 28, 29), it could be argued that this is merely a result of undershoot of the low tonal targets after a pitch accent on the indefinite article, in combination with perhaps an H*L or even HL* pitch accent on the tonic of the content word, but such an analysis, apart from leaving unexplained why an H*L or HL* pitch accent should occur in the same discourse contexts in which LH* otherwise occurs, is clearly insufficient for AZ23_Conc_ES_1851 (Figure 30). There, two pretonic syllables precede the tonic on the content word *millonario* “millionaire” and form a plateau that is as high as the highest value on the tonic, on which the beginning of the fall occurs. Because of examples like this one (cf. also Figure 27), instead of an

105 https://osf.io/r73wd/
106 https://osf.io/g5fs8/

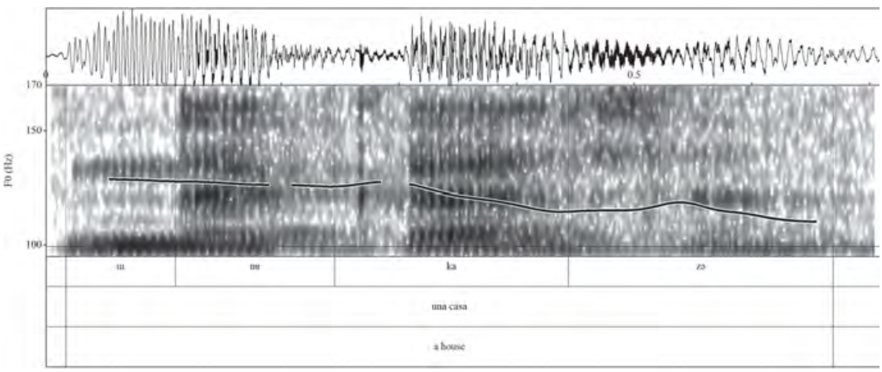


Figure 28: OZ14_Conc_ES_2242[107] (*una casa* 'a house').

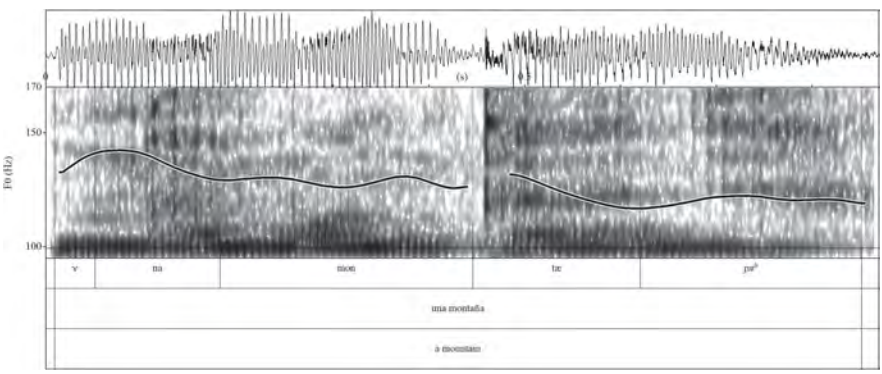


Figure 29: QZ13_Conc_ES_1567[108] (*una montaña* 'a mountain').

analysis e.g. in terms of stress shift, I would like to suggest that at least some of these cases are best understood as reflecting a prosodic configuration also encountered in the Quechua data (cf. section 6.1.2), where plateau-like realizations are common that extend from the initial boundary of a word to either the penult or the end of the phrase, and in which the tonic syllable serves at most for anchoring a tonal transition, here from H to L. In the Spanish *Conc* data discussed here, this type of realization was found on 38 of all 541 words (7%). Depending on where the fall took place and whether the tonic was still higher than the posttonic, representable broadly in ToBI as the difference between H*L and HL*, these cases were counted

107 https://osf.io/bhf3e/
108 https://osf.io/tr5j6/

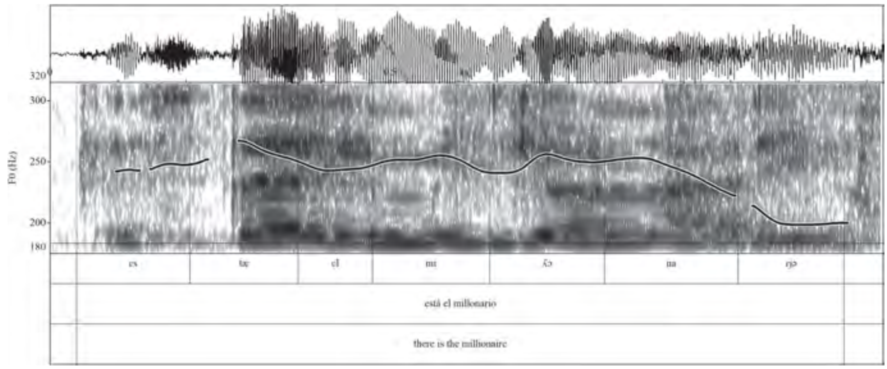


Figure 30: AZ23_Conc_ES_1851[109] (*está el millonario* '(there) is the millionaire').

as either accented (14 of 38) or unaccented (24 of 38). Table 8 sums up the counts of the categories discussed in this section.

The variable tendency for creating plateau-like realizations was also observed in phrase-final words with pitch ending high, i.e. presumably delimited by a high boundary tone. In a number of such cases, pitch was suspended at the same high level it had reached on the pitch accent (LH*) realized on the tonic until the end of the word. The resulting high plateau-like realization also extends to the intervening syllable in proparoxytones, as examples like LC34_Conc_ES_1298 (above in Figure 31) show. For comparison, consider the examples in Figure 32 without such a plateau-like realization.

Here, pitch drops again or stays level after the pitch accent on the tonic (even in the paroxytonic *naranja*), before then realizing a separate phrase-final rise that is also usually scaled higher than the pitch accent. The variant with the plateau-like realization is preferred in the seven *Conc* corpora by the speakers QZ13, OZ14, AZ23, ZZ24, XU31, OA32, and XQ33, who use it almost exclusively, while TP03, KP04, SG15, QF16, ZR29, HA30, and LC34 mostly prefer the realization with a separate rise (in LC34's case, the example given in Figure 31 is the only time he uses the plateau-like realization in *Conc*). No difference in discourse contexts could be made out that would differentiate the two variants functionally, they all seem to be cases of continuation rises (in most cases) or occasionally uncertainty. The plateau-like variant is quite similar to the realization with sustained pitch that Gabriel et al. (2011) describe as one possibility for realizing intermediate phrases with high boundary tones (H-) in Porteño Spanish. The one with a separate rise is comparable

109 https://osf.io/3w6sb/

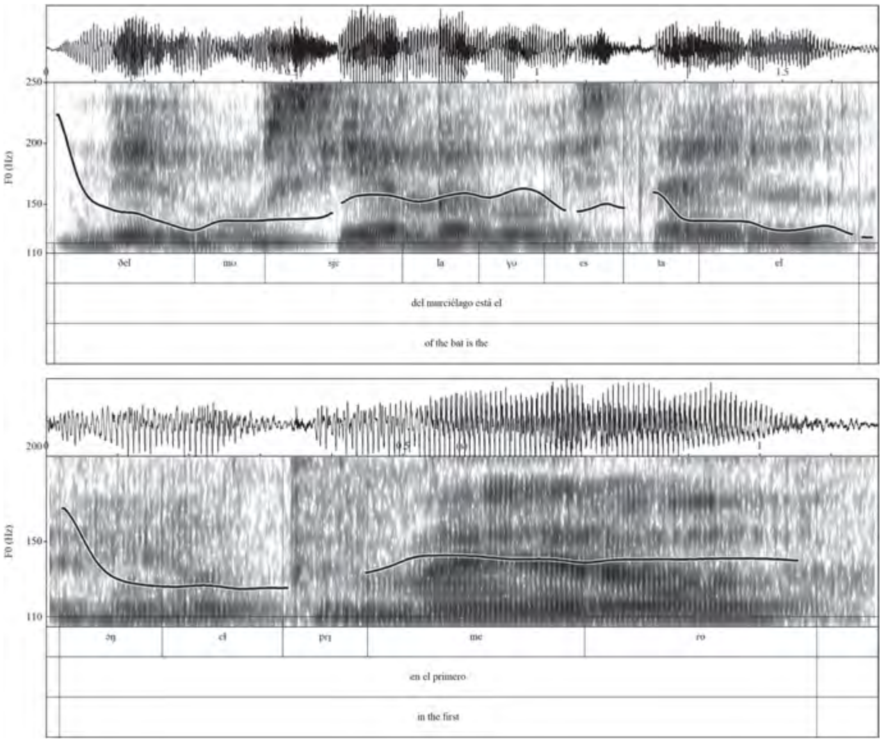


Figure 31: LC34_Conc_ES_1298[110] (*del murciélago está el* 'of the bat is the', top) and ZZ24_Conc_ES_0994[111] (*en el primero* 'in the first', bottom).

to what they describe as a realization with a "continuation rise", except that here, as we have seen, there isn't a "continuous f0 rise from the last stressed syllable until the break" (Gabriel et al. 2011: 163), but instead, f0 first either drops slightly or stays level after the stressed syllable to then produce a separate rise. The realizations where pitch drops slightly more resemble their realization with a "complex boundary tone", which is described as exhibiting "a small dip located between the pre-boundary pitch accent and the high F0 peak signaling the boundary" (Gabriel et al. 2011: 167). They propose that the dip is the effect of an additional L tone that forms a complex ip-level boundary tone together with the H-, but also consider the possibility that it is simply an effect of interpolation. I would tend to the latter interpretation, because otherwise two different boundary tone combinations

110 https://osf.io/cke6p/
111 https://osf.io/vabz2/

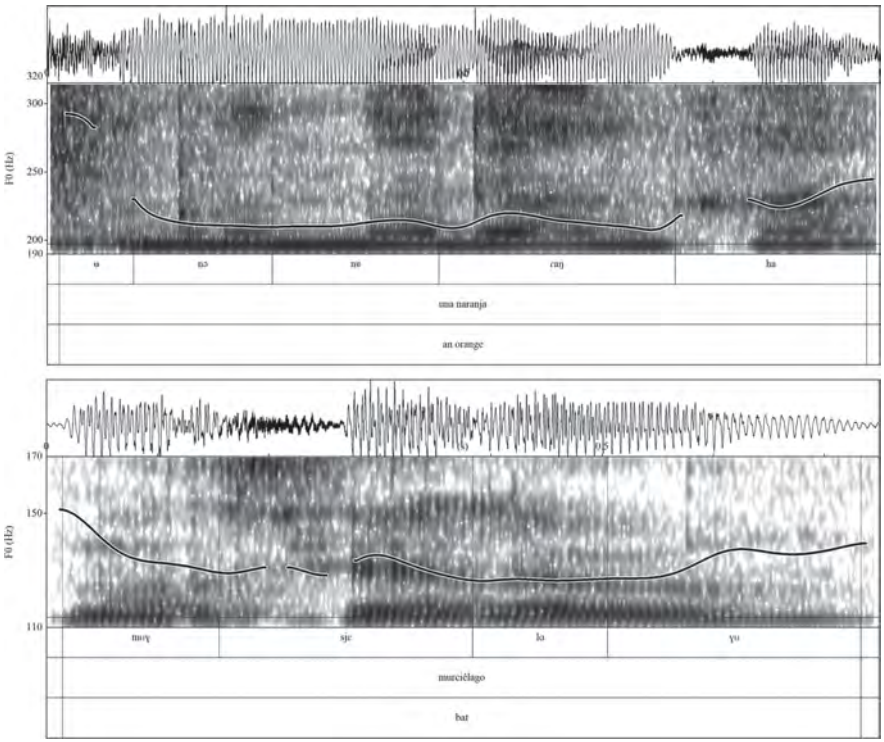


Figure 32: HA30_Conc_ES_0245[112] (*una naranja* ‘an orange’, top) and QF16_Conc_ES_0315[113] (*murciélago* ‘bat’, bottom).

would be posited without differentiating them contextually. Interestingly, Gabriel et al. (2011: 178) relate the relatively high frequency of cases of sustained pitch in their Porteño data (32%) to the influence of speakers of Italian on the speech of Buenos Aires, because Italian was found to have a far higher rate of sustained pitch (45.5%) compared to continuation rises (54.5%) than Peninsular Spanish (11.2% and 88.4%, respectively) in Frota et al. (2007). In our case, it should be noted that in the Quechua rising contour identified in data by the same speakers (cf. section 6.1.1), pitch regularly forms high plateau-like realizations, and the multiple alignment of the H tone responsible for these realizations is an important variable factor in the OT analysis of the Quechua data (cf. section 6.3).

112 https://osf.io/rezpf/
113 https://osf.io/z7eb2/

5.1.2 Interrogatives

This work is mainly concerned with declaratives. However, before picking up the the discussion of variant intonational realizations of them, I will provide a description of some of the intonational variation in the Huari Spanish data that is due to a difference in utterance type, in particular a description of interrogatives. It has been claimed that interrogatives display a larger array of both prosodic and pragmatic variability than declaratives (Cangemi & Grice 2016: 11–12). While the discussion will showcase some of these fine-grained pragmatic differences, we will encounter a comparative absence of formal differentiation.

5.1.2.1 (Neutral) polar questions

Neutral polar questions, in which one speaker requests information from another without a bias for either answer, are in the majority produced with final rises in our Spanish data.[114] At first sight, this makes them in principle very similar to declaratives with continuation rises. At least for some speakers, however, they are formally clearly differentiated: while in declaratives with continuation rise, the last accented paroxytonic word in the phrase before the final rise realizes an LH* pitch accent, in polar questions it does not form a peak but instead, a valley or trough is realized on it which extends until the final rise on the phrase-final syllable. This can be illustrated by the near-minimal pair of XU31_MT_ES_0871 (Figure 33) and OA32_MT_ES_0890 (Figure 34). Both are formed on the segmental material *debajo del árbol*. Their context[115] is given in (41): it shows that the first (at 87.1) is an unbiased request for information by the speaker who does not have the path on the

114 For some small quantification: of the 43 neutral polar questions in Spanish *Quién*, the corpus with the highest relative occurrence of questions, 36 have a final rise. Of the 7 that don't (and are basically flat throughout the utterance), 4 are produced by a single speaker. Three of the rising polar questions do not have a final rise but rise throughout the utterance without marking prominent positions, but they are also fairly short utterances.

115 The numbers in the leftmost column give the time in seconds at the beginning of the utterance in the same line. Consecutive turns by the same speaker are given separate lines and times when there are sufficiently long pauses (silences) between them. Overlap between speakers is indicated using square brackets, [for where it begins and] for where it ends (adopting a convention from GAT 2, cf. Selting et al. 2009). If only] occurs in a line, it means the utterance begins with overlap; if only [occurs, the overlap continues until the end of the utterance. Disfluencies (false starts, self-repairs and the like) are marked by a dash after the transcription of a word up to the point of interruption: *pas-*, *quier-*, *y-*, etc. Transcription is based on standard orthography for Spanish. Where this is not the case it indicates a particular pronunciation or lexical item and is explained in a footnote. The transcription of interjections and hesitation tokens has also been conventionalized: *ah* is an update token, *mhm* the nasalized version of the confirmation/update token *ajá*, *uh* a filled pause, *uhm* its nasalized version.

map in the *maptask*, taking the form of a polar interrogative. The other speaker understands this to be a request (seen from his uttering the acknowledgment token *ya*) and then produces the same material as a first part of further instructions for how to proceed (at 89.5), indicating turn maintenance and that the instructions are not complete via the final continuation rise. At least for the speakers that do make this difference, the analysis for the nuclear configuration in non-biased polar questions therefore could be L* H% (but see below), while for the declaratives with continuation rise it is LH* H%. Note also the pitch accent on the prefinal *debajo* in both examples, which in both cases is identifiable as LH*. This is the case also in other polar questions with pre-final accented words, suggesting that prenuclear pitch accents are of the form LH* both in polar questions and declaratives.

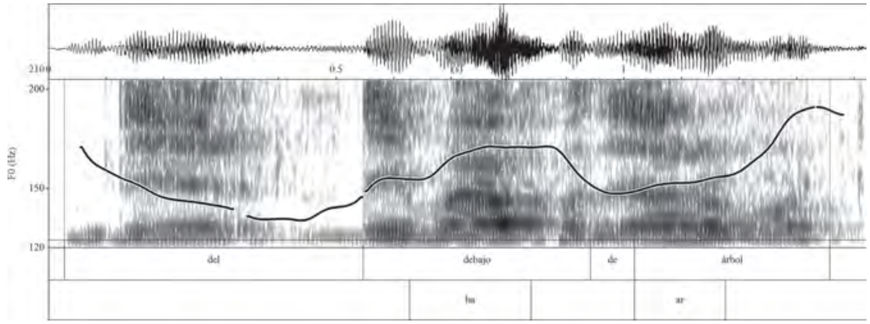


Figure 33: XU31_MT_ES_0871[116] (*debajo de árbol* 'below the tree'; neutral polar question with final paroxytone).

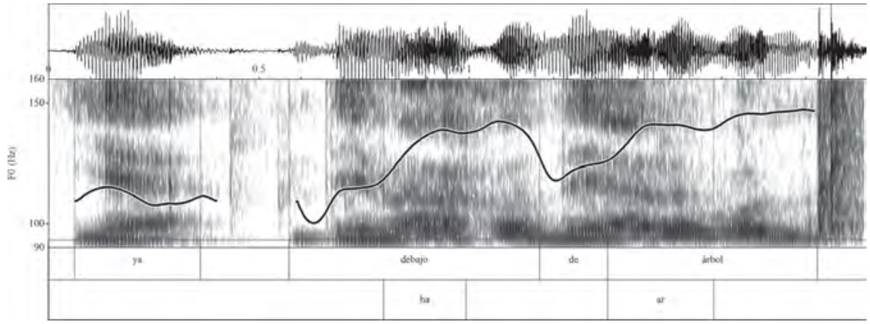


Figure 34: OA32_MT_ES_0890[117] (*ya debajo de árbol* 'yeah below the tree'; declarative with continuation rise, final paroxytone).

116 https://osf.io/tg5r2/
117 https://osf.io/cp8uk/

(41) XU31_OA32_MT_ES_0831–0969[118] (context for XU31_MT_ES_0871 and OA32_MT_ES_0890)

time	**OA32** *(the one with the path)*	**XU31** *(the one without the path)*
83.1	has encontrado un ovejita no	
87.1		**del- debajo de árbol**
89.0	ya	
89.5	**debajo de árbol**	
91.0	y da la vuelta	

A similar yet more subtle contrast can also be found with oxytones in final position. Examples for polar questions with final oxytones are fairly scarce (only 10 in the entire corpus studied here), but they do seem to support the following generalized description: in them, the low stretch before the final stressed syllable extends into it, forming a pitch elbow clearly within it, before then finally rising (cf. Figures 35 and 36). In continuation rises with a final oxytone, on the other hand, the same pitch elbow is formed before the final stressed syllable or at its very beginning (like in declaratives in general), beginning the rise with it or even before reaching it (cf. Figures 37 and 38). This is consistent with an analysis in which the elbows are low targets for an L tone. In polar questions, we could then assume that this low tone is associated and aligned with the stressed syllable, followed by a high boundary tone (L* H%), whereas in continuation rises it is the general declarative pitch accent LH* followed by the boundary tone (LH* H%). However, a slightly different analysis suggests itself when we also take polar questions with question tags into consideration.

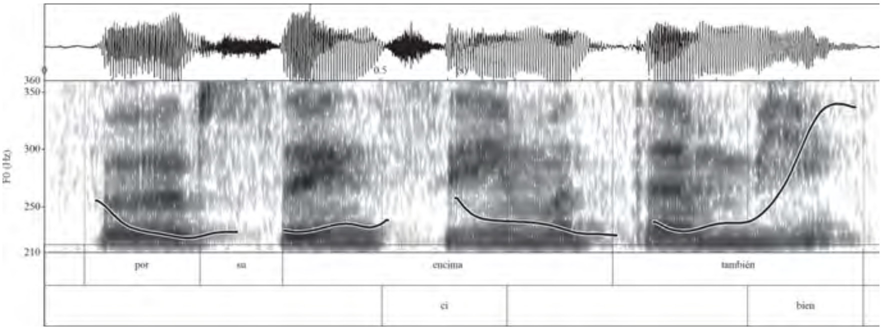


Figure 35: MD40_MT_ES_0324[119] (*por su encima también* 'also above it'; neutral polar question with final oxytone).

118 https://osf.io/s8vdr/
119 https://osf.io/zjfv3/

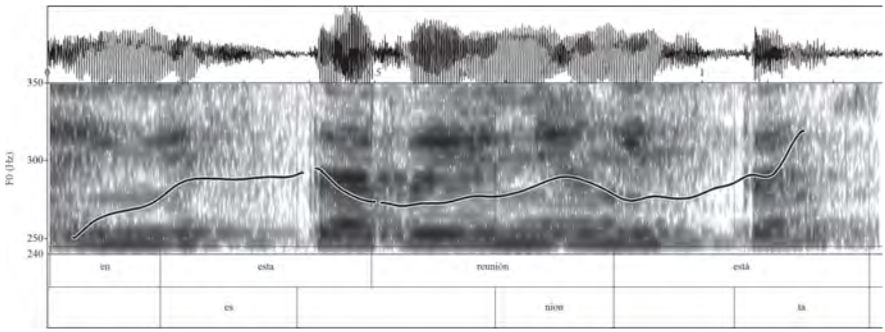


Figure 36: AZ23_Quien_ES_0747[120] (*en esta reunión está* ‘is s/he in this meeting’; neutral polar question with final oxytone).

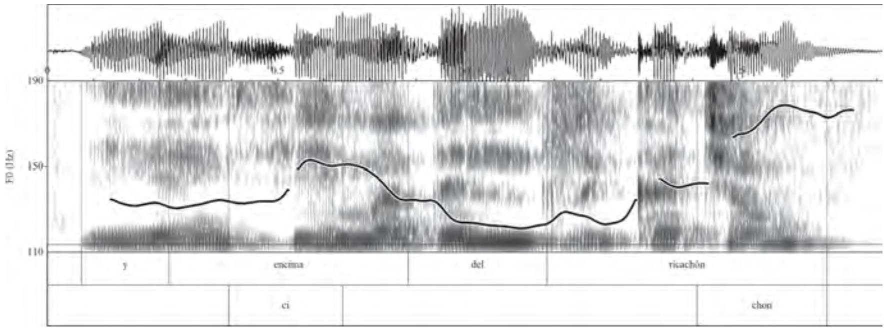


Figure 37: QZ13_MT_ES_0446[121] (*y encima del ricachón* ‘and above the rich guy’; declarative with continuation rise, final oxytone).

Polar questions with the question tag *no* such as *José viene mañana, no?* are usually taken to convey a confirmation bias, i.e. the speaker asks for a truth value on the proposition expressed by the question but with the expectation that this truth value will be of the same polarity as the proposition itself (Farkas & Bruce 2010). Confirmation-seeking polar questions (checks) have been found to have a prosodic form different from that employed for neutral polar questions in several Romance languages, including Bari Italian (Grice & Savino 1997; Grice & Savino 2004), Majorcan Catalan (Vanrell et al. 2013), Puerto Rican Spanish (Armstrong 2010), and somewhat tentatively, Madrid and Mexico D.F. Spanish (Hualde & Prieto 2015: 377). Thus we have to be cautious with using findings from polar questions with question tags

120 https://osf.io/8ph9d/
121 https://osf.io/nyqjr/

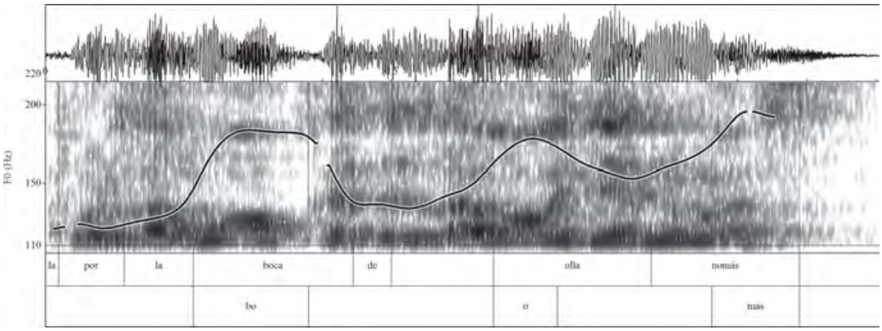


Figure 38: TP03_MT_ES_4425[122] (*por la boca de la olla nomás* 'just by the mouth of the pot'; declarative with continuation rise, final oxytone).

in order to argue a point about neutral polar questions. However, as we will see below (section 5.1.2.2), there is evidence for a particular intonational contour used for confirmation-seeking questions (a 'circumflex' or rising-falling contour) that is different from that of neutral polar questions in our data. This confirmation-seeking question contour however never occurs on the polar questions with tags, which instead seem to use the same contour as neutral polar questions (suggesting a kind of workshare relationship between the tag and the confirmation-seeking intonation). With this in mind, let us consider tag questions and what might be learned from them for the analysis of neutral polar questions in general. In such tag questions with *no* where the final word is a paroxytone, a peak forms on the final accented word, preceded by a low stretch with an elbow in a previous syllable, just as in declaratives; then after the peak, pitch returns to a low stretch on the posttonic syllable, apparently forming a low target, before then realizing the familiar final rise on the tag. See Figure 39 where the second valley, after the peak on the final stressed syllable, reaches even lower than preceding low stretches in the utterances, making it unlikely that this is due just to "sagging" interpolation. With a final oxytone, the same picture emerges: see Figure 40, where while the posttonic valley is not as deep as in Figure 39, the tag itself has a considerably longer duration than it has there.

The posttonic valley in the tag questions is straightforwardly analyzed as a low target indicating the presence of a low tone before the final H boundary tone, so that a final bitonal LH% is most plausible. In turn, this leads to two possible alternative analyses for neutral polar questions: in the first, neutral polar questions differ from both polar tag questions and declaratives with continuation rise in the iden-

122 https://osf.io/e97u3/

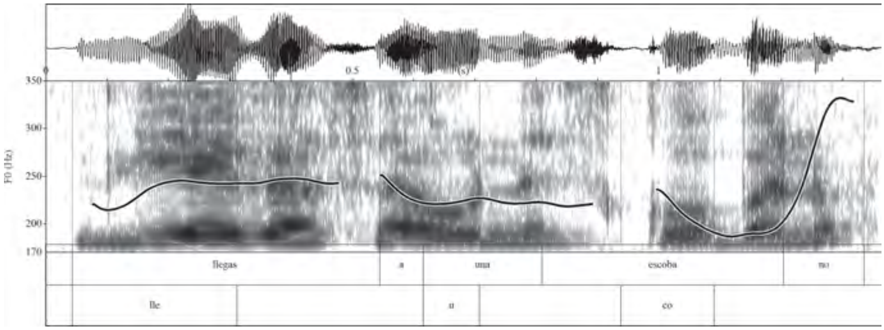


Figure 39: SO39_MT_ES_3237[123] (*llegas a una escoba, no* 'you get to a broom, right'; polar question with question tag after final paroxytone).

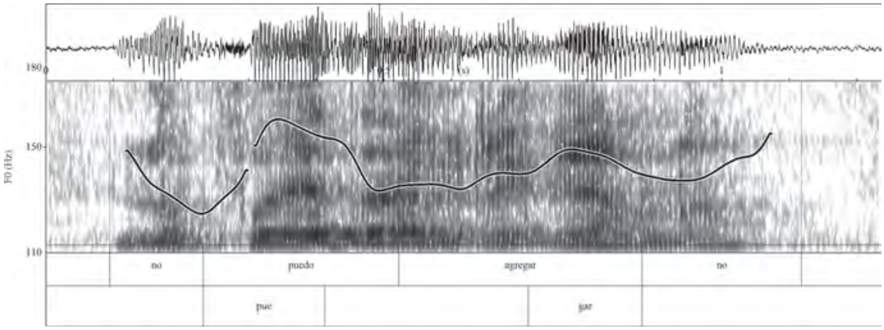


Figure 40: ZZ24_Quien_ES_0788[124] (*no puedo agregar, no* 'I can't add (to that), right'; polar question with question tag after final oxytone).

Table 9: Comparative analysis (alternatives I and II) between neutral polar questions, declaratives with continuation rise and tag questions.

Utterance type	Prefinal (prenuclear) accent	Final (nuclear) accent	Boundary tone(s)
declarative with cont. rise	LH*	LH*	H%
tag question	LH*	LH*	LH%
*neutral polar question (**Alternative I**)*	LH*	**L***	**H%**
*neutral polar question (**Alternative II**)*	LH*	**LH***	**LH%**

123 https://osf.io/3v8wy/
124 https://osf.io/hvw8e/

tity of their final (nuclear) accent, and tag questions differ from the other two via their boundary tone(s). This analysis is given in Table 9 as alternative I. Tag questions are then intonationally different from both neutral polar questions and from biased questions without tag (for those, see below).

Another possible analysis is that tag questions are intonationally the same as neutral polar questions, but that both are different from biased polar questions without tag. This requires the assumption that also in neutral polar questions without tag, the final accent is LH* and the final boundary tone LH%, but that when these tones come into conflict with one another due to time pressure (as with final oxytones and often also paroxytones), the boundary L wins out in the realization against the pitch accent H (which is truncated or severely compressed), precisely because in this way they are still differentiated from declaratives with continuation rise. A rather welcome consequence from the point of view of having a tidy system would be that in this way, phonologically both prefinal and final pitch accents are the same (LH*) across the board, i.e. in declaratives and all types of polar questions, while differences between them are implemented via boundary tones. This second analysis is summarized in Table 9 as alternative II. Neutral polar questions with a final proparoxytone would be an excellent testing ground for deciding between these two competing analyses: alternative I predicts a simple rise from the final stressed syllable to the end of the utterance across the intervening syllables, but under alternative II, we could expect to see the same rise-fall-rise pattern as in tag questions, fully realizing each tone of the LH* LH% configuration, since time pressure conditions are more relaxed.

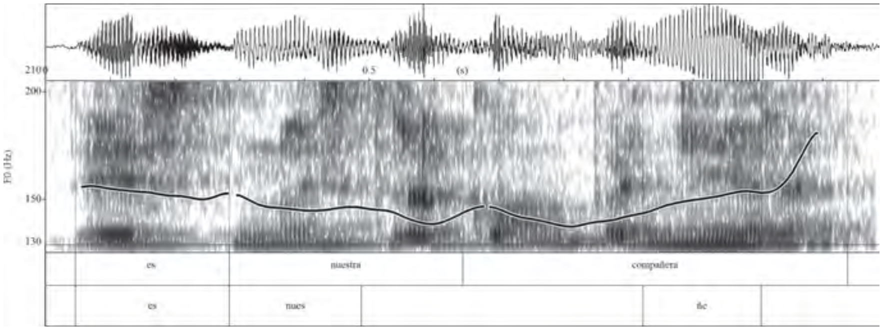


Figure 41: LC34_Quien_ES_1465[125] (*es nuestra compañera* 'is it our classmate'; neutral polar question with final paroxytone).

125 https://osf.io/q5f3z/

Unfortunately, no neutral polar questions with final proparoxytone have been found in the corpora used here. In their absence, we can still consider a polar question utterance like LC34_Quien_ES_1465, given in Figure 41: it has a final paroxytone where a first low elbow is formed in the pretonic at the end of the first syllable, then a slow rise follows that forms a peak or very short plateau at the end of the stressed syllable, and then a much steeper rise concludes the pitch movement towards the end of the utterance. Under alternative I, the first rise and the change in rise speed have no real explanation, since it predicts a low target in the final stressed syllable which would simply be a continuation of the low stretch that began after the last prefinal accent and which would therefore be at the same level as the first elbow in the final word. Under alternative II, the first rise is the realization of the LH* pitch accent on the final word, and the short plateau could be explained as the result of the L tone of the LH% boundary tones being undershot due to it competing with the H*, and once this competition is over, the final H% tone is the only factor affecting pitch level, going some way towards explaining the difference in rise times. Thus, this example (and others like it) seem to favour alternative II, but the evidence is somewhat inconclusive.

5.1.2.2 Polar questions biased towards confirmation (in the maptasks)

As mentioned above already, there are several examples of utterances in the data studied here that can be analysed as polar questions with a confirmation bias (checks) and have a distinguishing prosodic form. This form roughly corresponds to what has been called the "circumflex contour" for polar questions in the description of other varieties of Spanish (Hualde & Prieto 2015). In particular, there are frequent sequences in the *maptask* corpora where a certain type of these questions regularly occur. They can be schematized as follows (step **b** in bold being the confirmation-seeking question):

(42) Schematic sequence in the *maptasks* in which a type of check occurs
  a. Speaker with map: you move in relation *y* to landmark *x*
  b. Speaker without map: **relation *y* to landmark *x* ?**
  c. Speaker with map: [Confirms]
  d. Speaker with map: [Proceeds to next instruction]

These should be considered as questions with confirmation bias because 1) the epistemic imbalance between the speakers that is inherent in the game (one speaker has the path on their map, the other doesn't have it) effects a general bias for the speaker without the map to request information and the other to provide it, especially with regards to how to follow the path. 2) The information has previously been given (in step **a**), so it is not likely for the speaker without the map to intend the utterance produced in step **b** to be a completely neutral question.[126] 3) the speaker with the map crucially understands the utterance in step **b** to be some kind of request, as evidenced by them regularly giving a token of confirmation (step **c**), and only then (step **d**) moving on to giving further instructions. If step **b** were just a confirmation or acknowledgment itself (like "message received, I repeat: relation y to landmark x[127]"), then the speaker in the path could proceed to giving further instructions already at step **c**, omitting their own confirmation at that step. Sequences of that latter type also abound in the corpus, with the speaker without the map just uttering a confirmation or acknowledgment token (such as *sí*, *ya*, *hm*, or even *y*) at **b** and the other then moving on to the next instruction at step **c**, and they can thus be clearly differentiated from the cases discussed here.

Utterances performing such illocutionary functions have been called *clarification requests* (Ginzburg 2012; Łupkowski & Ginzburg 2016), and specifically in the context of map tasks, very similar utterances have been proposed to be called *OBJECT* in Grice & Savino (1997). Specifically, they contrast *OBJECT* moves with *ACKNOWLEDGE* moves, which simply acknowledge or confirm instructions given. On *OBJECT* moves, they state that they are "used to point out that there has been a break-down in communication, such that the game cannot continue until common ground is re-established" (Grice & Savino 1997: 30). It should be noted that Grice & Savino (1997), while agreeing with the interrogative nature of these *OBJECT* moves, explicitly separate them from what they classify as *CHECKS*, although the criteria are not quite clear: for *OBJECT* moves they state that they are what has been classified as 'echo questions' in the literature because they repeat what has been said

126 The move in step **b** might conceivably be performed because the information conveyed by it is incompatible with what the speaker without the map expects or believes to be in CG, i.e. as a kind of incredulity question. For the utterances discussed in this section, the context does not suggest this to be the case. Utterances where the context is compatible with a similar import of incredulity are briefly discussed in section 5.1.3.3 with the tentative conclusion that they have the same contour shape as the confirmation-seeking questions discussed here, but with the pitch peak aligned later.

127 This kind of utterance seems familiar from (narrative portrayals of) military-style dialogues. However, those are subject to normative control in following a conversational procedure that is designed to minimize risk of misunderstanding (especially in the context of less than ideal channels of communication, such as giving and receiving instructions over radio accompanied by the extreme background noise of armed altercations or heavy machinery) at the expense of increasing redundancy. In normal conversation, I argue that the maxim of quantity will discourage such verbatim repetitions of instructions for the sole purpose of confirmation and make them stand out (marked) as probably intended to mean more (e.g. being a request for confirmation; cf. also Farkas & Bruce 2010: 99 for the view that assertion confirmation is often implicit because it is the least marked response).

before by the interlocutor (Grice & Savino 1997: 30), while *CHECKS* are defined as confirmation-seeking questions asking for information “which the speaker believes has already been conveyed” (Grice & Savino 1997: 29). This is not a satisfying separation, since one definition is concerned with the form and the other with the meaning of the utterances (and both definitions could therefore be true about a token utterance of either type). Furthermore, *OBJECT* moves are said to be not “simple questions”, since “they could be responding within one game as well as initiating another (sub-) game” (Grice & Savino 1997: 30–31). It is not clear how this objection does not equally apply to those utterances classified by them as *CHECKS*.

It is clear that not all types of biased questions either fulfill exactly the same function or have the same form. They also often make it difficult to maintain a categorical separation between declaratives and interrogatives (but not between assertions and questions); but as argued above, the utterances encountered in the maptasks do seem to fulfill the criteria of being questions and of being biased in the sense that they ask for confirmation about information which the speaker has already received. In our corpus, the same intonational form is also used in utterances of the type occurring at step **b** even if there are other utterances intervening, i.e. if step **b** is not exactly a repetition of the previous utterance by the interlocutor, which speaks against an analysis as echo questions. (43) is an example.

(43) SG15_QF16_MT_ES_0453–0703[128]

time	**SG15** *(with the path)*	**QF16** *(without the path)*	
45.3	ya de ahí pasamos *so from there we move on*		= step a (part 1)
46.7	por		
47.5	de abajo del *below the*		
49.9	cómo se llama *what is it called*		= step a (part 2), solving the sub-QUD How IS X CALLED IN *RELATION Y TO LANDMARK X*?
52.1	mhm		
56.6	su nombre se me ha ido *I’ve lost the name*		
59.7		corderito *little lamb*	
60.4	corderito *little lamb*		

128 https://osf.io/jfk4a/

61.5		**debajo** *below*	= **step b**
62.0	mhm ya de abajo *yeah below*		= step c
65.0	de ahí pasamos *from there we pass*		= step d
66.5	por lado de los *alongside of the*		
69.4	gentes *people*		

In the sequence, SG15 starts producing the instruction corresponding to step **a**, but only gets to specify the relation y, before a small digression occurs in which the two of them solve the sub-QUD of how the landmark x is called. At 59.7 QF16 suggests *corderito*; note that this takes the form of a polar question with a final rise. SG15 confirms and then QF16 performs step **b**, but he does not produce an echo-question repeating the previous utterance. Instead, his question there is related only to the relation y that had been asserted before the digression occurred. This *debajo* in step **b** does not have the form of a neutral polar question with final rise, but the "circumflex contour" (cf. Figure 42). Yet, crucially, it is understood as a request for confirmation by SG15 in step **c**, when she gives that confirmation, and then proceeds to give new instructions in step **d**.

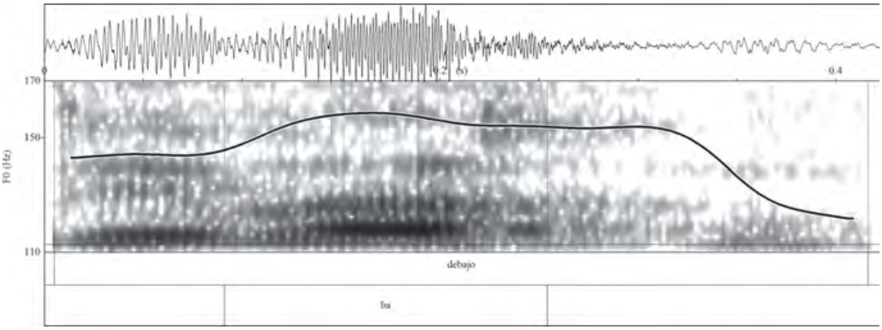


Figure 42: QF16_MT_ES_0615[129] (confirmation-seeking polar question). Cf. context (43).

129 https://osf.io/gby2a/

In another sequence, a confirmation-seeking question contrasts with a continuation rise where the two are intonational minimal pairs:

(44) QZ13_OZ14_MT_ES_0769–0971[130]

time	**QZ13** *(with the path)*	**OZ14** *(without the path)*	Intonational form
76.9	por el encima del difunto das una vuelta debajo del zorro *above the deceased you turn around below the fox*		
83.3		**debajo del zorro** *below the fox*	**"circumflex contour"**
86.1	estás *are you there*		neutral polar question
92.5		debajo del zorro *below the fox*	continuation rise
94.0	encima del nube que tiene truenos *above the cloud with the thunder*		
96.7		ya *right*	

Here, QZ13 gives instructions as step **a** at 76.9, followed by OZ14's first *debajo del zorro* at 83.3, which is here step **b** and has a circumflex contour. This is understood by QZ13 at least to be a request for suspension of further instructions because he waits for a while, then asks (*estás* at 86.1) whether OZ14 has now reached the point on the map at which *debajo del zorro* is a reasonable instruction (quite literally whether they have reached common ground on their maps[131]). At 92.5, OZ14 then

130 https://osf.io/rd3vm/

131 The commonality here with a sequence of a request for confirmation and then giving this confirmation might be explained thus: as Grice & Savino (1997: 30) point out, such moves asking for confirmation on already given information indicate some kind of break in the game which needs to be adressed until common ground is reestablished. Instead of treating OZ14's utterance at 83.3 as a request for confirmation (he doesn't give any), QZ13 seems to interpret it more temporal-spatially in terms of the game: OZ14 has not drawn the line on his map up to the point from where he can go *debajo del zorro*, and at which point the two of them have reached the same spot in their joint progress across the map (which is a very relevant part of their common ground for this game). A request for confirmation is partially a request for suspension of epistemic progression, here applied as a suspension of spatial progression.

repeats his previous utterance, but with the intonation of a continuation rise, and this is understood by QZ13 immediately as him being ready to continue, and certainly not as a request for further confirmation; he proceeds with the next instructions and OZ14 acknowledges them without a problem (94.0 and 96.7).

Regarding their form, these confirmation-seeking questions are very similar to what has been called “circumflex contours” in the description of other varieties of Spanish: on the last accented word, a peak is formed within the stressed syllable, with an elbow at the start of its rise in the pretonic and another elbow at the end of its rise in the posttonic, and low pitch utterance-finally. For Madrid Spanish, Hualde & Prieto (2015: 374) conclude after a review of the literature that the final-rise question contour is used for “pragmatically unmarked” polar questions, while the circumflex contour seems to be used for some others, including echo- and confirmation-seeking questions. This assessment seems broadly applicable here as well, with the difference that here (unlike in Madrid Spanish), it is difficult to separate the confirmation-seeking questions from simple declaratives with a final fall formally, since those also are characterized by peaks on each accented syllable followed by a fall to low after the final one (see section 5.1.1). As such, they contrast formally and meaning-wise with neutral polar questions, as we have seen above in (43), and also with declaratives with continuation rise, as in example (44) above. That is to say, they are used in different contexts with different responses following them than neutral polar questions and continuation rises. However, it is not quite clear whether they also formally differ from declaratives with final fall in any way. Another example is given in Figure 43. This utterance also does not repeat anything from the interlocutor’s previous utterance. Instead it expresses a conclusion that the speaker has reached from the preceding discussion (this is also indicated by the update token *ah*). It is understood by the other speaker to be a request for confirmation insofar as he responds to it by giving this confirmation. Here the final pitch accent has more excursion than the preceding ones, and it would be an interesting task for the future so see if this difference in excursion really distinguishes this type of confirmation-seeking polar questions intonationally from declaratives. That is what is proposed for Madrid Spanish for distinguishing one type[132] of confirma-

132 The situation for Madrid Spanish also seems less than fully clear: at one point, Hualde & Prieto (2015: 373) analyse an utterance described as a “marked confirmation yes/no question” with the upstepped nuclear configuration L+¡H* L%. However, in their tabular summary (Hualde & Prieto 2015: 389), they give the configuration L+H* HL% for confirmation-seeking questions (the contour is not discussed in the context of questions anywhere in the text itself) and L+¡H* L% for echo questions. In Estebas Vilaplana & Prieto (2010: 29), meanwhile, the nuclear contour H+L* L% is given for confirmation-seeking polar questions, which however in Hualde & Prieto (2015) is only discussed for questions in Puerto Rican varieties.

tion-seeking questions from declaratives with narrow focus (L¡H* L% vs. LH* L%, cf. Hualde & Prieto 2015: 373, 389). However, for our data here it is difficult to assert this because in both confirmation-seeking questions and declaratives, local pitch span varies on the final pitch accent. Solving this issue would require a quantitative analysis on more controlled data and probably also perception tests. Pending this, we cannot make the claim that the types of biased questions discussed here differ formally from declaratives with a final fall.

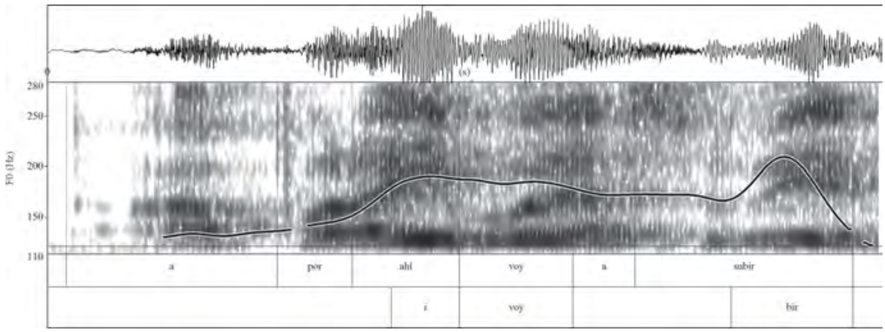


Figure 43: KP04_MT_ES_1975[133] (*a por ahí voy a subir* 'ah I'll go up there'; confirmation-seeking polar question with final oxytone).

The possibility seems to exist that with these kinds of utterances in the *maptasks*, the same intonational contour can be used for declaratives and for confirmation-seeking polar questions. Is it plausible that that is the case? Firstly, this contour is sufficiently different from that for continuation rises and that for neutral polar questions. It is therefore not likely to be misunderstood to signal a situation where the speaker without the map gives an acknowledgement (either via a continuation rise contour or with just a token) or where they want to request information on something that has not recently been discussed (which can be achieved using the neutral polar question intonation). Secondly, the epistemic situation between the speakers as given by the game makes assertions about where the path runs by the speaker without the path a relatively unlikely event (except for when that speaker describes what is on their map). Gunlogson (2001, 2008) discusses "declarative questions" in English[134] and develops a framework for understanding their pragmatics. Adapt-

133 https://osf.io/bu9tw/

134 In English, both word order and intonation are formal means involved in forming declaratives and interrogatives, and they sometimes seem to be at odds with one another: Gunlogson (2001, 2008) therefore differentiates between polar questions (with subject-verb inversion and finally ing from Gunlogson (2008: 113), the speaker without the path is not a source with regard to that information[135] and can only make a dependent commitment to any proposition about where the path runs. A dependent commitment is made if one speaker is a source for a proposition and the other also commits to it without being a source for it (this is what happens when a polar question in a neutral context is answered: the speaker asking the question proposes to make a dependent commitment on a proposition, not knowing the answer but expecting their interlocutor to know, cf. Gunlogson 2008: 121). In Gunlogson (2008)'s analysis, the formal device of a declarative signals a context in which a commitment has been made with regards to the proposition at-issue: either by the speaker themself with the utterance (an assertion), or previously in the discourse. Whether an utterance consisting of such a declarative is then interpreted as an assertion or a question is dependent upon context and upon whether the speaker can be a source for the proposition: assertions can only be made by a speaker who can reasonably act as a source for the proposition they assert (Gunlogson 2008: 116–117). In our situation, the speaker without the path is highly unlikely to be a source, whereas the speaker with the path is known to be one, to both speakers. Thus, for the speaker without the path to make an assertion about it is odd in the context, first because they would pose as being a source for it even though it is known to both speakers that they really cannot be, and second because in any case, the speaker with the path is known by both to know about the path, so that they do not need to be informed about it (cf. Gunlogson 2008: 118–120). This sets the odds in favour of an utterance about the path by the speaker without the path to be interpreted as something other than an assertion. In fact, it is highly likely to be interpreted as a question given this contextual epistemic constellation, but as a question in a context in which a commitment to the proposition already exists, which is signaled by the declarative form: this is

rising intonation), rising declaratives (without inversion, but with finally rising intonation) and falling declaratives (without inversion and finally falling intonation):

a. Is Bob home? (polar question, the ? symbolizing rising intonation)

b. Bob is home? (rising declarative)

c. Bob is home. (falling declarative, the symbolizing falling intonation)

Spanish does not normally differentiate polar interrogatives from declaratives by word order, but many of Gunlogsons's insights about the pragmatics involved are still applicable here.

135 A speaker is a source for a proposition if and only if they are commited to the proposition and if in the discourse context, their commitment does not depend on another speaker's stated commitment to the proposition (cf. Gunlogson 2008: 113). In our case, the speaker without the path can only be informed about the path via testimony of the speaker with the path, so they cannot be a source for it (and this is known to both speakers in the context).

felicitous, since a commitment has in fact already been made by the speaker with the path (often, but not always, in the directly preceding utterance). Thus the intonational contour under discussion can remain a declarative contour, unspecified as to whether it realizes an assertion or a question, with the possibility of being either where specific contexts allow or suggest it.

5.1.2.3 Alternative questions

Alternative questions provide two or more alternatives, as in a list, from which the adressee is to form a true answer (cf. Sadock & Zwicky 1985: 179; König & Siemund 2007: 291–292). The answer set is therefore not just {p, ¬p} as in polar questions, but may contain more elements.

(45) ¿Quieres agua o café?
'Do you want water or coffee?'

Krifka (2011: 1749) points out that without intonation, questions like (45) are formally indistinguishable from polar questions with a disjunctive constituent, *quieres (agua o café)?*, and (45) is furthermore ambiguous (without intonation) between an open-set and a closed-set interpretation: it might be interpreted to mean that the two options are water and coffee, and none else, or that water and coffee are just two of the available options. In the data discussed here, there are always just two named alternatives, and very often their semantics together with the context suggest that they are also the only two possible ones. For example, in a situation when the path in a *maptask* leads up to a landmark from the right side, and the speaker without the map asks whether they should go above or below it, these are the only relevant alternatives in the discourse. In principle, it would of course also be possible to go alongside it or around it, but since the position of the path up to that point is fixed already, and the path is thus coming towards the landmark from a given side on a two-dimensional map, for the purposes of the game only two alternatives are relevant.

Alternative questions are not overly frequent, even less so than polar and wh-questions (around 20 in all the corpora here taken together), and they are not evenly distributed across all speakers. Thus we cannot say whether the generalizations made here are valid for all speakers, but they do cover all of the examples encountered, and not just the ones presented here. The generalizations are: whether the final word of the first alternative is proparoxytone (Figure 44), paroxytone (Figure 45), or oxytone (Figure 46), a pitch accent that can once again uncontroversially be analyzed as LH* is associated with its stressed syllable, realized as a preceding low stretch and elbow before the stressed syllable and a peak within the stressed syllable. A high boundary tone is produced at the end of the phrase corresponding to the first alternative (with final oxytones, the peak of the pitch accent and that of the boundary tone are realized in one, just as in other utterance types). At the end of the second alternative, pitch is low, taken as evidence for a low boundary tone (L%). A further generalization (with the possible exception of KA36_Quien_0197, given in Figure 47) is that in the phrase corresponding to the second alternative, no visible or audible pitch accents are realized on stressed syllables, whether they are final or pre-final. Pitch falls steeply from the high boundary at the end of the first alternative, and then either continues to fall slowly or remains very low. The phrase of the second alternative is thus deaccented or realized with an extremely reduced pitch range. Material preceding the first alternative is normally accented (cf. Figures 44 and 46). As already mentioned, Figure 47 shows the only utterance where the second alternative could be seen as still forming pitch accents. However, it should be noted that this utterance is produced in a particularly low register and with only a relatively small pitch range overall (as is typical for this speaker). This means that the sustained pitch in the first syllable of *mujer* might also just be the product of elevation due to consonantal microprosody caused by the fricative at the beginning of the final stressed syllable (cf. a similar elevation at the beginning of *mujer* in Figure 46, where however it is very small compared to the intonational pitch movements).

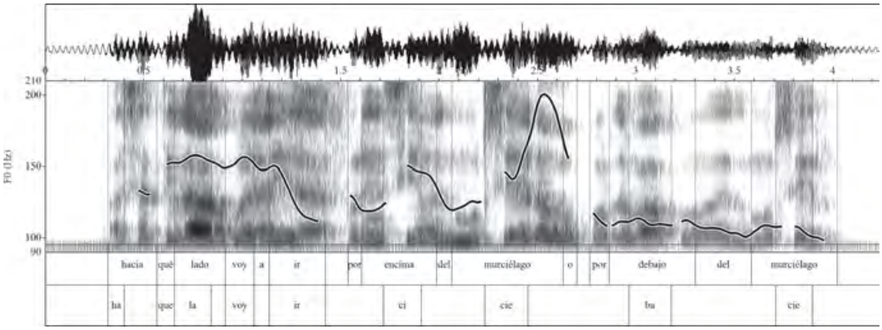


Figure 44: KP04_MT_ES_3219[136] (*hacia qué lado voy a ir por encima del murciélago o por debajo del murciélago* ‘which side am I going to go, above the bat or below the bat’; alternative question with two final proparoxytones).

136 https://osf.io/qmtew/

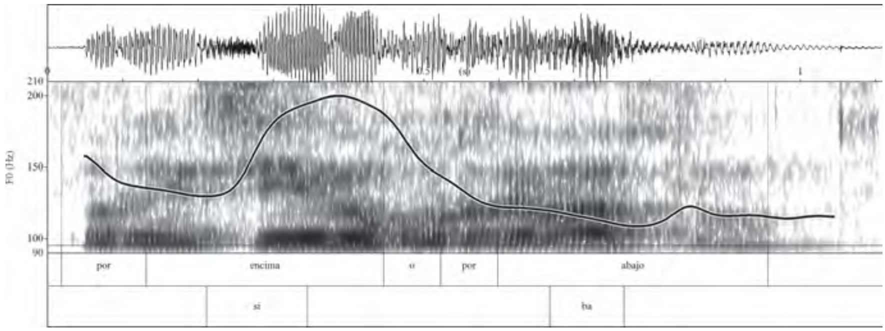


Figure 45: QF16_MT_ES_0136[137] (*por encima o por abajo* 'above or below'; alternative question with two final paroxytones).

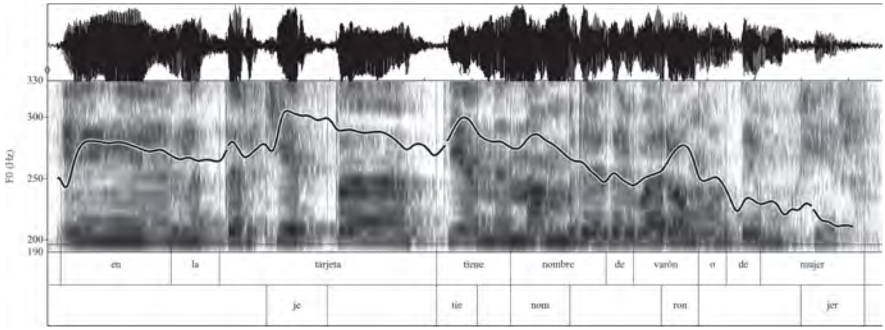


Figure 46: AZ23_Quien_ES_0020[138] (*en la tarjeta tiene nombre de varón o de mujer* 'on the card does it have the name of a man or woman'; alternative question with two final oxytones).

Deaccentuation or pitch compression will also be discussed in more detail in the context of declaratives in section 5.1.3.2. Its prevalence makes it necessary to include a reference to scaling in the intonational analysis of alternative questions here. From Figure 44, this reduced scaling can be seen to apply to a prosodic unit larger than an individual prosodic word, corresponding to the second alternative. The analysis is given in (46) in two versions, one with, one without, deaccentuation. The crossed-out notation is intended to convey the presence of accentable words that are deaccented or whose scaling is severely reduced, round brackets indicate

137 https://osf.io/xhdu7/
138 https://osf.io/gt423/

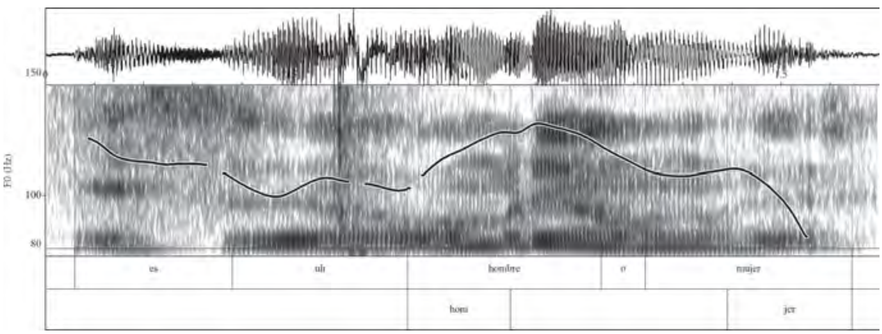


Figure 47: KA36_Quien_ES_0197[139] (*es uh hombre o mujer* 'is it uh a man or woman'; alternative question with a final paroxytone and a final oxytone).

optional (pre-nuclear) pitch accents. In section 5.2 the nature of the prosodic units involved will be discussed in more detail.

(46) Intonational analysis for alternative questions
  a. With deaccentuation
     [(LH*) LH* H-/%]$_{\text{alternative 1}}$ [~~(LH*) LH~~* L%]$_{\text{alternative 2}}$
  b. Without deaccentuation
     [(LH*) LH* H-/%]$_{\text{alternative 1}}$ [(LH*) LH* L%]$_{\text{alternative 2}}$

5.1.2.4 Wh-questions

The wh-questions attested in the corpora here present a heterogeneous picture intonationally. All of them form a pitch accent on the wh-word which is most likely to be analyzed once again as LH*. In all those where further accentable words precede the wh-word, they are also accented with an LH* pitch accent. Almost all wh-questions end on a final fall, but a small minority has a final rise. It is the part between the wh-word and the end of the utterance where the data is most ambiguous. If the wh-word is the last word in the utterance, then there is simply a steep fall after the peak on its stressed syllable (cf. Figure 48).

However, if more accentable words intervene, a variety of things can happen: firstly, after the wh-word, which has the highest pitch peak, pitch can fall progressively over the following accentable words, with or without compressed but identifiable pitch accents on each (cf. Figures 49 and 50).

139 https://osf.io/cdrwf/

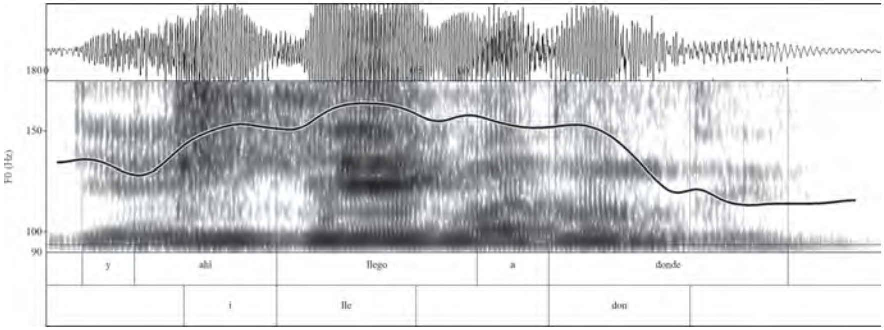


Figure 48: QF16_MT_ES_2110[140] (*y ahí llego a dónde* 'and there I get to where'; wh-question with the wh-word finally and preceding accented words).

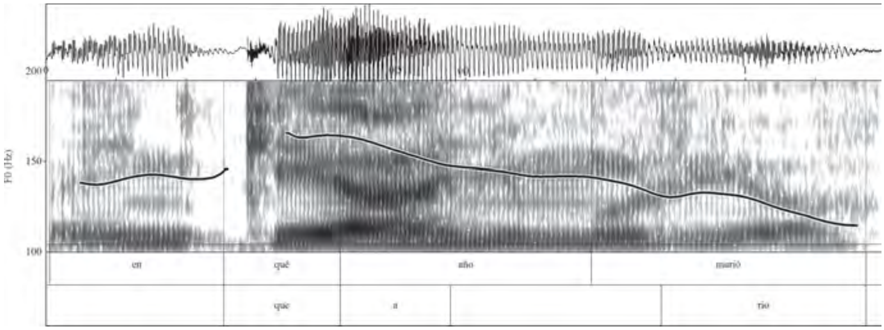


Figure 49: OV37_Quien_0155[141] (*en qué año murió* 'what year did s/he die'; wh-question with the wh-word followed by accentable words without identifiable pitch accents).

In several examples, the accentable words following the wh-word do not behave uniformly: up to a certain word from the wh-word, pitch is maintained at a relatively high plateau which can be taken as evidence that they are accented, then there is a steeper fall than in the previous examples, after which pitch is at a relatively low level and all following words are not accented. Where this is the case, the steep fall between the accented and the unaccented stretch does not occur at a uniform syntactic boundary: it can occur between an auxiliary and a verb (cf. Figure 51), between a verb phrase and a postposed subject (cf. Figure 52), or directly after the pitch accent on the wh-word (cf. Figure 53).

140 https://osf.io/39h2e/
141 https://osf.io/r4gsd/

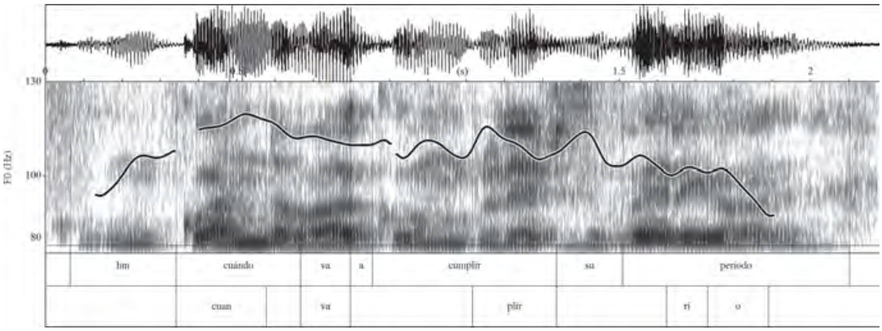


Figure 50: KA36_Quien_1075[142] (*cuándo va a cumplir su periodo* 'when is their term going to end'; wh-question with the wh-word followed by accentable words with pitch accents).

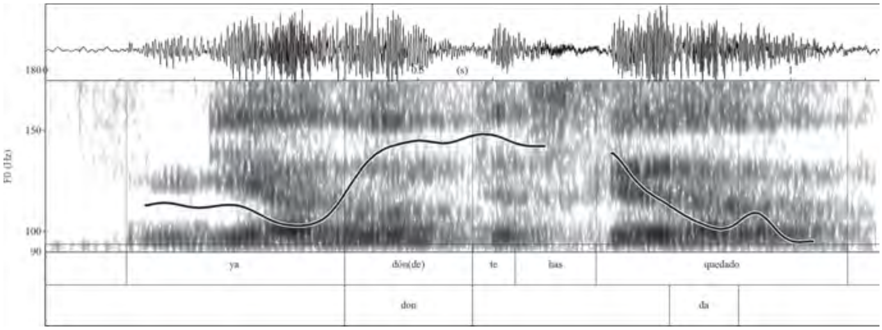


Figure 51: TP03_MT_ES_4108[143] (*ya dónde te has quedado* 'so where did you get left'; wh-question with a post-wh-word fall between an auxiliary and a verb).

That it is not always just one (the last) word which is deaccented/realized on the low pitch stretch (cf. Figure 53), makes it unlikely that the final pitch accent is L* while all preceding ones are LH*. The only possible generalization over all of these cases is one making use of optionality: at some point after the wh-word, deaccentuation (or severe compression) may occur. It seems reasonable to assume that deaccentuation happens after some kind of phrase boundary. However, the resulting phrases then do not correspond very well to syntactic or information-structural categories, since it is neither always just the wh-word, or the larger wh-NP, as in *cuántos años* (that could be said to be focused), nor something like the wh-NP + the following VP that

142 https://osf.io/ht96p/
143 https://osf.io/detz6/

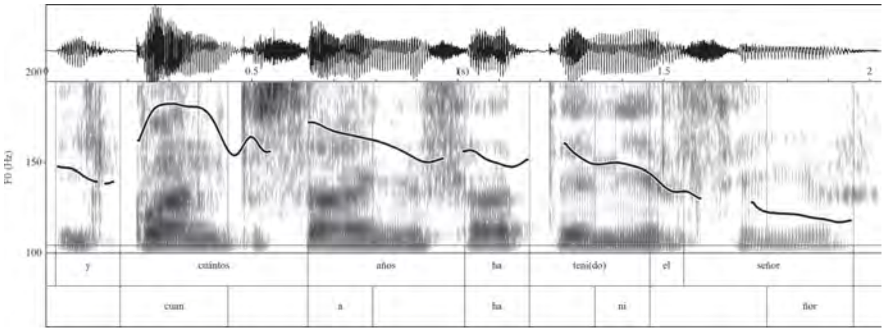


Figure 52: OV37_Quien_0446[144] (*y cuántos años ha tenido el señor* 'and how old was the gentleman'; wh-question with the wh-word followed by accentable words with and without accents).

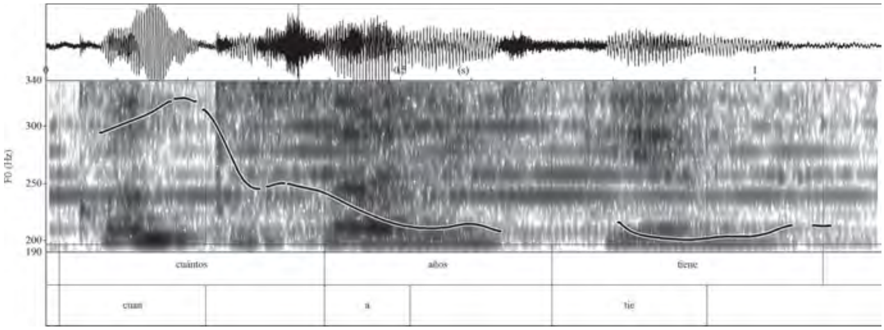


Figure 53: HA30_Quien_ES_0571[145] (*cuántos años tiene* 'how old is s/he'; wh-question with two deaccented words after the wh-word).

are separated as one from the following material, nor is the deaccented material itself homogeneous in this way. Nor is deaccentuation itself obligatory (cf. Figure 50).

(47) Analysis for wh-questions with optional deaccentuation
[LH*$_{\text{wh-word}}$ (LH*)] ([~~LH*~~]) L%

It is also questionable whether cases like Figure 49, where there is no steep fall after the wh-word but instead a slowly falling stretch that essentially seems to be interpolation, are phonologically different from the cases with steep falls. It could be that this really is just interpolation, then there would be no pitch accents on the accentable

144 https://osf.io/pcn4b/
145 https://osf.io/8geu6/

words after the wh-word and this is also a case of deaccentuation. It could, however, also be that this contour is the result of successive downstep and pitch compression on the pitch accents following the wh-word, then these cases would not be deaccented but just compressed. I will tend to the second explanation, since that allows us to unify deaccentuation as a phenomenon where the final L tone creates a low stretch aligned as far leftwards as possible, which then results in the typical steep falls followed by low stretches. However, the difference between compression and deaccentuation is clearly gradual and thus this picture somewhat oversimplifies (see also section 5.3.3). All of the prosodic variation in wh-questions concerning scaling here does not seem to correlate with any pragmatic differences. This is arguably different for those wh-questions that have a final rise. The few instances that can be found are all variations of asking *cómo*? “How/what?” that can be interpreted as a request for repetition because the speaker has not heard or understood well what the other was saying. Thus, these are not really asking for a variable to be filled in an open proposition, but essentially just for repetition of the preceding utterance, and they are understood as such (as evidenced by the reaction of the interlocutor, cf. (48) and Figure 54).

(48) AZ23_ZZ24_MT_ES_0102–0195 (context for AZ23_MT_ES_0153)[146]

time	**ZZ24** *(the one with the path)*	**AZ23** *(the one without the path)*
10.2	y por el centro del la quena y el cerro pasa el camino *and through the middle of the flute and the hill does the path go*	
15.3		**cómo va** *how do you mean*
16.6	por el cerro y la quena por la mitad pasa *it goes through the middle of the hill and the flute*	

5.1.2.5 Summary

The discussion of interrogative intonation has produced a number of interesting results. Despite a demonstrable range of pragmatic differentiation in the discourse contexts from which the data is drawn, the intonational form of these utterances seems comparatively unaffected by it, if not by prosodic context. The signaling of different utterance types and other pragmatic meanings seems to be achieved in large part by

146 https://osf.io/vn2c4/

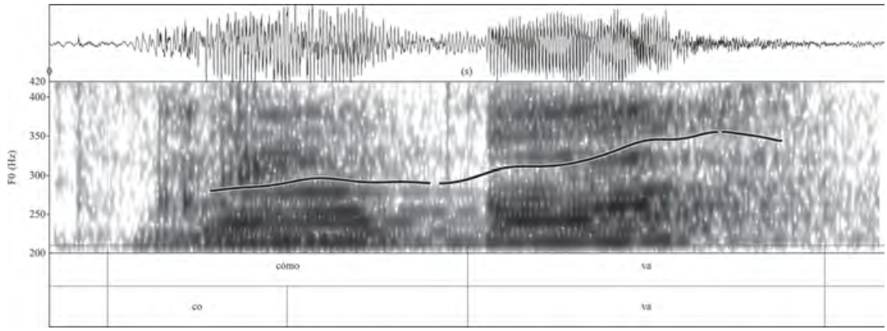


Figure 54: AZ23_MT_ES_0153 (wh-question with a final rise, a request for repetition).

other means, or is left to the context. In fact, if we take the analysis called "alternative II" for polar interrogatives and some form of the deaccentuation analysis for wh-questions to be correct, then so far pitch accent choice in the data as a whole has remained paradigmatically invariant, with a single pitch accent, LH*, occurring in both prenuclear and nuclear position in declaratives, all types of polar questions, alternative questions, and wh-questions. The edge tone inventory is only slightly larger, composed of two monotonal (L-/% and H-/%) and one bitonal (LH-/%) boundary tones. Certainly the results here cannot be called definitive in the sense that they can exclude the possibility of further pragmatically-conditioned variation in paradigmatic tone choice in Huari Spanish as a whole. Yet even so, these findings suggest a very regularized pitch accent system, especially when compared to the wealthy tonal inventory described for Madrid Spanish (cf. section 3.4.3). However, we found evidence that a more syntagmatic device aids in the cueing of utterance type prosodically in both wh-questions and alternative questions, namely deaccentuation or reduced scaling of material following the highest pitch accent, sometimes extending across entire phrases.

5.1.3 Patterns of (information structurally-conditioned) variation in declaratives

In what follows, variations on the main accentuation will be described and possible functional explanations (prosodic and information structural) for them explored.

5.1.3.1 Phrase accentuation

With the term "phrase accentuation" I want to label an intonational variant occurring in Huari Spanish declaratives whereby not every accentable word is pitch accented, but only the last one in a larger phrase (cf. Figures 55 and 56). This is in contrast to the "main" accentuation variant.

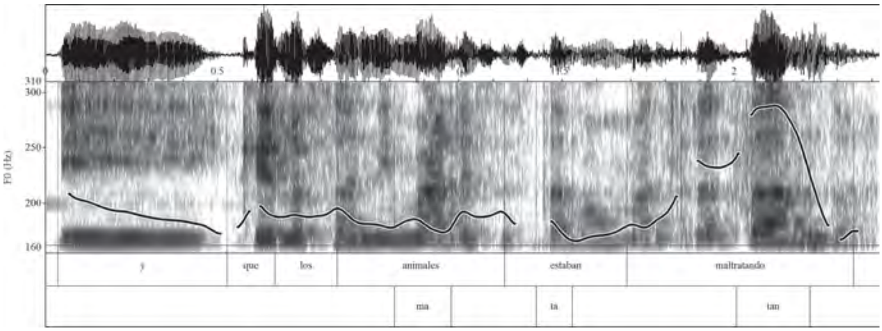


Figure 55: MS27_Cuent_ES_1520[147] (*y que los animales estaban maltratando* ‘and which the animals were mistreating’; declarative with strong phrase accentuation) Cf. context (53).

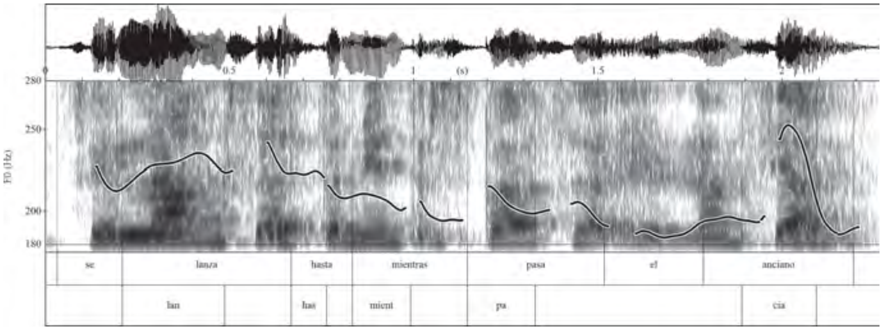


Figure 56: ZR29_Cuent_ES_1572[148] (*se lanza hasta mientras pasa el anciano* ‘it launches itself while the old man passes by’; declarative with strong phrase accentuation). Cf. (52) for context.

The two accentuation variants should not be seen as completely separate: it seems that a variation space exists in which the “main” accentuation variant is at one end, and a complete “phrase accentuation” at the other, and the distance between them is navigated by gradual degrees of pitch scaling. In between we find examples of utterances where each word clearly has a pitch accent but the last one has the largest excursion (cf. Figure 57), and others where the pitch accents on the non-final words have very small excursion locally, while only that on the last one is readily identifiable (cf. Figure 58).

147 https://osf.io/u2bzp/

148 https://osf.io/xqjkr/

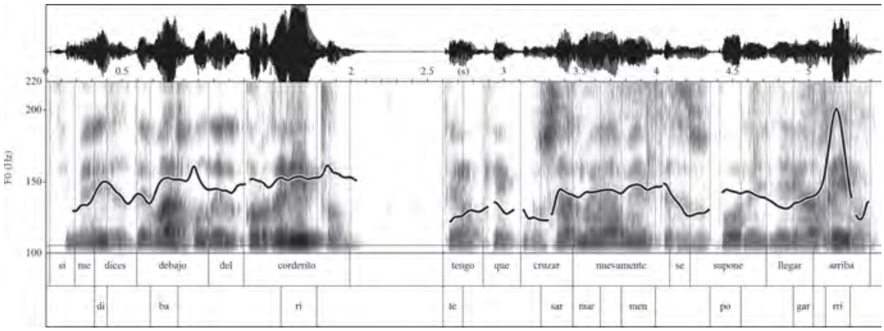


Figure 57: QF16_MT_ES_1011[149] (*si me dices debajo del corderito tengo que cruzar nuevamente se supone llegar arriba* 'if you tell me below the little lamb I have to cross again supposedly get to above'; declarative with weak phrase accentuation).

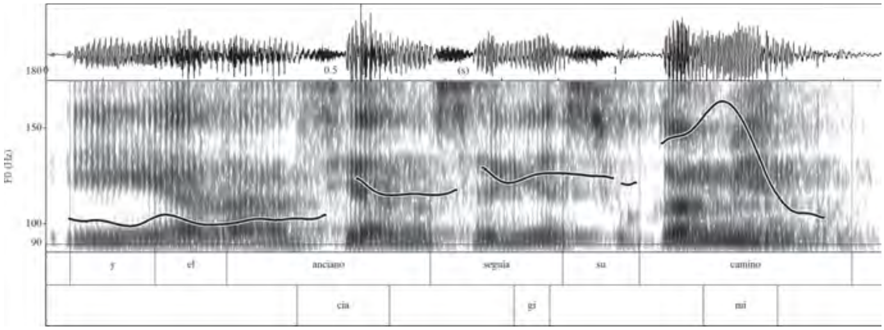


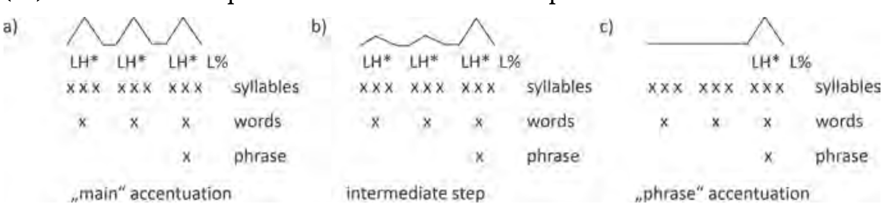
Figure 58: TP03_Cuent_ES_0561[150] (*y el anciano seguía su camino* 'and the old man continued on his way'; declarative with weak "phrase accentuation"). Cf. context (51).

The peak on the final word in these phrase accentuation examples is still aligned within the stressed syllable. But it is also followed by a final fall, indicating the presence of a low boundary tone. Thus, the domain over which "phrase accentuation" occurs in these cases seems to be identifiable with the same for which boundary tones mark the right edge. This should therefore be either the phonological phrase, the intermediate phrase or the intonational phrase, depending on which other evidence there is for an independent existence of these domains. In section 5.2, a more definite analysis will be provided. For our description now, it suffices to say that it is a phrase larger than an individual prosodic word. In the strongest cases of

149 https://osf.io/y75en/
150 https://osf.io/m47pq/

"phrase accentuation", we thus have only one pitch accent per such a phrase, associated with the final word in it. Thus pitch scaling is here used to signal a position of relative strength of the last accentable word in such a phrase compared to all the preceding words. This rightmost prominence relation is in general assumed to hold for phrases in Spanish at least in so-called broad-focus contexts (Hualde & Colina 2014: 266–268; Ladd 2008, cf. also section 3.7.3). It is also assumed to hold in a phrase in the "main" accentuation mode (with all pitch accents having more or less the same scaling) if nothing else changes. The phrase accentuation is thus an optimized expression of this relation via pitch scaling: whereas in the "main" accentuation, final prominence has to be either signaled by other means or is simply the default expectation (Ladd 2008: 257–259, cf. section 3.7.3), in the "phrase accentuation" it is openly signaled. A very schematic comparison between "main" and "phrase" accentuation is given in (49), with more gradually differing intermediate steps omitted.

(49) Schematic comparison between "main" and "phrase" accentuation



The "main" accentuation (49)a) and full "phrase" accentuation (49)c) versions of the same hypothetical utterance here do not differ in their metrical representation, in which the final word in the phrase is assigned the highest prominence in the phrase. This predicts that both versions should cue ambiguously between a reading in which focus is on the final constituent and a "broad-focus" reading. The main accentuation is furthermore partially ambiguous also in respect to a reading where one of the pre-final constituents is in focus, if no deaccentuation occurs (cf. Figures 66 and 67 in section 5.1.3.2). The phrase accentuation would certainly be less frequently expected in such contexts because it cues rightmost prominence more clearly. Thus, the "phrase accentuation" cues a metrical structure which is ambiguous between two information structural readings (focus on the final constituent vs. broad focus), while the "main accentuation" is in principle ambiguous between three readings (focus on the final constituent vs. focus on a pre-final constituent vs. broad focus) and metrical structures with either final or prefinal prominence.

The phrase accentuation optimizes the phrase as well as its final accent, because it treats the phrase as the domain of pitch accent culminativity, implemented partially gradually via scaling. Its pitch movement is also delimitative in that it clearly marks the right edge of the phrase in which the strongest word is final. In this it differs from the main accentuation, which does not signal the extent of the phrase via the pitch contour, but instead optimizes each prosodic word within it. These different optimization strategies are further exemplified in Table 10. It describes the continuum from main accentuation (a), most word-optimizing, to the different modes of phrase accentuation (b, c,and d). The rightmost case (d) is one where all pitch accents associated with stressed syllables are gone and only a pitch movement associated with the phrase edge occurs. This optimizes the phrase by pitch event culminativity, and by delimitation, but the pitch event is no longer located at the strongest position. It does not signal individual prosodic words at all. Some examples for this mode were already given in section 5.1.1.2 and we will also see some more later.[151]

Table 10: Optimization of different prosodic domains by differences in pitch accentuation and excursion.

more word-optimizing* <--> *more phrase-optimizing			
a) each word in a phrase with rightmost prominence has a pitch accent on its stressed syllable with similar excursion	b) each word in a phrase with rightmost prominence has a pitch accent on its stressed syllable, but excursion on the last one is larger than on the preceding ones	c) only the last word in a phrase with rightmost prominence has a pitch accent on its stressed syllable	d) pitch movement corresponding to a boundary tone on the phrase but not necessarily on the stressed syllable
			
most word-optimizing (pitch accent culminative on prosodic word, equal scaling)	both word optimizing and prominent word-optimizing (pitch accent culminative on prosodic word, unequal scaling)	optimizing both the most prominent word and the phrase (pitch accent culminative on phonological phrase and located on strongest position, pitch event delimitative for the phrase)	most phrase-optimizing (pitch movement culminative and delimitative for phonological phrase)

The phrase accentuation variant is not only different from the main accentuation with regards to how it signals the prominence of the final word (via pitch accent

151 Option (d) is not merely a case of option (c) without an additional low boundary tone, as the comparison between speakers ZE55, on the one hand, and XJ45 and NQ01, on the other, in section 5.2.4 will demonstrate.

culminativity and delimitation), but also in how it signals the non-prominence of the non-final words (via a reduction in pitch scaling up to what amounts to deaccentuation). It would be hard to reconcile such a realization with a reading where one of the pre-final constituents is at-issue. Taking this prediction as a point of departure, is it possible to find functional factors that favour the phrase accentuation? In the following, we will explore the possibility of identifying contexts in which it preferably occurs. The phrase accentuation does not occur in the speech of all the speakers studied here. In the data considered here, only 15 speakers produce something like it, and for all of them it is a minority mode of accentuation, while the majority of their utterances is produced in the main accentuation (cf. also its low occurrence in the *Conc* data discussed in 5.1.1.2). It is unknown whether this reflects a true distribution of this accentuation phenomenon among the speakers, or whether it is a sampling artefact, but overall it speaks to the variability that is at the speakers' disposal. As far as contexts favouring phrase accentuation can be identified, they are only potential loci for it, not categorical ones. The following discussion will argue that there are identifiable aspects of contexts that increase the overall likelihood of occurrence of the phrase accentuation, under the assumption that the phrase accentuation is felicitous when no particular highlighting of pre-final constituents is intended or plausible from the context, and when instead either the final constituent or the structure of the higher prosodic unit (the phrase) is to be emphasized. The analysis here is exploratory and looks at individual examples in context. For the future, a more controlled quantitative investigation would be desirable to confirm its results.

5.1.3.1.1 Broad or final narrow focus on an utterance

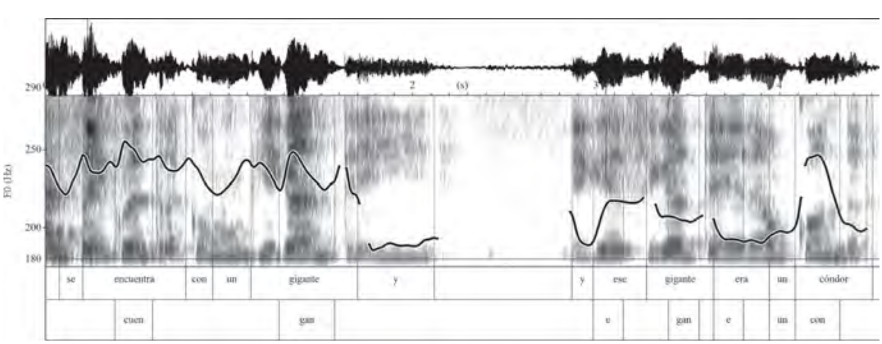


Figure 59: HA30_Cuent_ES_2285[152] (two declaratives, the first with with main, the second with phrase accentuation).

152 https://osf.io/zudme/

(50) ZR29_HA30_Cuent_ES_2268–2361 (context for HA30_Cuent_ES_2285)[153]
*General context: they are telling a story about an old man (*señor*), who wants to go visit his granddaughter and meets several obstacles on the way. ZR29 already told the story, now it is HA30's turn.*

time	**ZR29** *(the one who tells the story first)*	**HA30** *(the one who re-tells the story)*
226.8		en eso luego *so then*
227.7	se encuentra con un gi[gante *he encounters a giant*	
228.5		ah **se] encuentra con un gigante y**
231.6		**y ese gigante era un cóndor** *ah he encounters a giant and that giant was a condor*
233.7	*(they both laugh)*	

(50) is the context for HA30_Cuent_ES_2285 (Figure 59). In the excerpt given in the figure, HA30 first reproduces ZR29's preceding assertion that the old man met a giant (*se encuentra con un gigante y*). Both HA30's and ZR29's utterances here are answers to the QUD WHAT HAPPENED THEN TO THE OLD MAN?, overtly expressed by HA30 at 226.8. The giant (*gigante*) is introduced here for the first time in the story.[154] All accentable words are pitch accented. In the second utterance (*y ese gigante era un cóndor*), only the final word *cóndor* is pitch accented; notably, no relevant pitch movement takes place on *gigante.* The most plausible QUD here is something like WHO OR WHAT WAS THAT GIANT?, with *gigante* given and not at-issue. The giant (*gigante*) is overtly mentioned because it is the topic switched to here (the preceding utterances were predications on the discourse topic *el señor* (the old man), omitted as subject), but it is given and not intended to continue as discourse topic (the story continues with the old man as protagonist, now meeting the condor). Arguably, all of this contributes to the difference in accentuation variants: at 228.5, *gigante* introduces a new discourse referent and is both phrase-final and pitch accented, at 231.6 it is given and not intended to continue as topic so that the whole utterance is something of an aside. It serves to introduce the next relevant

153 https://osf.io/st6d8/

154 Strictly speaking, it was already introduced as a discourse referent when ZR29 first told the story to HA30. But it is now HA30's turn, and her asking for a prompt at 226.8 betrays her ignorance of what comes next, so *gigante* is not salient for her. And even though it is ZR29 who then introduces the giant, it seems likely that at 227.7, HA30 does not treat *gigante* as given because it is now her role in the game to tell the story to a third person.

discourse referent, the condor, which is placed phrase-finally and thus in the right position to be accented even in phrase accentuation.

(51) TP03_KP04_Cuent_ES_0455–0618 (context for Figure 58 (also Figure 21))[155]
*General context: they are telling a story about an old man (*anciano*), who wants to go visit his granddaughter and meets several obstacles on the way. TP03 is telling the story for the first time.*

time	**TP03** *(the one who first tells the story)*
45.5	en eso el anciano *so the old man*
46.9	sacó su lápiz *took out his pencil*
48.8	empezó a dibujar *began to draw*
50.4	una sombra *a shadow*
51.3	tan grande como el gigante *as large as the giant*
52.5	que le dio miedo al gigante y el gigante huyó *so that it frightened the giant and the giant fled*
56.1	**y el anciano seguía su camino** *and the old man continued on his way*
58.5	y durante el trayecto en su camino *and while he was on his way*
60.8	apareció un cóndor *a condor appeared*

A somewhat similar situation obtains in the case of TP03_Cuent_ES_0561 (Figure 58), with the context in (51). TP03 tells a version of the same story as HA30 does in (50), but as the one telling it for the first time. 45.5–52.5 recount how the old man (*anciano*) draws a shadow on the ground that is so large that it frightens the giant into fleeing. 45.5–50.4 answer the QUD WHAT DID THE OLD MAN DO?, with the old man as the discourse topic. In 51.3–52.5, a sub-QUD to that is answered, namely WHAT KIND OF SHADOW WAS IT? (continued in 52.5 with a further sub-QUD, WHAT EFFECT DID THE SHADOW BEING SO LARGE HAVE?), with the shadow (*sombra*) being the discourse topic. With the end of the utterance at 52.5, the sub-QUDs about what the old man did about the giant are answered and that part of the story is concluded.

155 https://osf.io/xp67y/

In 56.1, there is a return to the old man as (given) discourse topic which he remains afterwards (as evidenced by his subsequent pronominalization in 58.5), but the utterance itself is probably best interpreted as answering the QUD WHAT HAPPENED THEN?, which is a sub-QUD only to the overall QUD WHAT HAPPENS IN THE STORY?. An alternative, which would have *el anciano* as contrastive topic, would be to assume a QUD WHAT DID THE OLD MAN DO? as directly subordinate to a QUD asking after a set, WHAT DID THE SHADOW AND THE OLD MAN DO?, which is then answered by answering the two subquestions WHAT DID THE SHADOW DO? and WHAT DID THE OLD MAN DO?. Nothing in the context makes it plausible that such a parallelism between the shadow and the old man (that are two entirely different entities in terms of animacy, prototypical agenthood (cf. Dowty 1991 and relevance to the overall story) is intended here. Another alternative, resulting in narrow focus on *el anciano*, can equally be excluded, because that would assume a QUD WHO WENT ON HIS WAY?, but it is not presupposed in the context that someone went on their way. Yet another, with narrow focus on *camino*, would imply a QUD like WHAT DID THE OLD MAN CONTINUE ON?, but nothing suggests that it is presupposed that he continued on something. In sum, the most parsimonious QUD WHAT HAPPENED THEN?[156] makes the use of the phrase accentuation here plausible, since none of the individual referents or events expressed as accentable words (*anciano, seguía, camino*) are relatively more relevant for answering it than the others, but the utterance as a whole answers it.

(52) ZR29_HA30_Cuent_ES_1415–1645 (context for Figure 56)[157]

*General context: they are telling a story about an old man (*señor*), who wants to go visit his granddaughter and meets several obstacles on the way. ZR29 is telling the story for the first time.*

time	**ZR29** *(the one who tells the story first)*
141.5	en eso de nuevo se toca con el *so then once more he meets the*
145.2	cun-
145.8	cómo se llama *how is it called*
146.3	con el cóndor *with the condor*

156 This QUD is possibly also favoured by the use of the imperfect tense *seguía* here, while the surrounding narrative is told using the perfect tense. Compared to the perfect, the imperfect has been argued to be used to convey narratively backgrounded information (Hopper 1979; López-Ortega 2000). That would arguably make an internal information structural partition (with narrow focus on one of the elements) less likely.

157 https://osf.io/nvc23/

147.9 y
148.4 coge s- el anciano saca su maní se lo lanza
and the old man grabs- takes out his peanut throws it to it
151.8 como
because
152.5 al-
154.0 al condor le gusta tanto los- sus- (0.16) los manís
the condor likes the peanuts so much
157.2 **se lanza hasta mientras pasa el anciano**
it launches itself [at the peanuts] while the old man passes by
160.9 llega a su casa de su nieta de- de (0.13) saca el na- la naranja
arrives at the house of his granddaughter of of (0.13) takes out the orange

Again a similar argument can be made for ZR29_Cuent_ES_1572 (Figure 56), with context (52). Here, ZR29 tells the story for the first time, and describes the encounter of the old man with the condor: he takes out the peanuts he carries with him and throws them to the condor, who likes peanuts very much. This leaves him distracted so that the old man can continue on his way. The utterances from 141.5–145.2 as well as 148.4 have the old man as discourse topic, answering a QUD like WHAT DOES THE OLD MAN DO NEXT?. 151.8–154.0 as well as the first phrase in 157.2 then have the condor as topic, answering first WHAT IS THE CONDOR'S MOTIVATION? and then WHAT DOES THE CONDOR DO?. This last QUD is answered only in the first phrase in 157.2, by *se lanza* 'he throws himself [at the peanuts]'. As can be seen in Figure 56, this first phrase is separated from the rest by a high boundary tone after *lanza;* the rest of the utterance taking place in a separate phrase, *hasta mientras pasa el anciano* 'while the old man moves along' is then phrase accentuated. This might at first sight be taken as a topic switch back to the old man, *el anciano*, which is placed rightmost and thus receives the strongest pitch accent. With this QUD structure, *el anciano* would then be a contrastive topic (cf. Büring 2003; Roberts 2012b). However, such sentence-final subject NPs as *el anciano* here have been proposed to signal decreased salience or continued topics in Spanish (Ocampo 2003, 2010), they are called *antitopics* by Lambrecht (1994) and are not contrastive topics. The lack of parallelism in the utterances about the condor and the old man[158] also does not

158 Syntactically, 151.8–157.2 form a complex sentence together:
(i) [[como al condor$_i$ le gusta tanto los manís] *i* se lanza [mientras pasa el anciano]]

This sentence has an empty subject referring to the condor, realized only in the causal subordinate clause *como al condor le gusta tanto los manís*, while *el anciano* is the realized subject in the other subordinate clause headed by *hasta mientras*. A parallel QUD structure like WHAT DID THE CONDOR support an interpretation as contrastive topics. It seems much more likely that *el anciano* here simply realizes a given referent that only subsequently will be continued as a topic in the next utterance (160.9), meaning that the phrase answers a QUD like "what happened meanwhile?", which is not a sister to, but subordinate to the current QUD, "what does the condor do?", and thus is not at-issue in this utterance.[159] An alternative QUD asking only after the old man ("who passed along meanwhile?") is out because the action of passing along is not presupposed.

As an intermediate result, the three context examples have shown the use of phrase accentuation in contexts where only the final element in the phrase is foregrounded (50), where the entire phrase as a whole is the answer to the current QUD (51), and where the entire phrase as a whole is arguably the answer to a QUD that is not current, i.e. backgrounded (52).

AND THE OLD MAN DO? would thus be answered once by this empty subject in the main clause ([el cóndor] *se lanza*) and once by the finally placed but realized subject in the subordinate clause (*mientras pasa el anciano*). If this were the intended QUD structure, the syntax would therefore be maximally unhelpful in exposing its parallelism.

159 Riester (2019: 180–181) points out that many types of non-at-issue material such as adjuncts, evidentials and evaluatives only aren't at-issue with respect to the current QUD, but pose an answer to a different (subordinate) QUD. In many cases where the non-at-issue material is utterance-final, it can be difficult to establish whether it is really not at-issue or simply asserting the answer to the next QUD. In our case, it seems decidable:

(i) Q1: What does the condor do?
A1: *se lanza*
Q2*: What happens meanwhile?*
A2: hasta mientras pasa el anciano

(ii) Q1: What happens meanwhile?
A1: *hasta mientras pasa el anciano*
Q2: What does the condor do?
A2: *se lanza*

While (i) seems a plausible question-answer structure for ZR29_Cuent_ES_1572, (ii) is incoherent. This would also be the case if the utterance was ordered as *hasta mientras pasa el anciano (el condor) se lanza*. That's because the clause headed by *hasta mientras* is not just syntactically subordinate to another, but also denotes an action or event that must be interpreted in relation to another action or event (with the nature of the relation here being their temporal coincidence), but not the other way around. It would also be possible to utter *se lanza hasta mientras pasa el anciano* with the *hasta mientras* part at-issue, but only as answer to a current QUD such as WHEN DOES HE THROW HIMSELF FORWARD?, which implies a context set that has already been sufficiently restricted such that in all possible worlds contained in it, the proposition "the condor throws himself forward" is true; in other words, when the QUD DOES HE THROW HIMSELF FORWARD? has previously been answered affirmatively and it is thus still superordinate to the current QUD.

5.1.3.1.2 Coherent phrasing: Complements to relative clauses and others

The phrase accentuation also occurs on phrases that realize complements to relative clauses, such as in MS27_Cuent_ES_1520 (Figure 55), with context (53). MS27 is re-telling the story after being told it by CF28.

(53) MS27_CF28_Cuent_ES_1409–1582 (context for Figure 55)[160]
*General context: they are telling a story about a giant (*gigante*), who meets several animals. MS27 is now re-telling the story, after she was told it for the first time by CF28.*

time	**MS27** *(the one who re-tells the story)*
140.9	y el gigante
142.3	cogió su sombrero y siguió su camino
	and the giant grabbed his hat and continued on his way
145.5	en eso
147.4	ve
148.4	un funeral
	so then he sees a funeral
149.8	donde estaban las flores
	where the flowers were
152.0	**y que los animales estaban maltratando**
	and which the animals were mistreating
155.3	eh el gigante decide matar a los animales
	uh the giant decides to kill the animals

In 145.5–148.4 she recounts how the giant gets to a funeral (*un funeral*). In 149.8 and 152.0 then follow two utterances that realize relative clause complements, first (149.8) to *un funeral* itself (un funeral [donde estaban las flores]) and then (152.0) to the NP *las flores* that is part of the first relative clause (las flores [(y) que los animales estaban maltratando]). Both of these relative clause phrases are realized in phrase accentuation. It seems plausible that this is due to an effort to prosodically optimize such complement clauses as coherent (single) prosodic phrases in order to signal their relation as a whole to the preceding relative head; this is in principle better achieved with a phrase that has only a single pitch accent that also serves to delimit the phrase than by one with several pitch accents. It seems possible to prefer achieving such coherence over accenting each accentable word, even

160 https://osf.io/uqptz/

if those words introduce new referents.[161] This seems at least a contributing factor here. However, looking at more examples from this corpus and other corpora, we see that phrase accentuation is used (in particular, but not only) by MS27 and CF28 in a wide variety of utterances in their *Cuento* corpus that are certainly not all relative clause complements. For instance, it also occurs in the utterance following the two relative clause complements just discussed, in 155.3. It also appears in the initial utterances by both of them, i.e. when CF28 begins telling the story for the first time (cf. Figure 60) and at the beginning of MS27's retelling of it (cf. Figure 61). Both of these display a somewhat remarkable initial intonation: a pitch peak with strong excursion is realized on *había un* or *había*, then a steep fall occurs which ends on the first syllable of *gigante* and pitch then continues at a lower overall level. This pronounced pitch movement could result from an LH* pitch accent followed by a high boundary tone, and the following steep fall is then due to the leftward alignment of the next upcoming L tone that is already familiar from section 5.1.1.

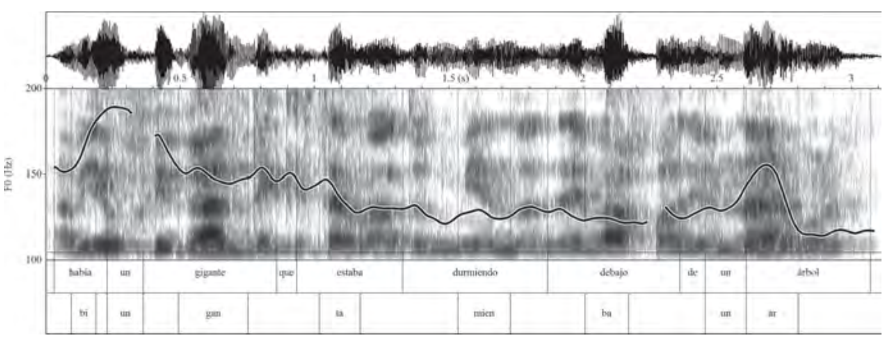


Figure 60: CF28_Cuent_ES_0012[162] (*había un gigante que estaba durmiendo debajo de un árbol* 'there was a giant who was sleeping under a tree'; declarative with phrase accentuation).

161 Although all the discourse referents mentioned in these utterances are introduced for the first time in this re-telling of the story, they are realized as definite NPs, as if given. It is possible that the speaker treats the re-telling of the story less like her own telling, pretending that the listeners don't know it, and instead more like a memory task she has to perform: how good she is at remembering the story she has been told, an ability which the listeners (the experimenters) are assessing by comparing it to the original version known to them. Great care was always taken by the experimenters to avoid the impression that the productions by the speakers would be assessed in such a way; we usually said at the beginning of the experiment session that this was explicitly not the case and also on occasion during the course of the session. However, since some speakers (like MS27 and CF28 here) were students at the time of recording, it is not impossible they associated the experimental tasks with similar tasks known from education contexts, at which they would be assessed in their performance.
162 https://osf.io/mx2k7/

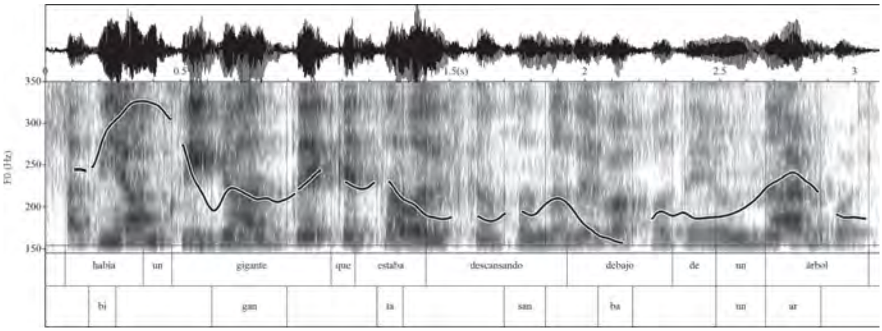


Figure 61: MS27_Cuent_1207[163] (*había un gigante que estaba descansando debajo de un árbol* 'there was a giant who was resting under a tree'; declarative with phrase accentuation).

Such a phrasing is odd in terms of an expected correspondence to syntactic or information structural categories. It might more readily be expected to separate *había un gigante* from the rest, mapping prosody to a topic-comment structure. As it is, the proposed phrasing goes against a mapping of prosody with information structure and syntax, if the boundary really occurs between the determiner *un* and the noun *gigante*. Yet in both utterances under discussion here, the picture seems to unambiguously point to this conclusion. I will call such increased scaling on the initial word in a large phrase an "initial boost". It will reoccur in the double topic-utterances treated in section 5.2.

The remainder of these utterances exhibit phrase accentuation. It is unambiguous that both these utterances end with a (LH*) pitch accent on the final word *árbol*, followed by a low final boundary tone, L%. In CF28's version, the excursion of none of the pitch movements on any of the stressed syllables between *gigante* and *árbol* really exceeds the fluctuation due to microprosody happening also on all the other syllables, and globally, pitch follows a decaying trajectory until a lowest point shortly before the final stressed syllable on *árbol*. On *gigante*, no pitch accent on the stressed syllable can be identified, but pitch is suspended perhaps due to a high boundary tone at its end, which auditory inspection supports. Thus, here at least the stretch *que estaba durmiendo debajo de un árbol* seems to form a single phrase with only one final pitch accent. MS27's utterance is an intermediate version where compressed intermittent pitch movements likely due to intonational tones can be more readily identified. At the end of both *gigante* and of *descansando*, final rises occur that are not due to microprosody, but are also not aligned with the stressed syllables of these words. These are cases of the rightmost type (d) of phrase accen-

163 https://osf.io/9c6nt/

tuation in Table 10, without pitch accents associated with prominent positions and only phrase-delimiting tonal movement. They can be seen as a type of continuation rise; in final position (at the end of the utterance), they do not occur here because of the final L%. Cf. the analysis of rising and falling contours in the Quechua data in section 6.1.1. A further example is Figure 62.

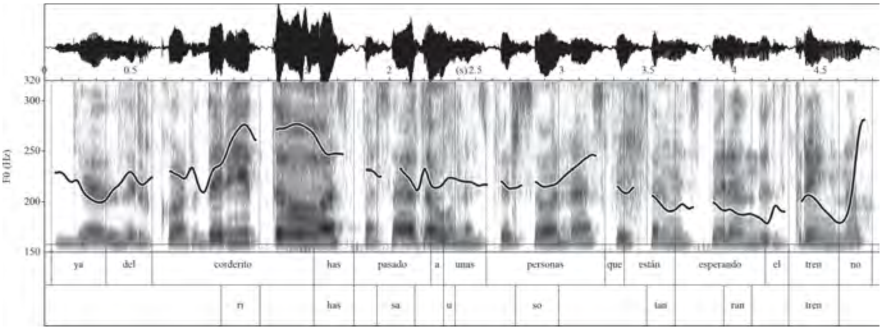


Figure 62: SO39_MT_ES_3351[164] (*ya del corderito has pasado a unas personas que están esperando el tren, no* 'ok from the little lamb you passed some people that are waiting for the train, right'; confirmation-seeking question with "phrase accentuation").

Here there is a pitch accent with strong excursion followed by a final high boundary tone on *corderito*, then the stretch *has pasado a una personas* is phrase accented only with a final high boundary tone, all other movement being due to microprosody, and the final stretch *que están esperando el tren no* is phrase accented with a pitch accent only on *tren*, followed by the LH% final boundary tone for polar tag questions (see section 5.1.2.1). The phrasing produced in these two examples (MS27_Cuent_ES_1207, Figure 61 and SO39_MT_ES_3351, Figure 62) and others like them further corroborates the suggestion that phrase accentuation can occur when it is worthwhile to signal a prosodic structure that is above the level of the individual prosodic word as a whole instead of signaling its internal structure. In SO39_MT_ES_3351, it helps signaling the separation between an NP refering to an anchoring landmark (*del corderito*), and a proposition that serves as an instruction for how to proceed from that landmark onwards (similar to topic-comment structures), and within this instruction it further helps signalling the relative complement clause (*que están esperando el tren*) as belonging as a whole to its preceding head noun (*unas personas*). In MS27_Cuent_ES_1207 it also separates a head noun (*un gigante*) from the complement clause (*que estaba descansando debajo de*

164 https://osf.io/eut2x/

un árbol), and within this complement clause it further separates the verb phrase from the locative adverbial (see (54) and (55), where square brackets indicate right phrase boundaries, syllables in italic are stressed and syllables in bold also accented[165]).

(54)	ya del corde**ri**to]	has pa*sa*do a *u*nas per*so*nas]	que es*tán* espe*ran*do el **tren** no]
	LH*H-	H-	LH* LH%

(55)	ha**bí**a]	*un* gi*gan*te]	que es*ta*ba descan*san*do]	de*ba*jo de un **ár**bol]
	LH*H-	H-	H-	LH* L%

I argue that this is then the generalization we can make: the phrase accentuation is preferentially used whenever a structure with rightmost prominence needs to be signaled that is intermediate between that of the level of the individual prosodic words and the whole utterance, and if this can be done at the expense of deemphasizing individual prosodic words and the discourse referents they encode (up to producing only phrase-edge tones and no pitch accents at all) without losing contextually relevant information. This entails that the phrase accentuation will preferentially be used only in phrases where either final focus or broad focus interpretations are allowed from the context, as observed. It also strikes a parallel to the phenomenon of 'dephrasing' described for 'edge-prominent' languages like Japanese and Korean (Pierrehumbert & Beckman 1988; Jun 1993, 2005b; Venditti et al. 1996; Ladd 2008; Igarashi 2015; cf. section 3.4.6): words that are less prominent are not phrased separately but in a larger phrase together with a more prominent word. In comparison with the Japanese case, it is mainly the position of the prominent word within the larger phrase that differs: there, the most prominent word is leftmost in the phrase, whereas here it is rightmost. It also ties in with another relevant observation: the phrase accentuation occurs more frequently in the *Cuento* than in the *Maptask* corpora, and this is partly because utterances are generally longer in *Cuento* than in *Maptask*.[166] In those longer utterances the phrase

165 In sections 5.2 and 5.3 it will be shown how the alignment of the H as a phrasal boundary tone nonfinally and with the stressed syllable finally results naturally from the additional presence of an IP-level boundary tone. There, the categories of the prosodic hierarchy involved will also be discussed in more detail.

166 Most likely, this is in itself due to better possibilities for planning ahead. In *Maptask*, speakers interact with each other constantly, must adapt to changing epistemic conditions, and can update the common ground only as a result of constant negotiation with each other. In *Cuento*, on the other hand, the conversational mode is much more monological. The first speaker in particular, but also the second, have all the time they need to tell the story. Unless they make mistakes, their right to speak and to update CG is in no danger of being contested, the stage is theirs. This opens up accentuation is more widespread, and this makes sense if it helps to signal intermediate phrasing structures that are mapped to information structure and syntax: they are simply absent in shorter and less structurally complex utterances. Note that in Quechua *Cuento* (cf. section 6.4.1), very similar conditions obtain, and they are also argued to be mainly responsible for the observed phrasing of larger speech sequences, with a rise-falling pitch contour that is in fact very similar to that of phrase-final phrase accentuation seen here.

Optionality obviously plays a large role in the use of the phrase accentuation in these spontaneous data. Not only regarding its occurrence at all, but also regarding the phrasal separations it creates: on virtually the same sentence under very similar context conditions, CF28_Cuent_ES_0012 (Figure 60) produces the entire relative clause complement in one phrase, while MS27_Cuent_ES_1207 (Figure 61) divides it further in two. In section 5.2, more complex and somewhat more controlled data will be analyzed. The results will suggest that as here, while certain phrasing divisions due to information structure are quite general, there is still a considerable space for individual variation.

5.1.3.1.3 Phrase accentuation as (individual) default?

So far, we discussed phrase accentuation (with either a single pitch accent or only a boundary tone) as a phenomenon whose occurrence is influenced by discourse contextual and information structural factors, and that is preferentially used when units of discourse that are larger than single referents and the expressions they encode (at the size of prosodic words) are to be emphasized at the expense of these smaller units. Such a view does not predict phrase accentuation to occur on single words. This is true for the *Maptask* and *Cuento* corpora studied here. However, there is some evidence in the *Elqud* corpus that for some speakers, in particular NQ01, XJ45 and ZE55, phrase accentuation has become almost generalized. Especially XJ45 produces phrases with only boundary tones almost as a default in *elqud*, and this not only on groups of several words, but also occasionally on single prosodic words. He realizes several words together in one such phrase even when from the discourse context it is clear that not the final but a prefinal word in this phrase bears the highest information load (i.e., with pre-final narrow instead of final narrow or

the possibility for planning and producing utterances containing larger coherent chunks of information than in *Maptask*. Cf. section 6.4 where a similar argument is made to account for prosodic differences between utterances from Quechua *Cuento* and *Maptask*, with absence of utterance-internal IS-partition in *Cuento* there resulting in contours that bear many similarities with the phrase accentuation utterances here.

broad focus). This can be seen exemplarily in his utterances XJ45_ELQUD_ES_16 (Figure 63), XJ45_ELQUD_ES_19B (Figure 64), and XJ45_ELQUD_ES_17 (Figure 65).

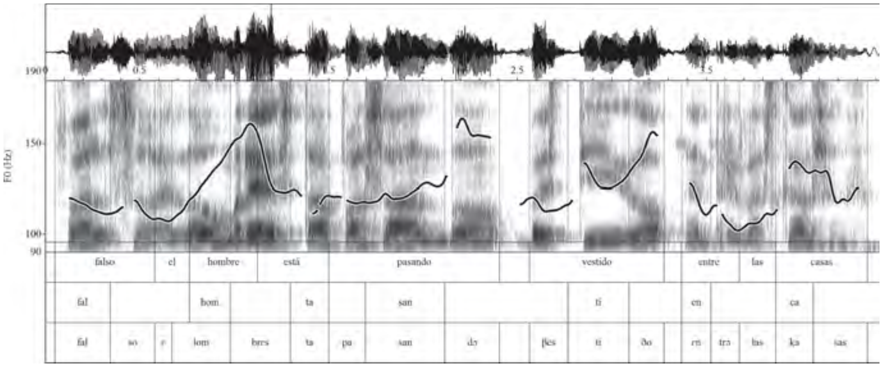


Figure 63: XJ45_ELQUD_ES_16[167] (*falso el hombre está pasando vestido entre las casas* 'wrong the man is passing between the houses with clothes on'; declarative with phrase accentuation).

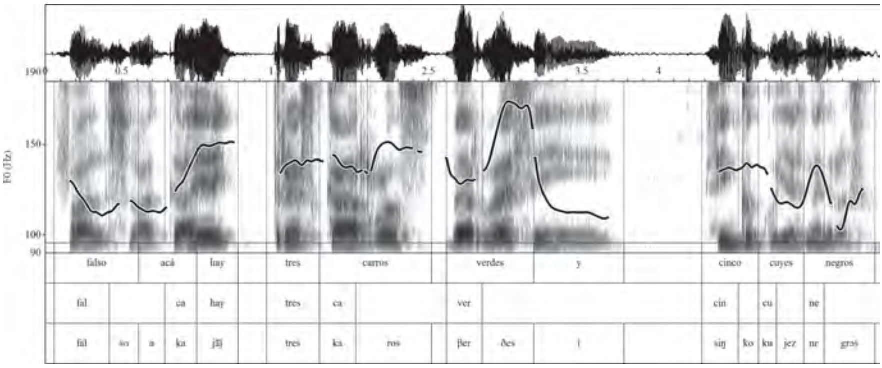


Figure 64: XJ45_ELQUD_ES_19B[168] (*falso acá hay tres carros verdes y cinco cuyes negros* 'wrong here there are three green cars and five black guinea pigs'; declarative with "phrase accentuation").

In XJ45_ELQUD_ES_16 (Figure 63), the only word that has a pitch peak aligned with its stressed syllable is the final one, *casas*.[169] All other words are either produced entirely as part of a low pitch stretch (*falso, está, entre*), or a final boundary rise

167 https://osf.io/h6wjb/

168 https://osf.io/ayp93/

169 As stated above, this will be analyzed as a result of the additional presence of the L% in sections 5.2 and 5.3.

is realized on their final posttonic syllable, with the preceding stressed syllable clearly not the target of the peak (*hombre, pasando, vestido*). Instead, the elbow starting the rise is placed somewhere on the stressed syllable. Even on the single word *vestido* such a phrasal contour is realized. *Vestido* here encodes the predicate that constitutes the contrast to the experimental provocation,[170] and it is likely that producing it in its own separate phrase is a means for signalling this.

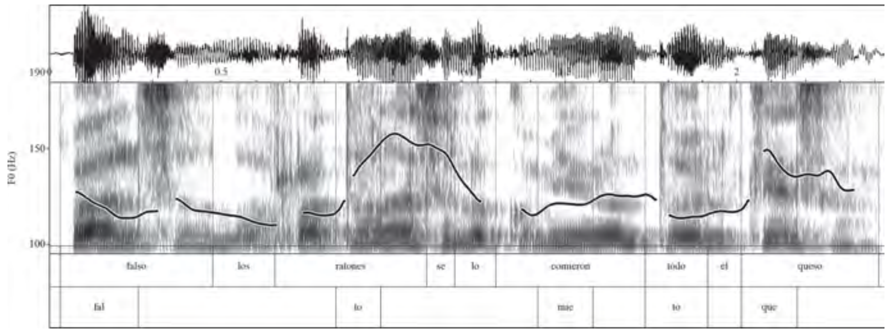


Figure 65: XJ45_ELQUD_ES_17[171] (*falso los ratones se lo comieron todo el queso* 'wrong the mice ate all the cheese'; declarative with "phrase accentuation").

In XJ45_ELQUD_ES_19B (Figure 64), *falso acá hay*, *tres carros*, *verdes* and *(y) cinco* each are realized as a separate phrase with a finally rising phrase accentuation contour, whereas in the last phrase, *cuyes negros*, the peak is again located on the final stressed syllable. Note that the final peak on *verdes* is scaled much higher than that at the end of the other phrases, coinciding with the separation between the two conjoined sentences that make up the utterance together. Section 5.2 will discuss evidence that the scaling of pitch peaks, both of accents and boundary tones, is used to signal boundaries at different levels of the prosodic hierarchy. Although they should both be the locus of the correction, the numerals *tres* and *cinco* are treated differently by the prosody here.[172] While *tres* is clearly realized as part of the low stretch in the phrase *tres carros* (with final peak on the posttonic syl-

170 The provocation is *El hombre está pasando calato entre las casas* "the man walks naked between the houses", while the animated image shows a man in formal attire and with a bunch of flowers moving between houses. The expectation was that a correction would concentrate on the state of dress of the man.
171 https://osf.io/7znmr/
172 The provocation is *aquí hay un carro verde y tres cuyes negros* "here there is one green car and three black guinea pigs", while the image shows three green cars and five black guinea pigs. The expectation was that the correction would concentrate on the number of cars and guinea pigs.

lable of *carros*), *cinco* is realized as the rightmost word in such a phrase, with a final peak (and a new phrase clearly beginning with *cuyes*). That is to say, *cinco* is aligned prominently at the right edge of the phrase, while *tres* is realized at its left edge. It seems that in *Elqud* in general, the relation between what logically is the locus of the correction and its prosodic realization is not straightforward, but instead mediated by both prosodic and syntactic structure as well as dependent upon interpretation of the context by the speaker, and even then the result is still subject to further variation (see section 5.2 for a detailed analysis of a subset of the *Elqud* utterances).

In XJ45_ELQUD_ES_17 (Figure 65), the relevant section concerns the last noun phrase, *todo el queso*. It is realized with phrase accentuation, with the quantifier *todo* realized completely on the low stretch before the pitch peak. Yet it is at the same time the unique location for the correction relative to the provocation, which is *los ratones se han comido un poco del queso* "the mice have eaten a bit of the cheese", while the animated visual stimulus shows them to have the whole cheese. This is a further point in case that the mapping between information structure, metrical structure, and its cueing via pitch should best be described in preferential, but not categorical, terms. All of the examples here demonstrate how for XJ45, the phrase accentuation variant is a virtual default, which seems to also mean that it lacks some of the functionality it has for other speakers. XJ45's *Elqud* examples represent a shift to a edge-prominent prosody instead of the head-prominent one exemplified by the main accentuation (in Jun 2005d, 2014b's terminology, see also sections 3.4 and 7.4.1), but using the same underlying tone sequence. In sections 5.2 and 5.3, this shift will be analyzed also in the context of what it implies for the prosodic hierarchy, and section 7.4 will also establish further connections to Quechua. In the following we will discuss utterances in which context conditions that impose an internal IS-partition on utterances in which focus is not final (like in XJ45's utterances here) do have an effect on prosody.

5.1.3.2 Deaccentuation in declaratives

In the Huari Spanish data, deaccentuation does not only occur in alternative questions or wh-questions (cf. 5.1.2.3 and 5.1.2.4), but also in declaratives. It may occur on postnuclear material, i.e. when the highest prominence is realized nonfinally within a larger phrase or utterance. Typically, as in other Spanish varieties, this can happen when context allows for an interpretation whereby an utterance is partitioned internally so that at-issue material precedes non-at-issue material. It is clearly an optional process. As just seen, in XJ45's phrase accentuation examples, these context conditions do not effect deaccentuation. To demonstrate that this optionality also exists in utterances with main accentuation, I first provide two examples within their contexts in which a given or backgrounded constitutent is preceded by one which is narrowly focussed and where no deaccentuation occurs.

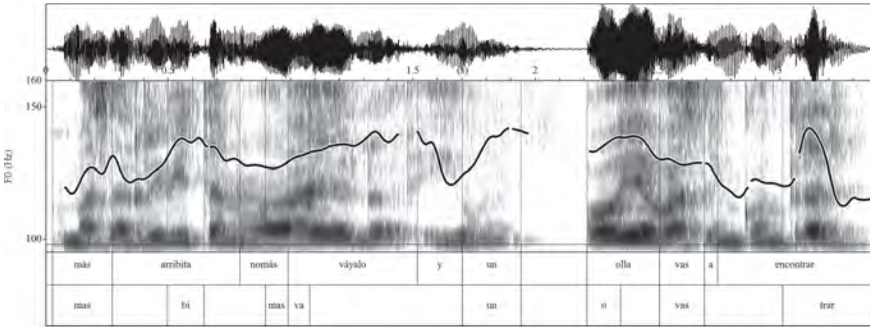


Figure 66: OA32_MT_ES_1443[173] (declarative with main accentuation). Cf. context (56).

(56) XU31_OA32_MT_ES_1344–1476 (context for Figure 66)[174]

time	**OA32** *(the one with the path)*
134.4	y un zorro vas a encontrar
	and a fox you'll find
136.9	y por el
	and by the
138.2	ahí vas a encontrar zorro y
	there you'll find fox and
139.7	debajo del zorro nomás ahí hay y ahí encontrar va
	just below the fox there it is and there you will find
142.0	un nube
	a cloud
143.2	de nube
144.3	más arribita nomás váyalo y un
	from the cloud just a little further above go and a
146.4	LH* H- LH* L%
	olla vas a encontrar
	pot you'll find

The first is OA32_MT_ES_1443 (Figure 66), with context (56). The relevant part is *y un olla vas a encontrar*. The context shows that *vas a encontrar* has been uttered

173 https://osf.io/m7vuf/
174 https://osf.io/udz93/

before and that OA32 formulates the introduction of each new successive landmark referent in the maptask with a variation upon that phrase. It is therefore plausible to take *vas a encontrar* as backgrounded here because it is already part of the corresponding QUD (i.e. not at-issue) which asks for each landmark, something like "what are you going to find?". The answer to this is *olla*,[175] which correspondingly would be in narrow focus. Despite this, there is a very clearly identifiable LH* pitch accent on *encontrar*.

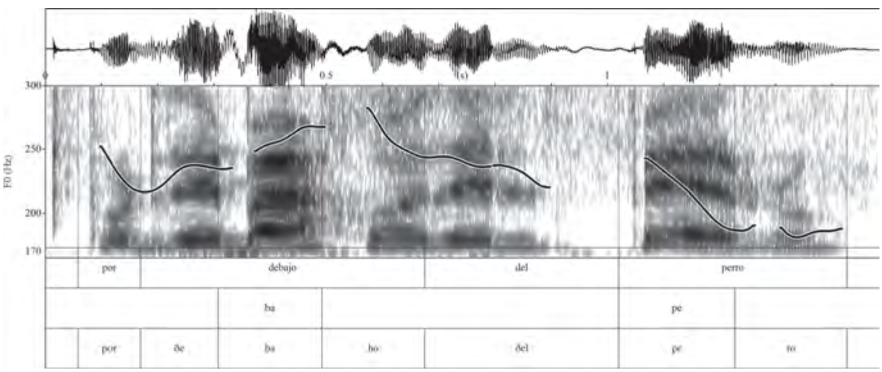


Figure 67: SO39_MT_ES_0743[176] (declarative with main accentuation). Cf. (57).

(57) SO39_MD40_MT_ES_0562–0775 (context for Figure 67)[177]

time	**SO39** *(the one with the path)*	**MD40** *(the one without the path)*
56.2	con por la persona que está agarrando su bolsa *with by the person holding their bag*	

175 Note the rise on the preceding article *un* as well as the short break before *olla*. This seems quite different to hesitations in these corpora, where pitch characteristically drops to low on the element before the "moment of interruption" (cf. Ginzburg et al. 2014, i.e., here on *un*). That element is usually severely lengthened and then optionally followed by a silent break or filled pause before the "continuation", i.e. the element that continues the "normal" flow of speech. See *del murciélago* in Figure 23 for an example. In contrast, here *un* is produced with a strong pitch rise, indicating a high boundary tone, and it is not lengthened to the degree expectable in hesitations. An interesting interpretation might be that here a constituent is cued as aligned with the left edge of a phrase to signal that it is in focus.

176 https://osf.io/uqa68/

177 https://osf.io/qnjp6/

60.4	por su [encima *above them*	
60.6		por de]bajo tá[178] *it's below*
61.7	no por [en]cima *no above*	
62.0		[por en-] *abo-*
62.7	*(laughter, whispering (9.4 s))*	
72.1		por encima del perro *above the dog*
74.3	LH* LH* L% **por debajo del perro** *below the dog*	
75.9		ya por debajo del perro *right below the dog*

The second example is SO39_MT_ES_0743 (Figure 67), with context (57). It shows that it is possible to realize pitch accents not only on words following prefinal narrowly focused constituents, but also following narrow corrective focus (that is, if the answer to the current QUD asks narrowly for a constituent and the assertion made by the speaker bears the [REVERSE] operation with regard to the proposition on the table, cf. Farkas & Bruce 2010). At 72.1, MD40 asks whether the path leads above the dog. This puts the proposition "above the dog" on the table, with a bias for confirmation, according to Farkas & Bruce (2010). By accepting this proposition as on the table, SO39 also accomodates to the presuppositions MD40's utterance makes: the existence proposition that the dog exists (as a landmark) and a proposition that the dog is the relevant upcoming landmark. When SO39 responds at 74.3 without challenging the presupposition, *el perro* is thus backgrounded, because the discourse referent it refers to is already given and because the current QUD must already include it: *por debajo del perro* is a correction to the proposition on the table, "above the dog?", but at-issue as the locus of the correction (or the [REVERSE] relation) is only the locative relation, not the landmark referent "the dog" itself, since it stays the same in the question and the correcting response and is presupposed.

178 A shortened form of "está". There is no acoustic trace of the first syllable. As can be seen from several other examples, this is not a one-time occurrence but seems to be the usual realization of this verb form for several of the younger speakers (e.g. OA32, OV37, MD40).

Thus from the context *por debajo* here has corrective focus, while *del perro* is backgrounded. Nevertheless, *perro* is pitch accented here with an LH* (with the L truncated, cf. section 5.1.1.2). An utterance with pitch accents on each prosodic word, i.e. in the main accentuation, is therefore ambiguous between different information structural configurations. This suggests that broadly, a similar relation between prosodic form and information structure holds for Huari Spanish as has been discussed for other varieties of Spanish in section 3.7.3.1. This is a relation mediated via metrical structure and probability or preference. In what follows we will consider cases of similar contexts where deaccentuation does take place.

5.1.3.2.1 Deaccentuation in reversals vs. on given material

The contexts for the two utterances in which we just observed that deaccentuation did not take place have been treated differently in the literature. Deaccentuation after corrective focus (as in SO39_MT_ES_0743, cf. below for what corrective focus is in terms of the model by Farkas & Bruce 2010) has been attested previously for Spanish in Hualde (2002); Gabriel (2007); Vanrell & Fernández Soriano (2018), amongst others. It has to be kept apart from deaccentuation of repeated (given) material in the absence of correction (as in OA32_MT_ES_1443), which Cruttenden (2006); Ladd (2008); Hualde & Colina (2014) all agree is at best marginal in Spanish, in contrast to e.g. English or German. In both Hualde (2002) and Vanrell & Fernández Soriano (2018), the phonetic variability between strongly reduced and fully absent pitch accents postfocally has also been noted. It is there treated as a single phenomenon with an explicitly agnostic stance as to whether it really constitutes deletion of accents phonologically. Here this holds as well: as with the phrase accentuation, some cases exist that are intermediate in their phonetic realization, where the nuclear pitch accent is still followed by further ones, but those are severely reduced in scaling. I will refer to the whole phenomenon here by deaccentuation for easier reference, but note that it seems in principle easier to derive a totally flat pitch contour as a variant realization for accents somehow marked as reduced in scaling phonologically than to explain why phonologically fully deleted accents should manifest a variant realization as severely compressed ones. Thus postnuclear reduction might be the better umbrella term. In the context of intonation in Germanic languages, where deaccentuation has longer been accepted than for Spanish, it has also been shown to variably manifest as more or less strong pitch compression/reduction (cf. Kügler & Féry 2017 for German), and that accented[179] positions are preserved by other prosodic means (cf. Beaver et al. 2007 for English).

179 Accented, not just stressed, positions. Beaver et al. (2007) compared the acoustic correlates of stressed syllables of words bearing second-occurrence focus (usually said to be deaccented when

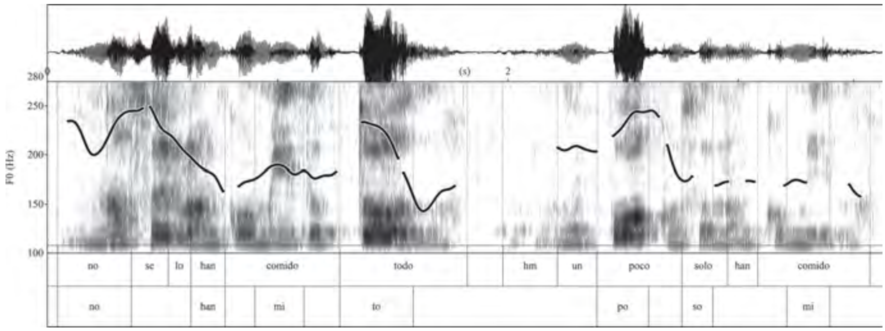


Figure 68: ZE55_ELQUD_ES_55[180] (*no se lo han comido todo hm un poco solo han comido* 'they haven't eaten it all hm only a bit they've eaten'; declarative with deaccentuation after *poco*).

It is thus quite possible that the absence of deaccentuation in contexts like that of OA32_MT_ES_1443 (56) is much less variable than in those of SO39_MT_ES_0743 (57), and that a difference in relative polarity between provocation and response is an important factor making deaccentuation more probable also in Huari Spanish. More specifically, a response might have to be a *partial reversal* (responding to a polar question) or *partial denial* (responding to an assertion; Farkas & Bruce 2010: 100–102, 105) of the provocation, with the contested material realized prefinally, and other material following, for this following material then possibly to be deaccentuated. Just as with the phrase accentuation, only some speakers deaccent in our data. In the Spanish *Maptask* and *Cuento* corpora, it's only found with TP03, KP04, SG15, QF16 and ZR29. As with the phrase accentuation, it is hard to say whether this means that this is really interspeaker variation, but as we just saw, at least the two speakers OA32 and SO39 have been found not to deaccentuate in contexts where it could be expected.

In comparison, consider ZE55_ELQUD_55 and SG15_MT_ES_1815 (Figures 68 and 69), the latter with context (58). ZE55_ELQUD_55 (Figure 68) is an example of a partial denial with deaccentuation from *Elqud*. The provocation stated that the mice had eaten all of the cheese while the visual stimulus showed them to have eaten only some of it. In the first part of the response, the provocation is denied (*no se lo han comido todo*). The second then specifies the alteration necessary for the

following the first focus) with those of words bearing no focus at all, and found significant difference not in pitch, but in duration and intensity between them. Stressed but un- or deaccented syllables in Spanish have also been found to preserve durational and intensity correlates of stress (Ortega-Llebaria & Prieto 2007, 2011; Torreira et al. 2014).

180 https://osf.io/vxceg/

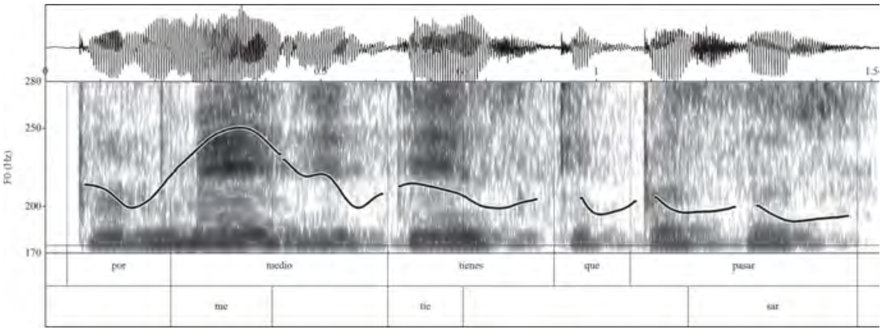


Figure 69: SG15_MT_ES_1815[181] (*por medio tienes que pasar* 'through the middle you have to go'; declarative with deaccentuation after *medio*). Cf. (58) for context.

speaker to accept the proposition. The material denoting the contested part of the proposition on the table is realized first and with an exhaustivity marker (*un poco solo*), and the following material denoting the uncontested part is deaccented.

(58) SG15_QF16_MT_ES_1723–1830 (context for Figure 69)[182]
General context (cf. Figures 190 and 191 in Appendix B*): the lamb, the millionaire, the rock, the bat (*murciélago*) and the dungheap are in different locations in the two maps. That is why QF16 cannot understand SG15's instructions well when told to go between the bat and the pot (*por el medio[183] del murciélago por lado de olla*) after passing underneath the skunk; that is*

181 https://osf.io/skfvr/

182 https://osf.io/ntxpa/

183 SG15 here consistently uses por el medio del murciélago on its own to mean "between the bat [and something else, in this case the pot]". Possibly, this is a calque from Quechua, where in the maptasks, X-pa (Y-wan) chawpi-n-pa (X-GEN (Y-INST) middle-3-GEN) is sometimes used to mean "between X and Y" (lit. "in the middle of X with Y") without uttering the Y-part in brackets if it was mentioned previously. Chawpi by itself means "center", "middle", "intermediary", "point of separation" (Parker & Chavez Reyes 1976: 52; Carranza Romero 2003: 50). It is not clear whether QF16 completely understands this usage, because at 203.3 he asks, entonces cruzo por la mitad del animal, del murciélago? "so then I cross through the middle of the animal, the bat?". SG15 gives a confirmation token then, but on her map, the path goes between the bat and the pot (not across the bat), and this is what she also states at other times. Thus, whether the relation expressed as por medio here by both is really also understood by both to mean the same thing or two different things ("between the bat [and something else]" vs. "through (the middle of) the bat") is not entirely clear. Fortunately, this is not important for the discussion here, since they treat it as if they were meaning the same thing.

*why he is astonished that he has to do a full turnaround (*una vueltaza*) at 172.3. They previously already tried to get past this part in the instructions when SG15 said to go between the bat, then they backtracked and have now arrived at it again.*

time	**SG15** (with the path)	**QF16** (without the path)
172.3		asu una vueltaza me tengo que dar *jeez a complete u-turn I have to do here*
173.9	mhm	
175.1	de ahí *from there*	
175.6		para eso tengo que pasar por encima del murciélago *for that I have to go above the bat*
178.5	no	
179.0	por su medio	
181.5	**por medio tienes que pasar** *no through it through the middle you have to go*	

In (58), QF16 states that he has to go above the bat at 175.6. This puts I HAVE TO GO ABOVE THE BAT on the table. SG15 objects to this proposition on the table in 178.5 and 179.0, partially denying it. When she utters *por medio tienes que pasar*, we can assume a QUD like WHERE IN RELATION TO THE BAT MUST YOU PASS?, where only *por medio* is at-issue as the contested part of the proposition and *tienes que pasar* is backgrounded because it is repeated and uncontested. The at-issue / non-at-issue separation here mirrors the separation between accented and deaccented material in the utterance.

Compare this to SG15_MT_ES_2272 (Figure 70), with context (28), which is segmentally nearly identical (58). A little later in the game at 225.1, QF16 is now putting the proposition denoted by *por medio del murciélago* on the table to be confirmed. SG15 confirms at 226.7 and 227.2, repeating *por medio tienes que pasar*. Here, there is no partial reversal interacting with a division between at-issue and non-at-issue material between *medio* and *tienes que pasar*. *Tienes que pasar* can be seen as given because the action of having to pass a landmark object is accessible from the context and almost the same phrase has been uttered earlier already, it being basically the default action in this game. The context is therefore quite similar to that seen for OA32_MT_ES_1443. That SG15 does not deaccent *tienes que pasar* in Figure 70 even though she was seen to deaccent in Figure 69 supports the hypothesis that even though there is also an information structural division between *por el medio* and *tienes que pasar* here, such a division has to interact with the difference in relative polarity between provocation and response that is part of a partial denial or partial reversal, but not of a confirmation. Deaccentuation seems more likely to take place on the uncontested material in such a reversal or denial following earlier contested material, than just on given material following new material.

(59) SG15_QF16_MT_ES_2222–2290 (context for Figures 70 and 71)[184]
General context: same as for (58), but somewhat later. They are once again going through the instructions on how to pass the bat.

time	**SG15** (the one with the path)	**QF16** (the one without the path)
222.2		ya
222.7		estamos en-
223.6		en la mitad de *right so we are in- in the middle of*
225.1		**por el medio del murciélago me has dicho** *through the middle of the bat you said*
226.7	mhm	
227.2	**por medio tienes que pasar** *through the middle you have to go*	
228.6		ya

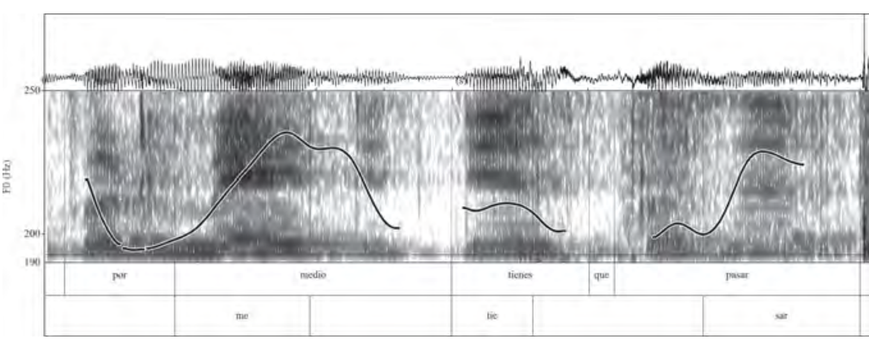


Figure 70: SG15_MT_ES_2272[185] (declarative with no deaccentuation). Cf. context (59).

184 https://osf.io/4xtd2/
185 https://osf.io/vun95/

5.1.3.2.2 Deaccentuation on "parentheticals" and evaluative additions

Neither the presence of a denial/reversal (generalized to a [Reverse] feature in the response in Roelofsen & Farkas 2015) nor givenness, however, actually are a necessary precondition for deaccentuation. There are cases of what can be described as "parentheticals" that fulfill neither of these conditions, as the following examples will show.

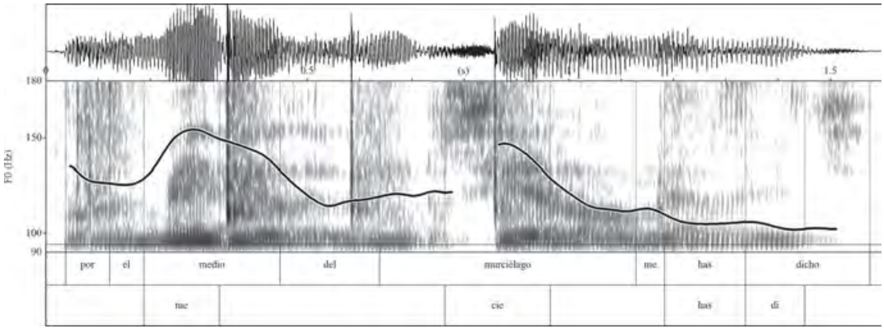


Figure 71: QF16_MT_ES_2251[186] (declarative with deaccentuation after *murciélago*). Cf. context (59).

In QF16_MT_ES_2251 (Figure 71, in the same context (59) as before), *me has dicho* is realized with flat and low pitch following a steep fall after the normally accented *por el medio del murciélago*. This example represents a number of cases where a verb phrase with an evaluative, epistemic or reportative verb that takes a sentential complement and is semantically superordinate to the proposition encoded by the material preceding it, but is expressed utterance-finally like an adverbial, is deaccented. The utterance here seeks confirmation for the instruction *por el medio del murciélago* (this part is what SG15 responds to), while the proposition that SG15 told QF16 about this, expressed by *me has dicho*, is not at-issue. As Roberts (2015, 2017) argues in the context of how the meaning of expressions of belief relates to the meaning of the propositions that are the target of these beliefs, such evaluative predicates can be at-issue if they are encoded with a full lexical expression,[187] but

186 https://osf.io/vht2m/

187 As opposed to encoded by a morphological affix, a particle, or intonation. The concurrent claim is that such doxastic, epistemic or evidential attitudes to a proposition, if expressed by affixation, a particle, or intonation, can never be at-issue; a conclusion which is supported by Faller (2014) for the Cuzco Quechua reportative evidential and which also seems to square with our own Quechua data.

usually, they are not.[188] Ortega-Llebaria & Prieto (2007, 2011) study the acoustic correlates of stress and accent in Spanish (and Catalan), making use of the difference in intonation between what they call "declaratives", on the one hand, and "parentheticals" (2007) or "reporting clauses" (2011), on the other. In their experimental sentences, the latter are instances of exactly such utterance-final evaluative adjuncts. Ortega-Llebaria & Prieto (2007, 2011) take it as a matter of course that what they call "parentheticals" or "reporting clauses" are deaccented, after a statement to that effect by Navarro Tomás (1968 [1944]: 115–116), where these clauses are defined as "la intercalación de un elemento incidental, con carácter propio, ajeno a la estructura melódica de la frase en que se encuentra".[189] This is the definition that Ortega-Llebaria & Prieto (2011: 78) quote, notably a definition of which intonational form (namely, the melodic structure being different from that of the rest of the utterance) is already a part. While Ortega-Llebaria & Prieto (2007, 2011) state that such parentheticals/reporting clauses are produced in a low monotone pitch without excursions, Navarro Tomás (1968 [1944]: 116) actually points out that there are also cases of such parentheses that are not deaccented, but where instead accent and pronunciation are strengthened relative to the rest of the utterance, and that this serves

188 Simply because the contents of beliefs or evaluations seem to be a more frequent topic of conversations than the nature of these beliefs or evaluations as opposed to their content, which is what needs to be the case for them to be at-issue. However, they can certainly be at-issue, which can be seen from the following examples (adapted from Roberts (2015: 47–48):

(i) Context: Why hasn't Louise been coming to our meetings recently?
  a. Henry believes she left town.
  b. She's left town, Henry thinks.

Possible replies:
  c. But she hasn't. I saw her at a supermarket yesterday. [targeting the content of the belief]
  d. No he doesn't. He told me he saw her at a supermarket yesterday. [targeting the belief state as opposed to the content]

The fact that (i d) is perfectly fine as response to (i a, b) shows that these belief states can be made to be at-issue. Equally, the speaker uttering (i a) or (i b) should be able to make the belief state at-issue in their own alternative utterances (ii a, b), and this should be accompanied by more accentuation on the postposed verbal complex *Henry thinks* in (ii b) than in a prototypical utterance of (i b).

(ii) Same context
  a. Henry thinks she left town, but usually what he says are just random guesses.
  b. She's left town, Henry thinks, but usually what he says are just random guesses.

It seems to me that Spanish is in general no different in this regard, but this would have to be shown from controlled speaker judgments.

189 "the insertion of an incidental element with its own character outside of the melodic structure of the phrase in which it is found", my translation.

"para subrayar expresiones especialmente intencionadas e importantes".[190] This suggests the possibility that such parentheticals can also occasionally be at-issue in Spanish, and that this might change their accentuation behaviour. It would be nice to unify these accounts and to analyze the absence of accentuation in utterance-final evaluative verb phrases like in QF16_MT_ES_2251 and other examples like it in our data together with the "parentheticals" observed by Navarro Tomás (1968 [1944]) and Ortega-Llebaria & Prieto (2007, 2011). However, this would require actually showing that they can be accented when they *are* at-issue, for which there is no data so far (see footnote 187).

Returning to our own data, these examples are curious because their contexts do not suggest the presence of a [REVERSE] feature, i.e. no difference in relative polarity between provocation and response, and there is no single identifiable opposite alternative salient in the context, against which the at-issue content is asserted (cf. Roelofsen & Farkas 2015: 385). And yet the deaccentuation in such cases even extends to entirely new information, as in KP04_Cuent_ES_2317 (see Figure 72).

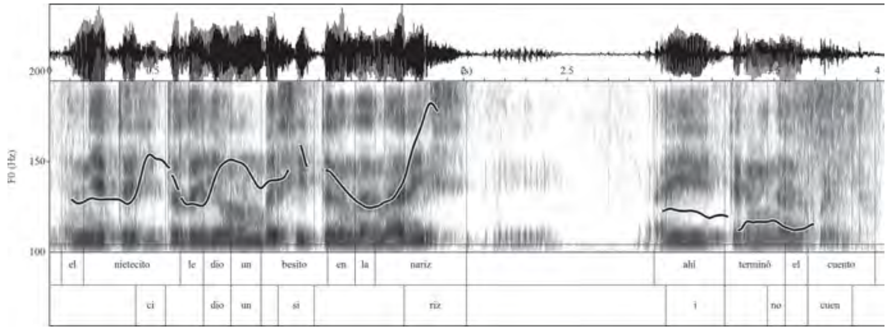


Figure 72: KP04_Cuent_ES_2317[191] (*el nietecito le dio un besito en la naríz / ahí terminó el cuento* 'the little grandson gave him a peck on the nose / there the story ended'; declarative with deaccentuation after *naríz*).

In this example, KP04 is finishing his re-telling of the story in *Cuento* (after TP03 first told it to him). The contents of the story itself end with *naríz*, up to which every accentable word is accented. This is followed by a silent pause (nearly 1 s), and then the additional phrase *ahí terminó el cuento* follows, which is deaccented. Note that both examples of deaccentuation of "parentheticals" seen here are not responses but provocations themselves, so that there can't be a [REVERSE] feature present in

190 "to highlight particularly intended and important expressions", my translation.
191 https://osf.io/j2mpv/

the context. KP04’s is an interesting example not only because the information contained in the deaccented part is new (albeit inferrable and somewhat formulaic), but also because the deaccented material is separated from the accented material by such a long break. This stands in quite a marked contrast to what happens in SG15_MT_ES_2396 (Figure 73 with context (60)).

(60) SG15_QF16_MT_ES_2291–2416 (context for SG15_MT_ES_2396)[192]
General context: directly following (59).

time	**SG15** (the one with the path)	**QF16** (the one without the path)
229.1	ya por encima de la olla tienes que bajar	
231.8		(6.4)
238.2		y a la quena ya no llego
239.6	**sí tienes que llegar**	
241.6	**tienes que bajar**	

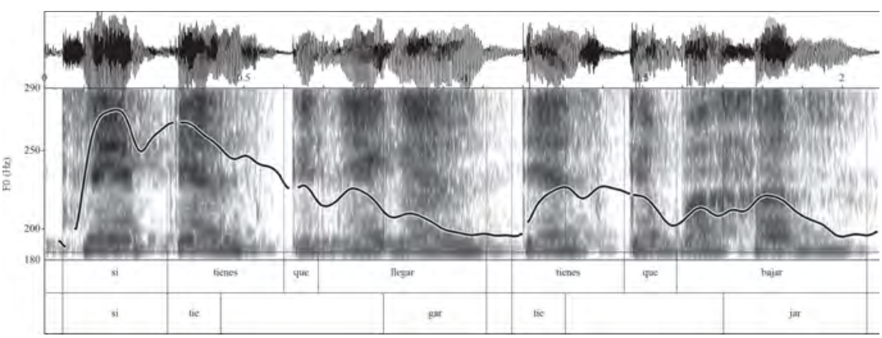


Figure 73: SG15_MT_ES_2396[193] (declarative with accentuation reset after *llegar*).

Here, QF16 asks whether the path does not let him reach the landmark of the flute, using a confirmation-seeking question with negative polarity, *y a la quena ya no llego* (238.2). This is then the proposition on the table, with a confirmation bias for the negative polarity of the proposition (Farkas & Bruce 2010; Roelofsen & Farkas 2015). At 239.6, SG15 reverses this bias with the polarity particle *sí*, which is pitch accented with very strong pitch excursion, followed by *tienes que llegar*, which is an at-issue addition and accented with less excursion, pitch falling globally throughout its realization. Nearly without a pause, this is then followed by *tienes que bajar*

192 https://osf.io/9fkwu/
193 https://osf.io/xdz3q/

(at 241.6), which is clearly independently accented and also answering a separate QUD, namely something like WHERE DO YOU HAVE TO GO? or possibly HOW DO YOU GET THERE?. However, with regards to the explicit QUD of 239.6, DO I NOT GET TO THE FLUTE ANYMORE?, *tienes que bajar* is just as non-at-issue as *ahí terminó el cuento* is to the QUD that is answered by the first part in KP04_Cuent_ES_2317, in addition to having a much shorter break in between. Yet the former is not deaccented, while the latter is. Whether this is because such parenthetical phrases are just much more likely to be not at-issue, or because the difference simply lies in individual variation, any analysis of deaccentuation using a division between at-issue and non-at-issue material must take into account the interaction between what is prosodically a separation between utterances (accompanied by pitch reset) and what constitutes a separate move in the conversational game, as well as the difference in relative polarity between a provocation and its response.

Summarizing, deaccentuation (variably realized as severely reduced pitch scaling) has been found to occur in the Huari Spanish data in various contexts. While it seems nearly categorical in alternative questions and wh-questions, in declaratives its occurrence is quite variable. According to this qualitative investigation, the presence of a [REVERSE] relation between the utterance and a provocation seems to be a far stronger contributing factor than only a contrast between new and given material. The sole presence of the latter was not found in examples with deaccentuation. Deaccentuation was also found outside of reversals/denials on material denoting evaluative additions or parentheticals. Deaccentuation does not categorically cooccur with any of these contexts, with individual speaker preference being a possible additional factor. While a thorough quantitative exploration of the contexts of deaccentuation remains a task for the future, prosodically it can be assumed that the necessary (but not sufficient) condition for deaccentuation is that the highest metrical position in a phrase at least of iP-size is not final in that phrase but followed by further accentable material.

5.1.3.3 Tonal target placement and epistemic biases

As a last item included in the intonational phenomena observed in simple Huari Spanish utterances, I will here briefly describe a shift in pitch accent peak alignment and its possible pragmatic function. Because the examples are isolated, the discussion will have to remain exploratory. There are a number of utterances where the temporal alignment of both the final peak and its preceding elbow with regards to the stressed syllable is divergent from that in the majority of the Spanish data. This seems to correlate with contexts that suggest, in the broadest terms, an attitude of epistemic bias with regards to the proposition expressed on behalf of the speaker.

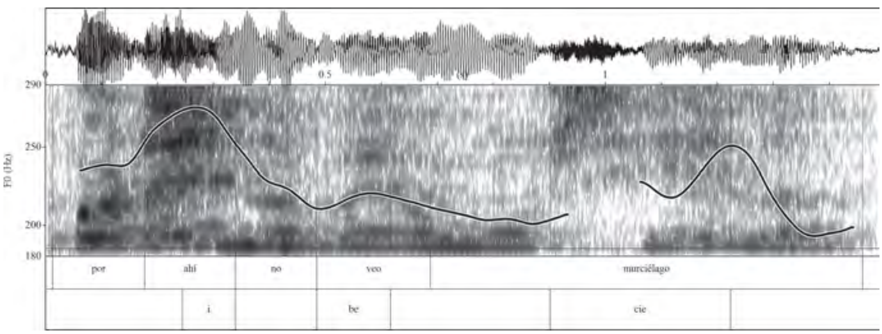


Figure 74: HA30_MT_ES_1372[194] (declarative with delayed peak on *murciélago*).

(61) ZR29_HA30_MT_ES_1197–1405 (context for Figure 74)[195]
*Global map context: the bat (*murciélago*) is at different positions in the two maps, so that HA30 cannot move between bat and pot at the point when ZR29 instructs her to do so (cf. Figures 190 and 191 in* Appendix B*).*

time	**ZR29** *(with the path)*	**HA30** *(without the path)*
119.7	pasas	
120.7	por debajo del zorro *you pass below the fox*	
122.9		ya
123.4	encima de la nube *above the cloud*	
125.4	por medio de	
128.8	el murciélago y la olla vas pasar *between the bat and the pot you'll go*	
132.2		murciélago y la olla *bat and pot*
133.6	sí	
134.1	por el medio de esos dos pasas	
136.4	por encima *yes between those two you pass above*	
137.2		**por ahí no veo murciélago** *I don't see a bat there*
139.9	pues acá hay *well here it is there*	

194 https://osf.io/dkawc/
195 https://osf.io/pu9jg/

Figure 74 shows the utterance HA30_MT_ES_1372, *por ahí no veo murciélago*. The proparoxytonic *murciélago* is utterance-final here, like in (40)/Figure 23, but there, the pitch elbow is in the pretonic and the accent peak is reached shortly after the start of the vowel in the stressed syllable. In contrast, here the peak occurs only at the end of the vowel, at the boundary to the posttonic syllable, and it is the preceding elbow[196] which is formed right at the start of the vowel of the stressed syllable (after the heightened pitch due to consonantal microprosody in the fricative segment). The peak on *murciélago* also clearly has much more excursion than that on *veo*, and, more strikingly, both the word and its stressed syllable have much more duration relative to the preceding material in the utterance. The context for Figure 74 is given in (61). In the global context, the bat (*murciélago*) is one of the landmarks whose position differs between the two maps; therefore, HA30 cannot move between the pot (*olla*) and the bat when ZR29 instructs her to. Sequentially, at 122.9, HA30 signals acceptance of ZR29's preceding instructions; this is understood by ZR29 who progresses towards directing the way around the next landmark on the map, which includes that HA30's path should move between the bat and the pot (123.4–128.8). This is the first time the bat (*murciélago*) is mentioned. At 132.2, HA30 suspends acceptance of those directions by uttering a request for clarification which takes the form of a repetition of the landmarks she should pass through, *murciélago y la olla* (the bat and the pot). ZR29 understands this to be a request for clarification, she confirms that HA30 understood correctly (*sí* at 133.6) and gives a partial repetition of her previous instructions, but pronominalizing the landmark referents of bat and pot as *esos dos* (134.1–136.4). At 137.2 then, HA30 produces the utterance under discussion here. Considering the preceding context, this utterance serves the purpose of continuing the suspension of acceptance of ZR29's current instruction. It clarifies the reason for the suspension: while ZR29 accedes to the request for suspension and clarification in 133.6–136.4, she must assume that what is at-issue is the path between the referents *murciélago* and *olla*, not the referents themselves, as she pronominalizes them, thus presupposing their existence in the relevant part of the map. This presupposition is however what is contested by HA30 in 137.2 for the landmark referent *murciélago*. Her turn targets an answer not to the current QUD, but to one (IS THERE A BAT?) that lies far back in the discourse history and which was treated as

196 There is also another elbow at the boundary of the pretonic to the stressed syllable, and the elbow within the stressed syllable is located somewhat higher than this earlier one. It is not clear which of these really is the target for the L tone; however, if we were to draw a straight line from the earlier elbow to the later one, it would have a considerably lower slope than the rise that takes place within the stressed syllable itself, and also than a line drawn from where the voicing ends to where it starts again in the corresponding section in Figure 23. Therefore we can probably say that the elbow is relatively later in Figure 74 than in Figure 23.

settled (in CG) by both interlocutors previously. She thus brings it back on the table, making it at-issue once again and contesting the settlement. The referent of *murciélago* is thus not saliently contrasted with another available referent, but instead a contrast is evoked between presupposing the referent of the expression *murciélago* (and its position on the map) and being unable to find a fitting referent for it, which is what HA30 expressly communicates here (*por ahí no veo murciélago* meaning here something like "in the relevant part of the map, I cannot find a referent that would fit the description of a bat"). I suggest that this is achieved by including a [Reverse] feature in the context update effected by this utterance, which specifically targets the presupposition that the bat is there, not the provocation, in an extension of how the [Reverse] feature is used in Roelofsen & Farkas (2015). It seems plausible that this presupposition challenge, pointing to a discrepancy between the two speakers regarding the status of the referent of *murciélago* in the common ground, is cued by the marked prosodic realization of *murciélago* with delayed tonal alignment here.

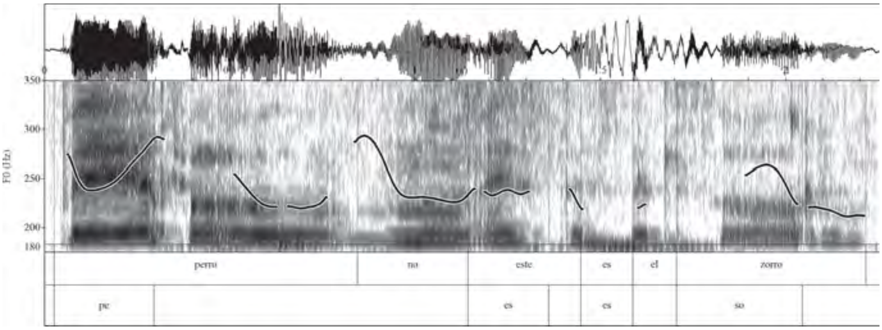


Figure 75: MD40_MT_ES_1542[197] (clarification question with delayed peak on *perro*, and declarative with normally aligned peak).

(62) SO39_MD40_MT_ES_1257–1580 (context for Figure 75)[198]
Global map context (cf. Figures 190 and 191 in Appendix B*): the positions of the lamb (*corderito*) are very different in the two maps: in SO39's map (with the path), it is in the lower half, whereas in MD40's (without it) it is at the top of the upper half, above the fox (*zorro*), while the dog (*perro*) is in the lower half. The position of the lady (*señora*) is also different between their maps: in SO39's it is at the top of the lower half on the right side, directly below the lamb, while in MD40's it is on the left upper side of the lower half.*

197 https://osf.io/evu35/
198 https://osf.io/94w2h/

time	**SO39** (with the path)	**MD40** (without the path)
125.7		del perro me doy vuelta por la señora *from the dog I turn around the lady*
128.5	mhm por encima por la mitad con el corderito lo es- que está por la mitad *yes above through the middle with the little lamb- that is in the middle*	
136.3		ahí no hay *it's not there*
137.1		tienes de por el perro *you gotta [say] via the dog*
141.6		del perro
142.8	del perro pasa a la señora no *from the dog it goes to the lady right*	
146.0		por debajo por [en- *below ab-*
147.4	[por en]cima *above*	
148.2		mhm ya
149.3	mhm ya pasar corderito por su debajo *yeah passing the little lamb from below*	
154.2		**perro**
155.0		**no este es el zorro** *dog no this is the fox*
158.8	*(laughter)*	

The next example of a delayed peak cueing an epistemic bias is MD40_MT_ES_1542, *perro no este es el zorro* (Figure 75), with context (62). The relevant element with marked tonal target placement is the first one, *perro*. As in the preceding example, there is a late tonal target realization, with the low pitch elbow realizing the L tone produced just at the beginning of the vowel in the stressed syllable, and the peak realizing the H tone at the end of the vowel. This is followed by a return to low in the posttonic and then a further rise at the end of it (visible only partially in Figure 75 because of some creakiness in the voice, but indirectly evidenced by the initial fall from high in the next word, *no*, and also clearly perceptible auditorily). The fall-rise movement in the posttonic is here taken to evidence the presence of LH% boundary tones, which were found associated with polar questions (see section 5.1.2.1). A conventional orthographic transcription might be something like "¿¡Perro?! No, este es el zorro." Note that in the segmentally similar *zorro* at the end of the example, the pitch peak is reached clearly earlier, just about after the midpoint of the stressed vowel, and with a considerable part of the following fall taking up its remainder (the auditory impression is also clearly very different). The context (62) helps interpret this example: globally, it has to be kept in mind that the lamb (*corderito*) is at the top of the lower half of SO39's map, but at the top of the upper half in MD40's. In both maps, the dog (*perro*) and the fox (*zorro*) are roughly in the middle of the lower and upper half, respectively. The lady (*señora*, actually an image of a cartoon figure with a bag of money) is in the upper right part of the lower half in SO39's map, but in the upper left part of it in MD40's map. At 125.7, MD40 suggests a path for how to continue from the dog, which is confirmed with a confirmation token (*mhm*) by SO39 who follows it up with a elaboration of instructions about how to move from there via the lamb at 128.5. MD40 then points out that the lamb is not there, and tells SO39 to repeat the instructions using the dog as starting point (136.3–141.6). SO39 obliges and asks to make sure that the immediate path from the dog to the lady is agreed upon (142.8). They then discuss whether to go above or below, and having settled this, SO39 continues the instructions on how to proceed, passing below the lamb (146–149.3). At this point, MD40 then produces the utterance under discussion. The *perro*-part, with severely delayed peak, here resembles the move performed by step **b** in the confirmation-seeking question sequence discussed in section 5.1.2.2. It is also a clarification request (one for intended content according to Ginzburg 2012: 149–150; Łupkowski & Ginzburg 2016: 250–251), but unlike the examples discussed there, it is not biased for confirmation, but for disconfirmation (an "incredulity question"): as the elaboration shows, MD40 intends to correct the identity of the referent they have been referring to as *perro*, committing to the proposition that the relevant referent is instead correctly labeled as *zorro*, the skunk (the instructions as given for her only make sense if they refer to the skunk). The peak on *zorro* in this assertion is not delayed, as expected. Thus we could again say that the context update conveyed by the utterance contains a [Reverse] feature, which does not target a proposition on the table but one that is presupposed, namely that there is a dog there at the relevant location or that the entity there is correctly referred to as a dog.

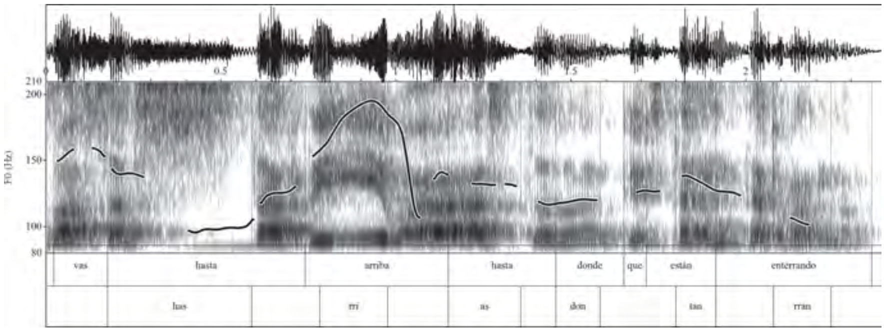


Figure 76: TP03_MT_ES_1312[199] (declarative with deaccentuation after and delayed peak on *arriba*).

(63) TP03_KP04_0906–1050 and 1211–1371 (context for TP03_MT_ES_1312)[200]
General context (cf. Figures 190 and 191 in Appendix B*): the lamb, the millionaire (*niño millionario*), the rock, the bat and the dungheap are in different locations in the two maps. They already had a conflict, because KP04 would have needed to do a complete circle (*círculo*) around the lamb back to the millionaire in order to follow TP03's instructions and objected. TP03 describes how to proceed from under the dog (*debajo del perro*). In the omitted part, they have a similar conflict about having to do a complete circle. TP03 then restarts by explaining how to move from the dog.*

time	**TP03** (with the path)	**KP04** (without the path)
90.6	pasa por este *goes by that*	
93.1		por los pies *the feet*
93.6	ajá por los pies *yeah the feet*	
94.8	ya ahí detente un rato *right stay there a while*	
97.0	y va ir *and it'll go*	
98.7	de lo que están ahí este *from the one they're there that*	
100.7	de lo que están enterrando *from the one they're burying*	

199 https://osf.io/8gk74/
200 https://osf.io/cgk7u/

103.7		de lo que están enterrando *from the one they're burying*
104.7	ajá *yeah*	
105.0–121.1	[...]	
121.1	ya por debajo del perro ha pasado *right below the dog it's passed*	
123.1		sí por encima del niño *yes above the boy*
125.0	ya	
125.6	por encima del niño está pasando no te vas a ningún círculo *right it's going above the boy don't go in any circle*	
128.0		ya
128.6	ya	
130.0	sin ce- sin cerrar nomás **vas hasta arriba hasta donde que están enterrando** *with- without closing [the circle] you go upwards up to where they're burying*	
133.8		ah ya paso por encima nomás *ah right I just go up*
135.5	ajá pasas por encima nomás pe *yeah right you just go up*	

The last example is the most complicated one. In (63), TP03 first mentions the landmark he calls *lo que están enterrando* ("that which they are burying", an image of a funeral) at 98.7–100.7, which is supposed to serve as departure point for his next instruction. This causes some confusion for KP04, who repeats that he then has to to do a full circle (in the omitted section). Doing a circle around the millionaire at this point means moving downwards on the map for TP03, for whom the current position on the path is directly above the millionaire (*encima del niño*) at 125.6. TP03 explicitly rejects going full circle in 125.6. In 131.2 (Figure 76), he then gives the proposition that corrects not only the circle path, but also what this implies for him, namely moving downwards. Since going full circle entails going downwards but not the other way around (from TP03's point of view at this moment), the corresponding QUD DOES THE PATH GO DOWNWARD? is superordinate to the one asking DOES THE PATH DO A FULL CIRCLE AROUND THE MILLIONAIRE? (cf. Roberts 2012b). TP03 must assume that KP04 presupposes that the path goes down here, and that this is the cause of their misunderstanding.[201] When uttering 131.2, he corrects this presupposition (the proposition that the path goes downwards which he believes KP04 to believe to be in the common ground), i.e. his move contains a [REVERSE] feature targeting it and it also asserts its complement, to go upwards. To this, going to where they are burying is a non-at-issue addendum, and deaccented as expectable from the preceding section. I argue that the delayed peak on *arriba* here cues the presence of a [REVERSE] feature in the context update targeting not the proposition on the table but a presupposition, as in the previous two examples.

In sum, I tentatively propose that what unites these three examples with delayed peaks is the presence of a [REVERSE] feature targeting a presupposition (a previous QUD), instead of the proposition on the table (the current QUD). Note that we dealt here with both assertions and a clarification request, suggesting this is some kind of (modal) non-at-issue meaning[202] component that is orthogonal to the difference between these speech acts. The clarification request by MD40 seems almost the direct opposite to those discussed in section 5.1.2.2 in terms of bias, while for assertions the difference is more complex.[203] However, formally the similarity between biased questions and assertions discussed there seems to be maintained here too, with the cue for the additional meaning being a delayed peak in

201 Previously at this location, KP04 had always made a circle here instead of doing what TP03 instructed him to do. Because the circle around the lamb from his position at this point could only be made by going downwards, he must assume that KP04 presupposes that the path goes downwards there (in reality, the relevant objects on their maps are differently placed).

202 In the form perhaps of a conventional implicature (cf. Bianchi et al. 2016; Fliessbach 2023), a presupposed modal meaning (Reich 2018), or an illocutionary operator (Faller 2014).

203 Assertions put a proposition p on the table and project its acceptance, while neutral polar questions of p put $\{p, \neg p\}$ on the table and only project the acceptance of either outcome (Farkas & Bruce 2010: 92, 95). Biased assertions and questions pose a more difficult case. Fliessbach (2023: 67–68) differentiates bias and commitment based on Farkas & Roelofsen (2017), and argues that both can be reacted to in disourse. However, he treats bias as an gradable but absolute value, while I would argue that the polarity of bias is also relevant, since questions with a bias for confirmation and ones with a bias for disconfirmation seem to have different forms. It is an open question how that would have to be modeled, and whether something like negative bias perhaps emerges out of a meaning like the one I propose. As argued above, the assertions and questions with negative bias discussed here should be seen as conveying an additional non-at-issue meaning component consisting at least of a [REVERSE] feature targeting not (only) the proposition on the table (at least TP03's utterance should be seen as a provocation, not a response, anyway), but a presupposition for a salient proposition in the context that is complementary to the proposition that is asserted.

both cases. The contrast conveyed here between a presupposition and the proposed context update is clearly different from that between two alternative referents or propositions (that obtains in some of the deaccentuation cases from the previous section), suggesting that a label of "contrastive focus" would be underspecifying in either case. There are further similar examples in the Huari Spanish data, but I suggest that the phenomenon should be done better justice by treating it as the main focus of separate further research.

5.1.4 Interim summary

Before moving on to the analysis of more complex utterances in the next section, I review the findings made in this section. I have described the intonation of Huari Spanish declaratives and interrogatives. In general, nearly every accentable syllable has been found to be pitch accented with an LH*, whose peak is regularly aligned within the stressed syllable, independent of position in a phrase (speech with these attributes has here been called the main accentuation variant). The analysis has revealed that there is comparatively little paradigmatic variability in tone choice. The bitonal LH* is possibly the only attested pitch accent, and equally, only three boundary tones (L-/%, H-/%, LH-/%) could be identified. Only the delayed peak found in the contexts discussed in section 5.1.3.3 might be analysed as a separate pitch accent, L*H or perhaps L+<H*, but it is still clearly relatable to LH*. On the other hand, however, two variant phenomena of pitch accentuation that cue information structural and discourse-pragmatic meanings were found, phrase accentuation and deaccentuation. Both of these accentuation modes are characterized by a marked syntagmatic contrast between accentuation of accentable positions in different parts of an utterance, with gradual steps between weaker and stronger tonal compression or even deletion. This also highlights the relevance of pitch scaling for intonational description. Their occurrence was found to be variably conditioned by discourse context, but is also subject to individual speaker preference, with one speaker showing signs of using the phrase accentuation even as a default. The phrase accentuation effectively constitutes a more phrase-optimizing intonation than the main accentuation or other familiar intonational systems of varieties of Spanish, which can be said to optimize prosodic words instead. In the final section, it was suggested that variability in peak alignment on stressed syllables might also be associated with differences in pragmatic meaning. In the next section, the results from this section will be built on in the discussion of more complex double topic-utterances, and integrated into a detailed analysis of the levels of the prosodic hierarchy that are evidenced in them. It will be shown that pitch scaling again plays a decisive role in signaling a prosodic structure with nonlocal dependencies.

5.2 Complex utterances: “Double topic”-constructions and hierarchical tonal scaling structure

This section expands the analysis of Huari Spanish to a specific type of complex utterances, the double topic-utterances from the *Elqud* experimental task. Their analysis will demonstrate that pitch scaling at the level of register heights is sensitive to a hierarchical prosodic structure that is recursive. In a second part, utterances of this type displaying the type of variation called “phrase accentuation” in the previous section will be considered separately.

5.2.1 Data

The utterances discussed here come from *Elqud* items 14, 26, 43, 52, 59, 65, and marginally 44. As described in section 2.4, *Elqud* is a corpus from an elicitation setup in which participants were asked to utter corrections to recorded utterances (the audio stimulus) as if the speaker of the audio stimuli were present in the conversation, based on differences between what the audio stimulus asserted and a simultaneously shown image or short animation (the visual stimulus). Although speakers were encouraged in general to respond in detail and expansively, using rather more than fewer words, no restrictions were enforced during the experiment on *how* the speakers should respond in each case. This led to some items by some speakers consisting only of generic objections such as *no es correcto* ‘that’s not correct’, *es falso* ‘that’s wrong’ or *no, es al revés* ‘no, it’s the other way round’, which are not useful for the analysis, but it also means that when speakers did respond completely, they did so in a way that was presumably more natural to them than if they had been told to respond observing a certain syntactic pattern, certain words, or suchlike. Such responses that were successful in terms of the experimental aim will be called ‘full responses’ in the following. Some responses also had to be excluded from the analysis because even though speakers did attempt to give expansive responses, they noticeably got confused, mixed up parts of their utterance or broke off after an incomplete attempt. This is a normal feature of uncontrolled speech under any conditions, and it happened with all speakers occasionally. In those cases, during the experiment, no attempt was made by the experimenters to ask them for another try, instead, they simply proceeded to the next item. Items 14, 26, 43, 52, 59, 65 and 44 are presented in Table 11 by giving the text of the recorded audio stimulus and a description of what differed in the visual stimulus from what the subjects heard.

These examples (both the recorded stimuli and the ‘full’ responses) have in common that they consist of two different predications made about two different referents. In terms of the QUD-model of discourse (cf. Roberts 2012b), they can be

Table 11: Stimuli for the Spanish ELQUD items 14,26, 43, 52, 59, 65, 44. Pitch accented syllables are underlined, words standing in a contrast relation with others in the utterance are capitalized.

Item number	Text of audio stimulus	Description of visual stimulus / mismatch
14	*El HOMBRE está COMIENDO y la MUJER DURMIENDO* ‘The man is eating and the woman is sleeping’	The man is sleeping while the woman is eating
26	*El PERRO está DEBAJO de la roca y el GATO está ENCIMA de la roca* ‘The dog is below the rock and the cat is above the rock’	The dog is above the rock and the cat is below the rock
43	*El colibrí ROJO está chupando la flor AMARILLA y el colibrí VERDE la flor ROJA* ‘The red hummingbird is drinking from the yellow flower and the green hummingbird from the red flower’	The red hummingbird is drinking from the red flower, the green hummingbird is drinking from the yellow flower’
52	*El colibrí VERDE está chupando la flor que está a la IZQUIERDA el colibrí AZUL está chupando la flor que está a la DERECHA* ‘The green hummingbird is drinking from the flower that is on the left, the blue hummingbird is drinking from the flower that is on the right’	The green hummingbird is drinking from the flower on the right, the blue hummingbird is drinking from the flower on the left
59	*El zorrillo PEQUEÑO está DETRÁS de la casa PEQUEÑA, y el zorrillo GRANDE está al FRENTE de la casa GRANDE* ‘The small skunk is behind the small house and the large skunk is in front of the large house’	The small skunk is in front of the large house, the large skunk is behind the small house
65	*El zorrillo GRANDE está al frente de la casa GRANDE, y el zorrillo PEQUEÑO está al frente de la casa PEQUEÑA* ‘The large skunk is in front of the large house and the small skunk is in front of the small house’	The large skunk is in front of the small house, the small skunk is in front of the large house
44	*El perro NEGRO está jugando con la pelota ROJA y el perro BLANCO está jugando con la pelota AZUL* ‘The black dog is playing with the red ball and the white dog is playing with the blue ball’	*No mismatch / filler*

understood as providing an answer to a superordinate QUD asking about what propositions hold of a set of referents, e.g. {*the people in the picture*}, such as ‘what are the people in the picture doing?’ by providing answers to subordinate QUDs that are formed by asking the superordinate question about each member of the set of referents individually, i.e. {*the man, the woman*}, such as *the man is sleeping* *and the woman is eating*. For English, the way in which this kind of information structural constellation is reflected in prosody has been discussed in terms of what has been called “A-accent” and “B-accent” since Jackendoff (1972) (cf. also Büring 2003; Ladd 2008; Roberts 2012b). We can also say that the referents about which the propositions are predicated are sentence topics in the sense described by Roberts (2011) as that of the entity in a sentence to which our attention is first drawn and about which we are then told something, a referential restriction upon the domain over which the proposition is to hold.[204] Complementarily, what is predicated of them will be called comments. In the corpus of utterances discussed in this section, the set of entities about which propositions are predicated always consists of two members (the man and the woman, the small skunk and the large skunk, etc.). They will therefore be called “double topic”-constructions. Since the topics are pairs (as members of the set of referents about which the superordinate QUD is asked) and the propositions that are predicated of them are also parallel to each other, there is a notion of contrast both between the two topics of each utterance and the two comments. This contrast manifests in at least one element that is different between the first and the second topic and comment, respectively. These contrasting elements are highlighted by capitalization in Table 11. A further notion of contrast comes into play when considering the discursive relation between the utterances produced by the experimental subjects and those that serve as the audio stimuli. In the experimental items, the utterances assert that a different state of affairs obtains between the members of the set of referents and the predications made about them than that which is asserted by the audio stimuli (the referents and the predications themselves are the same, nothing new is added in the corrections, but their relation is asserted to be different). They are thus partial denials in terms of Farkas & Bruce (2010). The way they are interpreted here, as reflecting a complex QUD structure that answers a question with two variables (WHO IS DOING WHAT?), the comments are at-issue with respect to the current QUD for each utterance half (TOPIC 1 IS DOING WHAT?, TOPIC 2 IS DOING WHAT?), while the topics are the answers to the question WHICH ARE THE TRUE MEMBERS OF THE SET OF AGENTS?, which is needed to answer the superordinate QUD (cf. Figure 77). Syntactically, the utterances consist of a conjunction of two sentences, the conjunction element either expressed by *y* ‘and’ or not (this is true both for the recorded stimuli and for the utterances produced by the speakers). The topics are realized as noun phrases consisting either of one (in the case of items 14 and 26) or two content words (a noun followed by an adjective,

204 The other sense of „topic“ that Roberts (2011) discusses is of course that of the ‘discourse topic’, what a larger portion of discourse is about. She herself identifies this more global sense of topic with the QUD (Roberts 2011: 1909), so that ‘discourse topics’ and ‘sentence topics’ in this conception can be equated with superordinate and subordinate, or more global and more local, QUDs.

in 43, 44, 52, 59, 65, with the adjective, never the noun, contrastive). The comments are realized as VPs of various complexity, containing several content words: while in 14, they consist just of the auxiliary *está* and the verb in the present participle, in all other examples they contain at least a further NP that is either a direct object of a transitive verb (in 43, 52), a prepositional object of a locative predicate (in 26, 59, 65) or an oblique prepositional object of a verb (in 44). In item 52, this NP is itself complex, containing a locative relative clause.

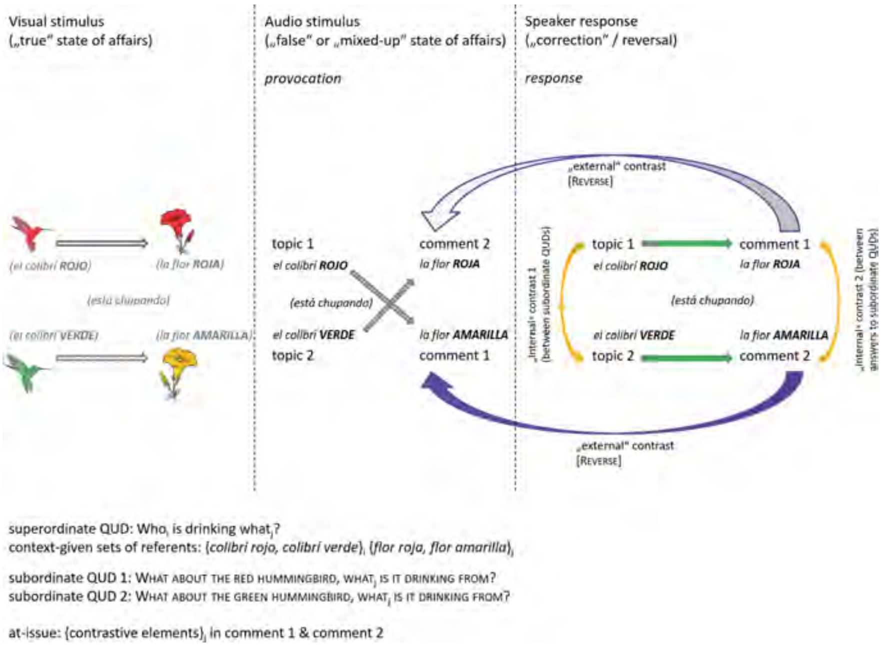


Figure 77: Schematic representation of information structural relations between the speaker responses and their stimuli and within the speaker responses themselves.

In the discussion of the actual utterances by the speakers, I will refer to the first element of the conjoined sentences, containing topic 1 and comment 1, as simply “the first part”, and the second “the second part”. Since speakers were free in how to specifically realize their responses to the stimuli, they sometimes produced utterances in which their first topic was the topic of the second part in the audio stimulus and vice versa. Here we will not speculate about whether this reflects different perceived complex information structures of the utterances (as in “no, the red flower is what the green hummingbird drinks from, not the yellow flower” vs. “no, it is the green hummingbird that drinks from the blue flower, not the blue hummingbird"). However, it is important to note that some of the stimuli allow several interpretations with respect to their implied information structure (definable as implicit QUDs), and consequently, the utterances produced by the speakers might also reflect different information structures. The most important one will here be briefly discussed in simple and largely atheoretical terms: item 59 might be interpreted as having either two or three locations of mismatch/contrast that could be realized in the response (here given by numbered bracketing):

(64) a. [The small skunk]$_1$ is [in front of]$_2$ [the large house]$_3$, [the large skunk]$_1$ is [behind]$_2$ [the small house]$_3$
b. [The small skunk]$_1$ is [in front of the large house]$_2$, [the large skunk]$_1$ is [behind the small house]$_2$

Utterance 65, on the other hand, does not allow for this contrast on the preposition in the comments, because the preposition is the same in both comments.

The audio stimuli and the full responses differ with regards to their length and the number of words on which we can expect pitch accents to occur. The default assumption here will be to expect a pitch accent on the lexically accented syllable of every accentable word, in accordance with the results of section 5.1.1.2. In Table 11, all accented syllables in the audio stimulus are underlined. From that we can see that item 14 has 5, 26 has 8, item 43 has 10, items 44, 59 and 65 each have 12, and item 52 has 14 accented syllables. In the responses by the speakers, we can similarly expect an increasing number of accentable (and accented) words across the items in this order, but due to the relatively free experimental setup not necessarily the exact same number as in the stimuli. In the following we will investigate systematic patterns in pitch accentuation and scaling in the utterances produced by the speakers as a reflection of differing degrees of complexity in the prosodic structure, which in turn can be correlated with information structure.

I will use the analysis of these double topic-utterances to provide evidence for the hypothesis that pitch scaling in these data reflects a hierarchical prosodic structure that is likely recursive, similar to what has been shown for English (Ladd 1988) and German (Féry & Truckenbrodt 2005; Truckenbrodt & Féry 2015) as well as other languages, as discussed in section 3.6.

In sections 5.2.2 and 5.2.3, I will establish the general pattern that the majority of these utterances follows via both quantification and analysis of individual examples. I argue that this majority represents the "main" variant of complex utterance accentuation in the Huari Spanish data, in parallel to what has been established in section 5.1 for simple utterances. In sections 5.2.2–5.2.5 we will then consider deviations from this pattern, especially those that run in parallel to the "phrase accentuation" variant also already observed in section 5.1. In the remainder of this section, I will discuss some more general observations characterizing "full" responses by the speakers. They differ from speaker to speaker and example to example in the degree in which they are "reduced", both syntactically and prosodically. Syntactically, an unreduced utterance would be one in which both parts are realized as full sentences, i.e. without ellipsis. Unreduced utterances are the majority here. Also frequent is ellipsis of given elements, with two extreme examples of such reduction being item 43 by LJ22 and TP03:

(65) LJ22_ELQUD_ES_43 (cf. Figure 98)
el colibrí verde flor amarilla y el colibrí rojo flor roja
'the green hummingbird yellow flower and the red hummingbird red flower'

(66) TP03_ELQUD_ES_43
el colibrí verde está chupando la flor amarilla y rojo rojo
'the green hummingbird is drinking from the yellow flower and red red'

While LJ22 does not produce any verbal elements in his utterance, presumably because they are already given in the audio stimulus, TP03 eliminates all elements in the second part that would be given from their occurrence in the first part, i.e. all elements that are not contrastive. Ellipsis can therefore occur on elements that are given through the (external) relation to the provocation (the audio stimulus), or through the (internal) relation to preceding parts of the utterance itself. Contrastive elements were preserved in most of these cases. Utterances where they were elided were not counted as full responses and excluded from the analysis. Ellipsis was observed to differ both across speakers and utterances, with TP03 and LJ22 and item 43 most prone to it, and e.g. speaker OZ14 and item 65 at the other end of the scale.

Prosodically, reduction here is above all a phenomenon affecting absolute pitch range within an utterance. This is largely speaker-dependent: some, as for example ZE55, employ a large range, others, such as ZZ24, seem almost to make an effort to reduce their pitch range as much as possible (see Table 12).

For that reason, in the quantitative analysis I will employ a transformed pitch measurement that relates an individual measured value to the maximum and minimum measured value in the utterance. The results will show that even though pitch range differs so much between speakers, overall this does not affect relative pitch height, i.e. the tonal scaling relationship between individual elements within each utterance. Besides a relatively small pitch range, the utterances by the speakers who exhibit it are also characterized by being quite reductionist in other ways: for LJ22, several assimilative processes take place in that for instance, unstressed vowels are often centralized towards schwa or even the whole syllable is elided,

Table 12: Mean pitch range in the double topic-utterances of ELQUD_ES according to speaker.

Speaker	Mean pitch range (Hz)	Pitch range (st from mean minimum)[205]	N (utt.) per speaker
OZ14	50.8	7	6
QP44	43	6.4	5
QZ13	59.8	7	6
SG15	87.8	7.6	3
TP03	44.4	6.5	3
XJ45	63.3	8.8	5
ZZ24	31.2	4.2	6
ZE55	141.5	11.4	5
NQ01	117.9	11.3	7
LJ22	46.9	6.9	5

nearly all consonants are produced voiced while some voiced ones are turned into approximants, and complex onsets are reduced, yielding something like e.g. [gʊləˈwi:] for *colibrí* in LJ22_ ELQUD_ES_43, [bəˈgẽ:jʊ] for *pequeño* in LJ22_ELQUD_ES_59. ZZ24 or TP03 mostly maintain unvoiced consonants but also frequently reduce unstressed syllables. Probably such "reduced" speech has a socioindexical component apart from perhaps expressing some boredom at the experimental task: note that all speakers having a mean range of below 7 st are male. Different local varieties of a language do seem to differ in the mean pitch excursion size they employ for the same tonal categories (e.g., Liverpool speakers of English use a very small pitch range compared to speakers of other British varieties for the encoding of largely similar tonal inventories, cf. Nance et al. 2018), so it is likely also socially meaningful below the level of regional varieties.

5.2.2 Analysis of individual examples / "main" variant

QZ13_ELQUD_ES_52 (Figure 78) is a good example to observe the hierarchical scaling relation found in the double topic-utterances. Experimental item 52 is most complex both in terms of number of accentable words and internal syntactic structure, as seen from Table 11.

205 Range in semitones was obtained by taking the bottom value as the one from which the difference to the top one is calculated using the *f2st* function from the *hqmisc* package in R (Quené 2014). The formula for conversion to st from two f0 values *a* and *b* in Hz is 12*(log2(a/b)), where b is the starting point from which the difference to a is to be calculated. See also Traunmüller (2005, 2017).

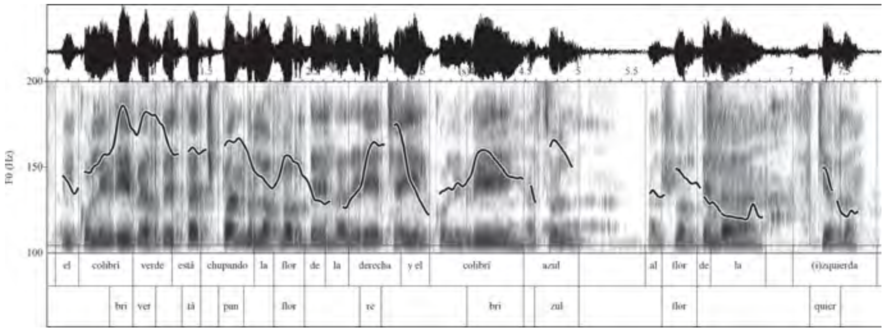


Figure 78: QZ13_ELQUD_ES_52[206] (*el colibrí verde está chupando la flor de la derecha y el colibrí azul al*[207] *flor de la izquierda* ‘the green hummingbird is drinking from the flower to the right and the blue hummingbird from the flower to the left’).

From Figure 78, it becomes clear that pitch height on each accented word is neither the same nor simply declining steadily, but instead seems to follow a more complex pattern. On the topics of both parts, the pitch accents on the noun and the following adjective are almost the same height or the second is slightly higher, no downstep takes place between them. Relative to the pitch height of the topics, the comments are downstepped: all pitch accents are scaled clearly lower than those of the topics. Within the first comment, there is further differentiation: in the verbal complex, the pitch accent on the auxiliary *está*, if accented at all, is scaled lower than on the verb *chupando*. The pitch accent on *flor* is downstepped relative to that on *chupando*, but it is very clearly realized nonetheless. The last pitch accent in the comment, on *derecha*, is upstepped relative to the preceding one: it reaches about the same height as the one on *chupando* again. The same overall scaling relation can be observed in the second part, although the verb here is not realized; but the topic is scaled higher than the comment overall, and the last element of the comment, *izquierda*, is not downstepped, but, if anything, upstepped. Literally on top of all this comes the scaling relation between the two parts of the utterance: part two is scaled to a lower pitch height, *overall*, than part one, while preserving the internal scaling relations that can also be observed in part one.

Figure 79 illustrates the downstep relations that obtain between the prosodic constituents of QZ13_ELQUD_ES_52 as described, via coloured reference lines:

206 https://osf.io/65yzm/

207 Similar occurrences of lacking gender and/or number agreement are frequent in the Huari Spanish data. This has often been attributed to Quechua contact influence (e.g. in Escobar 2000, 2011) and is one of the most frequently cited features of what has been labeled “Andean Spanish” (Andrade Ciudad 2021: 125).

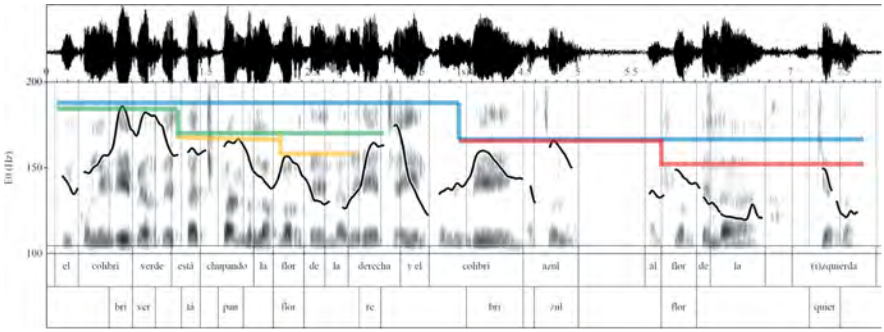


Figure 79: QZ13_ELQUD_ES_52 with reference lines added.

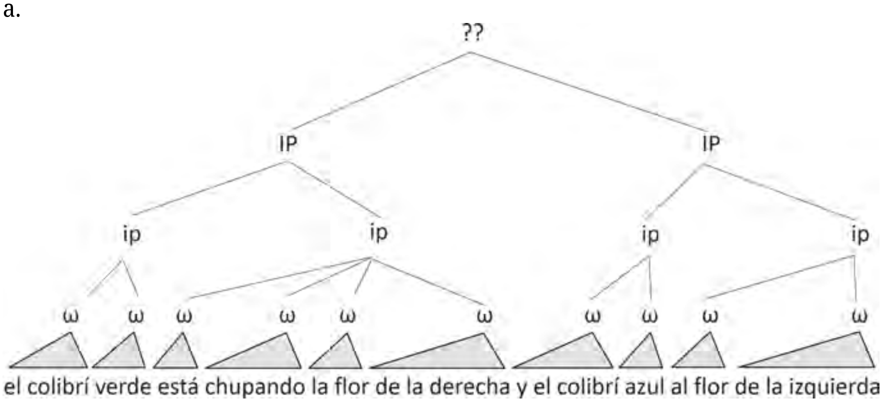
Figure 79 uses reference lines for register height to visualize the downstepping relations between the prosodic constituents in the example, a concept introduced in van den Berg et al. (1992) and developed further in Féry & Truckenbrodt (2005) and Truckenbrodt & Féry (2015). The blue reference line indicates overall H tone reference height for part one vs part two, the green line that for topic 1 vs comment 1, the yellow line that for the downstep relation between the main subcomponents of comment 1, the verbal complex and the object, and the red line that for topic 2 vs comment 2. The reference lines show very well that a downstep-within-downstep relation obtains here between the prosodic constituents; such a conception explains the partial reset that takes place on the second topic, where the pitch accents reach a height again that had previously already been passed below by some of the pitch accent in the first comment. A model using only global declination as a time-dependent effect taking place throughout the utterance or predicting pitch height of one pitch accent from the height of the preceding accent in an exponential model, the latter of which was found to best describe downstep between pitch accents in Mexican Spanish by Prieto et al. (1996), would be hard put to explain the downstep relations found here. Note that the reference lines describe only the downstep relations between prosodic constituents larger than the prosodic word; there is more than one prosodic word (deducible from the presence of pitch accents on accentable words) in many of the smallest parts defined by the lines. Considering this provides a way of reconciling the results in Prieto et al. (1996) with what is found here: in that study, only downstep within single NPs encompassing two to five accentable words bordering on the nonsensical and recorded in a reading task was investigated, so that no complex prosodic structure was assumed or predicted that could interact with scaling. Their results therefore are really applicable to downstep within one domain, and provide evidence that within that domain in Mexican Spanish, pitch decay on peaks does follow an exponential model, just as it has been shown to do in English (Liberman & Pierrehumbert 1984).[208] Coming back to the example at hand, this consideration opens an alternative analysis for the downstep described by the yellow reference line in the figure: we might also suppose that the downstep between *chupando* and *flor* is just downstep between two prosodic words, and not reflective of a downstep relation between larger constituents. At this point, we cannot really decide this issue, but it will be taken up again when incorporating upstep into the discussion: it might be more parsimonious to allow upstep only on nuclear pitch accents, i.e. those final in at least an iP/PhP (cf. section 3.4.4), instead of just any prosodic word; this would then favour the partition as described here by the yellow reference line, with both *está chupando* and *la flor de la derecha* forming separate constituents.

Regardless of how this issue is decided, it should be noted that the regularity in these downstep relations make a compelling argument against a fixed, non-recursive prosodic hierarchy: taking prosodic words as the domain at which exactly one pitch accent is assigned, and even leaving the downstep relation described by the yellow line aside, the downstep relations as described by the other lines force us to assign a phonological or intermediate phrase to the topic and comment each, whose downstep relation to each other are described by the green and red lines, and the whole extent of those lines, i.e. part one and part two each, will then have to be an intonational phrase. However, as described by the blue line, there is a downstep relation between those two, which could not be the case if the non-recursive IP was really the maximal prosodic constituent: it would then be the maximal domain of all phonological rule application and hence no phenomenon in one IP could make reference to a preceding (or upcoming) IP, as would here be the case. If we take the yellow line to describe a downstep relation between prosodic constituents larger than the prosodic word contained within a larger unit comprising the entire comment and then combining with the topic, the same problem arises already a step earlier in the rise through the constituents from bottom to top. These two alternatives, in the order they have been discussed, and a third, are schematized in (67)a-c.

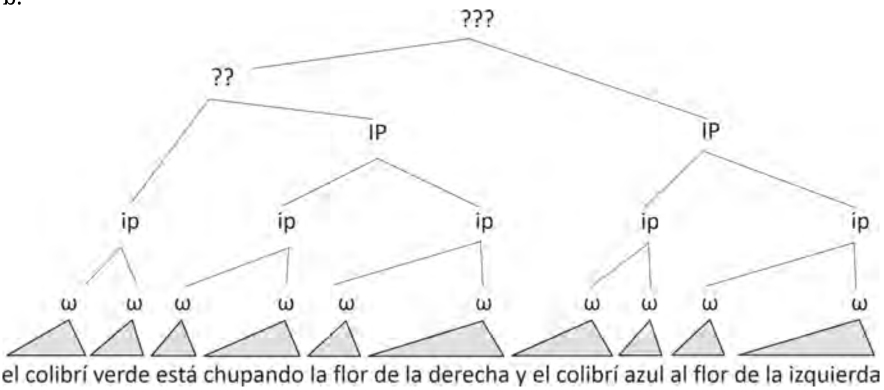
208 The reading stimuli for the Prieto et al. (1996) study expanded the noun phrase *rayo de luna* ‘ray of moonlight’ (two accents) to the maximal (and ambiguous in terms of parsing) *rayo de la luna de mi mayo de la gala de la Lola* ‘Lola’s gala’s moonlight ray of my May’ (five accents), while in the second experiment of Liberman & Pierrehumbert (1984), the stimuli consisted of lists with two to five names of berries, i.e. e.g. *blueberries and raspberries,* or *blueberries bayberries raspberries mulberries and brambleberries.*

(67) Three different prosodic structures for QZ13_ELQUD_ES_52

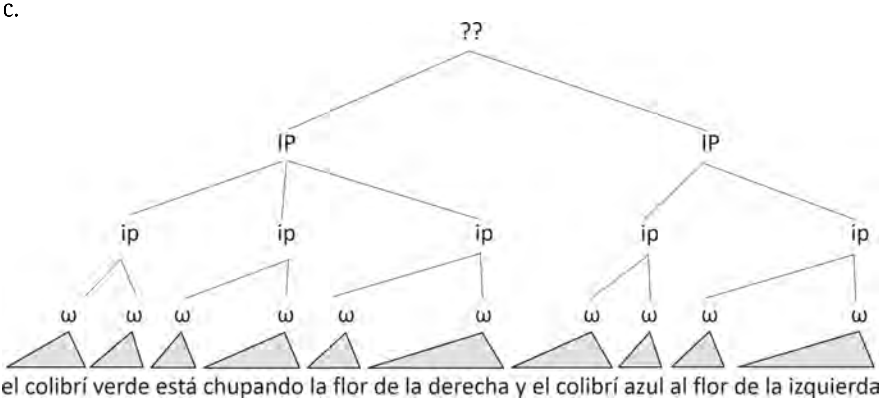
a.



b.



c.



(67)c is in a sense a compromise between (67)a and (67)b: it preserves the separate prosodic constituency of *está chupando* and *la flor de la derecho*, respectively, of the latter, but it does not add an intermediate prosodic level corresponding to the comment of part one. Assuming that some kind of downstep rule applies between all prosodic sisters under one node sequentially from left to right and that n-ary instead of just binary branching is allowed (as in the proposal for German in Féry & Truckenbrodt 2005: 234–236), then (67)c would just as well account for the observed downstep relation observed in the utterance as (67)a and (67)b. All three alternative structures are problematic for another reason: when they are assumed, intermediate phrases, just like intonation phrases, are usually taken to end on a boundary tone. In Figure 78, pitch movement relatable to boundary tones is only really observable at the boundary between part one and part two (a high boundary tone) and at the end of the entire utterance (a low boundary tone), but probably not iP-finally where the structures in (67) only assume intermediate phrases to end, e.g. at the end of the two topics. In principle, at least two options are available for explaining this discrepancy that preserve the concept that intermediate phrases have boundary tones:

1) There are boundary tones present at the right edges of all of the ips, but they are not as unambiguously identifiable because
   a. they occur at a lower level of embedding, which somehow makes them less conspicuous, **and/or**
   b. at the level of phonological representation, they are ambiguous between boundary tones and pitch accents (association to a prosodic edge or accented syllable), and since they would occur here in such close proximity to an already existing pitch accent on a lexically accented syllable, they secondarily associate with that syllable (on precisely this phenomenon in Eastern European languages, see Grice et al. (2000); Ladd (2008)), making them seem part of the pitch accent in the phonetic implementation, *and perhaps*, they are more likely to do so when the edge they would normally associate with is lower-level or more deeply embedded than when it is high-level or less embedded (which is why they surface between part one and part two); **or**
2) There are no boundary tones present at the right edges of these prosodic constituents, and they have been wrongly identified as intermediate phrases.

In the discussion of double topic-utterances exemplifying a type of “phrase accentuation” (section 5.2.4), we will see utterances where boundary tones are more clearly present also in some of the places where (67) takes iPs to end in.

An alternative explanation for the scaling difference in the pitch accents, namely that it is simply due to whether or not a pitch accent is adjacent to a boundary tone, with high boundaries causing pitch accents in their vicinity to be scaled high, and low ones causing them to be scaled low, can probably be discarded. On the one hand, this is because the final word in the second part, *izquierda*, is scaled as high as the preceding *flor*, even though only *izquierda* is followed by an L boundary tone, and it additionally comes after *flor*, so that it should be even lower due to declination or downstep. On the other, we might assume that if there are really iPs forming on smaller groupings here, that they are accompanied by their own boundary tones also when followed by an IP-level boundary tone, i.e. T-T%. As mentioned in section 3.4.7, this is a theoretical convention not usually followed in the literature on Spanish intonation, except for in Gabriel (2007). But if we were to follow this convention, it might be proposed that underlyingly, the tonal configuration on the final *izquierda* here is LH* H-L%, with the additional H- only surfacing in a "boost" to the scaling of the pitch accent. This approach would lead to problems, however. On the one hand it is somewhat circular because it couldn't independently postulate the presence of such "hidden" boundary tones only visible through their pitch boosting behaviour except where scaling behaviour was seemingly in need of an explanation. On the other, if these boundary tones are supposed to be present systematically, then at least the first and second topic, and *derecha* and *izquierda*, respectively, would be expected to be scaled at the same height. Since they aren't, the systematic relation in their scaling still needs to be explained, boundary tones or not, and parsimony would then suggest not stipulating a construct such as "hidden" boundary tones if they do not help to explain a phenomenon.

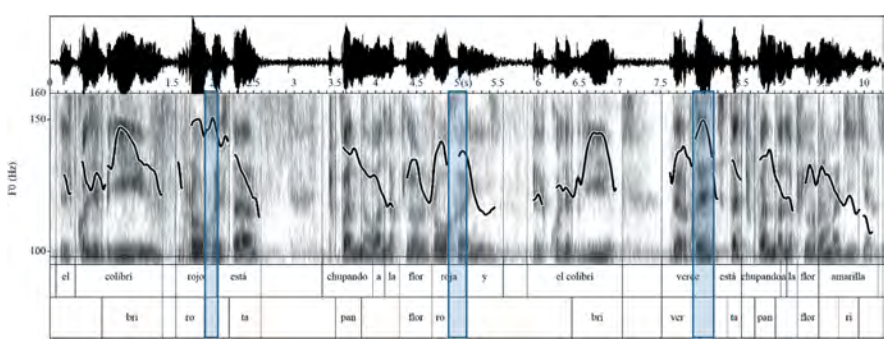


Figure 80: OZ14_ELQUD_ES_43[209] (*el colibrí rojo está chupando a la flor roja y el colibrí verde está chupando a la flor amarilla* 'the red hummingbird is drinking from the red flower and the green hummingbird is drinking from the yellow flower'). Blue marking highlights high posttonic pitch indicating the presence of H boundary tones.

209 https://osf.io/hp4fm/

This point is further corroborated when considering an additional example, Figure 80. Here, the scaling relation between pitch accents belonging to the different parts of the utterance is less easily visible because the differences in pitch range between accents lying on the same reference line is smaller, but it is still present. In addition, it is different in that while in Figure 78, H boundary tones between topic and comment could not really be identified, they are unambiguously present here, identifiable from high pitch on the posttonic of *rojo* and *verde*, in addition to that between the first and second part, identifiable from high pitch on the posttonic of *roja* (all three marked in blue in Figure 80). However, here the presence or absence of boundary tones also does not seem solely responsible for the scaling of the surrounding pitch context: the high boundary tone in the three positions is preceded by a very high first topic, a less higher first comment and a second topic that is in between these two heightwise, respectively. The peaks on both comments follow a general downward trend, although the first one is followed by a high boundary tone, and the second by a low one, and the topics do not show this downward trend, although they are followed by a high boundary tone, just like the first comment. All this means that the binary identity of a boundary tones (H or L) is insufficient to explain the observed scaling relation, which evidences more levels than just two and is partially independent of boundary tones. The quantitative analysis will corroborate this impression.

5.2.3 Quantified analysis of hierarchical scaling structure via relational measurements

For a broader view of things, let us look at how the scaling relations between individual units turn out under quantification. To that purpose, the double topic-utterances from *Elqud*, i.e. responses by the ten speakers OZ14, QP44, QZ13, SG15, TP03, XJ45, ZZ24, ZE55, NQ01, and LJ22 in the experimental items 14, 26, 43, 44, 52, 59, and 65 were annotated by hand and measurements were taken via a praat script (see Appendix C). Only those utterances qualifying as full responses were considered. The following annotations and measurements were taken: mean, minimum, and maximum F0 on every accentable syllable, representing scaling of pitch accents, and on every syllable that is the final voiced one in one of the four parts of each utterance, i.e. in topic 1, comment 1, topic 2, and comment 2, respectively, representing scaling of potential boundary tones there. The time stamps of all of these measurements were also taken. The word class of each word bearing an accentable syllable was also annotated, as well as whether the word was contrastive in the utterance or not. For each utterance, a highest pitch measurement ($maximum_{\text{utterance}}$) was determined that was not due to consonantal microprosody or noise. In the same way, a lowest pitch for each utterance was determined ($minimum_{utterance}$). These utterance maxima and minima served to derive normalized values for each measurement taken on accentable and part-final syllables, according to the following linear transformation adapted from Truckenbrodt & Féry (2015: 29):[210]

(68) Transformation for normalized pitch values

$$pitch\ value_{transformed} = \frac{(pitch\ value_{measured} - minimum_{utterance})}{(maximum_{utterance} - minimum_{utterance})}$$

Transformed pitch values are all on the same scale and thus become relative measurements that are comparable between speakers and utterances, even though individual items might differ substantially in their absolute values. Some further utterances had to be excluded from the final measurements because of insufficiently voiced pitch tracks or because they included too many false starts and hesitations or ellided contrastive words. One utterance was also excluded because it was produced in a different word order than all others. This leaves a total of 51 utterances whose measurements were considered in the final analysis in *R*, distributed across speakers and experimental items as shown in Table 13.

Table 13: Spanish Elqud utterances considered in the analysis, sorted according to speaker and experimental item.

	experimental item							
speaker	14	26	43	44	52	59	65	*total*
OZ14	✓	✓	✓	×	✓	✓	✓	6
QP44	×	✓	✓	×	✓	✓	✓	5
QZ13	✓	✓	✓	×	✓	✓	✓	6
SG15	✓	×	×	✓	×	✓	×	3
TP03	✓	✓	✓	×	×	×	×	3
XJ45	✓	×	✓	×	✓	✓	✓	5
ZZ24	✓	✓	✓	✓	✓	×	✓	6
ZE55	✓	✓	✓	×	×	✓	✓	5
NQ01	✓	✓	✓	✓	✓	✓	✓	7
LJ22	✓	✓	✓	×	×	✓	✓	5
total	9	8	9	3	6	8	8	51

210 Truckenbrodt & Féry (2015) use averages of values taken at predetermined positions for the minimum and maximum instead of utterance-specific measurements. Since the data in the present work are more spontaneous, the measurement were adapted in the way described.

As can be seen from the table, most speakers did not produce a full analyzable utterance in response to all experimental items. Since item 44 was a filler and did not contain a mismatch between the visual and the audio stimulus, most speakers only responded with short responses along the lines of *sí está bien* 'yes it's alright', leaving only 3 full responses.

In the following, I present pooled measurements from all utterances considered for each of the experimental items alongside an individual example. The measurements are the transformed values of the mean measurements for the accented and part-final syllables, of which again the means across all utterances were taken.

Figure 81 gives these pooled values for the experimental item 14, the shortest and least complex one. In all the barplots in section 5.2.3, bar height represents mean across all utterances, error bars represent one standard deviation above and below the mean, respectively. In sections 5.2.3.6 and 5.2.3.7 I will discuss some of the variation represented by the error bars separately. Represented for topic one and two are the mean values for the accented syllable and the part-final, boundary-adjacent measurement. For the comments in part one and two, the value for the accented syllable of the noncontrastive word (the copular auxiliary verb *está*) and of the contrastive word (the verb in its present participle) as well as of the part-final measurement are given. Bars are given in the order of the position whose mean values they represent. Error bars represent one standard deviation above and below the mean across all utterance for each bar. I will talk about some of the variability the error bars represent specifically in the discussion. Inferential statistical analysis over the entirety of the data also follows in section 5.2.3.6. The mean values here most notably show a considerable scaling difference between the first and the second part. They also indicate a general downward trend/downstep within each part that is counteracted by increased scaling/upstep on the contrastive pitch accented syllable which is also final. A scaling difference between topic and comment is less clearly represented, and only between topic and comment of the second part.

5.2.3.1 ELQUD-item 14

Figure 82 gives the utterance produced in response to this experimental item by speaker SG15. Overall it shows the scaling differences also observed in the pooled measurements. However, the contrastive verb in the second comment is here not produced with increased scaling/upstep, leading to a clearer differentiation between the second topic and comment via relative scaling.

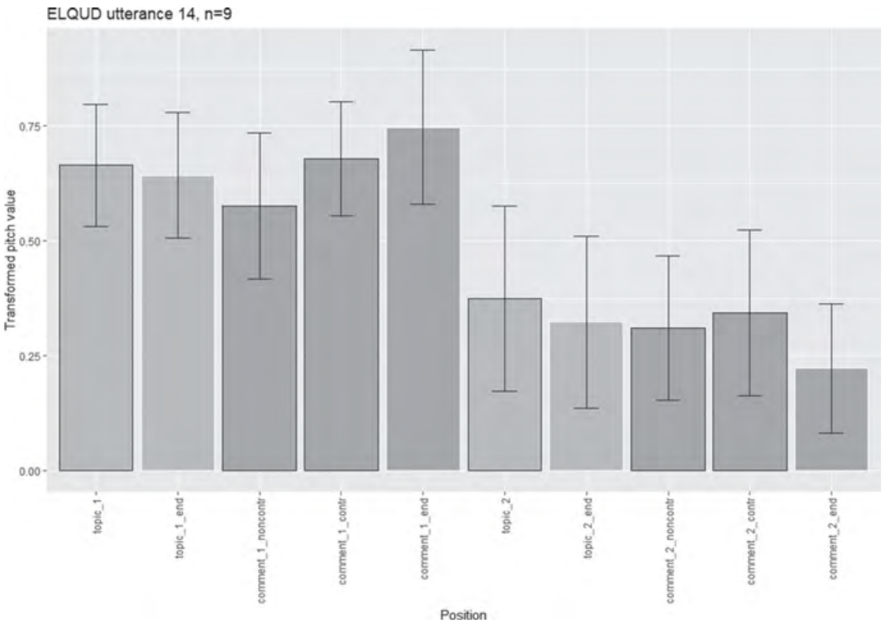


Figure 81: Barplot of pooled transformed mean pitch values from all utterances of experimental item 14 from Spanish *Elqud*.

Example: [*el* ***hom****bre*]$_{\text{topic1}}$ [*es****tá***]$_{\text{comment1\_noncontr}}$ [*dur****mien****do*]$_{\text{comment1\_contr}}$ [*la* ***ni****ña*]$_{\text{topic2}}$ [*es****tá***]$_{\text{comment2\_noncontr}}$ [*co****mien****do*]$_{\text{comment2\_contr}}$ (cf. Figure 82).

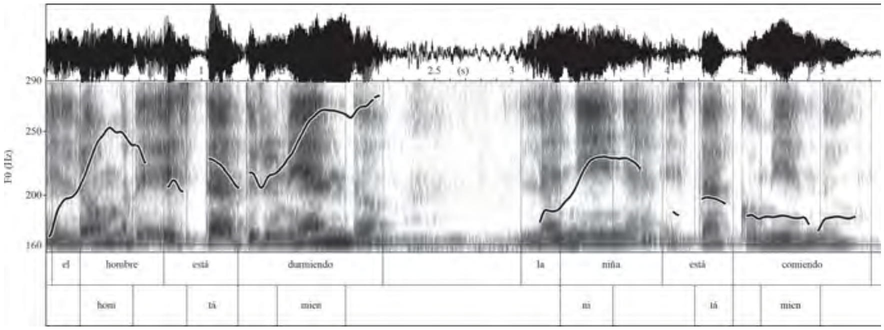


Figure 82: SG15_ELQUD_ES_14[211] (*el hombre está durmiendo la niña está comiendo* ‘the man is sleeping the girl is eating’).

211 https://osf.io/x9d4r/

5.2.3.2 ELQUD-item 26

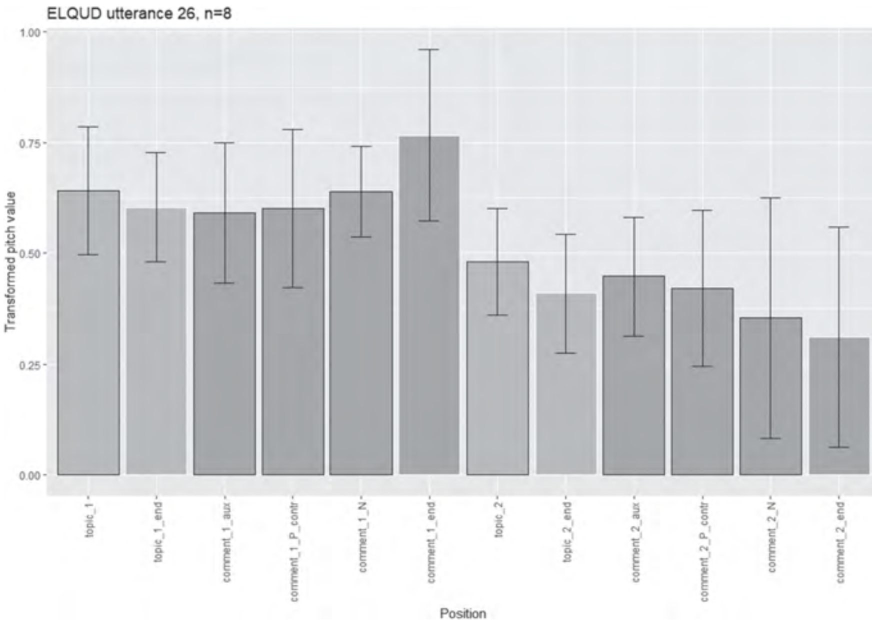


Figure 83: Barplot of pooled transformed mean pitch values from all utterances of experimental item 26 from Spanish *Elqud*.

Example: [*el **pe**rro*]$_{\text{topic1}}$[*es**tá***]$_{\text{comment1\_aux}}$[*en**ci**ma*]$_{\text{comment1\_P\_contr}}$[*de la **ro**ca*]$_{\text{comment1\_N}}$[*el **ga**to*]$_{\text{topic2}}$ [*es**tá***]$_{\text{comment2\_aux}}$[*de**ba**jo*]$_{\text{comment2\_P\_contr}}$[*de la **ro**ca*]$_{\text{comment2\_N}}$ (cf. Figure 84).

Figure 83 gives the pooled values for responses to experimental item 26. The topics remain simple, while the comments are more complex compared to 14, with bars giving the mean values for the auxiliary copular verb *está* (aux), a contrastive prepositional noun (P), and a noncontrastive noun (N) in both comments.[212] Note that the word bearing the contrast (*encima* "above" and *debajo* "below" in OZ14's response, cf. Figure 84) here is prefinal in the comments, unlike in 14, where it was final. As in item 14, there is a clear scaling difference between the two parts which are also separated by a high boundary rise, but the difference between topic and comment more clearly emerges only in the second part. The first comment is initially scaled just a little lower than the first topic, but its final element is higher scaled/upstepped again, to the same level as the topic. Note again that this final

212 Not all speakers produced each element in the second comment.

element is not the contrastive one in the first comment, which precedes it and is scaled lower on average.

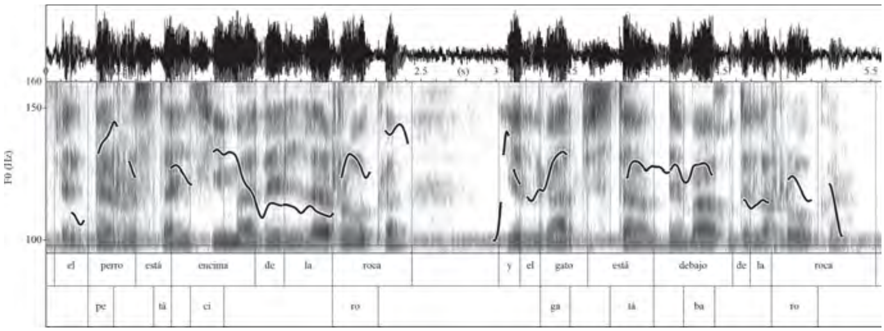


Figure 84: OZ14_ELQUD_ES_26[213] (*el perro está encima de la roca y el gato está debajo de la roca* 'the dog is above the rock and the cat is below the rock').

5.2.3.3 ELQUD-items 43 and 44

With experimental item 43, whose values are given in Figure 85, internal complexity increases yet again. While the topics in 14 and 26 consisted of a single noun, here they consist of a noncontrastive noun followed by an adjective that is contrastive. For comment 1, in addition to the auxiliary, values are given for the accentable syllable on the verb in the present participle, followed by a noncontrastive noun and its contrastive adjective. For comment 2, most speakers (except for OZ14 and QP44) only realized the noncontrastive noun and the contrastive adjective, so the values for the noncontrastive elements are represented in a single bar. In the first comment, ZE55 did not produce the verb, NQ01 the auxiliary, and LJ22 neither of them. Regarding scaling, in this item the differences between topic and comment in both parts are most pronounced. We also again see upstep or increased scaling on the contrastive or final (which cannot be determined here) accentable word in the subparts except for the last one. Unlike in the previous two items, comment 1 and topic 2 here are nearly at the same level, with topic 2 even scaled slightly higher. While this is a very slight difference, it is rather incompatible with a purely local downstep model. This is even more so when considering that the final boundary tones in both parts are scaled roughly to the same height, but after comment 1 this is followed by a continuation at the same or increased register level, while after topic 2 it is followed by a considerable drop and then decline. Their being at the

213 https://osf.io/uzwa4/

same level, and topic 2 being scaled slightly higher, is however fully compatible with a model assuming a nested prosodic structure cued via these scaling relations, in which a first sister at the same phrasal level is scaled higher than following ones, such as represented in Figure 86 and proposed by Truckenbrodt & Féry (2015) for German.

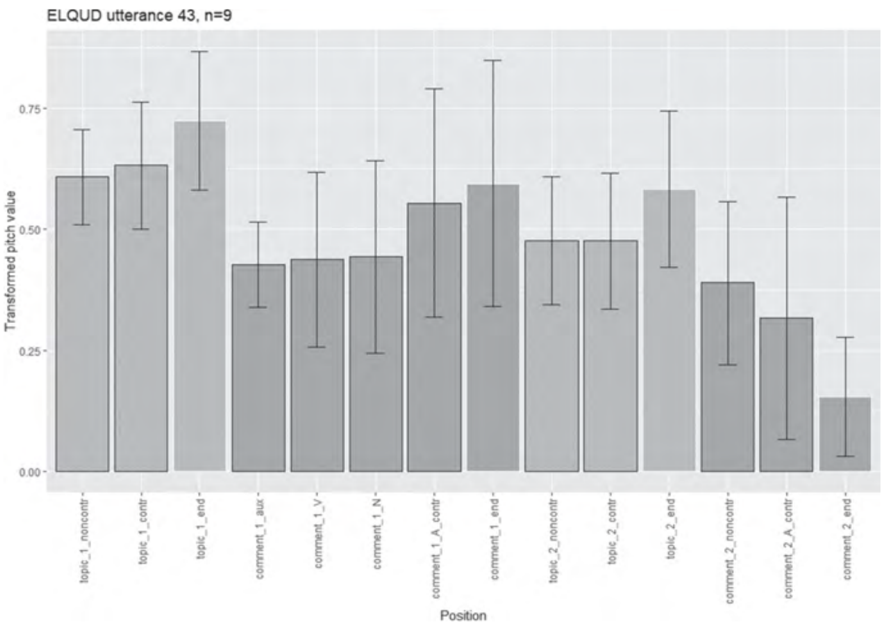


Figure 85: Barplot of pooled transformed mean pitch values from all utterances of experimental item 43 from Spanish *Elqud*.

Example: [*el coli**brí***]$_{\text{topic1\_noncontr}}$[***ro**jo*]$_{\text{topic1\_contr}}$[*es**tá***]$_{\text{comment1\_aux}}$[*chu**pan**do*]$_{\text{comment1\_V}}$[*la **flor***]$_{\text{comment1\_N}}$ [***ro**ja*]$_{\text{comment1\_A\_contr}}$[*el coli**brí***]$_{\text{topic2\_noncontr}}$[***ver**de*]$_{\text{topic2\_contr}}$ [*la **flor***]$_{\text{comment2\_noncontr}}$[*ama**ri**lla*]$_{\text{comment2\_A\_contr}}$ (cf. Figure 87).

If, as represented in Figure 86, the four information structurally determined subparts of the utterance are each assumed to form a phrase, and those in turn group in twos to form phrases corresponding to the two main parts which then come together to form the utterance as a whole, and this relation is reflected in tonal scaling as described, then this would perfectly explain the scaling difference observed. Note that this necessitates assuming more than two phrasal prosodic units of iP-level or above, that otherwise do not seem to be differentiated in their prosodic makeup. Thus such a model would require a recursive prosodic structure at the ip/IP-level, as discussed in section 3.6. Assuming a flat structure at the subpart-level, thus with quaternary branching, could not explain the observed different scaling relations between topic 1 and comment 1, and topic 2 and comment 2, on the one hand, and comment 1 and topic 2, on the other. A structure with ternary branching, taking comment 1 and topic 2 as a single iP, on the other hand, would clash with the attested boundary tones between these two parts. The reality of the largest phrasal unit, given as “iP3” in Figure 86, besides being necessary as the domain of full reset, is furthermore supported by the fact that it is closed by a low boundary tone, and that the last contrastive or final accent in it is not upstepped, as in all non-final phrases, but downstepped.

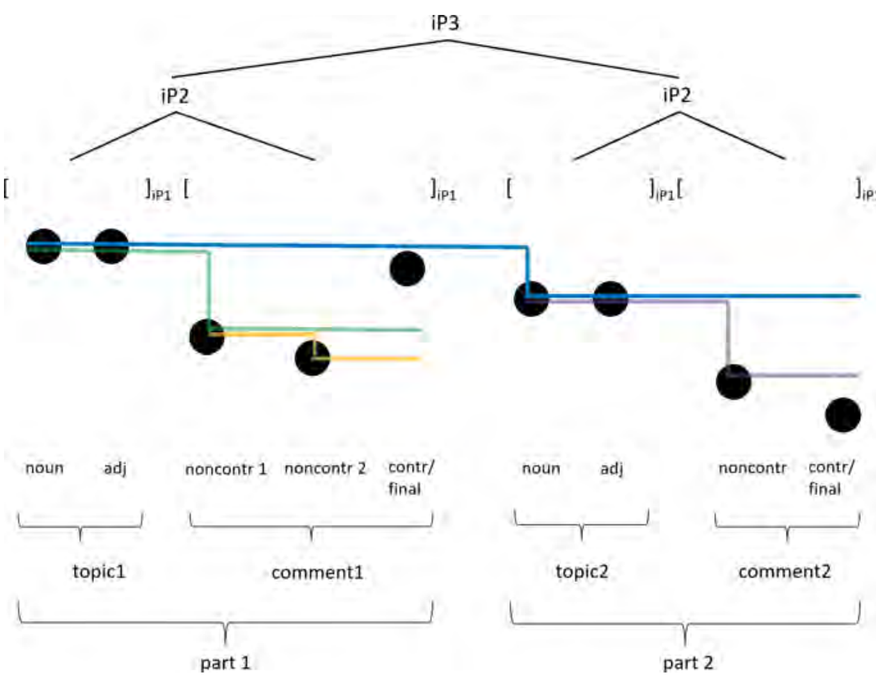


Figure 86: Schematized scaling model with register lines and phrasal prosodic structure for complex Spanish double topic utterances from *Elqud.*

We already saw an example of a speaker response to item 43 in Figure 80 above. Figure 87 gives another, by speaker XJ45. Note the somewhat unexpected increased scaling on the accented syllable of the noncontrastive noun of topic 1 and the boundary tone following the second topic.

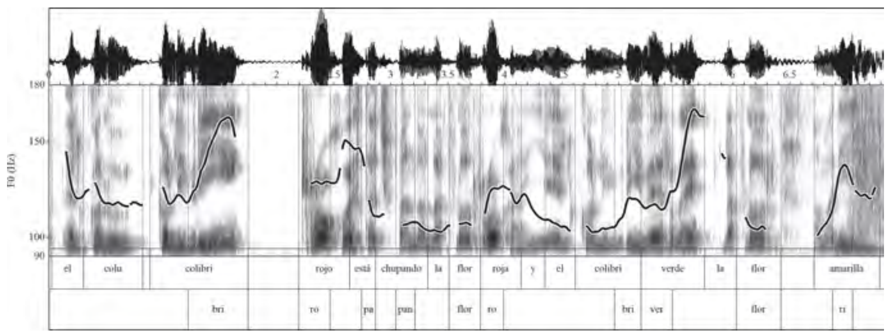


Figure 87: XJ45_ELQUD_ES_43[214] (*el colu- colibrí rojo está chupando la flor roja y el colibrí verde la flor amarilla* ‘the red hummingbird is drinking the red flower and the green hummingbird the yellow flower’).

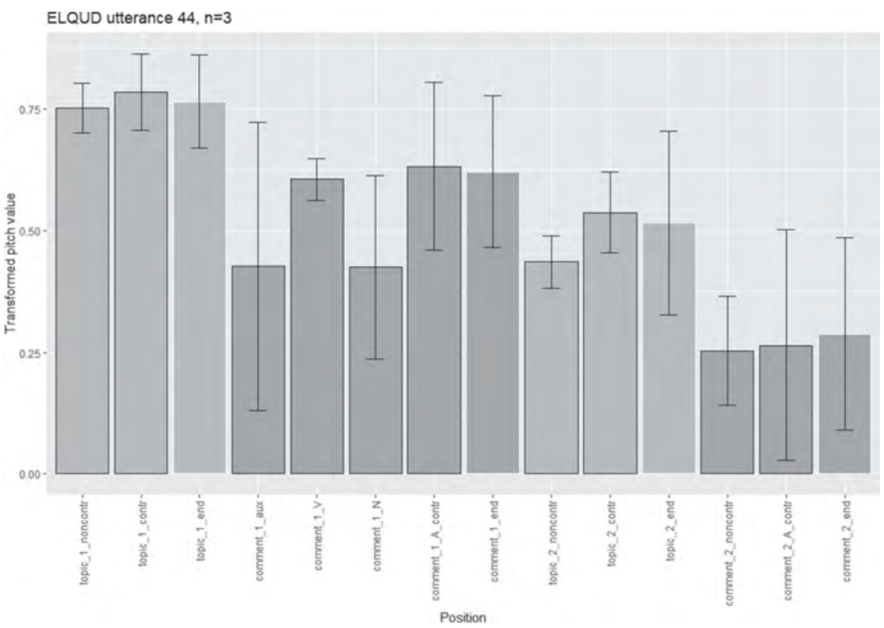


Figure 88: Barplot of pooled transformed mean pitch values from all utterances of experimental item 44 from Spanish *Elqud*.

Example: [***pe**rro*]$_{\text{topic1\_noncontr}}$[***ne**gro*]$_{\text{topic1\_contr}}$[*es**tá***]$_{\text{comment1\_aux}}$[*ju**gan**do*]$_{\text{comment1\_V}}$ [*con el pe**lo**ta*]$_{\text{comment1\_N}}$[***ne**gra*]$_{\text{comment1\_A\_contr}}$[***pe**rro*]$_{\text{topic2\_noncontr}}$[***blan**co*]$_{\text{topic2\_contr}}$ [*con la pe**lo**ta*]$_{\text{comment2\_noncontr}}$[*a**zul***]$_{\text{comment2\_A\_contr}}$ (cf. Figure 89).

214 https://osf.io/ugehf/

Figure 88 gives the pooled values for item 44. They come from only three utterances and so are prone to individual variation, but it is still interesting to see that essentially the same scaling relations obtain as for item 43, even though responses to 44 are not partial denials, since no mismatch between visual and audio stimulus exists. Figure 89 gives an example of a response to item 44 by speaker ZZ24.

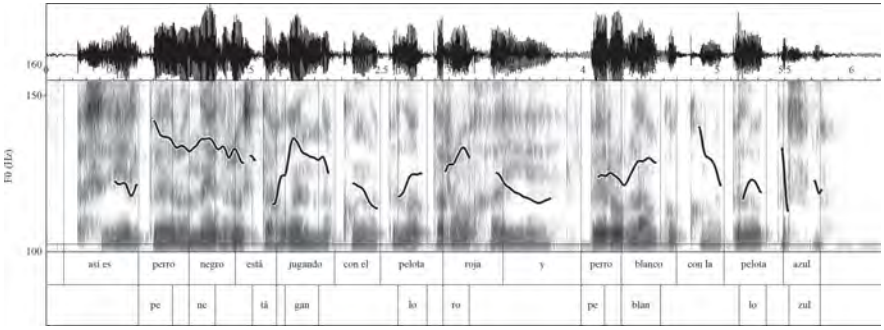


Figure 89: ZZ24_ELQUD_ES_44[215] (*así es perro negro está jugando con el pelota roja y perro blanco con la pelota azul* 'that's it black dog is playing with the red ball and white dog with the blue ball').

5.2.3.4 ELQUD-item 52

Figure 90 gives the pooled values for responses to experimental item 52. Here in the first comment, speakers varied between a realization with a relative clause as in the audio stimulus, e.g. *el picaflor verde está chupando la flor que está a la derecha* "the green hummingbird is drinking from the flower that is on the right" (NQ01), and a less complex version, e.g. as in QZ13's (Figure 78), which has *el colibrí verde está chupando la flor de la derecha* "the green hummingbird is drinking from the flower on the right". In order to treat these uniformly, the second instance of the auxiliary was not included in the values given in the barplot. In terms of scaling, the results are very similar to those of experimental item 43 and 44, further supporting the analytical consequences drawn from them. An interesting difference concerns the accent on the noncontrastive noun in the first topic, which is scaled higher than the one on the contrastive adjective following it. This occurs variably (cf. the error bars) in utterances produced in response to all of the double topic-items, e.g. also in XJ45's 43, seen in Figure 87. It could be something like a boost occurring only on the accent that is initial in the utterance as a whole, and would as such serve as an additional scaling cue to the overall prosodic structure. I discuss

215 https://osf.io/qskvf/

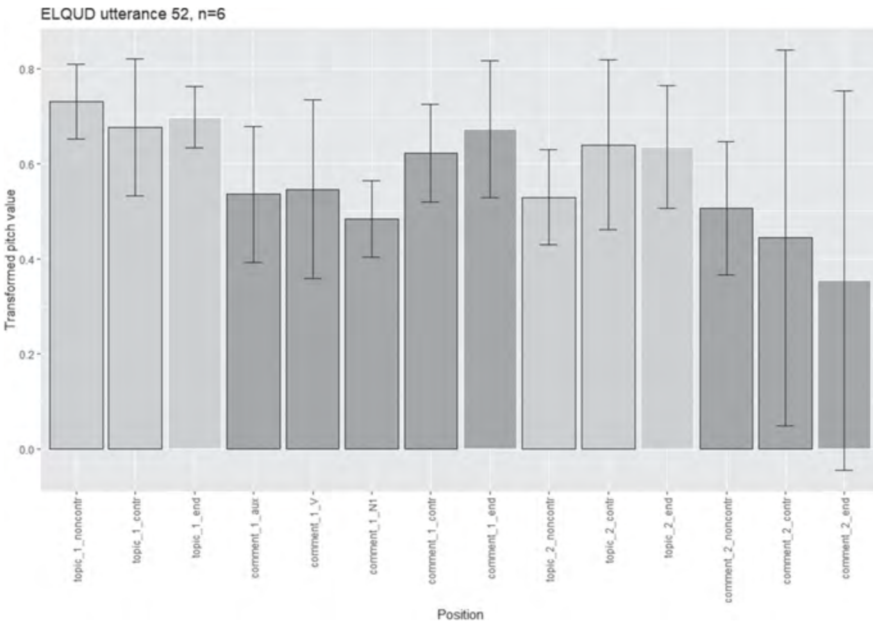


Figure 90: Barplot of pooled transformed mean pitch values from all utterances of experimental item 52 from Spanish *Elqud*.

Example: [*coli**brí***]$_{\text{topic1\_noncontr}}$[***ver**de*]$_{\text{topic1\_contr}}$[***está***]$_{\text{comment1\_aux}}$[*chu**pan**do*]$_{\text{comment1\_V}}$[*la **flor***]$_{\text{comment1\_N1}}$[*que está a a la de**re**cha*]$_{\text{comment1\_contr}}$[*coli**brí***]$_{\text{topic2\_noncontr}}$[***azul***]$_{\text{topic2\_contr}}$ [*está chupando la **flor***]$_{\text{comment2\_noncontr}}$[*que está a la iz**quier**da*]$_{\text{comment2\_A\_contr}}$ (cf. Figure 91).

this further in section 5.2.3.7. Above, we already saw QZ13’s utterance produced as response to this experimental item (Figure 78); Figure 91 is another example produced by speaker QP44.

5.2.3.5 ELQUD-items 59 and 65

Finally, we will consider the results from experimental items 59 (Figure 92) and 65 (Figure 93). Responses to these items were very similar, which reflects that this is also the case for their stimuli (cf. Table 11). The main difference in the stimuli is that for item 65, the location of the two skunks in both parts is the same (*el zorrillo grande está al frente de la casa grande, y el zorrillo pequeño está al frente de la casa pequeña* “the big skunk is in front of the big house and the small skunk is in front of the small house”) and not a mismatch between visual and audio stimulus, while in 59, it is (*el zorrillo pequeño está detrás de la casa pequeña, y el zorrillo grande está al frente de la casa grande* “the small skunk is behind the small house and the big skunk is in front of the big house”). Thus in 59, but not in 65, the prepositional

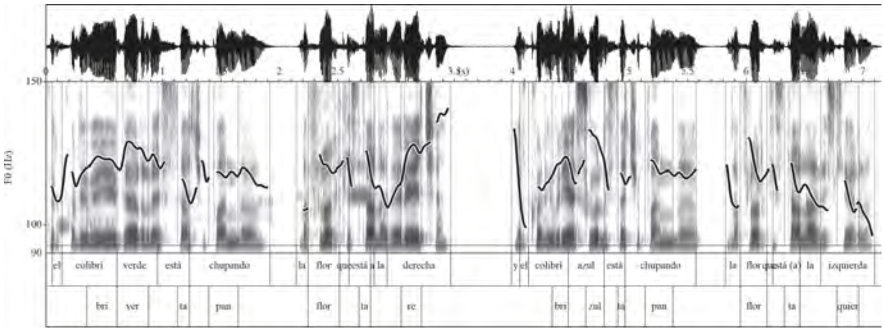


Figure 91: QP44_ELQUD_ES_52[216] (*el colibrí verde está chupando la flor que está a la derecha y el colibrí azul está chupando la flor que está a la izquierda* 'the green hummingbird is drinking from the flower that's on the right and the blue hummingbird is drinking from the flower that's on the left').

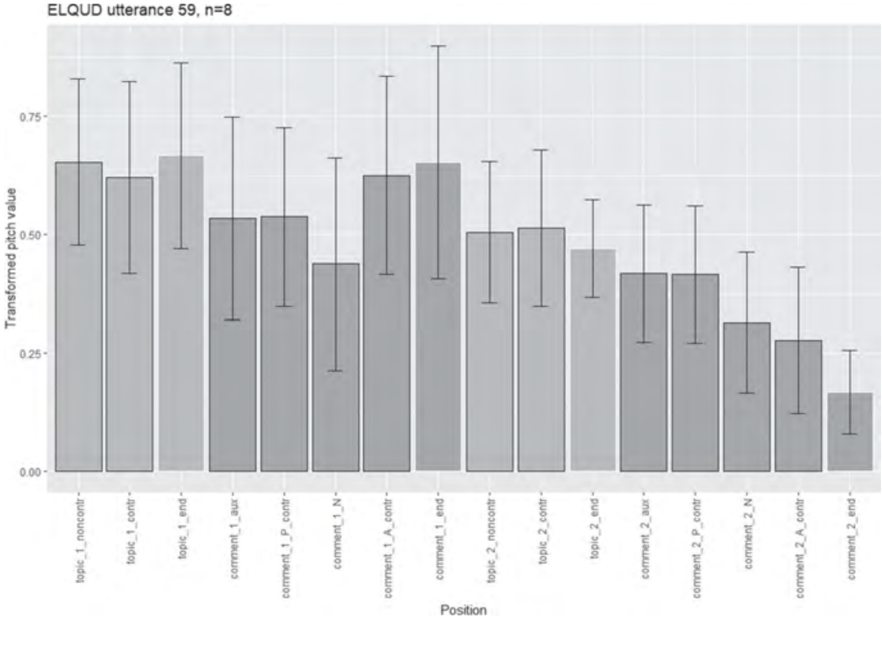


Figure 92: Barplot of pooled transformed mean pitch values from all utterances of experimental item 59 from Spanish *Elqud*.

Example: [*el zo***rri***llo*]$_{\text{topic1\_noncontr}}$[*pe***que***ño*]$_{\text{topic1\_contr}}$[*es***tá**]$_{\text{comment1\_aux}}$[*al* **fren***te*]$_{\text{comment1\_P\_contr}}$ [*de la* **ca***sa*]$_{\text{comment1\_N}}$[**gran***de*]$_{\text{comment1\_A\_contr}}$ [*el zo***rri***llo*]$_{\text{topic2\_noncontr}}$[**gran***de*]$_{\text{topic2\_contr}}$[*es***tá**]$_{\text{comment2\_aux}}$ [*de***trás**]$_{\text{comment2\_P\_contr}}$[*de la* **ca***sa*]$_{\text{comment2\_N}}$[*pe***que***ña*]$_{\text{comment2\_A\_contr}}$ (cf.Figure 94).

216 https://osf.io/5ema7/

noun (P in the figures) in both comments is potentially contrastive, in addition to the adjectives in both topics and comments. This is also marked in the labels below the bar giving the pooled value for the prepositional noun in the two figures.

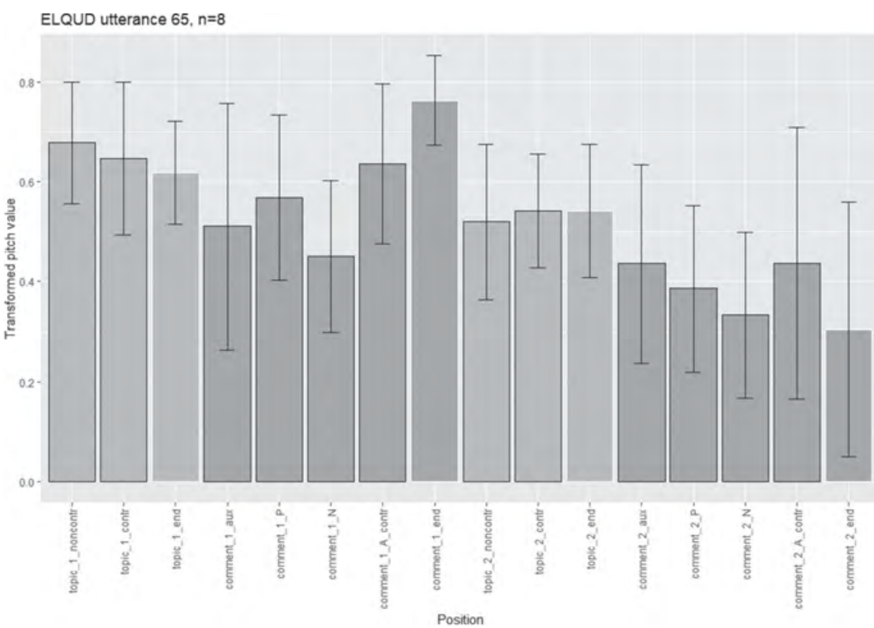


Figure 93: Barplot of pooled transformed mean pitch values from all utterances of experimental item 65 from Spanish *Elqud*.

Example: [*el zo**rri**llo*]$_{\text{topic1\_noncontr}}$[***gran**de*]$_{\text{topic1\_contr}}$[*es**tá***]$_{\text{comment1\_aux}}$[*al **fren**te*]$_{\text{comment1\_P}}$ [*de una **ca**sa*]$_{\text{comment1\_N}}$[*pe**que**ña*]$_{\text{comment1\_A\_contr}}$ [*el zo**rri**llo*]$_{\text{topic2\_noncontr}}$[*pe**que**ño*]$_{\text{topic2\_contr}}$[*es**tá***]$_{\text{comment2\_aux}}$ [*al **fren**te*]$_{\text{comment2\_P}}$[*de una **ca**sa*]$_{\text{comment2\_N}}$[***gran**de*]$_{\text{comment2\_A\_contr}}$ (cf. Figure 95).

Interestingly, this difference does not translate to an increased scaling on the prepositional nouns in the responses to 59 (Figure 92) compared to those to 65 (Figure 93). If anything, the opposite can be observed, namely that those values in the responses to 65 are actually scaled higher. This could suggest that it is not contrastiveness per se that causes increased scaling, but instead whether a pitch accent is final in a phrase. This would also be supported by the results from item 26 above. Apart from this somewhat unexpected result, the scaling relations between the main parts for responses to items 59 and 65 are very comparable to those for the other more complex items 43, 44, and 52. They are thus also very compatible with the hierarchical scaling model proposed for 43 and only insufficiently explained by models that seek to explain downstep/declination only locally and without making reference to a recursive prosodic structure. Figures 94 and 95 give examples for responses to item 59, and 65, respectively, by speakers LJ22 and QZ13.

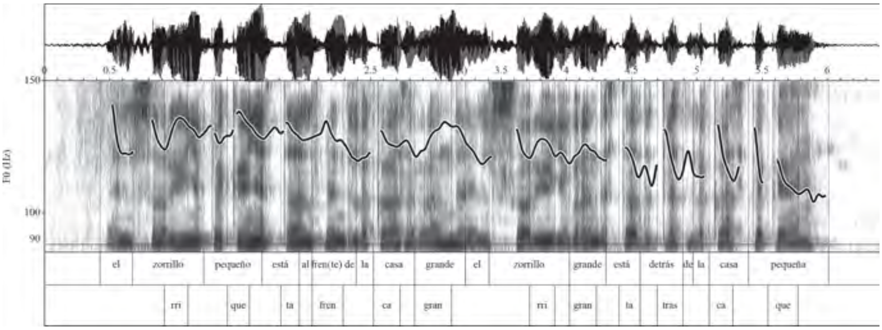


Figure 94: LJ22_ELQUD_ES_59[217] (*el zorrillo pequeño está al frente de la casa grande el zorrillo grande está detrás de la casa pequeña* ‘the small skunk is in front of the large house the large skunk is behind the small house’).

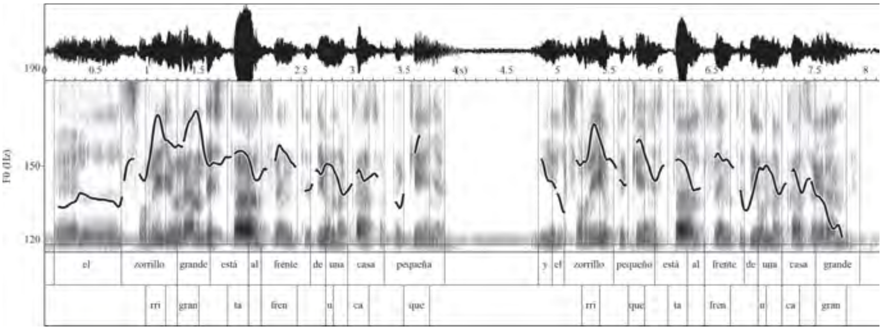


Figure 95: QZ13_ELQUD_ES_65[218] (*el zorrillo grande está al frente de una casa pequeña y el zorrillo pequeño está al frente de una casa grande* ‘the large skunk is in front of a small house and the small skunk is in front of a large house’).

5.2.3.6 Inferential statistical results

The foregoing descriptive results strongly suggest that scaling on pitch accented syllables in these complex utterances is affected by a hierarchical prosodic structure that itself reflects an information structural partition into two topics with their comments. To support this conclusion, inferential statistics were run on the trans-

217 https://osf.io/rg8zh/
218 https://osf.io/c3uv9/

formed pitch height data. They cannot replace or even fully replicate the linguistic accounts rivaling to explain the scaling phenomena observed in the data, but they can constitute evidence in favour or against them. The rivaling accounts here are: as a null hypothesis, pitch accent scaling is not actually affected by non-local relations and instead simply follows a declining course over the utterance, as originally proposed for English by Liberman & Pierrehumbert (1984) and for Mexican Spanish by Prieto et al. (1996). Against that I have proposed as alternative hypothesis a recursive prosodic structure which affects pitch accent scaling non-locally via register levels (mapping to information structural partitions), similar to what has been proposed for German by Truckenbrodt & Féry (2015). To approximate these diverging accounts, a multiple regression model was fitted to the data with two categorical variables and one continuous predictor variable. The model takes *subpart of the utterance* (with the four levels topic 1, comment 1, topic 2, and comment 2), *contrastiveness*, i.e. whether an accentable syllable was part of a contrastive word or not, and *normalized start time of the accentable syllable from the beginning of the utterance* as predictor variables, and transformed mean pitch height on accentable syllables as dependent variable. In this regression model, the predictor variables of *utterance subpart* and *contrastiveness* represent the hierarchical scaling account, while *syllable start time* as predictor variable approximates the purely local declination account. According to the model, transformed mean pitch value on accentable syllables is significantly or even highly significantly affected by which of the utterance subparts an accentable syllable belongs to. It is nearly significantly affected by whether the word is contrastive or not. It is not significantly affected, however, by the start time of the syllable.[219] The adjusted R^2 for the

219 These are the model results:

	estimate β	st. error β	t-value	p-value
intercept	0.51	0.02	26.42	< 0.001 ***
start time of accented syllable	0.01	0.00	1.47	0.142
comment2	-0.18	0.03	-7.11	< 0.001 ***
topic1	0.13	0.02	5.09	< 0.001 ***
topic2	-0.06	0.03	-2.51	0.0125 *
contrast on word: yes	0.03	0.02	1.61	0.107
$F(5, 489) = 30.4$, $p < 0.001$				
Adj. $R^2 = 0.229$				

The intercept in the model corresponds to the estimate for a syllable in a noncontrastive word in the first comment. While checking model assumptions, a few outliers were found that did not seem to be due to error and thus should be included. For that reason, a robust linear model using *rlm* in the R package *MASS* (Venables & Ripley 2007) with the same predictor and dependent values was model is only about 0.23, meaning that only 23% of the variation in the transformed pitch height data on accentable syllables is actually explained via the predictors included. That is not too surprising, since the recorded data are quite spontaneous and it is easy to see that there is a considerable amount of individual pitch height variation throughout the utterances from the examples we have seen so far. That the utterance subparts affect pitch height significantly despite this overall variability probably speaks to the robustness of this effect. In order to further explore the two rivaling accounts, a simple model was also constructed using only *start time of the syllable* as predictor. This yielded a significant result, but explained just 10% of variation in the data.[220] That this variable becomes insignificant in the more complex model suggests that its effect becomes superseded by the effect that *subpart* has on transformed pitch height. This is further confirmed by applying regression models to the pitch height values in each utterance subpart separately, again with *start time of the syllable* as predictor. Those models did not produce any significant results except in the second topic, where a later starting time actually implied a somewhat higher pitch value, against the expecations from the null hypothesis.[221] Figures 96 and 97 provide visualisations of the transformed mean

also executed. A Levene's test for homogeneity of variance also returned significant results, indicating variances are not equal across the predictor groups. Therefore, robust model estimates and p-values were also calculated using the *sandwich* package (Zeileis 2004; Zeileis et al. 2020) and the *coeftest* function from the *lmtest* package (Zeileis & Hothorn 2002). Since results from all of these robust approaches did not differ substantially from those of the non-robust original model (and in particular did not decrease significance of any of the predictors), only the results of the latter are reported here.

220 These are the results for this simplified model:

	estimate β	st. error β	t-value	p-value
intercept	0.61	0.02	38.41	< 0.001 ***
start time of accented syllable	-0.02	0.00	-7.4	< 0.001 ***
F(1, 493)= 54.8, p < 0.001				
R^2= 0.1				

221 These are the results for the model predicting transformed mean pitch height on accentable syllables via syllable start time only within the second topic:

	estimate β	st. error β	t-value	p-value
intercept	0.36	0.05	7.25	< 0.001 ***
start time of accented syllable	0.03	0.01	2.96	< 0.004 **
F(1, 82)= 8.74, p = 0.004				
R^2= 0.1				

Within the other three subparts of the utterance, the model did not produce significant results.

pitch value distributions per utterance subpart. The difference in the values per subpart of the utterances is clearly visible, with those for the first comment and the second topic basically at the same level in the longer utterances responding to item 43, 44, 52, 59, 65. Yet we saw in the previous sections that there is a clearly identifiable boundary pitch movement between those two subparts, so that they each must correspond to a prosodic unit. Together, these results firmly support the conclusion reached from the inspection of individual examples and the pooled values for each experimental item, namely that the null hypothesis account explaining pitch height via local downstep or declination is insufficient, and that instead the hierarchical scaling account is far more capable of explaining what is observed. That the pitch height differences between the subparts are scaled differently depending on which subparts they separate cannot simply be the effect of a purely local downstep progression, and is much more plausibly explained as due to a hierarchical structure with different boundary strengths. Regarding the result of contrastiveness not being a significant predictor of increased pitch height, this supports the suggestion already made in the discussion of items 26 and 65, that the factor responsible for the solidly observed upstepped values on accents in some words is perhaps less that they are contrastive, and instead that they are final in a phrase.

5.2.3.7 Discussion

We have now seen fairly solid evidence that pitch accent scaling in these Spanish double topic-utterances from *Elqud* follows a hierarchical prosodic structure. In the following, I would like to explore some further issues mostly related to variability. From looking at individual examples as well as the length of the error bars on some of the barplots and the relatively low adjusted R^2 of 0.23 in the regression model, it is evident that there is a considerable amount of individual variation and that not all individual utterances produce the same scaling relations. In the next section, we will look at a subset of utterances by a group of speakers in detail that display an accentuation pattern similar to what was described as “phrase accentuation” in section 5.1.3.1. Here I would like to concentrate instead on two other sources of variability. The variable tendency to scale the initial noun in the first topic particularly high, as an initial “boost”, was already mentioned. This is visible not only in individual examples but also in the barplots for items 52, 59, and 65, suggesting it is not an isolated quirk. It seems to stand in opposition to the tendency to increase relative scaling on the final accent in a subpart. The other variable tendency is to produce the very final accent, in the second comment, as either upstepped or downstepped relative to preceding ones. The particular large standard deviations on the pooled values for this position in virtually all the experimental items indicate this variability, as do individual examples. For instance, XJ45 and LJ22 produce

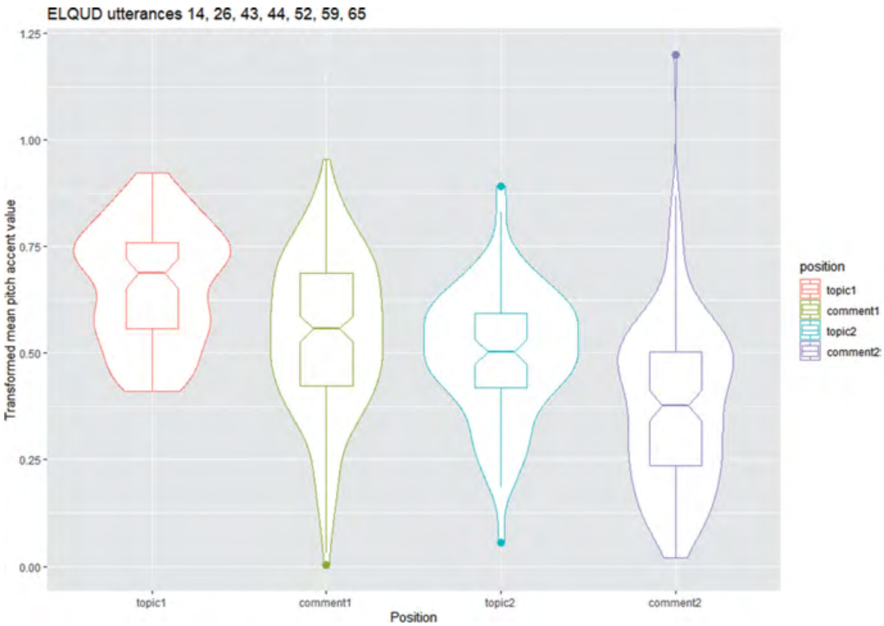


Figure 96: Combined violin- and boxplots for the transformed mean pitch value measured on accentable syllables in the double-topic utterances in Spanish *Elqud*, sorted according to utterance subpart. Violin shape indicates density. Horizontal bar in boxplot indicates median, notches indicate confidence intervals, box sixe interquartile range. Number of observations per utterance subpart: 84, 181, 84, 146.

the upstepped variant in their responses to items 43 and 65, respectively (Figures 87 and 95), while QP44 and QZ13 produce the downstepped one in their 52 and 59, respectively (Figures 91 and 94). I would like to suggest an explanation for both of these sources of scaling variation, the initial "boost" and the final upstep/downstep, in terms of conflicting functions. In both cases, upstepping the final element in the phrase corresponding to an information structurally relevant unit, especially when it is contrastive, arguably aids in cueing precisely this information structural status: not only that the item itself is contrastive, but also to delimit the particular subpart the accent is the final one of. In contrast, "boosting" the initial element in the utterance, and downstepping the final one, aids in cueing the delimitation of the utterance as a whole. This is perhaps particularly relevant in these long utterances with complex internal structure.[222] The cues for these two types of structural

222 Recall that we already saw two other utterances with a similar initial boost that seems to go against information structure, Figure 60 and Figure 61 in section 5.1.3.1.2. They are also comparatively long.

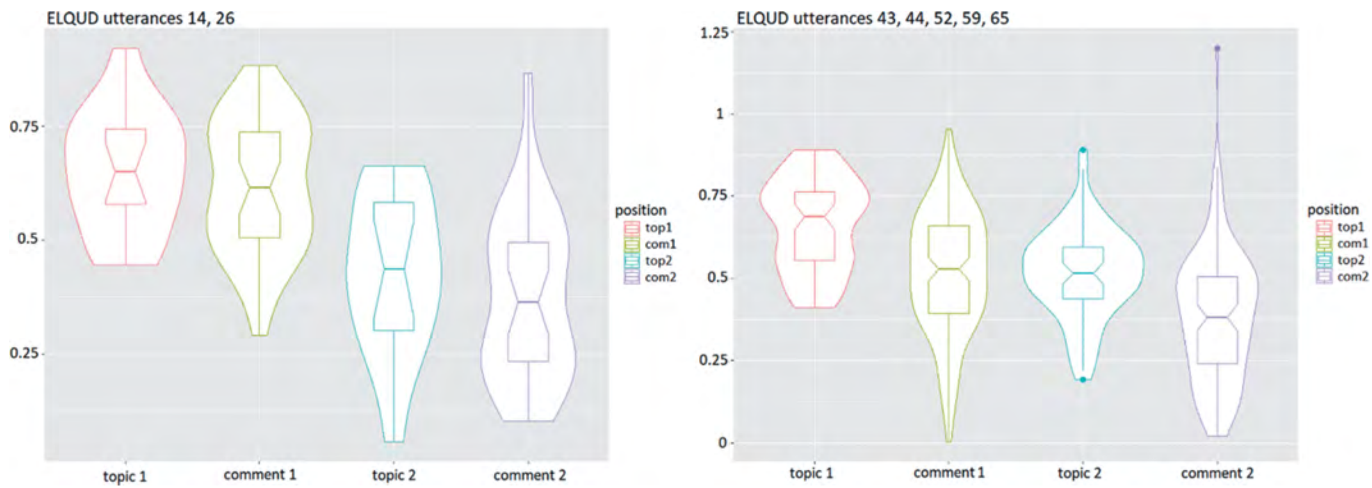


Figure 97: Same data as in Figure 96, but separated into the data from the shorter experimental items 14, 26 on the left, and those from the longer items 43, 44, 52, 59, 65 on the right. Number of observations per utterance subpart: 17, 41, 17, 35 (left); 67, 140, 67, 111 (right).

indication are naturally incompatible with each other, since scaling is always relative and they have opposite effects on the same position (down- or upstepping the final element in a subpart). It would be an interesting question for further research to explore whether this hypothesis of conflicting structural cues holds up in perception.

Another issue is the effect that item length seems to have. In the responses to items 14 and 26, the shortest and least complex utterances, we saw that the greatest difference in register height was observed between the two main parts, while that between the four subparts only really emerged in the longer responses, those to items 43, 44, 52, 59, and 65. This difference comes out very clearly in the comparison between the value distributions for the two item groups in Figure 97 (left vs. right). Furthermore, in 65 it was also the case that mean pitch height on the prepositional noun in the first comment seemed scaled up relative to the preceding auxiliary verb and the following noun, even though the prepositional noun here is neither contrastive (unlike in the responses to item 59) nor in final position in one of the four main subparts. That function words including the copulas are more likely not to be accented (cf. section 5.1.1.2) might also play a role here. But the upscaling of the noncontrastive prepositional noun nonetheless seems to be an indication that in a longer subpart, there is a tendency to form smaller units whose final elements are upstepped in order to create a somewhat alternating profile of prominence cued by pitch height, which would be rhythmically motivated. Both of these observations evidence an interplay of various factors influencing phrasing: there seems to be a conflict between a tendency not to create a large amount of embedded high-level phrases (embodied in NONRECURSIVITY, cf. section 3.6.1), and one to cue information structurally relevant units and relations via prosodic structure as expressed through hierarchical scaling. Their interaction seems mediated via rhythm: at short utterance lengths, the ban on recursivity mostly prevails, but at larger ones and with increased internal complexity, rhythmic considerations must also come into play that favour a greater number of prosodically distinguished units, because this leads to creating more eurhythmic prominence profiles (via regular contrasts in pitch height at both lower and higher levels of the prosodic structure). Together, this rhythmic demand and that for cueing information structural relations faithfully then seem to overcome the resistance against non-recursive prosodic phrasing. Recall that in section 3.6.4, results from earlier studies on Spanish phrasing also pointed to an interplay of mapping-like factors and more genuinely prosodic/rhythmic ones. The results here effectively indicate that this interaction also takes place at higher levels of the prosodic structure in addition to pointing to the relevance of scaling as a cue for phrasing.

The last issue I would like to raise here before moving on to the discussion of speakers responses that exhibit phrase accentuation concerns the nature of the categories that the prosodic units identifiable via register height scaling map onto. In sections 3.6.3 and 3.6.4, the discussion already showed that assuming a tight syntax-prosody mapping is not necessarily the way to go when looking for prosodic recursion. I would like to argue here that the prosodic structure indeed cues partitions relevant for information structural interpretation, and that it is difficult to claim that it maps onto syntax instead, using one of the speaker responses that are particularly elliptical. Figure 98 is LJ22’s response to item 43 which we already saw in (65), with reference lines to show that it portrays the hierarchical scaling structure even though overall pitch range is rather small.

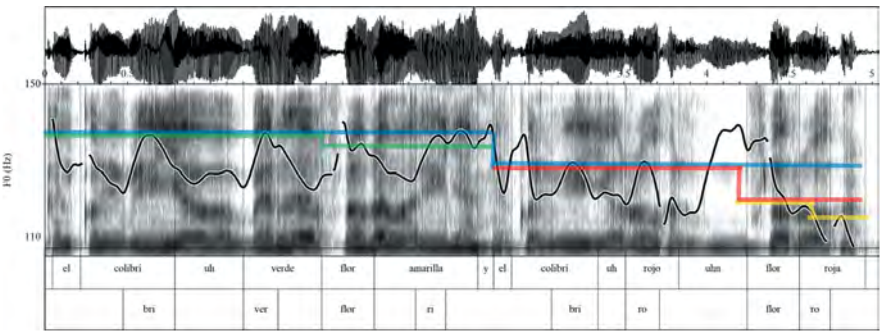


Figure 98: LJ22_ELQUD_ES_43[223] (*el colibrí uh verde flor amarilla y el colibrí uh rojo uhm flor roja* ‘the green hummingbird yellow flower and the red hummingbird red flower’). Cf. (65).

This is an utterance in which syntax above the level of the NP has essentially been done away with, and yet prosodic structure, as evidenced by scaling and partially also boundary tones, still cues the relevant information structural relations between the referents. In both parts of the utterance, there is no verb that could project any syntactic structure above the noun phrase. Neither could therefore qualify as a “standard clause” because they lack “a predicate, and a locus for Tense”, which are seen as necessary ingredients in one account (Selkirk 2011: 452). Hence, the Match constraints from Selkirk (2011) would predict phonological phrases to form on each of the noun phrases *colibrí verde, flor amarilla, colibrí rojo* and *flor roja*, but not ips or IPs (ι in Selkirk 2011) on either them or on the first and second part, respectively, since those are strictly matched only to clauses (Selkirk 2011: 455). However, there is a clear boundary tone (as evidenced by the presence of a sustained high pitch posttonically on *amarilla*) between the two parts, indicating that the units it separates must have ip or IP status. By the scaling evidence, the noun phrases consti-

223 https://osf.io/4seky/

tuting the topics and comments, respectively, are also phrased separately. It seems then that in the absence of higher-level syntax, the provision in the Match (clause, ι) constraint that it must be a clause that is matched to ι would itself have to be violable and can be outranked when there is a need to cue relevant information structural categories (such as the topics and comments here) that are necessary for the correct interpretation of the utterance. An utterance like this one is strictly speaking syntactically effectively defunct, but via the (scaling and other) cues provided by prosody, *el colibrí verde flor amarilla y el colibrí rojo flor rojo* becomes parseable as a sequence of two larger units that are again comprised of two units each. One likely interpretation for such a grouping is that of topic1-comment1, topic2-comment2, and as such, this grouping is distinct from other possible groupings (e.g. not likely to become interpreted as a list of items of equal status) even without a specific context. However, the meaning relation between the elements identifiable as topics and comments can only be fully filled by the context.[224] Even semantically, the utterance is only fully interpretable in a context like this one where the crucial variable specifying the relation in which the topical and comment elements stand to each other can actually be filled without a verb. That is to say, this utterance is only interpretable without ambiguity because the preceding turn in the discourse, the provocation (Farkas & Bruce 2010), can here be assumed to be known to the hearer by the speaker. Presumably, if one is to follow Rizzi (1997), which Selkirk (2011) does, then what the prosodic structures map onto in this utterance are not categories from information structure (conceived as generalizations about means by which we keep track of referents and their relations in conversations, not just isolated sentences), but syntactic categories in the left periphery (projected from functional heads in an expanded complementizer domain), which are responsible for building up information structure for each sentence (a view that, if taken literally, is not easily reconcilable with a view of information structure based purely on context as adopted here and detailed in section 3.7.2). In this view, these functional heads are there, whether overtly expressed or not, and their complements are clauses which satisfy the Match (clause, ι) constraint. However, this effectively comes down to saying that a severely elliptical sentence such as in the utterance

224 Utterances that should better be interpreted in terms of information structural relations instead of syntactic ones are actually not overly exceptional even in languages that have not usually been called "topic-prominent" in the somewhat dated typology by Li & Thompson (1976). To give one English example, a customer at a restaurant might tell the waiter *the Colcannon, that's me* in a situation where the waiter has arrived with several dishes at a table with several customers but does not remember who ordered what. In this situation everyone will understand the customer not to mean that they are a (sentient) stew, but merely that they ordered it.

under discussion is syntactically the same as one that has fully expressed verbal structure. Such a sentence-based view on information structure cannot explain that this utterance is perfectly understandable only in this discourse context (and not understandable in others). In such a view, the action that relates the topic referents to the comment referents would remain underspecified and the sentence would possibly have to be judged unintelligible because access to the discourse context were lacking. Thus claiming that the syntax here is the same as in a corresponding non-elliptical sentence would actually mean saying that syntax is largely irrelevant for interpretation. Instead, I prefer a view in which both prosody and syntax provide important cues for an information-structural interpretation of utterances in context, but where such an interpretation can also be shaped by only or mainly one of them in cases where the other is lacking.

5.2.4 Variant double-topic utterances with phrase accentuation

In this section we will specifically investigate a subset of the double-topic utterances that exhibit what was called “phrase accentuation” in section 5.1.3.1. In phrase accentuation, unlike in the main variant of accentuation for the Huari Spanish data considered here, only phrase-final stressed syllables are fully pitch accented while preceding stressed syllables are either completely deaccented or their pitch span is severely compressed. The double topic-utterances we looked at so far all exhibited the normal accentuation pattern in which nearly all accentable syllables are also indeed pitch accented. Phrase accentuation in the corpus of double topic-utterances is produced by three speakers, ZE55, NQ01, and XJ45. We will look at two of their longest utterances, the responses to items 59 and 65, separating the discussion by speaker. In particular, we will explore how their divergent accentuation pattern interacts with hierarchical pitch scaling as cue for a recursive prosodic structure.

5.2.4.1 ZE55’s 59 and 65

Speaker ZE55’s response to experimental item 59 is given in Figure 99. In the first topic, both the noncontrastive noun and the contrastive adjective seem to form phrases with a LH* pitch accent on the stressed syllable and a rising boundary tone, unambigously identifiable at least after the adjective. Between them, there is reset to a low level that stays the same throughout the utterance. In the first comment, only every second accentable word (*atrás* and *pequeña*) is actually pitch accented, while the preceding accentable words (*está* and *casa*) seem to only form part of the initial low stretch before the rise to the accented syllable, so that in both cases, something like LH* is formed on the two words together. Pitch scaling also

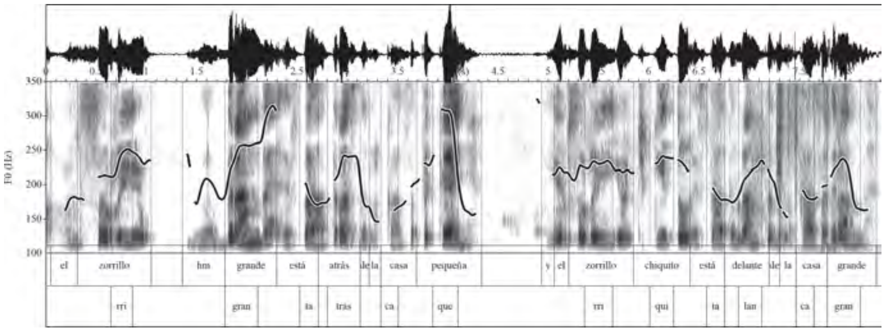


Figure 99: ZE55_ELQUD_ES_59[225] (*el zorrillo hm grande está atrás de la casa pequeña y el zorrillo chiquito está delante de la casa grande* 'the big skunk is behind the small house and the little skunk is in front of the big house').

follows an alternating pattern, but at a higher level: of the first two pitch accented elements (the two composing the topic), the boundary rise on the second is scaled much higher; similarly, of the two pitch accents composing the first comment, again the second (on *de la casa pequeña*) is scaled much higher, but here it seems to be additionally followed by a low boundary tone. This increased scaling corresponds to the division between the subparts, and the additional low boundary tone after the first comment seems to further demarcate the first main part as a whole. In contrast, pitch on the second topic (*y el zorrillo chiquito*) is basically level without distinguishing between stressed and unstressed syllables, but quite high overall, at the same level as the non-upstepped pitch accents in the utterance. This is comparable to the plateau realizations also already discussed in section 5.1.1.2. This kind of "suspended" realization clearly distinguishes the second from the first topic. The second comment is produced with effectively the same pattern as the first forming pitch accents on alternating accentable words, except that the final pitch accent (on *de la casa grande*) is not increased in scaling. Note that we cannot observe any global declination or downtrend: high targets stick to either of two reference lines, the higher or lower one, whose values remain more or less constant throughout the utterance, and similarly, low targets also seem to be relatively fixated on a horizontal reference line that remains constant throughout the utterance, without decline. Figure 100 illustrates the scaling relations in the utterance.

Figure 100 shows that register height here can only be said to distinguish between the two main parts: the highest reference line is reached only in the first main part. But apart from that, the subparts are additionally distinguished

225 https://osf.io/32k64/

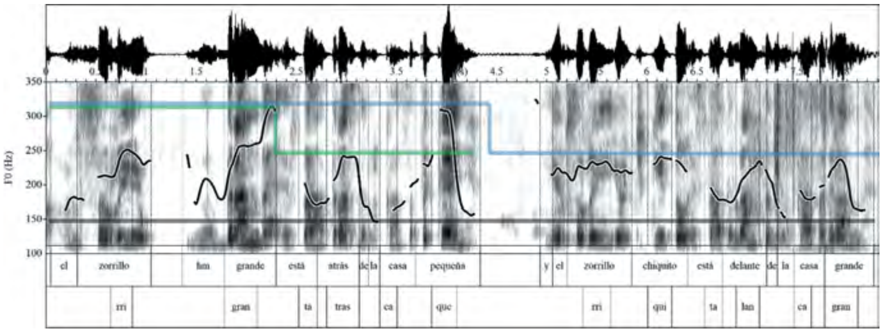


Figure 100: ZE55_ELQUD_ES_59 with added reference lines for levels of scaling.

by further cues: the first topic is realized with two pitch accents or possibly even two rising phrases at the end of which the boundary tone is scaled high; the first comment ends on an additional low boundary tone after reaching the high reference level; the second topic is realized as a plateau; and the second comment resumes the pattern of the first but without the upstep on the final accent. Thus a similarly complex prosodic structure is required for this utterance as for the double topic-utterances with “main” accentuation: minimally, each subpart must form an iP/IP-level phrase, the two main parts form one at a level above, and then the whole utterance forms another, even higher one. Additionally, a domain must be defined at which pitch accents are culminative. This is usually assumed to be the phonological phrase, to be distinguished from the prosodic word at which stress is assigned, thereby designating possible locations for pitch accentuation. Let us consider the distribution of boundary tones. A high boundary tone is unambiguously present after the first topic. Low boundary tones are present at the ends of the first and second part. It is not unlikely that there is also a high boundary tone at the end of the second topic, although this is not as unambiguous: the peak on *chiquito* here is reached already within the stressed syllable, but the high pitch extends into the posttonic, which could point to the existence of a high boundary tone, possibly secondarily associated with the metrically prominent syllable (cf. Grice et al. 2000; Gussenhoven 2000a, see also section 3.5.2). Assuming the presence of a boundary tone there has the advantage of providing us with a further argument for mapping each of the two topics and comments to prosodic units of the same level in a principled way, namely that of ip or IP, which then are the domains at whose edges boundary tones are assigned (cf. Gussenhoven 2004; Féry 2017). There are thus three levels of prosodic structure above the foot that are systematically distinguished in this utterance, the prosodic word (providing stress positions), the phonological phrase (at which pitch accents are assigned) and the ip/IP (at which boundary tones are assigned). However, as argued above, we then need more in order to represent all the contrasts that are present here. The utterance differentiates clearly between the first and the second part via scaling: the highest peaks in the second part are globally downstepped with respect to the first part. This necessitates a mapping of each part to a further level of prosodic structure, and of the whole utterance to another, because otherwise the downstep taking place would have no domain with respect to which it occurs. However, because these levels do not specify additional tones, they should be assumed to be recursive instantiations of ip/IP/ι instead of separate prosodic domains, following the same line of argumentation as employed in section 3.6 for other languages.

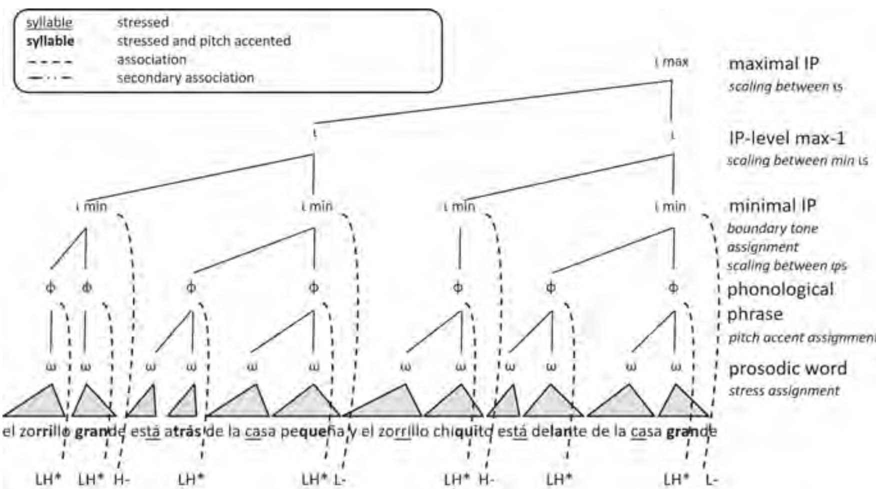


Figure 101: Proposed recursive prosodic structure for utterance ZE55_ELQUD_ES_59.

Figure 101 illustrates this proposed prosodic structure for the utterance ZE55_ ELQUD_ES_59. At the level of the prosodic word, stress is assigned, but this only surfaces via pitch movement when the prosodic word is the last one in its phonological phrase. At the next level, phonological phrases are joined together to form minimal IPs, whose edges are associated with boundary tones, and which are cueing information-structurally relevant partitions into topics and comments. A scaling relation of upstep obtains at this level between the subordinate phonological phrases. One level above, the minimal IPs are joined together to form intermediate IPs, between which a downstep relation obtains at the highest level, that of the maximal IP. Under this proposal, at least the same number of contrasts is thus produced prosodically as with the "main" variant: scaling in the form of downstep distinguishes between the two main parts, mapped to max IPs, and in the form of upstep it also distinguishes between the less and more prominent parts within the first topic and comment, those that create the contrast in reference between the first and second topic (*zorrillo* ***grande*** vs. ***chiquito***) and first and second comment (*casa* ***pequeña*** vs. ***grande***). Further contrasts are created by phrasing: the separation between topic and comment in both parts is achieved via the placement of boundary tones. And the phrasing of several prosodic words together into a single phonological phrase so that a pitch accent peak is only realized on the last one of them again creates a difference in prominence which cues information structure: the repeated parts of the verb and noun phrases in the comments (*está* and *casa*) are not separately accented and thus made less prominent than the ones which differ between the two parts (***atrás*** vs. ***delante*** and ***pequeña*** vs. ***grande***). Arguably, phrasing also sets the first topic apart, because it consists of two phonological phrases (one for each prosodic word), each with its own pitch accent, while everywhere else, also in the second topic, one phonological phrase encompasses two prosodic words. Contrast between the the first topic and the first comment, and also between the first and the second topic is thus signaled redundantly.

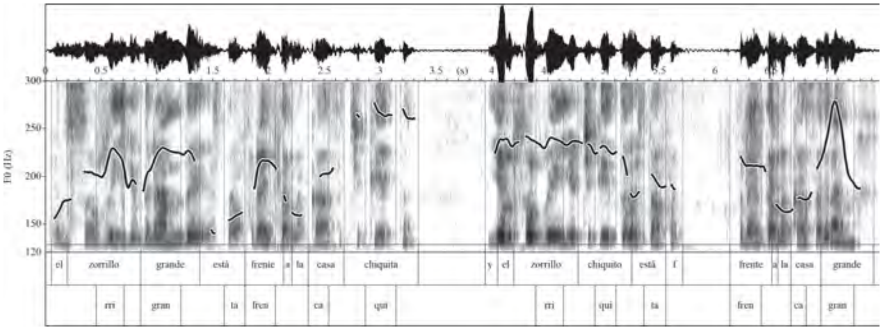


Figure 102: ZE55_ELQUD_ES_65[226] (*el zorrillo grande está frente a la casa chiquita y el zorrillo chiquito está f- frente a la casa grande* ‘the large skunk is in front of the small house and the small skunk is in front of the large house’).

ZE55’s utterance in response to item 65 (Figure 102) is similar in many respects to that to item 59: scaling distinguishes effectively between three reference levels, two for high tones, of which the higher one is upstepped, and a bottom level for low tones. Overall declination, with regards to either high or low targets, does not seem to take place at all. Phrase accentuation at the level of the phonological phrase

226 https://osf.io/xt2sb/

distinguishes between individually accented prosodic words and groups of them where only the final prosodic word has a pitch accent peak. The grouping is the same as in 59, even with respect to the difference between the first and second topic. While in the first, two phonological phrases are formed with distinct pitch accent peaks, the second consists of a single phrase, realized as a plateau with no pitch distinction between syllables. Phrasing at the level of the minimal IP is accomplished via boundary tones and separates the two topics and comments. Essentially the only difference to ZE55_ELQUD_59 lies in the scaling of the first topic and the noun phrase of the second comment: while the rise at the end of the first topic in 59 reached up to the highest reference line, here in 65 that line is marked only by the rises on the final noun phrases of the first, and also second, comment (which in 59 reached only to the downstepped reference line). Thus, here scaling does not distinguish between part one and part two. However, they are still contrasted overall with each other by the difference in their final boundary tones: part one here unambiguously ends in a high boundary tone, part two in a low one. This was not the case in ZE55_ELQUD_ES_59, where both parts ended in low boundary tones, but scaling clearly distinguished between them. In addition, the difference in phrasing between the first and the second topic, just as in 59, also highlights this contrast.[227] Without a representation of an overarching domain (the maximal IP) which can make reference to the units below it, such regular differences could not be created. If each part (one and two) were the maximal domain for the application of prosodic rules (or some kind of prosodic planning), the phonology would be blind to all of these phenomena. However, all three phenomena considered here (scaling difference, boundary tone difference, phrasing difference between the two topics) are not blind to that, but they do seem to be interchangeable or sharing the same functional load to a degree. The conclusion I would therefore like to draw is that regarding the levels of prosodic structure which they distinguish, the two utterances are essentially equivalent. But it seems that the means by which the prosodic groupings can be expressed are variable. ZE55's "phrase accentuation" variant is functionally equivalent to the "main" variant in that it encodes the same prosodic information, but by different and even variable means. The tonal makeup of a phonological phrase in both variants is the same, with a H tone associating with the pitch accented syllable and an L tone occurring before it.

227 The first and second topic in both ZE55_ELQUD_ES_59 and 65 are (for me) very distinct in the auditory impression that they produce, the presence (on the first) and absence (on the second) of pitch accents can be clearly heard in both versions, as well as some difference in rhythm.

5.2.4.2 NQ01’s 59 and 65

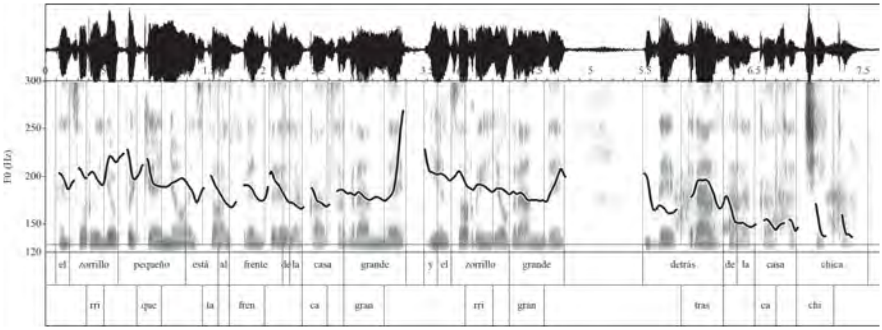


Figure 103: NQ01_ELQUD_ES_59[228] (*el zorrillo pequeño está al frente de la casa grande y el zorrillo grande detrás de la casa chica* ‘the small skunk is in front of the large house and the large skunk behind the small house’).

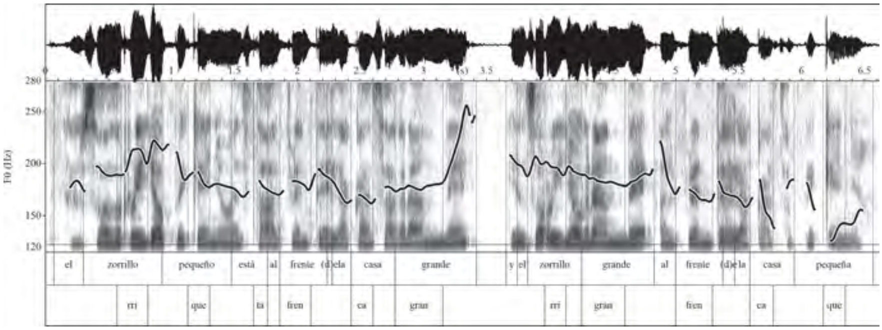


Figure 104: NQ01_ELQUD_ES_65[229] (*el zorrillo pequeño está al frente (d)e la casa grande y el zorrillo grande al frente (d)e la casa pequeña* ‘the small skunk is in from of the large house and the large skunk in front of the small house’).

Speaker NQ01’s utterances in response to experimental item 59 and 65 are given in Figures 103 and 104, respectively. These utterances (and equally 43 and 52 by the same speaker) share a number of similarities with those produced by ZE55, but they are also different in some respects. The central similarity is that they show pitch accenting behaviour where not every accentable word realizes a peak aligned with

228 https://osf.io/4n6ur/
229 https://osf.io/67ezm/

its stressed syllable, but where instead often several words are grouped together in a single pitch accenting event. Sometimes, single prosodic words are also phrased separately this way. NQ01, just like ZE55, produces the first topic as two phrases (a peak on each prosodic word) in both 59 and 65, while the second topic, in both utterances, is phrased into one. However, while ZE55 produces peaks aligned with the stressed syllable of the last prosodic word in those phonological phrases that contain several prosodic words, NQ01 here mostly produces such units without any pitch peaks aligned with the stressed syllables, and instead with only a final rise which often begins after the last stressed syllable (e.g. *casa grande* and *zorrillo grande* in both her 59 and 65). As with ZE55, the extent of the phrase that is realized in this way is variable, extending over several words or just one, as in the first topic in 59, where both *el zorrillo* and *pequeño* are each realized in a separate phrase. However, whenever this is the case, the highest pitch peak seems to be reached not on the stressed syllable, but on the last one. It seems most intuitive to analyze the phrase-final rises in these pitch accent-less phrase as evidence for the presence of boundary tones. This would seem to mean either assuming boundary tones at the edges of phonological phrases here, but not in ZE55, or taking each of these phrases to be immediately encompassed within a minimal IP, which then assigns boundary tones. The latter option, in addition to feeling somewhat unwieldy because it creates a lot of structure that does not do much, is also not desirable, because it still needs to explain why there are no pitch accents on the metrically strong syllables. I instead propose to adopt the former option and again employ the notion of secondary association. If we assume that the edges of phonological phrases are in principle available for association with high tones (i.e., also in ZE55), then the difference between ZE55's and NQ01's realization is reduced to a difference in whether it is this edge or the metrically strongest position at which an available tone is most likely to manifest. This competition can be modeled as two conflicting constraints, and their difference in ranking then decides which position is associated primarily, and which secondarily, with the available H tone (see section 5.3 for the OT analysis). In turn, this determines the pitch realization as mainly aligned within the metrically strong syllable or at the edge of the phrase.

Analysing the differences between the speakers in this way allows us to keep a lot of the prosodic structure the same: in particular, we can keep assuming that there *is* a metrical structure by which the stressed syllables in NQ01's utterances are assigned more prominence than their neighbours, even if it does not surface in the pitch realization; the difference between the first and second topic can also be modeled in the same way as for ZE55: as two separate versus one single phonological phrases, but in both cases equally within only one minimal IP.

Moreover, we can assume that the same tonal sequences exist on each phonological phrase here at a level of phonology at which they are still unassociated (cf. Torreira & Grice 2018), namely LH. For ZE55, these tones are associated as LH*, with the leading L sometimes aligning and possibly associating with the stressed syllable of the less prominent prosodic word.[230] For NQ01, they are associated as LH- or possibly L*H-. The existence of the L tone is here deducible from the presence of an elbow in the pitch contours, most clearly seen in the second topics in 59 and 65, where pitch falls steadily until the end of *y el zorrillo gran*, from where it then sharply rises on the last syllable. Here, it looks as if this elbow is aligned with the end of the stressed syllable of the second word (which is taken to be the metrically strongest in the phrase). If this alignment turns out to be consistent, it could be seen as evidence for a (secondary) association of the L tone with this position. The difference between the two patterns is exemplarily visualized in Figure 105, where (a) shows the proposed association pattern for ZE55, and (b) that for NQ01.

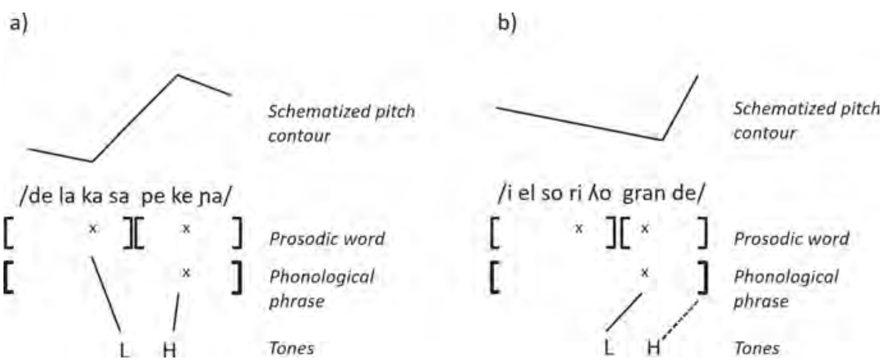


Figure 105: Proposed association model for „phrase accentuated“ realizations by ZE55 (***a***, e.g. Figure 99) and NQ01 (***b***, e.g. Figure 104), given exemplarily using the phrases *de la casa pequeña* for ZE55 and *y el zorrillo grande* for NQ01. Solid lines between tones and positions in the structure indicate association, dashed lines alignment.

A relevant question is whether the utterances by NQ01 can be argued to have the same complex prosodic structure as the “main variant” realizations and those by ZE55. As a matter of fact, just taking the mean values per accentable syllable in

230 This is motivated mostly theoretically because of the nature of the OT constraints involved. See section 5.3 for a short discussion. I’m claiming that associating the L is a possibility, but it is also possible that it just stays unassociated.

NQ01's utterances, unlike ZE55's, actually produces scaling-based differentiation between the subparts similar to that shown by the pooled values in section 5.2.3, albeit also with considerable drops in pitch within each subpart. This is potentially an indication that the differing levels of the L tones, if they are indeed associated with the end of the metrically strongest syllable, could cue the hierarchical scaling structure here. In fact, it is obvious that NQ01 treats the L tones differently from ZE55: the pitch values of the local minima clearly decline overall as the utterance progresses, with partial reset achieved after the break between the first and second part, whereas with ZE55, they are all very clearly at the same low reference line. As far as I am aware, scaling structures have only been proposed on the basis of peak, i.e. H tone, height (cf. Ladd 1988, 1990, 1993, 1994, 2008; van den Berg et al. 1992; Féry & Truckenbrodt 2005; Truckenbrodt & Féry 2015).

But even leaving this aside, there is other evidence for a prosodic structure here that divides the entire utterance in two main parts, and those again in subparts corresponding to the topic and comment, respectively. This evidence consists in the boundary tones and their relative height. In both 59 and 65, the rise accompanying the boundary between the two main parts produces the highest pitch in the utterance, and it is especially clearly at a higher level than that reached by the boundary tones separating the topic from the comment in both parts of both utterances. However, the division between part one and part two cannot be one between entirely independent domains, based on the argument already made that the topics in both parts are systematically phrased differently (and similarly for ZE55 and NQ01), and that their final boundary tones differ from each other, with the end of the first part being bounded by a final high boundary tone, while the second one ends on a low one. There is also clearly no full pitch reset between the two parts and pitch in the second part declines much more steeply overall than in the first one. Note that the strongest boundary within the first part in both utterances here seems to separate the noncontrastive noun from the following contrastive adjective in the topic, rather than separating the topic as a whole from the comment. This is possibly another instantiation of the "initial boost" discussed before.

There is a noteworthy difference between NQ01's 59 and 65, regarding phrasing and scaling within the second part. In 59, the second topic is followed by a high boundary tone and a short break, and then in the second comment, *detrás* is realized with what seems to be the finally rising movement produced on phonological phrases by this speaker. *Detrás* is thus realized in its own separate phonological phrase, standing out in the utterance, since otherwise, only in the first topic are individual prosodic words realized in separate phonological phrases. After *detrás*, the rest of the comment is realized as a single phrase, so that the phrasing here is (*detrás*)(*de la casa chica*). In 65, on the other hand, the corresponding stretch, *al frente de la casa pequeña*, which is also separated from the preceding topic by a high boundary tone but not a pause, seems to be phrased as (*al frente de la casa*)(*pequeña*), as evidenced by the final boundary rise at the end of *casa*, and the realization of the phonological phrase contour on *pequeña*. This difference in phrasing corresponds to the difference in information structure as evoked by the difference in stimulus mismatch between the two experimental items, with *detrás* contrastive in 59, but not *al frente* in 65 (cf. Table 11 in section 5.2.1). NQ01 here aligns the focused elements with right phrase edges, apparently determining that the best way to signal the correction against the audio stimulus (the presence of a [REVERSE] feature in the context, in Farkas & Bruce 2010; Roelofsen & Farkas 2015’s terms) is by cueing *detrás* to have highest prominence in the final comment in 59, even to the detriment of the following (also contrastive) *chica*, which is compressed, deaccented, or dephrased here. Aligning focus with phonological phrase edges is a strategy that is of course far more effective in the “phrase accentuation” variant than in the “main variant”, where the right edge of virtually every prosodic word is also aligned with a phonological phrase edge. Additionally, we saw that increased scaling on a prefinal contrastive word is perhaps dispreferred because this conflicts with upstepping the accent on the final word in a prefinal iP/IP. Thus, NQ01 here perhaps makes use of a possibility for cueing information structure that is less available in the main variant, indicating that the different variants are possibly not equally well equipped for prosodically expressing the same complex information structures. However, the difference is not always expressed in “phrase accentuation”, as seen from ZE55, who does not differentiate between 59 and 65 in this way.

5.2.4.3 XJ45’s 59 and 65

XJ45 is the third speaker who exhibits phrase accentuation. We already saw his utterance in response to experimental item 43 (Figure 87), which exhibits only some of the features of phrase accentuation, and in section 5.1.3.1.3 he was shown to use phrase accentuation independent of the context conditions under which it occurred with other speakers. Yet he varies the most between accentuation patterns in the double topic-utterances: while ZE55 and NQ01 produce all their examples with “phrase accentuation”, and all other speakers do not, XJ45 does both, some utterances exhibiting features of both patterns. Here, we will look at 59 (Figure 106) and 65 (Figure 107), and see what makes them phrase accentuated and yet different from the ones produced by ZE55 and NQ01.

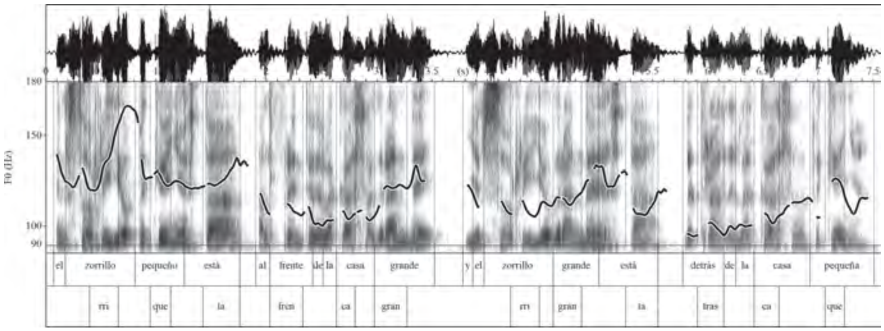


Figure 106: XJ45_ELQUD_ES_59[231] (*el zorrillo pequeño está al frente de la casa grande y el zorrillo grande está detrás de la casa pequeña* 'the small skunk is in front of the big house and the large skunk is behind the small house').

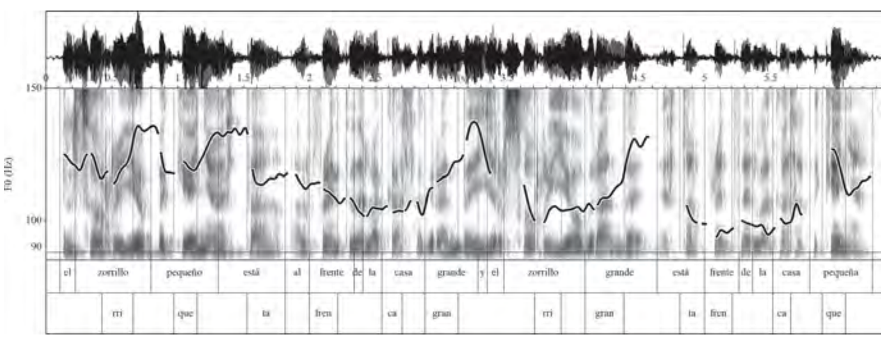


Figure 107: XJ45_ELQUD_ES_65[232] (*el zorrillo pequeño está al frente de la casa grande y el zorrillo grande está frente de la casa pequeña* 'the small skunk is in front of the big house and the large skunk is in front of the small house').

As with ZE55 and NQ01, in XJ45's utterances the two topics are realized differently, with the second one being clearly realized with a single rising pitch movement, while in the first one, there is a remarkable final rise on the noncontrastive *el zorrillo* in both utterances, indicating that the topic is here produced with a phrase boundary after the first prosodic word, apparently in two further instances of the "initial boost" phenomenon. In general, the pitch movements realized on the various parts seem similar to those observed for NQ01: pitch is low for most of the phrase and then rises to a peak at the end of the last syllable, with an elbow realized at or very close to the stressed syllable of the last word.

231 https://osf.io/h4wgn/
232 https://osf.io/sb3v9/

This seems to be the case regardless of whether the last syllable is stressed or not. I therefore take these movements to reflect the same association pattern as that analysed for NQ01 and exemplified in Figure 105 b) above, with an H tone associating with the right edge of the phonological phrase and an L tone preceding it. The extent of these phonological phrases is also similar: they cover at least one prosodic word, but also more than that: arguably, in 59, *al frente de la casa grande* in the first comment and *detrás de la casa pequeña* in the second are each realized within a single phonological phrase which would then extend over three prosodic words each. Although XJ45 seems to phrase somewhat larger chunks together than NQ01, he also has in common with her how he separates topic from comment and part one from part two by phrasing, with high boundary tones at the end of each of these units. However, not everything is as with NQ01: while she consistently produces final peaks whose rises do not even begin in the penultimate syllable even if that syllable is stressed, with XJ45, the stressed penult sometimes clearly takes part in the rise or even forms its own peak. Notably in both utterances there is a shoulder on the stressed syllable of *grande* at the end of the first comment and a peak on the stressed syllable of *pequeño* in the second, where it is then followed by a fall to a final low target, indicating a final low boundary tone. In this respect then, so to speak, XJ45 seems somewhat in between NQ01 and ZE55. Based on these intermediate cases, a more flexible association model seems to be at work here, one that allows the H tone assigned by the phonological phrase to associate both with the edge and the most prominent syllable. Where a separate peak is reached in the stressed syllable (on *grande* at the end of the first comment and on *pequeña* at the end of the second) there is also very likely a boundary tone following belonging to a higher level (at least a minimal IP; H at the end of the first comment and L at the end of the second). That suggests that the constraint aligning the higher-level tone with the phrase edge outranks that aligning the lower-level tone with it, so that the latter tone has to move to the next closest available TBU, the final stressed syllable. This situation is schematically illustrated in Figure 108. Here the unassociated tonal sequence is LHL or LHH, with LH being the tones assigned by the phonological phrase, and the final L/H being an addition, assigned by the minimal IP.

In this proposal, the presence of a final L_ι or H_ι has the effect on the preceding H_ϕ that it now aligns and associates with the strongest syllable. This means that the tones in the sequence associate and align in a ranked hierarchy, with the higher-level tone having precedence. In a non-IP-final phonological phrase, for XJ45 this leads to the H_ϕ associating with the boundary and the L_ϕ with the preceding prominent syllable, but in this final context, because there is an L_ι or H_ι to its right that belongs to a higher level and occupies the position at the phrase edge, the H_ϕ has to stand down, so to speak, and now associates with the strongest syllable. For ZE55, the issue mostly does not arise, because here, the association with the strongest

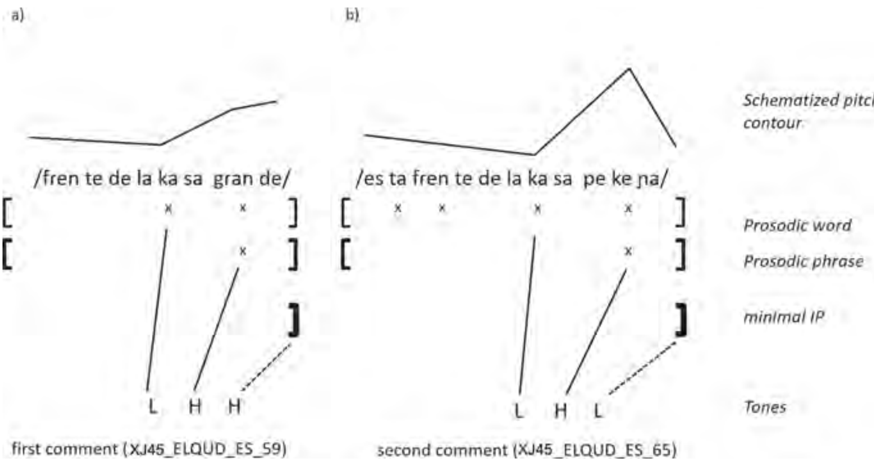


Figure 108: Proposed alignment and association model for the last phrase in the first comment of 59 (a) and the second of 65 (b) by XJ45. Solid lines between tones and positions in the structure indicate association, dashed lines alignment.

metrical position outranks association with the edge of the phonological phrase for the H_ϕ tone. This leads to the observed association pattern in non-final contexts, and in final contexts at least on (pro-)paroxytones, no conflict arises, because the additional L_ι or H_ι to the right of the H_ϕ aligns with the TBU at the right edge of the minimal IP which is free. The alignment of the T_ι is still predicted to outrank the association of T_ϕ with the strongest position but this would only have observable effects on final oxytones, for which see section 5.3.3. For NQ01, on the other hand, the situation is a little different again: in her 59, the final phrase of the utterance, *de la casa chica*, seems to realize just a final fall, so that a final L_ι tone, aligned with the right edge of the IP, seems likely. In her 65, however, the final phrase is produced like any non-final one, with a final rise on the last syllable. Here, there does not seem to be a final L_ι tone (instead, the phrase is produced in the lowest pitch register) aligned with the edge of the IP. I propose that tones assigned by the phonological phrase have the same alignment and association ranking for NQ01 and XJ45: the H_ϕ tone seeks to be rightmost. The difference lies in what happens when additional boundary tones turn up. For XJ45, their constraints outrank those of the phonological phrase tones. For NQ01 in 59, the faithful realization of the IP-final tone also seems to win out, with the phonological phrase tones not surfacing, while in 65, they instead seem to prevail. In both cases, the H_ϕ tone is not realized further to the left, associated with the strongest metrical position, as in XJ45' case. Why this difference occurs between NQ01's 59 and 65 could perhaps be answered by referring to the different information structures the two utterances have. The elements in 59 after the separately phrased and contrastive *detrás* might be dephrased, suppressing any following phonological phrase tones. This would allow the L_ι to surface. In this way, no constraint rerankings are needed between the utterances: in both 59 and 65, the association and alignment of the H_ϕ outrank that of iP/IP-level tones, resulting in the final rise in 65, but in 59, the PhP-tones are simply not present. All of this will be analyzed in OT in section 5.3.

Another point of divergence between XJ45 and NQ01 concerns the scaling of boundary tones. As seen before, NQ01 scales the boundary tones separating the two main parts highest, and those separating between topics and comments lower. XJ45, on the other hand, does not seem to do this. In 59, the boundary separating the first topic noun from its adjective is highest, and the following boundaries, except for the final one, seem more or less scaled at the same level. In 65, all boundaries except the final one are at roughly the same level. Some kind of downstep-like scaling does seem to be taking place, but it occurs on the stretches between the boundary tones: in 65, this results in a downstep progression of roughly equal height between the four subparts. In 59, it looks as if this downward progression reflects the embedded structure in a similar way to the one seen in the pooled analysis as well: the first comment is lower than the first topic, but the second topic is somewhat higher again but not as high as the first, and the second comment is then lowered again relatively to that.

5.2.5 Comparison between the variants

This section compares the variation between the variants discussed individually. The variation assumed in the prosodic structures is given in Figure 109. It provides the prosodic structure for three utterances, 65 by QZ13 (A), 59 by ZE55 (B) and 65 by XJ45 (C), which are supposed to exemplify the range of variation in the most relevant aspect, namely the grouping of prosodic words to phonological phrases. The structures A-C demonstrate that at the level of prosodic structure, the "main" variant is not very different from the "phrase accentuation" variants. The only real difference lies in the grouping within the comments, and here QZ13's and ZE55's versions stand together against that by XJ45, cutting across the divide between "main" and "phrase accentuation" variants. For QZ13's 65 and ZE55's 59, I have analysed the verb together with the preposition (*está al frente/atrás/delante*) as phrased into one phonological phrase, and the prepositional complement (*de una/la casa pequeña/grande*) into another. This is because for the main variant it is suggested by the scaling in QZ13's 65 and 52 (cf. Figures 95 and 78, respectively): the peak on *está* is slightly lower than that on *al frente* (or *chupando* in 52), while the first peak of the prepositional complement is lowered with respect to both of them, hinting at a further level of embedded scaling. For ZE55, it is suggested straightforwardly by the presence of a pitch accent on the preposition (*atrás/delante*) but not on the verb (just as on the following adjective but not its preceding noun), suggesting their being grouped together (cf. Figure 99). No evidence for any such differentiation is found in XJ45's 65, where both comments are entirely phrased into one phonological phrase (and then directly into a minimal IP), as evidenced by there being only one prominent pitch movement over that whole stretch, the final rise (cf. Figure 107). I do not necessarily want to claim that the more complex phrasing as analysed for QZ13's utterances here is valid for all utterances in the "main" variant or for all except the ones by XJ45. I think it is quite likely that some of the other speakers also produce the more simple grouping. In this sense, the structures A-C just serve to exemplify these two options.

The treatment of the relationship between prosodic word and phonological phrase deserves some justification. Another analytical option would have been to take the prosodic word as the domain at which pitch accents are assigned, as is proposed by Hellmuth (2007) for Egyptian Arabic, which realizes pitch accents on nearly every content word (similar to the "main" variant here). However, for the "phrase accentuation" variant, the consequence would be either to treat the much larger units in which only a single pitch event occurs (the entire comments in the case of XJ45, for example) as a single prosodic word, lumping together several content and function words in any order, or to say that pitch assignment happens at the prosodic word level only in the "main" variant, but at the phonological phrase level in the "phrase accentuation" variant. Both options do not seem appealing, the first because we could not then associate the L tone in the "phrase accentuation" variant with the stressed position of a non-final prosodic word without giving up on culminativity or making the prosodic word again recursive, and because a prosodic word would then consist of an accentable content word plus clitics in the "main" variant, but of several accentable content words plus clitics in the "phrase accentuation" variant, without any good reason. The latter option, on the other hand, seems (at least to me) inelegant and somewhat ad-hoc. What is more, for other varieties of Spanish that normally also pitch accentuate nearly every word (i.e., like the "main" variant here), it has been shown that deaccentuation, in the sense of an accentable content word not being realized with its own pitch accent, does happen in certain contexts, but it is strongly dispreferred when that would mean deaccenting an entire phonological phrase (Rao 2009: 18), and that deaccented words in the majority still maintain longer duration and/or higher intensity on the stressed syllable (Ortega-Llebaria & Prieto 2011; Torreira et al. 2014). This supports the established view that stress is a property at the level of the prosodic word, distinct from pitch accent as a property of the phonological phrase, and that this difference particularly holds out in such contexts where indi-

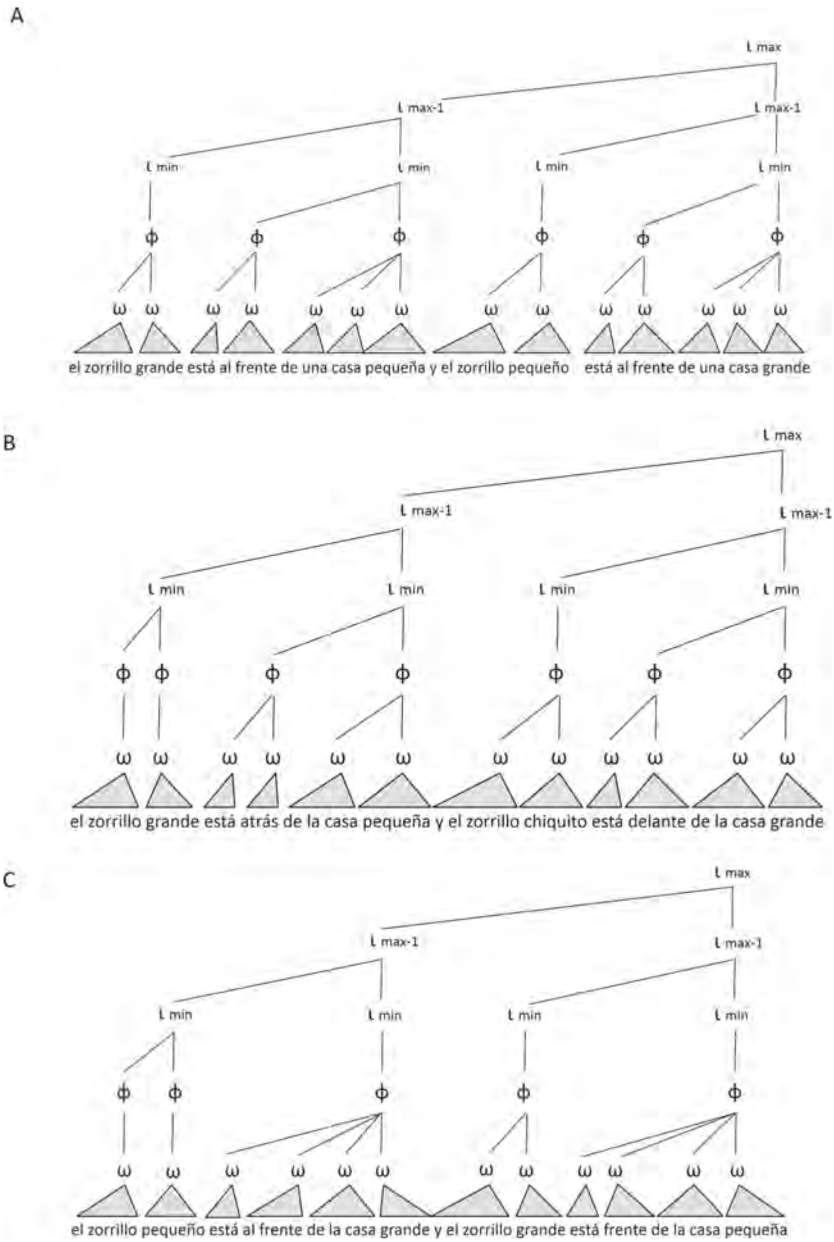


Figure 109: Prosodic structures for QZ13_ELQUD_ES_65 (A), ZE55_ELQUD_59 (B), and XJ45_ELQUD_65 (C), intended to exemplify the range of variation in these structures for the Spanish "double topic" constructions, including the "main" variant (A) and the different variants within the "phrase accentuation" variant (B, C).

vidual words are not accented. Stress is culminative at the level of the prosodic word as defined in section 3.3.2: at least one and only one syllable must be stressed in a prosodic word. For accent we can at least assume a minimality condition that there must be at least one pitch accent in a phonological phrase, but possibly more than one. I would like to follow this view here and therefore assume that also in the "main" variant, pitch accent assignment happens at the phonological phrase level.

This leads to the analysis of what is the main point of difference between the "main" and "phrase accentuation" variants. At the level of the phonological phrase, tones are assigned. Their identity is determined by the utterance type, so that based on the evidence gathered in this work, I propose that for declarative IPs, the basic tonal sequence is LH. How many LH sequences are assigned per phonological phrase is determined by reference to the metrical structure, and I propose that this is the main difference between the "main" and "phrase accentuation" variants. In the "main" variant, the relevant grid level is that of the prosodic word: the number of grid marks at that level determine how many LH sequences are provided for each phonological phrase. In the "phrase accentuation" variant, the level selected by tonal assignment is the next higher, that of the phonological phrase, making pitch accent culminative at that level in this variant in addition to being minimal. Hence, only one LH sequence is assigned for each phonological phrase in the "phrase accentuation" variant, while potentially several are assigned in the "main" variant. How these sequences are then aligned and associated is a difference again cutting across the divide between the variants: for the "main" variant as well as for ZE55, the most highly ranked association position for the H is the strongest metrical position (and then the subsequent next-strongest metrical positions, i.e. all the stressed syllables, in the "main" variant), while for NQ01 and XJ45, alignment with the edge of the phonological phrase itself is ranked higher (see the discussion in section 5.2.4). Figure 110 gives a schematic overview over this difference in tonal assignment.

The first topics in the "phrase accentuation" variant take up a somewhat special role here. As already discussed, in the "phrase accentuation" utterances, they are produced with prominent pitch movements on both prosodic words (the noun and the adjective), unlike the second topics, that are produced within a single pitch movement. As can be seen from Figure 109 (B & C), this has been analysed as the first topic consisting of two phonological phrases, one for each prosodic word, while in the second topic, the two prosodic words are contained in one phonological phrase. This is similar to the "initial boost" on the noncontrastive noun of the first topic that we saw also in the pooled data. This would further support the suggestion that there is a functional equivalence between scaling-based cues and alignment-based cues to prosodic phrasing.

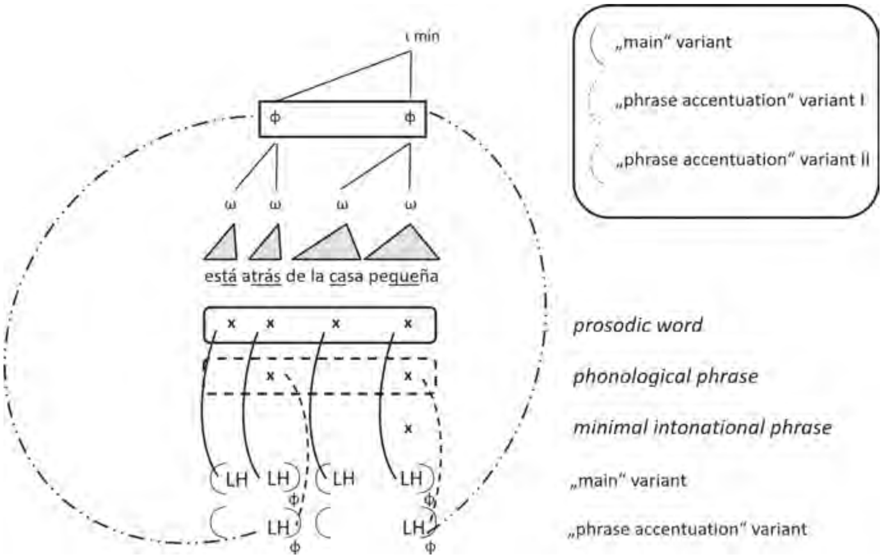


Figure 110: Difference in tone assignment between the „main" and the „phrase accentuation" variants. Phrase accentuation variant I is the one exemplified by ZE55, variant II that by NQ01 & XJ45.

Table 14 summarizes the different cues for the prosodic structure across the variants discussed. The cues by which intonation separates the double-topic utterances into chunks corresponding to the information structure are given, ordered after variants and speakers. Additionally, the table lists the tone assignments and associations at different levels of the prosodic structure. It is claimed that effectively very similar prosodic and metrical structures underlie all the variants here, and that what the table summarizes are just the different cues by which these structures are expressed in the signal. The structures are so similar because the prosodic phrasing in all cases serves the function of cueing information structural partitions: in the responses to the more complex items (43, 44, 52, 59, 65), material corresponding to the two topic domains, and each of the two comment domains, is phrased as a minimal IP. Each of the two main parts comprising a topic-comment sequence corresponding to an individual assertion answering sub-QUD is mapped to an IP one level below maximal (max-1), and the entire complex assertion answering the superordinate QUD is mapped to a maximal IP domain. Table 14 summarizes the evidence that all of these levels are signaled prosodically in all variants observed, although the means by which they are signaled vary. The responses to the less complex items 14 and 26 are disregarded here. As discussed in section 5.2.3.7, it seems likely that constraints against too complex prosodic structures win out in them because of their relative shortness.

In the upper part of Table 14 (white background), the phonetic cues that are used in the different variants to signal the prosodic structure(s) given in Figure 109 and discussed just before, are given and compared. Since the table attempts to summarize a sizeable chunk of the results of this chapter, it necessarily uses somewhat compressed language. It should therefore be read as follows: **phrasing** refers to prosodic grouping of segmental material, in the case of prosodic constituents at or above the minimal IP this might be made observable via **boundary tones** (in the case of NQ01 and XJ45 also at the level of the phonological phrase); in the case of phonological phrases, it mostly occurs via an **alignment** of the locally most prominent constituent (the most prominent prosodic word in the phrase) with the right edge of the phonological phrase, which in turn might be visible either via upstep on the peak of that prosodic word (sometimes in the "main" variant) or by having a pitch peak realized only on the stressed syllable of that prosodic word, but not others, in the phrase (the "phrase accentuation" variant). In a particular case, phrasing also refers to the systematically differing realizations (with two vs with one pitch peak) of the two topics in the "phrase accentuation" variant. **Scaling** is of course also seen as cueing phrasing in the sense of prosodic grouping and separation, but since it is of particular interest here, it is treated separately in the table. Scaling as systematically employed differences in relative pitch height is seen as reflecting pitch height orientation of a prosodic constituent along an abstract reference line and is expressed either via the relative height of pitch accent peaks (in the "main" variant) or of boundary tones (occasionally in the "phrase accentuation" variant).

Table 14: Prosodic means employed for the separation of the utterances ELQUD_ES_43, 44, 52, 59 and 65 into parts corresponding to their discourse and information structure (upper part, white background), and tone assignment, alignment and association behaviour (lower part, grey background), separated according to variants.

	"main" variant	**"phrase accentuation"**		
		ZE55	**NQ01**	**XJ45**
Separation into part I and II	**Scaling; phrasing**; **boundary tones**	**phrasing** via **L boundary tones**; **Split topic 1 vs. unsplit topic 2**	**Phrasing** via highest **scaled boundary tones**; **Split topic 1 vs. unsplit topic 2**	**Phrasing**; **Split topic 1 vs. unsplit topic 2**
Separation between topic and comment	**Scaling** (also **boundary tones** for some speakers)	**Phrasing** via H **boundary tones**	**Phrasing** via next-highest **scaled boundary tones**	**Phrasing** via **boundary tones** (also **scaling**)

Table 14 (continued)

	"main" variant	**"phrase accentuation"**		
		ZE55	**NQ01**	**XJ45**
Separation within topic and comment	**Scaling** (downstep within minimal IP, upstep on last ω); **Alignment** of strongest ω with right edge of ϕ	**Phrasing** into separate ϕ; **Alignment** of strongest ω with right edge of ϕ; **Scaling** (upstep on last ϕ in topic 1 and comment 1)	**Boundary tones** of ϕ; **Alignment** of strongest ω with right edge of ϕ	
ϕ-level tones	**LH** for **each ω in ϕ**	**LH once** per **ϕ**		
	H associated as H* with σ' in each ω; L as leading tone (unassociated)	H associated as H* with σ' in strongest ω in ϕ; L as leading tone or associated to an available prefinal σ'	H aligned with right edge of ϕ; L as leading tone or associated with σ' in strongest ω in ϕ	H aligned with right edge of ϕ or associated as H* with σ' in strongest ω in ϕ; L as leading tone or associated to an available prefinal σ'
IP-level tones	**H** or **L**			
	aligned as L% or H% with right edge of IP	aligned as L% or H% with right edge of IP	L aligned as L% with right edge IP if it wins out; H aligned as H% with right edge of IP	L aligned as L% with right edge IP if it wins out; H aligned as H% with right edge of IP

We can see from the table that the use of scaling decreases when moving from left to right: in the "main" variant, represented in the first column, scaling is present as a cue for the prosodic separation and grouping of information structurally relevant units in all three rows, as evidenced by the analysis of the pooled data in section 5.2.3. It separates successively smaller levels moving downwards in the table: the separation between the two main parts, corresponding to the two sub-QUDs; between the topic and comment in each part, and between the different constituent parts within the comment. For the speakers of the "phrase accentuation" variant, scaling is used for fewer of these separations: ZE55 uses it only within the comment; NQ01 only for the separation between the main parts and the topic and comment, but applied to the boundary tones that are also present at the edges of these parts; XJ45 in his "phrase accentuation" utterances, at the very right, does not scale his boundary tones differentially (but perhaps the register level of the L tones in each subpart, which is not covered here). However, the table also shows that these speakers use other means to make up for this lack of scaling: phrasing in the form of different boundary tones associated with the right edges of phonological phrases as well as the prosodic levels above it, together with alignment and the phrasing difference between the first and second topic, does all the work for XJ45's "phrase accentuation" mode, it does part of the work for the other two "phrase accentuation" speakers, but it is much less used in the "main" variant. In the middle columns (ZE55 and NQ01), both phrasing by boundary tones and scaling is used to a certain degree redundantly, cueing the same boundaries, but overall, it seems warranted to say that some kind of trade-off between these cues takes place. In addition, a "vertical" variation (i.e., moving up or down along the axis of larger or smaller information structural constituents) can clearly be observed that interacts with the horizontal one: in the "main" variant, boundary tones mostly only separate the highest levels between the two main parts (except for speakers OZ14 and maybe ZZ24); for the "phrase accentuation" variant, also the level below, between topics and comments; for the two speakers that produce the phonological phrase H tone as an edge tone, NQ01 and XJ45, even at the lowest level. However, even NQ01, who uses scaling for the boundary tones of the levels of minimal IP and upwards, does not use scaling with regards to the boundary tones at the level of the phonological phrase. This variation in cues employed for prosodic phrasing is somewhat reminiscent of that observed in Gabriel et al. (2011) for Argentinian Spanish. They also observe that boundary tones do not always get realized as posttonic pitch rises and peaks (continuation rises), but instead also sometimes as plateaux, pitch reset, or preboundary upstep. However, there are also some differences to their study. For the cases they discuss, they "assume an underlying intermediate phrasal boundary tone H-, which can be phonetically realized in different ways" (Gabriel et al. 2011: 162). It is this underlying H- tone which they explicitly make out to be responsible for "modifications of the scaling of the pitch accents located in the close surroundings of the boundary" (ibid.). However, in the data analyzed here, there are also sometimes clear cases in which this explanation does not suffice. As already discussed in section 5.2.2, the binary difference between presence or absence of a boundary tone at a boundary is not enough to explain the systematically observable scaling differences, which do not follow a simple separation between "high" and "low", but instead reflect a more complex structure of embedded reference lines, as we have seen throughout. Additionally, scaling differences can be observed whether a clear boundary tone (in the sense of a posttonic pitch peak at a break) is *also* present or not, and for NQ01, the height of the boundary tone peaks themselves differs with regard to whether they occur at the break between phonological phrases or between higher-level prosodic constituents. I therefore argue that here, scaling is not merely a variant realization of a boundary tone, but via its less categorical nature can convey additional and independent cues to prosodic structure. Scaling and boundary tones can occur complementarily or together, but it is not the boundary tone that is underlying, but the prosodic structure itself. The other difference to the findings in Gabriel et al. (2011) is that I make this variation in cues out to be largely systematic between speakers. Although the amount of data does not really suffice to say this with more certainty, in principle it might point to an inter-speaker variation in the use of these intonational cues within the speaker community that could be similar to that found by Niebuhr et al. (2011) between pitch accent “shapers” and “aligners” for German and Italian.

In the lower part of Table 14 (grey background), the tone sequences provided at each level of prosodic structure and their association properties are given. Here, the most essential *shared* property emerges: each tonal sequence assigned at the level of the phonological phrase, whether realized as pitch accents on each prosodic word, only once per phonological phrase, or partially as boundary tones, can be analyzed as LH, which was also by far the most prevalent pitch accent identified for less complex declarative utterances in section 5.1.1. I propose that this LH sequence is an essential part of what makes up a “declarative” in the Spanish of these speakers. I follow Beckman & Pierrehumbert (1986); Hayes & Lahiri (1991); Heusinger (2007) and others in assuming that the intonational phrase is the level at which a somehow ‘full’ tune is associated to a text; like Hayes & Lahiri (1991) for Bengali as well as in the works of the Grup d’Estudis de Prosòdia and others on Spanish, collected in Prieto & Roseano (2010, 2009–2013); Hualde & Prieto (2015), I assume that a part of this tune is determined by the utterance type of the IP. LH is also the tone sequence taken to be associated with declaratives in most works on Spanish. The analysis is therefore that the information that the sentence type is “declarative” is encoded at the level of each IP by making this the tone sequence that is minimally produced in each phonological phrase whenever a tone is required. Note that the tone sequence associated as a pitch accent in the variants where there are pitch accents in this analysis does not change between those constituents that form part of the topic and those that form part of the comment, as is proposed by Steedman (1991, 2000) for English. The difference between topics and comments is here assumed to be conveyed only via a probabilistic relation between prosodic structure and information structurally relevant partitions, as detailed in section 3.7.3. The fact that the prevailing tone sequence can successfully be identified to be LH, while its association and alignment properties (as pitch accents or boundary tones) vary systematically between the “main” and “phrase accentuation” variants, can be seen as evidence for assuming an autosegmental tonal tier at which tones are *not* associated (in line with such work as Grice et al. 2000; Gussenhoven 2000a, 2004; Torreira & Grice 2018, see also the discussion in Ladd 2008: 285–288), that association comes later or separately,[233] and that even though different varieties and contexts have various effects on their association and alignment, there is some essential unity to the pitch accents nearly always analyzed as forming part of Spanish declaratives, namely that they consist of precisely this tone sequence, LH, whether it is associated as LH*, L*H, or L+>H* (cf. Hualde 2002; Gabriel 2007 for similar proposals for unity).

I assume a recursive iP/IP domain here for these Spanish utterances for the same reasons that have been brought forth for them for other languages in section 3.6. When there is evidence, as there is here, for a number of prosodic groupings that are only differentiated via the varying strength of a continuous cue, i.e. scaling here, but not via a different tonal make-up or other discrete properties, then it is most parsimonious to take these groupings as a recursive instantiation of a prosodic domain. Incidentally, it is interesting that it is the iP/IP domain which shows evidence of being recursive, since arguments for or against the existence of intermediate phrases not only in Spanish often hinge on whether "lesser" instantiations of IP-level boundaries should be taken as evidence of a separate category or not. This problem would not arise anymore with the assumption of a recursive IP category. Assuming a recursive IP domain does not mean that a recursive IP structure is always present in all utterances. As Féry (2017: 78) also argues, it is only utterances exceeding a certain length and complexity that show evidence of recursive phrasing, as we also saw in the analysis of the responses to items 14 and 26 in comparison to the longer ones.

The proposed recursive structure is of a kind that has been called *compound phrasing* in Frota (2000, 2012, 2014) and *compound prosodic domains* in Ladd (2008). These are "balanced" recursive prosodic structures in the sense of van der Hulst (2010: 319–320), where a recursive prosodic constituent may dominate two constituents of the same category (one level below, (69)a), but not one of the same category and one of the one below ((69)b).

233 I do not want to imply any actual temporal sequence in a processing model here. The point (also in the works cited) is just that it makes sense to conceive of tone sequences for a certain prosodic domain as separate from their concrete associations, because this allows for a variety of generalizations on empirical observations otherwise not possible. This is not very far from the original concept of tones as autosegments (Goldsmith 1976), enriched by an explicit formulation about at which level in the prosodic hierarchy tone sequences are provided. Gussenhoven (2004: 146–147), working in an OT framework, also reveals a conception of tones having an existence independent of their associations, when he argues that faithfulness constraints preserve "phonological substance", amongst which he counts tones, but not "a relation, like an association or an alignment".

(69) a. balanced / compound recursive structure

[$]_{IP\_n}$

[$]_{IP\_n\text{-}1}$ [$]_{IP\_n\text{-}1}$

b. unbalanced recursive structure

*[$]_{IP\_n}$

[$]_{\phi}$ [$]_{IP\_n\text{-}1}$

As discussed in section 3.6.2.4, Frota proposes a recursive IP for European Portuguese on the basis of sandhi and intonational phenomena, because the boundary cues are of the same type at each edge (fricative voicing, preboundary lengthening, boundary tones), but those that are gradient are notably stronger at those boundaries with higher-level edges. Because the boundaries only differ by phonetic strength, but not by type, Frota argues that the recursive analysis is superior to one involving two different prosodic units, e.g. the IP and the ip (Frota (2014: 12–14)). This is the argument also used here. The structures analysed here as recursive are also of the type of ((69)a), where no IP (or any other prosodic unit of any category) is ever analysed as dominating units of different categories, but always ones of equal category (either of the same category, but a level below, or of the category immediately below). Only NONRECURSIVITY is violated in such structures, while the other components of the SLH are all kept intact (cf. section 3.6.1). With such structures, the categories of the prosodic hierarchy stay strictly ranked, and prosody remains essentially flatter than syntax, but proliferation of prosodic categories is somewhat hemmed in and phenomena such as boundaries that are of different degree but the same type can be elegantly accommodated (Ladd 2008: 298–299). In other words, exactly those empirical observations that we have made throughout this chapter can be theoretically accounted for without all hell breaking loose.

5.3 OT-Analysis of the “main” and “phrase” accentuations and their relation to Quechua

This section develops an OT-analysis for the main and phrase accentuation variants of Spanish declarative utterances as described in the previous sections. In section 5.1, we saw that the phrase accentuation is likely influenced by information structure, whereas in section 5.2 it was also shown to be a variant characteristic of individual speakers. Here, only the phonological aspects of the difference between the variants will be treated. Although the difference between the two is the main strand of variation observed in the data here, concentrating on it necessarily leaves out other interesting aspects. However, as will become clear from the analysis, similar prosodic properties are at play that determine the difference between these

Spanish variants here as the ones that are also relevant for the differentiation between the Quechua intonational variants described and analysed in chapter 6. In this sense, the OT-analysis in both sections constitutes a proposal for how the prosodic space of possibility available to the bilingual speakers of the Huari speech community comprising both Spanish and Quechua can be delineated via what I suggest are some of its most important dimensions. Such a notion has to be considered carefully and formulated with a high degree of differentiation. Note that the analysis for Quechua proposes that word stress plays only a marginal role; mostly what is responsible for the characteristic pitch movement in Quechua phrases is therefore analysed as phrasal tones aligned with the edges of phonological phrases (see sections 6.1, 6.3). In contrast, for the Spanish phrase accentuation examples we have seen that at least in some cases, pitch peaks are realized on the stressed syllable of the last word in a phrase, meaning that the Spanish word stress is in fact preserved. In addition, the tonal sequence assigned at the level of the phonological phrase for Quechua declaratives is LH, HL, or LHL, depending on the type of phrase, whereas the analysis for Spanish has shown that it is LH, also in the "phrase accentuation" variant. Any theory proposing some kind of intrusive implementation of "Quechua" grammar wholesale into what is otherwise "Spanish" grammar would therefore fall short of the facts here. The "phrase accentuation" variant is neither just "Quechua in Spanish" nor "Spanish in Quechua". The analysis will instead demonstrate that it is possible to account for the observed similarities and differences in a granular fashion, allowing for intermediate differentiation.[234] As stated in section 4, the OT-analyses in both main chapters is intended to tackle the third group of research questions (36) that ask about which prosodic properties of Huari Spanish and Quechua are specific to each language, and which are shared. I use OT for this, in particular in answer to subquestion (36)b, because an OT-analysis describes the observed prosodic behaviour as a ranked set of separately identifiable constraints. It thus allows for a fine-grained analysis of what the minimal structural differences between individual observed variants are, and can thus lay bare how far apart or close together the different variants are along several dimensions in a prosodic possibility space independent of which language they belong to.

234 Note that the version of OT used here is not really fine-grained enough for example to capture the continuous scaling differences between a pitch accent that is "fully accented", "slightly compressed", "very compressed", and "deaccented". I take it for granted that a fully realistic model of both the phonetics and phonology involved here would require recourse to articulatory, perceptual, and neuronal explanations and specifications, e.g. as in Katsika (2012); Katsika et al. (2014); Tilsen (2013, 2019); Boersma et al. (2020). However, I think that the OT analysis here can still demonstrate those relations between the variants that are most relevant, in part precisely because it abstracts away from many of the more continuous aspects.

5.3.1 Variation in tone distribution for the phonological phrase and the intonational phrase

Some assumptions to start with: in non-tonal languages such as Spanish, the tones that are associated with stressed syllables change their identity, if at all, only for the expression of postlexical meaning and not according to lexical specification. So, no tonal specification takes place at the level of the prosodic word, and instead only at the level of the phonological phrase and the intonational phrase. I take the tone sequence LH for Spanish declaratives to “belong” to the phonological phrase level (with the possible number of tone sequences per PhP differing between “main” and “phrase accentuation”), and further boundary tones to belong to the intonational phrase level. By saying that tones “belong” to the PhP I mean that they can be taken to be indexed, T_{ϕ}, and thus that one set of association and alignment constraints refers to them, while another could refer e.g. for tones belonging to the IP, differently indexed (T_{ι}). This is the practice also followed in Gussenhoven (2004). At the same time I assume with some of the literature that tones must be unassociated at some level of the phonological representation, and that alignment and association are processes separate from tonal identity. Thus at the phonological phrase level, tones are provided whose identity is determined by such meaning-related factors as utterance type (or e.g. modality). These tones are not directly associated to a specific position and their number is determined by the number of relevant prosodic units. As argued above, I assume that the relevant unit in the case of the “main” variant in Spanish is that of the prosodic word: for each prosodic word, an LH tone sequence is assigned at the level of the phonological phrase. For the “phrase accentuation” variant of Spanish (and for Quechua), the relevant unit is the phonological phrase itself: for each phonological phrase, an LH tone sequence is assigned for Spanish, and an LH/HL/LHL sequence for Quechua. For easy reference, I will call those tones that are assigned once per prosodic word in the “main” variant and once per phonological phrase the “minimal tone sequence”.[235] In the “main” variant, several such minimal tone sequences can then make up the tone sequence

235 Cf. the observations by Ladd (2008: 285–287) on the obligatoriness of at least one pitch accent in a phonological phrase (intermediate phrase in his terms), associated with either the strongest metrical position in that domain or an edge, and on his concurring assumption that the other, pre-“nuclear” pitch accents, if they occur, are somehow subordinate to the main pitch accent, and that their tonal make-up is “a single linguistic choice”, while their number depends on the metrical structure and how tones are assigned to positions in it. The claim is that whether there are one or four prenuclear pitch accents, they will all consist of the same tone sequence. Based on the analysis in the previous sections, I take the tone sequence LH in Huari Spanish to also be the same for the final and obligatory nuclear accent, so that it is fully general.

for a phonological phrase, whereas in the “phrase accentuation” variants only one occurs per phonological phrase. In all variants, at the level of the IP, additional boundary tones might then be added to make up the IP tone sequence. Thus, the first main point of variation to be modeled is concerned with how many minimal tone sequences are allowed per phonological phrase. This is not totally straightforward. I base many mechanisms of the analysis on the intonational OT-model by Gussenhoven (2004). Although Gussenhoven (2004: 145) asserts that “[b]roadly, alignment constraints are responsible for creating the underlying tone string”, his subsequent discussion of tonal alignment constraints is not concerned with the problem at hand here, namely the amount of tones that are available for association and alignment within a given prosodic domain. In his discussion of English compounds, he does cover the distribution of pitch accents, but it is a somewhat different problem: there, the problem is determining where in units of similar size the main stress falls. Consequently it is a question of determining where the *one* pitch accent in such a unit will occur, not how many pitch accents can occur in a unit. That is to say, the English compound stress problem is concerned with how information from morphology can work on metrical structure so that the main stresses, at which the pitch accents occur, shift (cf. (70)), whereas in our case, the metrical structure between the main and phrase accentuation variants is the same in the relevant points, but the distribution of pitch accents per domain changes (cf. (71)).

(70) (from Gussenhoven 2004: 276)
  a. TOM Paine Big BAND (i.e. a Big Band led by Tom Paine)
  b. Tom PAINE Street BLUES (i.e. blues induced by Tom Paine Street)
  c. TOMcat-free ROOF

(71) difference in pitch accent distribution between main and phrase accentuation
  a. de una casa pequeña (QZ13_Elqud_ES_65)
```
x  xx  x  x  x  x  x
   x    x        x
                 x
 LH*  LH*      LH*
```
  b. de la casa pequeña (ZE55_Elqud_ES_59)
```
x  x  x  x  xx  x
      x       x
              x
            LH*
```

Equally, he uses the example of the different respective orders of the intonational phrase boundary tone and the lexical tone in Venlo Dutch vs. Roermond Dutch (cf.

Gussenhoven & van der Vliet 1999 and Gussenhoven 2000a, respectively) to argue that phonological alignment[236] and association are two independently needed processes to describe tonal behaviour, and to claim that “since alignment constraints are generally used as the mechanism to determine the order of morphemes, it would be undesirable to devise a different one just for tones” (Gussenhoven 2004: 150). In (72)a, the right-alignment of the lexical H tone (given as H_{Lex}), associating with the final mora of the intonational phrase, takes precedence over that of the L boundary tone (given as L_{bound}), while in (72)b, a different ranking results in the reversed tone order. However, this case cannot simply be transferred to our situation: the H and L tones in the Dutch example come from two different sources: one is lexical, the other is an intonation phrase-level boundary tone, while in our case, the order LH for our tones is again a “single linguistic choice”, that of choosing one (intonational) morpheme over another. To come back to the segmental comparison, using alignment for the determination of the LH sequence would be like saying that not only the order of suffixes is determined by alignment constraints, but also their internal make-up. In fact, Gussenhoven (2004: 151, footnote 7) points out that the different order of the tones in the two dialects of Dutch is actually coming from a “different sequencing of the tones”, confirming that tone sequence is not purely determined by alignment, but still leaving the question unanswered that is pertinent here.

(72) Adapted from Gussenhoven (2004: 150)

a. Representing Roermond Dutch (schematized)

b. Representing Venlo Dutch (schematized)

The fact that Gussenhoven (2004) does not seem to be concerned with the number and identity of the minimal tone sequences per a given prosodic domain as an issue for the constraint-based analysis suggests a seemingly easy solution for the problem here: these can simply be taken to be part of the input and are therefore not the result of constraint interaction, and we would then only need faithfulness

236 Not meant here is phonetic alignment, cf. the introduction to the terms in section 3.5.

constraints to keep them the way they are. In a way, this does make sense: the tonal processes that Gussenhoven addresses and proposes to treat with constraints for alignment and association are "phonological adjustments" (2004: 145) that take the properties we are here concerned with as previously established. However, this is problematic: the input is usually taken to be an underlying lexical representation (Kager 1999, who is only concerned with segmental phonology, takes the main purpose of faithfulness constraints to be the preservation of lexical contrasts), but the tones in an intonation language are postlexical. Alternatively, they could be part of a postlexical constraint-based process that takes place before alignment and association (intuitively, this seems very plausible). But then, and not just for our purposes here, it would still be important to understand the constraint interaction in this part of the process, since the number of tones per prosodic domain is clearly subject to systematic variation, not just in the Huari data, but in intonational typology in general.[237] It should therefore be undertaken to be analysed via constraint interaction, and not just be seen as something given. However, even though such an ordering of processes seems quite likely and appealing, it goes against the general assumption[238] of parallelism in OT, meaning that all structure really is evaluated (and built up) in parallel. It seems, however, as if the strictness of this assumption has been somewhat relaxed in practice, also in Gussenhoven (2004). A further

237 As „frequency or domain of pitch accents/AP/word tones", it forms criterion iii for determining the *macro-rhythm* of a language in Jun (2014b: 526) prosodic typology: "[l]anguages where every word receives a pitch accent or AP/word boundary tone are more macro-rhythmic than those with less or more frequent pitch accents or AP/word boundaries per word". This is used to group Spanish, with a pitch accent on nearly every content word, as a language with strong macro-rhythm apart from English, where pitch accents occur less frequently per phonological phrase, and which has medium macro-rhythm. Cf. Cole et al. (2019: 115) who repeat this grouping.

238 Kiparsky (2015: 3) calls it „the central principle of OT", but argues (11–12) that any kind of modularity assumption, also one separating between syntax and phonology in grammar, is strictly speaking a violation of it, and that a less radical alternative, modularization or stratification, has in fact been explicitly practiced in a number of works, including such early ones as McCarthy & Prince (1995). Prince & Smolensky (2004 [1993]: 7, 25) assert that the majority of their analyses is based on the assumption of parallelism, but also admit that deciding whether parallel or serial approaches should be favored is "a challenge of considerable subtlety". Gussenhoven (2004: 276–278) treats the mechanisms responsible for the different stress patterns of (70) as the result of a constraint-based version of Lexical Phonology (Kiparsky 1982), which then "present themselves to the postlexical grammar of English". That is to say, he tacitly also assumes some kind of modularity or serialism within morphophonology, with each module being governed by OT-like constraints, but obviously with different constraint rankings at each of them. This is the essence of what Kiparsky (2015) more explicitly proposes as "Stratal OT". A different modification of OT that is also intended to overcome the problems that arise from a strict application of parallelism is Harmonic Serialism (McCarthy 2016), where the results of a one-step-only modification by Eval are fed back as input to Gen in a loop, until no further changes are effected by the constraints (*convergence* is reached).

option, in any case, is just as problematic: if the tones are actually assigned by some constraint on the input, then it will not be possible to have any faithfulness constraints act upon them, since faithfulness constraints relate what is in the output with what is in the input. In fact, Gussenhoven (2004: 146–147) explicitly assumes that tones are part of the input.[239] In the following, I adopt an approach similar to the one Kiparsky (2015) proposes. I assume some kind of serialisation or modularization whereby the output of one process (tone sequence distribution) becomes the input of another (alignment and association) – because this is crucial for the difference between main and phrase accentuation. The tones that are aligned and associated have to be the output of a previous process of constraint evaluation, because they are clearly postlexical, and not assigned by markedness constraints. One way to arrive at different amounts of minimal tone sequences in the different variants would be to assume the maximal number of them in the input and then have a standard faithfulness constraints like MAXIO(T) (73) preserve them for the “main” variant, while for the “phrase accentuation” variant, a constraint that deletes all minimal tone sequences that do not belong to prosodic words that are heads at the level of the phonological phrase would have to be ranked higher, resulting in leaving only one minimal tone sequence per phonological phrase under pitch accent culminativity. However, here the question arises how the input “knows” how many minimal tone sequences (equaling the number of prosodic words) it should provide, if all structure really is evaluated (and built up) in parallel (i.e., this would have to run parallel to processes e.g. of cliticization that might result in altering the number of prosodic words). Making the basic assumption even more general opens up a path to pursue: we can assume that the input actually invariably provides a minimal tone sequence for each TBU contained in the domain at which tones are assigned (i.e. one tone sequence per mora or syllable in the PhP or IP), and that a family of markedness constraints that delete minimal tone sequences exist, whose ranking relative to MAXIO(T) would then determine how many minimal tone sequences there actually are in a larger domain such as the phonological phrase.

239 “Faithfulness is expressed in terms of correspondences between the elements in the input and elements in the output (McCarthy and Prince 1995). By ‘element’ I mean any phonological substance, like a feature, a segment, a tone, an accent, a constituent like ϕ,but not a relation, like an association or an alignment.” At a later point, he contradicts this definition by proposing a faithfulness constraint FAITH(ASSOC) that is intended to preserve the difference between Accent 1 and 2 in Swedish (Gussenhoven 2004: 216). It is quite telling that this slip occurs in relation with lexical instead of intonational tones, because it indicates an overall grammar model that is modular, in which certain (lexical) relations are established before others (postlexical ones) can follow.

(73) MaxIO (T): every tone in the input has a correspondent in the output (adapted from Gussenhoven 2004: 147)

The desired output results can be produced using several constraints employed in Gussenhoven (2004). He has NoContour[240] and OCP (obligatory contour principle) as a family of constraints ordered with respect to the level of structure at which each member applies. We also need another faithfulness constraint, Ident, which preserves the identity of elements from input to output (cf. Gussenhoven 2004: 147), but adapted specifically to preserve the minimal tone sequence (TS) for the phonological phrase.

(74) OCP (T, α): no adjacent tones may be the same if they are within the same domain α.

(75) NoContour (T, α): no adjacent tones may be different if they are within the same domain α.
   a. NoRise (T, α): no adjacent tones may be LH if they are within the same domain α.
   b. NoFall (T, α): no adjacent tones may be HL if they are within the same domain α.
   c. NoCrowd (T, TBU): no two tones may be associated to the same TBU.
   (all adapted from Gussenhoven 2004: 146)

(76) Ident (TS, ϕ): Disallows tone sequences in the phonological phrase that are not whole multiples of the minimal tone sequence.

Appropriately different relative ordering of these constraints could then result in allowing only tone sequences to survive once per syllable (as in Mandarin-like tone languages), once per prosodic word (as in the "main" variant here and other varieties of Spanish), or once per phonological phrase (as in the "phrase accentuation" variant here as well as in Quechua and other languages, e.g. Bengali (Hayes &

240 The NoContour group of constraints are motivated articulatorily in Gussenhoven (2004: 146): overly complex (tonal) configurations are demanding articulatorily and therefore avoided. Although this looks at first sight like a clear-cut conflict between a perceptual, hearer-oriented tendency for dissimilation (OCP) and an articulatory, speaker-oriented tendency for assimilation, Boersma (1998: 416) also describes a tendency against the repetition of similar articulatory gestures, which he takes to be part of the articulatory functional motivation for the OCP.

Lahiri 1991)). The approach therefore has the merit of providing some typological comparability while not postulating any language-specific constraints.[241]

Table 15: OT-tableau with the constraint rankings to arrive at the correct tone distribution for the “main” variant of Spanish. NoContour abbreviated to NoCo.

$(((LH)_\sigma(LH)_\sigma)_\omega((LH)_\sigma(LH)_\sigma)_\omega)_\phi$	OCP (T, σ)	NoCo (T, σ)	Ident (TS, ϕ)	MaxIO (T)	OCP (T, ω)	NoCo (T, ω)	OCP (T, ϕ)	NoCo (T, ϕ)
a. $(((LH)_\sigma(LH)_\sigma)_\omega((LH)_\sigma(LH)_\sigma)_\omega)_\phi$		****!				6*		7*
b. $(((LL)_\sigma(LL)_\sigma)_\omega((LL)_\sigma(LL)_\sigma)_\omega)_\phi$	****!		7*		6*		7*	
☞ c. $(((L)_\sigma(H)_\sigma)_\omega((L)_\sigma(H)_\sigma)_\omega)_\phi$				****		**		***
d. $((()_\sigma(H)_\sigma)_\omega((L)_\sigma()_\sigma)_\omega)_\phi$			*!	6*				*
e. $((()_\sigma(L)_\sigma)_\omega(()_\sigma(H)_\sigma)_\omega)_\phi$				6*!				*
f. $((()_\sigma(L)_\sigma)_\omega(()_\sigma(L)_\sigma)_\omega)_\phi$			*!	6*			*	
g. $(((H)_\sigma(L)_\sigma)_\omega((H)_\sigma(L)_\sigma)_\omega)_\phi$			**!	****		**		**

Table 16: OT-tableau with the constraint rankings to arrive at the correct tone distribution for the “phrase accentuation” variant of Spanish. NoContour abbreviated to NoCo.

$(((LH)_\sigma(LH)_\sigma)_\omega((LH)_\sigma(LH)_\sigma)_\omega)_\phi$	OCP (T, σ)	NoCo (T, σ)	OCP (T, ω)	NoCo (T, ω)	Ident (TS, ϕ)	MaxIO (T)	OCP (T, ϕ)	NoCo (T, ϕ)
a. $(((LH)_\sigma(LH)_\sigma)_\omega((LH)_\sigma(LH)_\sigma)_\omega)_\phi$		****!		6*				7*
b. $(((LL)_\sigma(LL)_\sigma)_\omega((LL)_\sigma(LL)_\sigma)_\omega)_\phi$	****!		6*		7*		7*	
c. $(((L)_\sigma(H)_\sigma)_\omega((L)_\sigma(H)_\sigma)_\omega)_\phi$				**!		****		***
d. $((()_\sigma(H)_\sigma)_\omega((L)_\sigma()_\sigma)_\omega)_\phi$					*!	6*		*
☞ e. $((()_\sigma(L)_\sigma)_\omega(()_\sigma(H)_\sigma)_\omega)_\phi$						6*		*
f. $((()_\sigma(L)_\sigma)_\omega(()_\sigma(L)_\sigma)_\omega)_\phi$					*!	6*	*	
g. $(((H)_\sigma(L)_\sigma)_\omega((H)_\sigma(L)_\sigma)_\omega)_\phi$				**!	**	****		***

In Tables 15 and 16, the tableaux showing the correct rankings for producing the “main” and “phrase accentuation” variant tone distribution, respectively, are given:[242] the separate OCP and NoContour constraints are “inherently” ordered

241 For many language varieties, it would of course also need to be enriched by some mechanism determining the tonal identity for each minimal sequence where it is not the same across the board (e.g. in many tone languages or when a nuclear tone sequence is different from a prenuclear one).

242 Note that all tones resulting from this process are indexed as belonging to the PhP (Tϕ). In Table 15 and Table 16, the indexation on the brackets only refers to the boundaries of the corresponding prosodic units.

(cf. Gussenhoven (2004: 146)): their least demanding versions (disallowing like or unlike adjacent tones within a mora or syllable) are ranked higher than their more demanding versions, since if a language does not allow like adjacent tones within a word, it will automatically also ban them within a syllable. They are, however, ranked so that they are interwoven with each other: the NoContour constraint for the syllable level is ranked below the OCP constraint for the syllable, but higher than the OCP constraint for the prosodic word, and so on. This will have the result that as long as nothing else intervenes, an output with no tones at all will be the optimal candidate, since in that way, both OCP and NoContour as defined in (74) and (75) are maximally satisfied (resulting in the elimination of candidates a and b in both tableaux). The crucial difference between Tables 15 and 16, and therefore between the tones output for the "main" and the "phrase accentuation" variant, respectively, is the ranking of the faithfulness constraints MaxIO(T) and Ident(TS, ϕ) with respect to these two markedness constraints: if they are ranked between the syllable- and the prosodic word-level versions of OCP and NoContour, as in Table 15, then the candidate with exactly one LH tone sequence per prosodic word (c) will incur the smallest number of high-ranking violations, which is the correct result for the "main" variant of Spanish. If, on the other hand, they are ranked between the prosodic word- and phonological phrase-level versions of the two markedness constraints, then the candidate with one LH sequence per phonological phrase (e) will be optimal, just as in the "phrase accentuation" variant. The relative ranking of Ident(TS, ϕ) over MaxIO(T) ensures that candidates d and f are not selected. At this point it has to be noted that the constraint model here creates a somewhat more categorical picture than what is found in reality. As section 5.1.3.1 showed, utterances with "phrase accentuation" often do not fully eliminate all traces of prefinal pitch accents but merely drastically reduce their scaling relative to that of the final one. In reality therefore, the OCP- and NoContour constraints here should probably thought of as having an effect that is gradual in this way. In the following I will leave this issue aside, but consider it a very worthwhile topic for future research.

Having acknowledged these restrictions, the approach seems viable overall. It has a consequence that at first sight might seem somewhat strange for a language like Spanish: the most "faithful" output would be one in which there are LH-rises on each syllable or even mora, a situation that looks quite like an (admittedly strange) tone language. However, this is what should be the case if we take the idea seriously that it is mainly the different constraints rankings that create the typological differences between languages. And since the markedness constraints of OCP and NoContour will eradicate all of these tones up to the point where the two faithfulness constraints intervene, this will result in a tonal optimization of either the word ("main" variant) or the phrase ("phrase accentuation" variant), which is the desired result. For the second leg of the variation journey we need to complete in order to move from “main” variant Spanish to the phrase accentuation, and thence to Quechua, I now make the assumption of modularization (cf. Kiparsky 2015). Because the association constraints in Gussenhoven (2004) all take the identity and number of underlying tones to be a settled issue, they would run into problems when acting at the same time as the ones we have just discussed. I assume that part of the difference between the “main” and the “phrase accentuation” variant, that which concerns the number of minimal tone sequences provided, can be modeled using the approach detailed above and given in Tables 15 and 16. The tones provided by them are then available to be aligned and associated according to constraints treated in what follows. This will reveal further variation, adding to the dimensions along which we can describe the variants to differ.

5.3.2 Variation in association and alignment

The second dimension of variation that this OT-analysis intends to cover has to do with the different association and alignment behaviours the different variants have. In the “main” variant and in ZE55’s “phrase accentuation” variant of Spanish, the H tone of the LH minimal tone sequence is associated with the metrically strongest position in the prosodic word or the phonological phrase, respectively. Even though in ZE55’s case it is only the metrically strongest position in the phonological phrase and not the stressed syllables of non-final prosodic words (like in the “main” variant) that are associated with the H tone, there is no need for a difference in constraint ranking between the “main” variant and ZE55’s one, as long as different numbers of minimal tone sequences are provided. The constraints we use here are all taken or adapted from Gussenhoven (2000a, 2004). Gussenhoven (2004: 149) has two groups of association constraints, one mandating that a certain TBU be associated with a tone, and another mandating that a certain tone be associated with a TBU. I take TBUs to be syllables here. The constraints associating TBUs with tones are inherently ordered, according to Gussenhoven (2004: 149), in that a language that associates tones to all syllables will also associate tones to stressed syllables, but not the other way around. Thus the constraints (77)-(79) decrease in the specific demand they make on the input. NoAssoc is a constraint ranked somewhere in between these inherently ordered constraints, making sure that not all languages need to associate tones to all possible TBUs. NoCrowd makes sure that only one tone can associate with a given TBU. Mostly this is sufficient, but in section 5.3.3, cases will be discussed showing that NoCrowd does not in fact capture the phenomena related to tonal crowding realistically.

(77) $(\sigma')_{\omega'} \leftarrow$ T: associate a tone with the stressed syllable of the metrically strong prosodic word in a phonological phrase.

(78) $\sigma' \leftarrow$ T: associate a tone with a stressed syllable.

(79) $\sigma \leftarrow$ T: associate a tone with a syllable.

(80) NoAssoc: TBUs are not associated with tones.

(81) NoCrowd: a TBU has only one tone.

The constraints associating tones to TBUs are specified according to which tone they refer to. Here, non-subscripted tones refer to the tones assigned by the phonological phrase, and tones subscripted with a small ι refer to tones assigned by the intonational phrase.

(82) H $\rightarrow$ TBU: the H tone is associated with a TBU.

(83) L $\rightarrow$ TBU: the L tone is associated with a TBU.

In our case, a relative ranking of $(\sigma')_{\omega'} \leftarrow$ T >> $\sigma' \leftarrow$ T >> H $\rightarrow$ TBU >> NoAssoc >> $\sigma \leftarrow$ T >> L $\rightarrow$ TBU will ensure that stressed syllables in strong prosodic words (and preferably also all stressed syllables) are associated with a tone (either H or L), but not all syllables. Additionally, while if a H tone should end up in a position where it can only associate with an unstressed syllable, it will still do that, an L tone in the same position will remain unassociated.

The alignment constraints (84)-(90) make reference to the position of a tone within a constituent or relative to another tone. Like the tonal association constraints, they make reference to a specific tone, L or H, indexed as belonging to the level of the phonological or intonational phrase (in this section, all constraints refer to PhP tones; explicit indexation will be used in section 5.3.3 when IP-level tones are added). They are violated incrementally in a stepwise fashion: a right-edge alignment constraint, for example, is satisfied if the tone it aligns is the rightmost tone in the constituent, and if its right edge is aligned with the right edge of the constituent (phonetically, if the pitch minimum or maximum takes place within the rightmost sonorant segment of the constituent, cf. Gussenhoven 2000a: 133); it incurs one violation for each tone that interferes with its alignment and for each TBU further away from the rightmost one (cf. Gussenhoven 2004: 151). Each directed alignment constraint has a directly opposed counterpart (i.e. a constraint right-aligning a tone has a counterpart left-aligning the same tone); below, we will list only those that are relevant for the analysis here. If their counterpart is not shown, it is taken to be ranked far below. If a tone A is associated with a particular position, it is also aligned there, i.e. it will block the full alignment of another tone B if that tone is to one side of it but seeks to align to the other and both the association constraint of tone A and the faithfulness constraint LINEARITY are ranked higher than B's alignment constraint (cf. Gussenhoven 2004: 155). Alignment constraints can also align tones with each other, or with a position in the prosodic structure, e.g. a stressed syllable (Gussenhoven 2004: 156–159). I follow Gussenhoven (2004: 156) in assuming that edges of prosodic constituents are not TBUs and tones can therefore not associate with them (cf. also section 3.5.2); alignment constraints thus suffice to position tones at constituent edges. However, if a tone is aligned with an edge, there is a free TBU available there (i.e., the final syllable) and if a constraint is active that demands for that tone to be associated, then it will not only align with the edge but also associate with the TBU (in our case this holds for H but not for L).

(84) ALIGN (L, Lt_{ϕ}): Align (the left edge of) L with the left edge (of the leftmost syllable) of the phonological phrase.

(85) ALIGN (L, Rt_{ϕ}): Align (the right edge of) L with the right edge (of the rightmost syllable) of the phonological phrase.

(86) ALIGN (H, Lt_{ϕ}): Align (the left edge of) H with the left edge (of the leftmost syllable) of the phonological phrase.

(87) ALIGN (H, Rt_{ϕ}): Align (the right edge of) H with the right edge (of the rightmost syllable) of the phonological phrase.

(88) ALIGN (H, σ'): Align (the right edge of) H with the right edge of a stressed syllable.

(89) LINEARITY: The sequence of tones in the output is the same as in the input.

(90) NOTARGET: A tone does not form a target (cf. Gussenhoven 2004: 155).

NOTARGET functions in parallel to NOASSOC: it prevents a tone from satisfying two or more alignment constraints (multiple alignment) if it is ranked between the constraints. Thus, ALIGN (T, σ') >> NOTARGET >> ALIGN (T, Rt_{ϕ}) will ensure that even if T does align with σ', it will not form a stretch extending to the right edge of the phrase. The prevention of multiple alignment can also be achieved by restricting the alignment of one tone via that of another. The two faithfulness constraints LINEARITY and MAXIO(T) as well as NOCROWD are here ranked above all the other constraints.

5.3.2.1 Main variant Spanish

With the constraints introduced, we can now proceed to show that the same ranking of them produces the correct result for both the "main" variant and the "phrase association" variant that still forms peaks aligned with stressed syllables, i.e. the one schematically represented by c) in Table 10 from section 5.1.3.1, and exemplified by the behaviour of speaker ZE55 in the double topic-utterances. I will refer to it as "ZE55's phrase accentuation" as a shorthand.

(91) Association and alignment constraint ranking for "main" variant and ZE55's "phrase accentuation"
ALIGN (H, σ') >> (σ')$_{\omega'}$← T >> H → TBU >> σ'← T >> ALIGN (L, Rt$_{\phi}$) >> ALIGN (L, Lt$_{\phi}$) >> ALIGN (H, Lt$_{\phi}$) >> NOASSOC >> NOTARGET >> ALIGN (H, Rt$_{\phi}$)

In this constraint ranking and with the tone sequences we are dealing with here, ALIGN (L, Rt$_{\phi}$) and ALIGN (L, Lt$_{\phi}$) are in direct competition and cause L to be as right-aligned as possible (i.e. always directly next to the next H), but also forming low stretches towards the left. Since multiple alignment is not constrained for L and σ'← T is ranked above NOASSOC, the L could in principle associate with available stressed TBUs it thus reaches (this is spreading[243] according to Gussenhoven 2000a, 2004: 153–155). This doesn't happen because there is one H for every stressed syllable and the high-ranked ALIGN (H, σ') together with NOCROWD ensures that an H associates with a stressed syllable, rather than an L. ALIGN (H, Rt$_{\phi}$), which will become important in the phrase accentuation variants by NQ01 and XJ45, is here ranked below NOTARGET, rendering it ineffective. Because the edge-alignment constraints for L are both higher-ranked than any edge-aligning constraints for the H tone, they effectively prevent the H from multiple alignment and spreading which would result in plateau realizations (see candidate (f) in the tableau for ZE55's phrase accentuation (Table 18)). While low stretches due to L aligning multiply occur very regularly, high plateaus extending between two accented syllables

243 The terminology is slightly confusing, because what enables spreading is the simultaneous satisfaction of opposing *alignment* constraint, but technically a tonal phenomenon can only be called spreading when the tone *associates* with more than one TBU (Gussenhoven 2004: 217). Consequently, NOTARGET is separate from NOASSOC. In many of the contours observed in this work, multiple alignment of tones is active, but there is less evidence that this also involves spreading (the association of the tone with more than one TBU it thus covers).

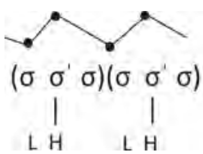
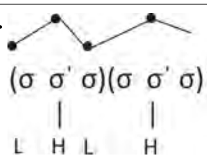
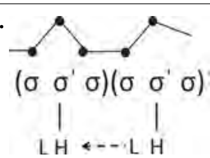
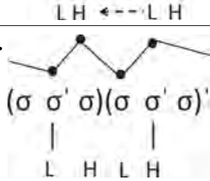
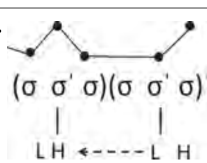
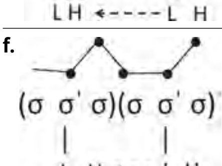
were indeed far less frequently found (cf. section 5.1.1.2). They are therefore not the main focus here, but of some interest because the plateau-forming behaviour is very frequent in the Quechua data. Whether they are allowed or disallowed is decided by the relative ranking of ALIGN (L, Rt_{ϕ}) and ALIGN (H, Lt_{ϕ}), respectively (as long as they're both above NOTARGET). If ALIGN (L, Rt_{ϕ}) is ranked above ALIGN (H, Lt_{ϕ}), then it will be the L that forms a low stretch, whereas if they swap places, high plateaus will be formed. This is only possible when there is only a single minimal tone sequence in the phonological phrase, because with several, as in the “main” variant, the faithfulness and association constraints would prevent a single H from aligning multiply across several stressed syllables at the expense of other tones. As an interesting side effect, the plateau-like realization (as found in section 5.1.1.2) thus turns out to be connected with the phrase accentuation. This seems intuitively plausible since they both are a step closer to the prosodic configuration most prevalent in the Quechua data than the “main” variant. However, since the plateau realization is quite rare in the Spanish data, I will continue the further analysis with the assumption that ALIGN (L, Rt_{ϕ}) is ranked above ALIGN (H, Lt_{ϕ}), disallowing its formation. The full ranking is then as follows:

(92) Full constraint ranking for “main” variant and ZE55's “phrase accentuation”
LINEARITY >> MAXIO(T) >> NOCROWD >> **ALIGN (H, σ')** >> **$(\sigma')_{\omega'}$← T** >> **H → TBU** >> **σ'← T** >> **ALIGN (L, Rt_{ϕ})** >> **ALIGN (L, Lt_{ϕ})** >> **ALIGN (H, Lt_{ϕ})** >> **NOASSOC** >> **NOTARGET** >> **ALIGN (H, Rt_{ϕ})** >> L → TBU >> σ← T

The constraints not given in bold in (92), i.e. the high-ranking first three constraints (LINEARITY, MAXIO(T), and NOCROWD) and the two low-ranking last ones (σ ← T and L → TBU) will not normally be given in the tableaus below, in order to make them less crowded. The tableaus show that this ranking produces the expected results both when there are as many LH sequences as there are prosodic words, in the “main” variant (Table 17) and when there is only one LH sequence per phonological phrase in the “phrase accentuation” variant by speaker ZE55 (Table 18).

In Table 17, six candidates are given for the “main” variant. The first three (a-c) all satisfy the four highest-ranking constraints ALIGN (H, σ'), $(\sigma')_{\omega'}$← T, H → TBU and σ'← T. They differ in the way they satisfy ALIGN (L, Lt_{ϕ}) and ALIGN (L, Rt_{ϕ}): candidate a) satisfies ALIGN (L, Rt_{ϕ}) at the expense of ALIGN (L, Lt_{ϕ}), producing a low target only on the syllable before the final stressed syllable, candidate b) the other way around, with only one low target directly after the first stressed syllable. Candidate c), the optimal candidate, satisfies both as much as possible: two low targets are produced, directly after the first and directly before the final stressed syllable, with a low stretch between them: this is indeed what we observe in the “main” variant, with pitch normally falling rapidly after the peak on a stressed syllable and staying

Table 17: OT-tableau with the constraint rankings to arrive at the correct association and alignment behaviour in the phonological phrase for the "main" variant of Spanish. Brackets are around prosodic words, accents mark the strongest element within its domain (syllable within prosodic word and prosodic word within phonological phrase); lines signify association between a tone and a syllable, dashed arrows indicate multiple alignment of a tone. Black dots are tonal targets.

[(σσ'σ)$_\omega$ (σσ'σ)$_\omega$']$_\phi$ LHLH	**ALIGN (H, σ')**	**(σ')$_\omega$ ← T**	**H → TBU**	**σ' ← T**	**ALIGN (L, Rt$_\phi$)**	**ALIGN (L, Lt$_\phi$)**	**ALIGN (H, Lt$_\phi$)**	**NOASSOC**	**NOTARGET**	**ALIGN (H, Rt$_\phi$)**
a. 					**7***	***!		**	****	*****
b. 					**8*!**	**		**	****	*****
☞ **c.** 					**7***	**		**	*****	*****
d. 	*!		*		**6***	****		**	****	****
e. 	*!		*		**6***	**		**	**6***	****
f. 	**!		**		*****	****		**	*****	***

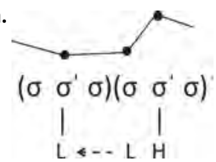
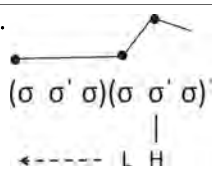
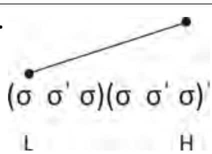
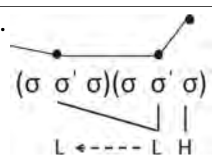
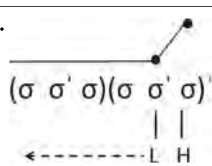
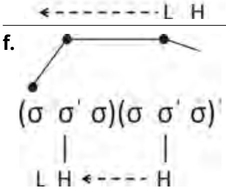
low until right before the next one. The last three candidates (d-f) all fail to be selected because they violate Align (H, σ') and H → TBU, thus preventing a realization with the middle L and H as floating central tones (d), H as right-aligned boundary tone (e) or one that could be annotated as L*H (f).

5.3.2.2 Phrase accentuation Spanish variant 1 (exemplified by speaker ZE55)

For the “phrase accentuation” variant produced by ZE55, corresponding to c) in Table 10, the same constraint ranking holds, as shown in Table 18. Candidate a) is the optimal one, with the H associating with the strongest syllable and the L again forming two elbow targets, one closely aligned before the H, and another on the other stressed syllable, with which it associates. This makes the assumption that an association constraint prefers an alignment of at least one edge of a tone with the corresponding edge of the TBU it associates with. Without this assumption, candidate b) would actually be preferred, because Align (L, Lt) would incur one violation mark less. In fact, neither for the realizations by ZE55 nor the phrase accentuation utterances discussed in section 5.1.3.1 is the contour represented by candidate a) the only one attested, the one represented by candidate b) also seems to occur. This is difficult to assess with certainty also because L tone targets are notoriously difficult to identify and because it then becomes partially a matter of theoretical preference whether the L should be seen as associated or not. It seems plausible that the L tone associates variably, due to σ'← T having a ranking distribution that here significantly overlaps with that of Align (L, Lt_{ϕ}) (cf. Boersma & Hayes 2001). The tableau thus displays the ranking which is probably selected more frequently.[244] Candidate (f), as mentioned above, displays the plateau realization which is here prevented by the ranking of Align (L, Rt_{ϕ}) over Align (H, Lt_{ϕ}). If their ranking were reversed, candidate (f) would win, generating the realization observed occasionally in section 5.1.1.2 and also likely on the second topics in the double topic-utterances by ZE55, NQ01, and XJ45.

244 This is modeled in Boersma & Hayes (2001: 47–49) by assuming that the position of a constraint on a continuous scale has a random perturbance range so that at the moment of production, a *selection point* on the scale for the constraint is generated that is different from its position. If two constraints C1 and C2 are closely adjacent on the scale, this leads to a majority of outcomes exhibiting C1 >> C2 (if C1 is ranked higher), while a minority exhibit C2 >> C1, with their frequency ratio determined by their proximity on the scale.

Table 18: OT-tableau with the constraint rankings to arrive at the correct association and alignment behaviour in the phonological phrase for ZE55's phrase accentuation variant of Spanish. Brackets are around prosodic words, accents mark the strongest element within its domain (syllable within prosodic word and prosodic word within phonological phrase); lines signify association between a tone and a syllable, dashed arrows indicate multiple alignment. Black dots are tonal targets.

$[(\sigma\sigma'\sigma)_\omega\ (\sigma\sigma'\sigma)_\omega{}']_\phi$ **LH**	ALIGN (H, σ')	$(\sigma')_\omega \leftarrow$ T	H → TBU	σ' ← T	ALIGN (L, Rt_ϕ)	ALIGN (L, Lt_ϕ)	ALIGN (H, Lt_ϕ)	NOASSOC	NOTARGET	ALIGN (H, Rt_ϕ)
☞ **a.** 					**	*	****	**	****	*
b. 				*!	**		****	*	*****	*
c. 	*!	*	*	**	*****		*****		**	
d. 	*!				*	*	*****	***	*****	
e. 	*!			*	*		*****	**	6*	
f. 					*****!		*	**		*

5.3.2.3 Phrase accentuation Spanish variant 2 (exemplified by speakers XJ45 and NQ01)

As a next step, we proceed to the next “phrase accentuation” variant that XJ45 displays in his double topic-utterances, and which in non-final phrases corresponds to d) in Table 10 from section 5.1.3.1. In order to generate the correct output for the phrase accentuation variant by XJ45 and NQ01, a crucial reranking has to be undertaken. This variant regularly has final rises at the end of the phonological phrase instead of peaks on the metrically strongest syllable. This central difference is achieved by moving ALIGN (H, Rt_{ϕ}) from its position below NOASSOC and NOTARGET to the top of the ranking as shown, i.e. above $(\sigma')_{\omega'} \leftarrow T$ (but still below LINEARITY, MAXIO(T), NOCROWD not shown in the tableaus), and by moving ALIGN (H, σ') to a position further below. This results in the ranking (93).

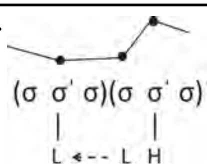
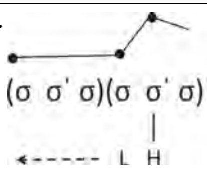
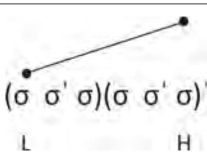
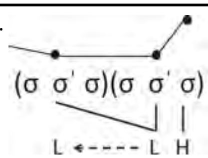
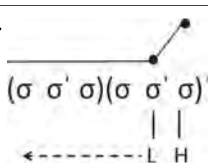
(93) Association and alignment constraint ranking for XJ45’s “phrase accentuation” variant
ALIGN (H, Rt_{ϕ}) >> $(\sigma')_{\omega'} \leftarrow$ T >> H → TBU >> $\sigma' \leftarrow$ T >> ALIGN (H, σ') >> ALIGN (L, Rt_{ϕ}) >> ALIGN (L, Lt_{ϕ}) >> ALIGN (H, Lt_{ϕ}) >> NOASSOC >> NOTARGET

Table 19 gives a tableau with this ranking[245] showing that this will indeed pick out the correct candidate: candidates a) and b), which were very good in ZE55’s variant, are now out because they violate the high-ranking ALIGN (H, Rt_{ϕ}). Candidate c) does not do that, but since it does not associate any of the tones, it is discarded because of violation of the still high-ranking $(\sigma')_{\omega'} \leftarrow$ T and $\sigma' \leftarrow$ T. Candidate d), on the other hand is optimal: it aligns the H tone at the right edge of the phonological phrase, allowing it to associate with the available TBU there and leaving the preceding L tone to associate with the strongest syllable and spread leftwards, also associating with the other stressed syllable and forming a second target there. Candidate d) is preferred over candidate e) again because of the assumption we made above about associations preferring tonal targets; however, also for XJ45 deciding between the contours of d) and e) is not always straightforward from the data.

We can now model the final step attested step for Spanish, namely the kind of “phrase accentuation” that NQ01 produces in her double topic-utterances and elsewhere. Like XJ45, she does not produce peaks on strong syllables in non-final phonological phrases, but instead final rises. Apart from the behaviour in IP-final phonological phrases, which will be treated in the next section, there is only one noticeable difference between XJ45 and NQ01 in the ranking of the constraints we have looked at so far: NQ01 never seems to align elbows with any of the pre-final

245 NOTARGET is not displayed in the tableaus from here on.

Table 19: OT-tableau with the constraint rankings to arrive at the correct association and alignment behaviour in the phonological phrase for speaker XJ45's phrase accentuation variant of Spanish. Brackets are around prosodic words, accents mark the strongest element within its domain (syllable within prosodic word and prosodic word within phonological phrase); lines signify association between a tone and a syllable, dashed arrows indicate multiple alignment. Black dots are tonal targets.

[$(\sigma\sigma'\sigma)_\omega$ $(\sigma\sigma'\sigma)_\omega$' $]_\phi$ **LH**	**ALIGN (H, Rt_ϕ)**	**$(\sigma')_\omega \leftarrow T$**	**$\sigma' \leftarrow T$**	**H → TBU**	**ALIGN (H, σ')**	**ALIGN (L, Rt_ϕ)**	**ALIGN (L, Lt_ϕ)**	**ALIGN (H, Lt_ϕ)**	**NoAssoc**
a. 	*!					**	*	****	**
b. 	*!		*			**		****	*
c. 		*!	**	*	*	*****		*****	
☞ d. 					*	*	*	*****	***
e. 			*!		*	*		*****	**

stressed syllables. Instead, the low stretch in her phonological phrases spreads uniformly from their left edges to directly before the rise on the final syllable. I therefore assume that candidate e) in Table 19, which is dispreferred under the assumption of association positions preferring targets, is actually the normal outcome candidate for her. In order to arrive at a ranking that will uniformly produce that output while still upholding that assumption, $\sigma' \leftarrow$ T is moved further down the ranking, below NoAssoc. This positions this ranking even closer to Quechua, where $\sigma' \leftarrow$ T plays no role in any variant (see section 6.3).

(94) Association and alignment constraint ranking for NQ01's "phrase accentuation" variant
ALIGN (H, Rt$_\phi$) >> $(\sigma')_{\omega'} \leftarrow$ T >> H $\rightarrow$ TBU >> ALIGN (H, σ') >> ALIGN (L, Rt$_\phi$) >> ALIGN (L, Lt$_\phi$) >> ALIGN (H, Lt$_\phi$) >> NoAssoc >> NoTarget >> $\sigma' \leftarrow$ T

With this ranking, the selection of candidate e) from Table 19 is assured, whether the assumption holds or not. Note that contours generated by this constraint ranking would be very hard to differentiate from ones in which association constraints play no role at all: the position of the tones is entirely determined by alignment constraints. From a Spanish perspective, it makes sense to maintain the notion of association here, since we have good reasons to assume that H tones associate with stressed TBUs in "main" variant Spanish. The theoretical consequence is therefore to retain the assumption of association in the "phrase accentuation" variants, especially if we want to trace the difference between the variants in terms of minimally necessary steps. However at this stage, the matter becomes less easy to determine empirically, since differences would only emerge in quantified positional measurements of the turning points and effects on other acoustic correlates of the kind done in section 6.1.6 on Quechua rise-falls and Spanish main variant paroxytones, and the results, as they are there, would still be open to some interpretation. Here I will assume that the above ranking with association is valid for at least some of the Spanish utterances. In section 7.4 we will then consider what happens when we remove the association constraints.

5.3.3 IP-final behaviour

Things become slightly more complex once IP-boundary tones get thrown into the mix. The limits of what the discrete OT-model can faithfully represent will become evident. However, we will also see that it allows us to generalize between cases that might seem disparate at first. First of all, there is the case where the IP-boundary tone is realized at the right edge of the phrase while the H tone of the phonological phrase forms a separate peak on the preceding stressed syllable (independent of whether the IP tone is an L or a H itself): we can assume that this is achieved by a constraint ALIGN (T_ι, Rt_ι), which aligns a tone provided by the intonational phrase with the right edge of that phrase (which is also the right edge of the rightmost phonological phrase). If this constraint is ranked above the other association and alignment constraints, then in cases of IP-final paroxytones, this will lead to the observed outcomes, not only for the main variant but also for the phrase accentuation variants: all the other constraints so far discussed staying as they are, this will result in the T_ι aligning at the right edge of the phrase. Since the H tone of the phonological phrase cannot now align with the last TBU anymore (the edge of the IP and that of the rightmost phonological phrase being the same), it will move one syllable to the left onto the strongest metrical syllable with which it associates, and the preceding L is also "pushed" further back, possibly associating with one of the available other stressed syllables and satisfying its alignment constraints as well as possible. For ZE55 and the "main" variant, since ALIGN (H, Rt_ϕ) is not ranked high anyway, no competition takes place between the H tone of the phonological phrase and the boundary tone of the IP (as long as the final word is paroxytonic, see below), but the result is the same for the more edge-prominent variants of "phrase accentuation" (cf. e.g. XJ45's ELQUD_ES_16 (Figure 63) and 19B (Figure 64) in section 5.1.3.1.3). This shows that in the phrase accentuation variant creating final rises in non-final PhPs, the H tone of the PhP is not really fixed at either the phrase edge or the strongest syllable: when no other tone intervenes it will align rightmost, but if it cannot, it will still align and associate with the strongest position. This provides support for the idea that tones are really autonomous on their tier and not predestined for any kind of association. It also demonstrates that pitch contours appearing to be the same or very similar can be generated from various rankings. The same contour also occurs in the Quechua data, where it is again generated from different rankings. ALIGN (T_ι, Rt_ι) also has a counterpart, ALIGN (T_ι, Lt_ι). This is ranked below the other constraints, but its effect emerges in cases of deaccentuation (see below). The effect that the competition between the two tones for the rightmost position has is necessarily enabled by NOCROWD (although it doesn't capture the full phenomenon, see below). The ranking is given in (95).

(95) Constraint ranking including IP tones
NOCROWD >> ALIGN (T_ι, Rt_ι) >> Association and alignment ranking for main / phrase accentuation >> ALIGN (T_ι, Lt_ι)

The same constraint ranking also generates the attested outcomes in utterances with final oxytones and with deaccentuation. In utterances with IP-final oxytonic or monosyllabic words, the same competition between the IP boundary tone L_ι and the last H tone of the phonological phrase takes place also for the main variant that we just discussed for IP-final paroxytones in the phrase accentuation variant. An example in an utterance that is in the main variant is (38) / Figure 21 from section 5.1.1; an example from an utterance with phrase accentuation by NQ01 is given in Figure 111.

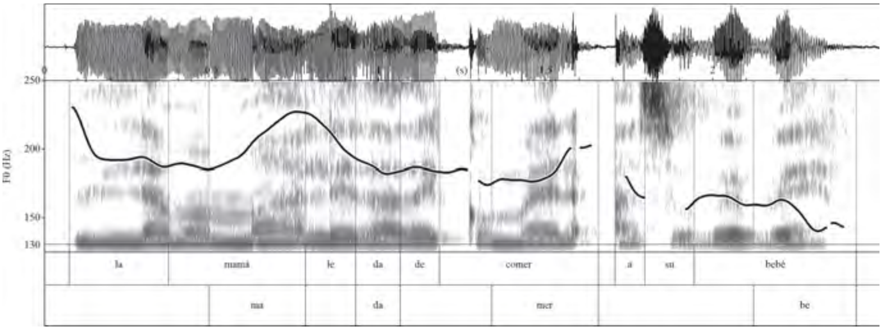


Figure 111: NQ01_ELQUD_ES_60[246] (*la mamá le da de comer a su bebé* ‘the mother gives her baby to eat’).

What we observe in both cases is that the IP-final stressed syllable realizes a pitch peak, followed by a fall. This suggests that both tones are realized on this last syllable, with the right-alignment of the L_ι winning out against that of the H. This is exactly what the constraint ranking would predict for oxytonic or monosyllabic final words: Align (H, σ') and Align (H, Rt_ϕ) now seek alignment for the H to the same TBU and are both equally outstripped by Align (T_ι, Rt_ι). But H will still align as far to the right as possible. Both $(\sigma')_{\omega'}\leftarrow$ T and H $\rightarrow$ TBU effect an association of the H with the stressed syllable, while L_ι remains an unassociated boundary tone, there being no constraint forcing its association. The two utterances exemplify however that this account leaves out relevant aspects of what is observable here. In Figure 21, the H tone clearly wins out over the L_ι in terms of scaling and (impressionistically) perceptual salience, while it is the other way around in Figure 111. The tones are clearly in competition under adverse temporal conditions.

246 https://osf.io/d46st/

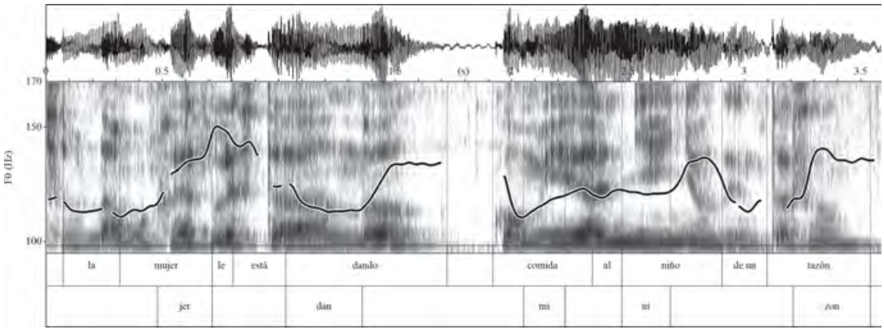


Figure 112: XJ45_ELQUD_ES_60[247] (*la mujer le está dando comida al niño de un tazón* 'the woman is giving food to the child from a mug').

XJ45'S ELQUD_ES_60 (Figure 112) is an even more radical example for the H tone winning out over the L_ι which here arguably does not get realized at all. For utterances like this one (recall also NQ01's ELQUD_ES_65/Figure 104), we could assume that the relative ranking between ALIGN (H, Rt_ϕ) and ALIGN (T_ι, Rt_ι) is actually reversed, with the former now outranking the latter, so that the L_ι aligns to the left of the H, with its realization then becoming difficult to detect in the low stretch that is realized anyway before the peak of the H. This is parallel to Gussenhoven (2000a)'s proposal for Roermond Dutch (cf. (70)). The nonperipheral realization of boundary tones is perhaps not so strange now that we have witnessed the malleable roles tones can assume throughout the discussion. To my mind, a much graver problem is that modelling this via a discrete reranking of the two constraints obscures the fact that many of the phenomena involved are essentially gradual. Two of the resulting processes when several tones compete for the same temporal window have been called truncation and compression in the literature, truncation signifying a pitch movement that is incompletely realized, and compression one that is completely realized but with reduced temporal span. Arguably in Figure 21 the H is compressed and the L_ι slightly truncated; in Figure 111, it is the H whose scaling is considerably truncated, and in Figure 112 the L_ι is fully truncated. The two processes have often been seen as mutually exclusive strategies that languages adopt wholesale, i.e. as a typological dichotomy that separates languages into 'truncating' and 'compressing' ones. This view is espoused by Ladd (2008: 182), although it is already noted there that "the distinctions are by no means clear-cut". Indeed, Prieto & Ortega-Llebaria (2009)'s results indicate that individual differences seem to play a substantial role: one of their four peninsular Spanish speakers shows a

247 https://osf.io/rv8u5/

much greater preference for truncating the LH*L% contour in IP-final oxytones under narrow focus than the other three (who prefer compensatory lengthening of the final syllable). Two further studies shed a more discriminating light on the phenomena. Cho & Flemming (2015) show for Seoul Korean that on the one hand, compression can gradually lead to truncation: with increasing time pressure (faster speech rate), the second L tone of the LHLH tune in the accentual phrase (AP) will first be increasingly undershot and then not realized at all even in four-syllable APs. On the other hand, they also attest that there is a form of categorical truncation: in two- or three-syllable APs, the same second L tone will never be realized, no matter how slow the speech rate. Although both phenomena target the same tone, they assert that only the latter, but not the former, constitutes a case of phonological tone deletion (Cho & Flemming 2015: 379). Rathcke (2016) compares German and Russian in situations where pitch accents HL*, H*, and L*H on IP-final paroxytones and oxytones with a variably voiced or unvoiced coda in the stressed CVC syllable were followed by an L% boundary tone. Her results demonstrate that the two languages, previously both classified simply as 'truncating', employ both compressing and truncating strategies as well as temporal re-alignment of the tones, but do so differently from each other, and to different degrees (up to and including categoriality) in relation to the experimental conditions (cf. Rathcke 2016: 221–223). She also shows that in both German and Russian the phonological status of a tone is relevant for which strategy is employed: in the H* L% and L*H L% tunes under the severest time pressure conditions, the final fall was extremely undershot, truncating the target of the L boundary tone, while in the HL* L% tune, it was mostly preserved or only slightly undershot, leading to the conclusion that L tones that are part of pitch accents are preferentially preserved over those that are boundary tones (Rathcke 2016: 223). Both Cho & Flemming (2015) and Rathcke (2016) paint a very nuanced picture of truncation and compression, and concur on the point that these processes do not affect all tones equally: if a tone is affected in Seoul Korean, it is the AP's second L tone, and in German and Russian it is the L tone that signals the IP boundary. Recall that the most likely analysis proposed for neutral polar questions in Huari Spanish (section 5.1.2.1) in contrast requires that the realization of boundary tones wins out over that of the final pitch accent under crowding conditions, so that the formal differentiation of that utterance type is preserved. This kind of targeting behaviour in principle squares well with a constraint-based approach, in which the faithful preservation of the different tones is ranked against each other, so that under adverse conditions, one of them will be preferentially reduced or eliminated.

Yet the preferential and gradual nature is not really captured in our constraints here, demonstrating the limitations of this essentially discrete modelling. Equally, the existing competition between the two tones is only badly captured with a constraint like NoCrowd, which only forbids the association of two tones with the same TBU (cf. Gussenhoven 2004: 149), but does not say anything about a case like this one, where the tones are in competition even though probably only one of them (the H) seeks to associate, and yet can be seen to lose out in some cases in terms of relative contrast. This matter will not be resolved here, but it should be understood that realistic versions of the constraints involved here probably would need to refer to gradual differences in articulatory and perceptual contrastiveness of pitch movements under differing temporal conditions. Yet at the same time there is also an element of categoriality involved, since adverse temporal conditions are not simply those with increased speech rate: in Cho & Flemming (2015), increase in speech rate does increases the undershoot for the second L tone, but it is the more phonologically-mediated increase in time pressure in the form of having to realize four tones in less than four syllables that categorically deletes it.

Keeping this in mind, the same constraint ranking (with Align (T_ι, Rt_ι) dominating Align (H, Rt_ϕ)) can also be maintained for deaccentuation. We saw that deaccentuation occurs occasionally in the Huari Spanish data, and even though its occurrence is likely somewhat subject to preference, it seems most likely to occur on postnuclear material, i.e. when the strongest metrical position is prefinal in a structure. An attempt at a definition is given in (96), based on a similar formulation in Féry (2017: 155):

(96) If a non-final constituent of a (minimal) IP receives highest prominence, then any constituents of the same level following it are not assigned any tones at the level of the phonological phrase (deaccentuation)

Just as in the phrase accentuation, the realization of the postnuclear pitch accents in cases of deaccentuation varies gradually between reduction in scaling and deletion (realization as flat). In section 5.1.3.2, I noted that it seems more likely that a totally flat pitch contour is a variant realization for accents with reduced scaling than to assume that phonologically fully deleted accents should manifest a variant realization where tones are only somewhat reduced in scaling. To capture this in a general account, the competition between the constraints aligning the L tones leftwards here and faithfulness constraints preserving the pitch accents would have to be allowed to have gradually variable outcomes.[248] The constraints here are not

248 Note however that such a gradual competition could very well accommodate the notion resulting from the tone suppression constraints in section 5.3.1, that there are tones in principle available even for each syllable. Effectively this is a conception in which the different levels of the prosodic hierarchy are in constant competition with each other about the creation of articulatory and perceptual contrasts, with in principle gradual outcomes.

fully equipped to deal with this gradual variation in scaling. However, assuming that (96) captures (an aspect of) the phenomenon truthfully at some (abstract) level, this allows for a generalized treatment of deaccented and non-deaccented utterances in the following fashion:[249] in the postnuclear stretch, there won’t be any phonological phrase tones available. However, since this is still part of an IP, there will be an IP tone available (in most cases an L_ι), aligning to the right. Earlier we said that its counterpart, ALIGN (T_ι, Lt_ι), needs to also be included in the ranking, although below all the other constraints considered. Because all other constraints ranking above it that block its effect refer to phonological phrase tones, it is now free to act here and its presence then accounts for the observable formation of a low elbow in these cases directly following the nuclear syllable on which the last H tone is realized, often causing quite abrupt falls (but cf. Barnes et al. (2010) for an alternative account). And there is in fact evidence that this constraint is also active in utterances without deaccentuation, coming from utterances with IP-final proparoxytones, such as Figure 23 in section 5.1.1 or Figure 113, where the pitch on the last word forms a peak on the stressed syllable and then drops abruptly at the beginning of the penult as in the cases with deaccentuation, forming a low stretch from there to the end of the phrase.

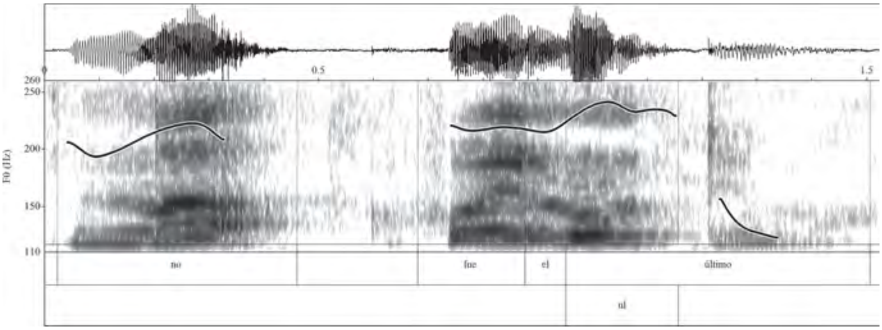


Figure 113: ZE55_ELQUD_ES_51[250] (*no fue el último* ‘no it was the last one’).

249 The way it is formulated, (96) also applies equally well to cases of both deaccentuation and dephrasing, unifiying them somewhat (cf. section 3.4.6). It also preserves the notion that the strongest (nuclear) accent in any phrase from the minimal IP-level upwards is always rightmost (that is what deaccentuation makes sure of), but it assumes a recursive prosodic structure in order to do so.
250 https://osf.io/uxenr/

Summing up, this section has used OT to demonstrate that the observed intonational variants of Huari Spanish can be analyzed as differing prominently along two dimensions, that of tone distribution per domain and that of tonal alignment and association. It was shown that via a systematic reranking of a limited number of constraints, Huari Spanish utterances move from one end of the spectrum, in which prosodic words are optimized by realizing an LH* pitch accent on each stressed syllable, to the other, in which phrases comprising several words are optimized via edge-seeking tones; with several attested positions in between also falling out from this stepwise reranking process. The analysis did not extend to two additional (and related) dimensions, tonal scaling and crowding, even though it was found that they are very relevant to the intonational characterization of Huari Spanish. Yet their continuous nature does not lend itself to an essentially categorical (even if stepwise gradable) analysis as the one employed. This is in itself a noteworthy result. Still, the analysis has provided the Spanish half to answering the third group of research questions laid out in section 4, especially (36)b. In the next chapter (6), we will explore the intonation of Huari Quechua, including with an OT-analysis (section 6.3) that provides the other half of the answer to that question. More comprehensive conclusions will then follow in the final chapter (7).

6 Huari Quechua

This chapter describes the intonation of Huari Quechua. It is separated into three sections. The first section describes the intonational contours observed, and argues that variation exists in the tonal alignment patterns underlying them, of which only a subset makes reference to a regularly determined word stress on the penult, while others only refer to phonological phrase and prosodic word boundaries. Section 6.2 explores the distribution of these alignment variants and relates them quantitatively to aspects of meaning and information structure. Section 6.3 relates the alignment variants with each other and the variation found in the Spanish data via an OT-analysis. Section 6.4 qualitatively explores further aspects of how information structure is cued in interaction between prosody and morphosyntax.

6.1 Description of Huari Quechua tonal contours

In section 3.3.3, it emerged from the discussion on the literature about prosody in Ancash Quechua that it is doubtful whether the existing accounts all describe the same phenomenon of word stress, and that the pitch behaviour described is perhaps more indicative of the behaviour of phrasal tones seeking alignment with prosodic edges. In this section I will provide an analysis of the intonational patterns observed in the Huari Quechua data. I will bring evidence to bear against the assumption that tonal events in this Quechua variety are homogeneously affected by a stress position determined at the word level. Instead, it will be argued that Quechua intonation is based on phrasal tones assigned at the level of the Phonological Phrase (corresponding to the Accentual Phrase in the descriptions of some languages). Three phrasal contours will be identified: a rising, an only-falling, and a rise-falling one (section 6.1.1). Some of the tones forming these contours always seek alignment with the phrase boundary (sections 6.1.3, 6.1.5). The variation in alignment of the others follows two main patterns: alignment with the word boundary (section 6.1.2), and with the word penult (section 6.1.4). There is also a marked pattern only observed on loanwords of Spanish origin (section 6.1.8). Even though the word penult is determined to be regularly metrically prominent (as argued in many existing accounts), crucially, its influence on prosody is found to be small overall and varying depending on the alignment pattern: in the word-boundary variant, it has no influence, while in the word-penult variant it serves as a landmark for tonal alignment but nothing else. This is shown to result in quantifiable alignment differences between phenotypically similar contours of Huari Quechua and Spanish (section 6.1.6). The results in this section overall suggest that Huari

https://doi.org/10.1515/9783111304595-006

Quechua as represented by this data ranks very low on a typology of how much of the prosodic phonology is affected by stress position, as proposed by Hyman (2014). This is further corroborated by the marked (and overall infrequent) behaviour of loanwords from Spanish: the stress position they are sensitive to is treated quite differently by the prosody. I suggest that the observed variability is at least partially responsible for the heterogeneous descriptions found in the existing literature.

The qualitative analysis in this section will treat declaratives, but section 6.2.1 demonstrates that utterance type is not an overly relevant factor influencing contour choice in Huari Quechua, similar to findings on other varieties (cf. Cole 1982; O'Rourke 2005). This is also confirmed in section 6.1.6, where peak alignment in the same contour shows no evidence of differing between utterance types.

6.1.1 Introduction to the rising and falling contours

Most generally speaking, there are two different kinds of phrase-level tonal contours: falling and rising ones. Falling contours minimally include a fall from high, possibly followed by a low stretch, as *(only-)falling contours*. Optionally, the fall is also preceded by a rise and/or a high plateau, in *rising-falling contours*. *Rising contours* minimally consists of a rise, optionally preceded by a low stretch, and optionally followed by a suspended high plateau and/or a further final rise.

(97) AZ23_Conc_Q_0960

segunda	fila-chaw	segundu-chaw	kiru
second	row-LOC	second-LOC	tooth

"in the second row, the second [one is the] tooth"[251]

Both a rising and a rising-falling contour are produced in AZ23_Conc_Q_0960[252] ((97)/Figure 114). The utterance consists of two phrases, *segunda filachaw* and

251 This is an example from *Conc* (cf. section 2.4 for a description). Thus it should be understood as "in the second row (*segunda filachaw*), the card in the second column (*segunduchaw*) is the one with the tooth (*kiru*)".

252 Figures with Quechua examples include five tiers in the textgrid: a tier with an IPA transcription aligned with syllable boundaries (1), a tier separating morphological words in an orthographic transcription and aligned with the word boundaries (2), a tier separating the words into morphemes (3) and another with the English glosses (4), and one giving an approximate English translation (5). The syllabification in tier (1) is based on the Quechua phonotactics described in Parker (1976), according to which both onsets and codas cannot be complex and onsets are preferred over codas (VCVCV will be syllabified as V.CV.CV and not VC.VC.V). The IPA transcription is phonetic and aims to be as faithful as possible, using both the auditive percept and the spectral information provided

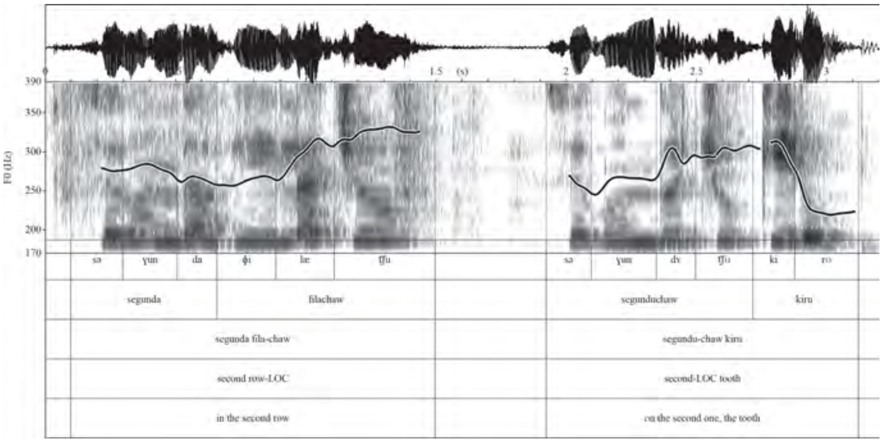


Figure 114: AZ23_Conc_Q_0960[253] (declarative with a rising and a falling contour).

segunduchaw kiru, the first one with a rising contour and the second with a falling one. In the first one, *segunda* is basically flat, then pitch rises on the penultimate

by *praat* as basis. Occasionally, this results in the transcription of additional vowels or consonants (such as epenthetic vowels breaking up consonant clusters, glides between vowel segments, or nasal segments before voiced onsets) that are not represented in the morphological word form. The transcription also shows that many consonants are often subject to processes of lenition or place assimilation, and occasionally to dissimilations; and that especially word- and phrase-final vowels are often realized as centralized, with a breathy quality, or partially or completely unvoiced, with a considerable variability of realization in general. The extent of this varies between speakers. In cases of extreme vowel reduction I have tended to preserve Quechua phonotactics as described above, e.g. transcribing a realization of /hawanchaw/ with a severely reduced final vowel as [ha.wan.t͡ʃʉ̥], while a transcription approach ignoring any preconceived notion about supposed underlying forms might have chosen [ha.want͡ʃ]. When there was absolutely no trace of a vowel segment, I also sometimes chose a transcription violating the phonotactics. In any case, the spectrogramme and the interval boundaries in the figures allow readers to inspect the relative duration and amount of energy in the vowel segments for themselves. I have not used the length diacritic [ː] at all, because it seems there is no principled way of using it here: phonologically long vowels (based on existing descriptions) are not always consistently produced with longer duration than their short counterparts, even in the same utterance. The length of the syllable interval in the figure, compared with the morphological form given in tiers (2) and (3), where long vowels are represented by a repetition of the vowel symbol, provides information about actual temporal realization. I do not think phonologically contrastive vowel length is nonexistent here, just that there is no principled way of using the length diacritic in a phonetic transcription of these data. Parker (1976) describes processes responsible for the shortening of (morphologically) long segments, e.g. one where a syllable that is morphologically CVːC is produced as CVC, because maximal syllables are either CVː or CVC. In the data here, vowels in these positions seem to show the same durational variability as all others.

253 https://osf.io/acp7y/

syllable of the phrase (*la* of *filachaw*) and stays high on the final syllable. In the second one, pitch is flat on the first two syllables, then rises on the penult of the first word *segunduchaw* and stays high until the first syllable/penult of *kiru*, then falls steeply, forming a low stretch on the final syllable. Which contour is realized does not depend on the lexical items involved,[254] as (98)/Figure 115 shows.

(98) LC34_MT_Q_1037 and XQ33_MT_Q_1053
hirka hawa-n-pa
hill below-3-GEN
"below the hill"

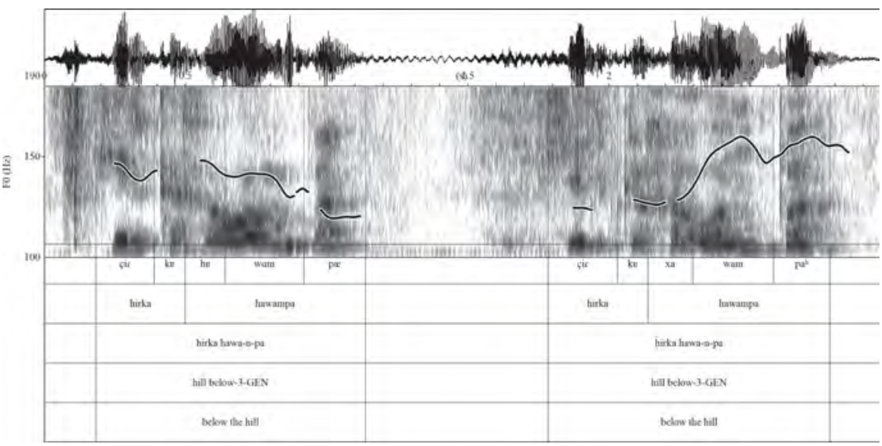


Figure 115: LC34_MT_Q_1037[255] (declarative realized with a falling contour) and XQ33_MT_Q_1053 (confirmation-seeking polar question realized with a rising contour).

Here, two utterances that are lexically and syntactically identical but pragmatically different are realized by two speakers in direct succession, first with a falling and then with a rising contour.[256] The same word *hawanpa* realizes a fall on its penult in the first case and a rise in the second, while no tonal event occurs on *hirka*, which is overall high in the first, and overall low in the second utterance. *Hirka* does not

254 Except to the extent that additonal alignment patterns are optionally available for loanwords from Spanish, for which see section 6.1.8. In the preceding sections, any examples containing Spanish loanwords show intonational behaviour also available to "native" Quechua words, which is their majority behaviour (cf. 6.1.8.3).
255 https://osf.io/mfbse/
256 In a very similar context as the one described for Huari Spanish confirmation-seeking questions, cf. 5.1.2.2.

belong to a different lexical class from *hawa*, e.g. of unaccented vs. accented words, respectively, as might be suggested from the perspective of languages like Japanese or Basque (cf. section 6.1.8.2). This is demonstrated by (99)/Figure 116, where the same lexical item is realized with a rising-falling contour.

(99) XQ33_MT_Q_1213
hatun ka-q hirka-pa
big COP-AG hill-GEN
"the hill that is big"

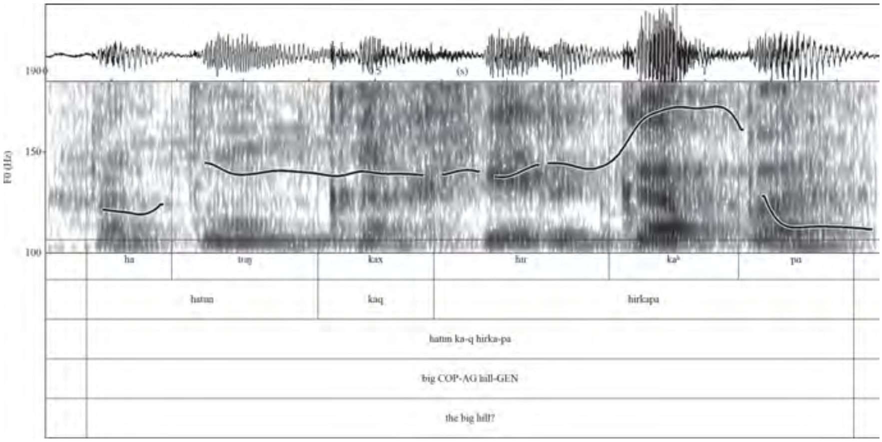


Figure 116: XQ33_MT_Q_1213[257] (confirmation-seeking polar question with falling contour).

No evidence was found for a lexical distinction affecting the type of pitch event that can be realized, apart from when words of Spanish origin are involved, for which see section 6.1.8. Nor does a certain class of suffix such as e.g. the genitive –*pa* here effect the presence or absence of pitch accenting on a word (as is the case in some varieties of Basque, cf. e.g. Elordieta 1998; Hualde 1999; Hualde et al. 2002). That can be seen from (100)/Figure 117, where *qiru-pa* is just as lacking in pitch events as the two instances of *hirka* above, while *waqta-n-pa* does realize a pitch event.

(100) LC34_MT_Q_0095
qiru-pa waqta-n-pa
wood-GEN behind-3-GEN
"behind the wood"

257 https://osf.io/rquxg/

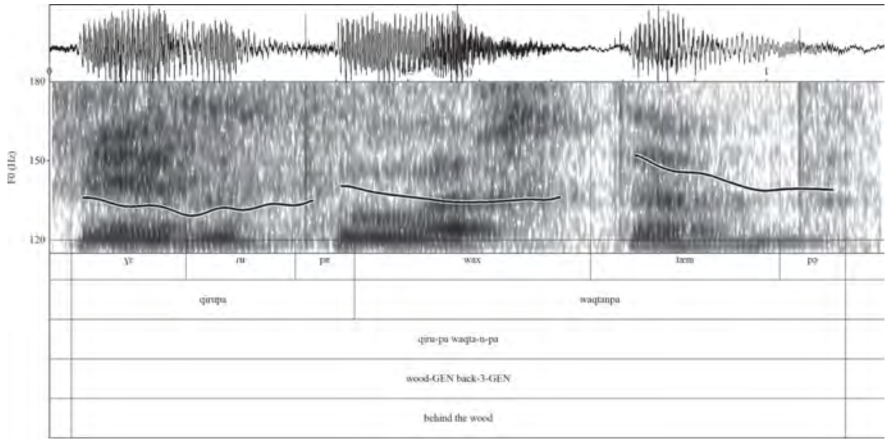


Figure 117: LC34_MT_Q_0095[258] (part of a declarative with falling contour).

That there are no categorical lexical restrictions on which type of word the low or high stretch or the tonal transition can be realized is also valid for the difference between lexical and function words. In XQ33's Conc_Q_0690 ((101)/Figure 118), a rising contour is realized on two nouns, with the rise taking place only on the penult of the second, the spatial noun *hawanchaw* "in its below[259]", and the first noun realized low without any pitch event.

(101) XQ33_Conc_Q_0690
qillay hawa-n-chaw krus
money below-3-LOC cross
"below the money [is] the cross"

258 https://osf.io/y97cx/

259 As can be seen from the gloss in (101), this is analyzed as *hawa-n-chaw* "below-3-LOC", with the 3rd person suffix referring to the preceding noun *qillay* "money", thus "below the money". I label as "spatial nouns" a group Quechua words that behave syntactically just like any noun, but refer to spatial relations (cf. Parker (1976: 87–88)). To give an example, in a possessive construction like *qillay hawa-n-chaw*, marking the possessor additionally (glossed as genitive) is also possible, yielding *qillay-pa hawa-n-chaw* "money-GEN below-3-LOC", with essentially the same meaning. The marking can also be reversed, i.e. *hawa-pa qillay-nin-chaw* "below-GEN money-3-LOC" (with the 3rd person marker in the form –*nin* because of the phonotactic restriction to simple codas), so that the meaning becomes "in the money from below" (e.g. as opposed to money from above). The fact that these spatial nouns regularly take person markers in Quechua is adduced by several authors to account for the frequency of parallel constructions, such as *en su debajo del dinero*, (instead of *debajo del dinero*), which are also attested in abundance in our data, in the Spanish varieties of Quechua-speaking regions (cf. Escobar 2000).

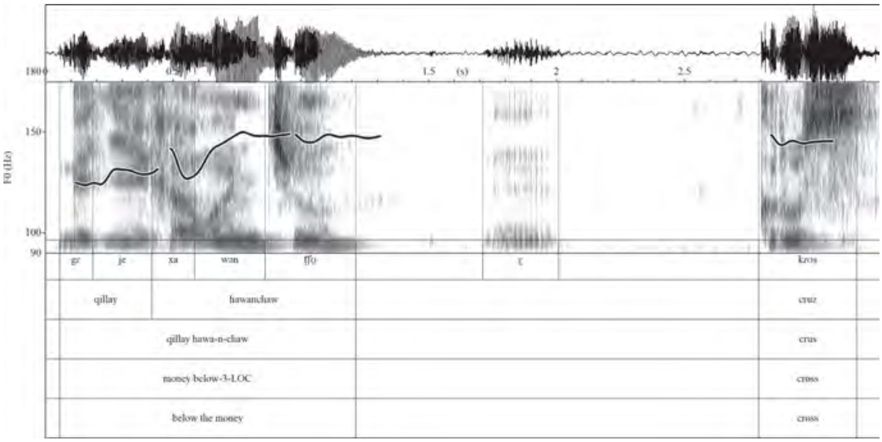


Figure 118: XQ33_Conc_Q_0690[260] (declarative with a rising and a high flat contour).

Examples like TP03_Conc_Q_1372, QF16_Conc_Q_1737, and XU31_MT_Q_1855 ((102)/ Figure 119, (103)/ Figure 120, (104)/Figure 121, respectively) show in turn that even the function words *tsay* and *kay*, the distal and proximal demonstratives, can realize a rising or falling pitch event, at least if they function as pronouns, expandeable by suffixes just like any other noun, instead of determiners.

(102) TP03_Conc_Q_1372

kanan	allawka	ladu-yki-pa	aja	tsay-chaw-mi
now	right	side-2-GEN	yes	DEM.DIST-LOC-ASS

"now, to your right side, yes, there"

(103) QF16_Conc_Q_1737

arash	kay-chaw
lizard	DEM.PROX-LOC

"[the] lizard, here"

260 https://osf.io/xe759/

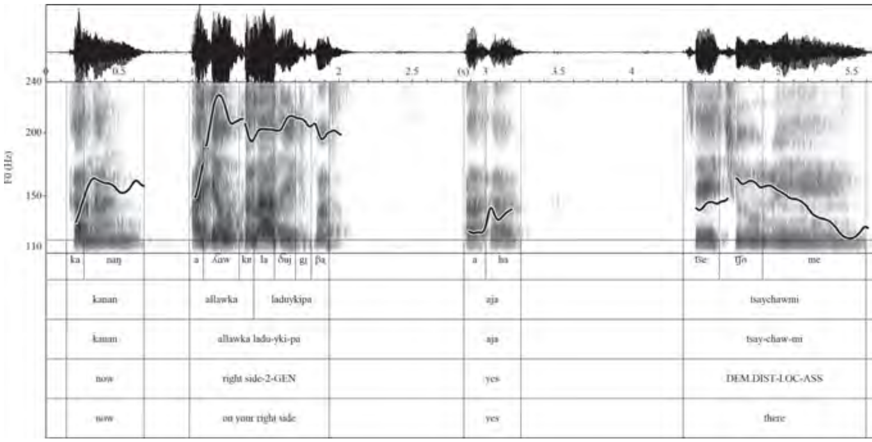


Figure 119: TP03_Conc_Q_1372[261] (declaratives consisting of rising and falling contours).

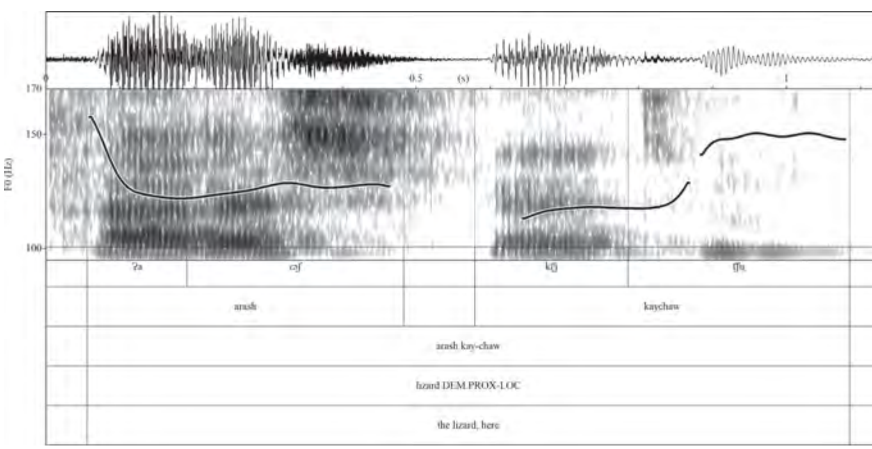


Figure 120: QF16_Conc_Q_1737[262] (declarative consisting of two rising contours).

(104) XU31_MT_Q_1856

tsay-pita	ultu-ta	rikaa-nki	yaku-pita	yarqa-mu-sha-ta
DEM.PROX-ABL	tadpole-OBJ	see-2	water-ABL	leave-CIS-PRTCP-OBJ

"then you see a tadpole coming out of the water"

261 https://osf.io/g3f56/
262 https://osf.io/bruj9/

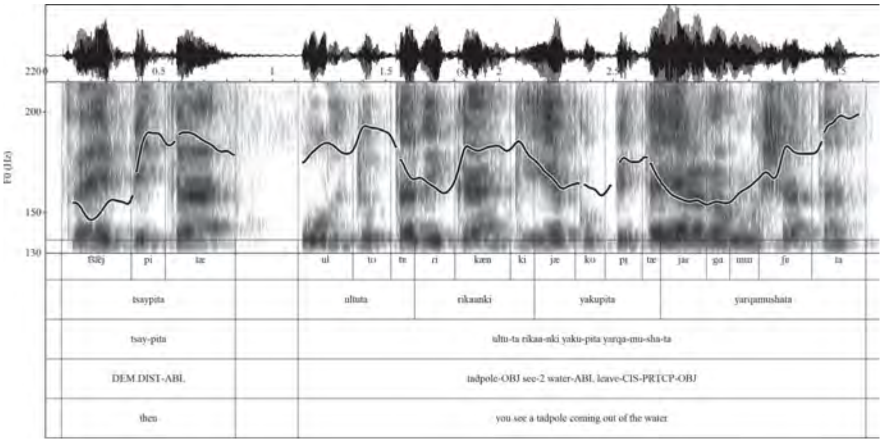


Figure 121: XU31_MT_Q_1855[263] (declarative consisting of five rising contours).

Examples like XQ33_Conc_Q_0531 ((105)/Figure 122), HA30_MT_Q_1241 ((106)/Figure 123), QF16_MT_Q_1348 (0/Figure 124), or HA30_Conc_Q_0211 ((108)/Figure 125) demonstrate that the low stretch following the fall in falling contours is not limited to only the final syllable of the phrase but can in fact encompass one or several full words (although such cases as HA30's MT_Q_1241, where it seems to cover the last 3–4 words, are rare). The fall also does not have to take place in the last word of the phrase.

(105) XQ33_Conc_Q_0531

quena	ladu-n-chaw	chuqllu-pa	aqtsayku-n
flute	side-3-LOC	corncob-GEN	corn.hair-3

"beside the flute is the hair of the corncob"

(106) HA30_MT_Q_1241

tsay	naani	hawa-n-chaw-na-m	huk	runa
DEM.DIST	path	below-3-LOC-DISC-ASS	one	person

ichira-ykaa-n	qillay-ni-n	haya-ku-shaa
stand-PROG-3	money-FON-3	carry-MID-PRTCP

"below the path there now a man is standing holding money"

263 https://osf.io/8tybp/

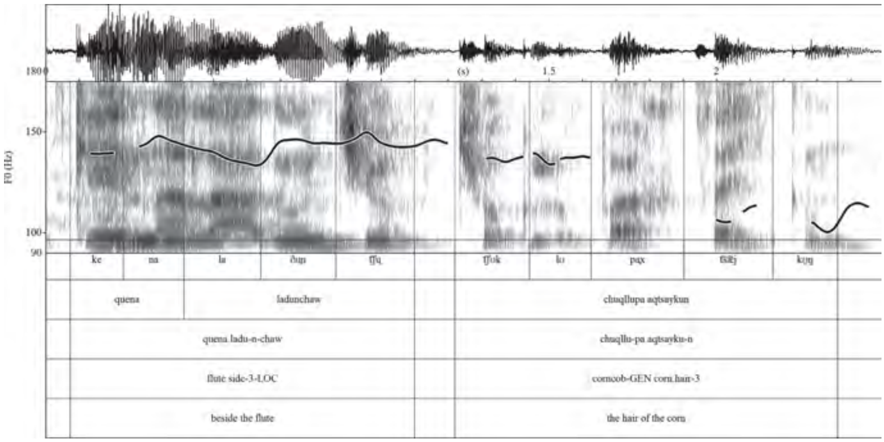


Figure 122: XQ33_Conc_Q_0531[264] (declarative with a rising and a falling contour).

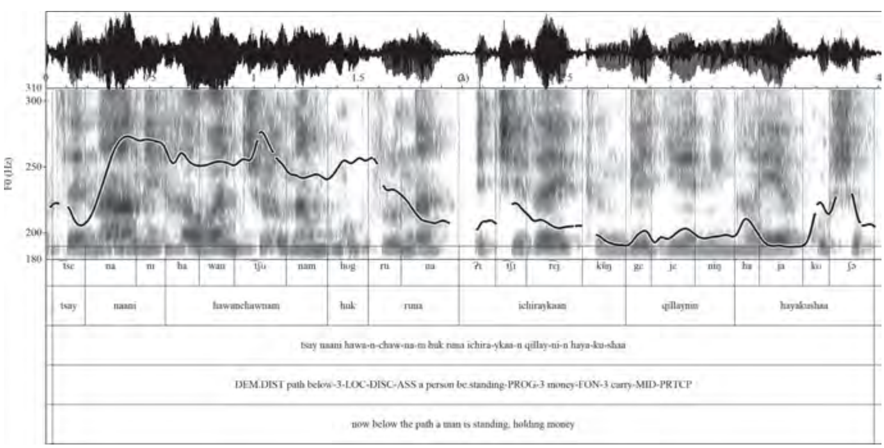


Figure 123: HA30_MT_Q_1241[265] (declarative with a rise-falling contour).

(107) QF16_MT_Q_1348

atuq	ka-q-man	chaa-tsi-y
fox	COP-AG-DEST	arrive-CAUS-INF

"let it get to the fox[266]"

---

264 https://osf.io/ujxny/

265 https://osf.io/dvc76/

266 This utterance clearly has imperative force. Prosodically, this is not expressed (imperatives having no prosodic form distinct from declaratives). Morphologically, this is here (and regularly)

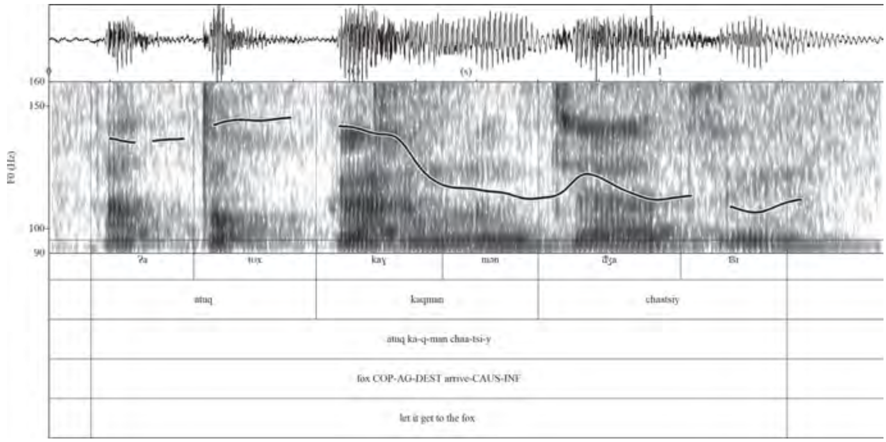


Figure 124: QF16_MT_Q_1348[267] (imperative consisting of a falling contour).

(108)	HA30_Conc_Q_0211			
	washa	kantu-chaw	ukush-mi	ka-ykaa-n
	DEM.DIST	edge-LOC	mouse-ASS	COP-PROG-3
	"at that edge there is a mouse"			

Equally, examples like (110)/Figure 127 (111)/Figure 128 (in the next section) show that the low stretch preceding the rise in contours that have it is not restricted to a single word.

From examples like (98), where the first *hirka hawampa* is uttered by the speaker who has the path on the map in the maptask, and the second by the one without the path, it could be thought that the falling contour corresponds to declaratives, and the rising to (polar) interrogatives. In fact, both types of contours occur on utterances where the context points to a declarative function as well as in those with an interrogative function. (99) already shows that a confirmation-seeking polar question, realized with a rising contour in (98), can also be realized with a

indicated by the morpheme –*y*, which is glossed as –INF, in accordance with the glossing conventions of Bendezú Araujo et al. (2019). This is because it is formally identical to the "infinitive" suffix –*y*. It can't be said whether these infinitive imperative constructions are therefore parallel to the use of infinitive constructions for directive functions in languages such as Spanish (*hablar de otras cosas* "(let's) speak of other things!") or German (*jetzt mal die Klappe halten* "shut your mouth now"), where there are additional other means of marking imperatives. In Quechua, –*y* is the only morpheme for positive imperatives; negative imperatives (vetatives) are formed with the special vetative particle *ama* sentence-initially together with the negation suffix –*tsu* attached to one of the words in the sentence.

267 https://osf.io/wrjpn/

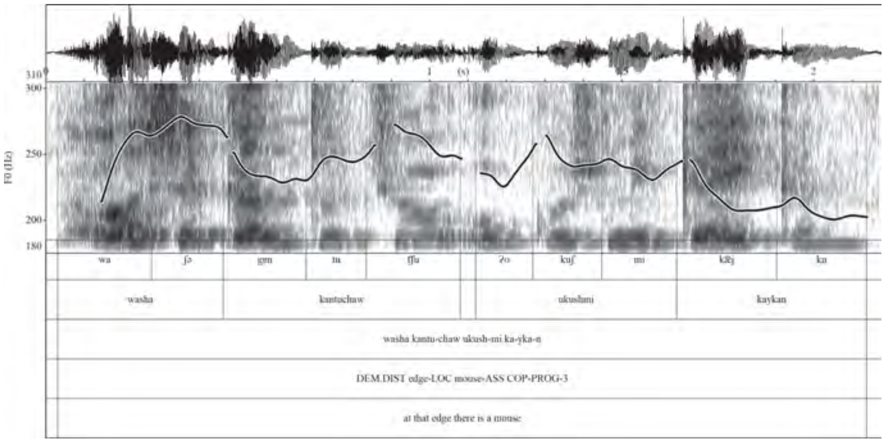


Figure 125: HA30_Conc_Q_0211[268] (declarative with two rising and one falling contours).

falling contour. Section 6.2.1 will delve a little further into the relation between contour type and interrogativity.

This section has shown that no lexical or morphological distinction has a categorical influence on whether a pitch event takes place on a word or not and which form it takes, and that a single tonal contour often encompasses several words, regardless of its shape (rising, falling, or rising-falling). A further general observation is that in many utterances, pitch contours consist of relatively flat and long stretches sometimes extending over more than a single word, whether pitch is high overall (plateau-like realization), or low. The process generating these pitch contours is clearly phrase-based, not word-based. The functional differentiation of the contours will be discussed in detail in section 6.2, where it will be shown that rising contours can broadly be characterized as continuing or incomplete and falling ones as closing or complete (6.2.2). It will also be argued there (6.2.3) that it is the prominence and position of the words in a phrase that determine the location of the low and high stretches in it. In the following sections, I will provide a description of the alignment patterns in these contours regarding possible locations for the rise and fall in multi- and single-word phrases, using examples from the tasks *Conc*, *Maptask* and *Cuento*.

268 https://osf.io/m4ec2/

6.1.2 Alignment of the tonal transition with the word boundary

(109) KP04_Conc_Q_1532

tsay	tikra-nqa-yki	ka-q	tsuqllu
DEM.DIST	turn-NMLZ-2	COP-AG	corn

"the one you turned around [is] corn"[269]

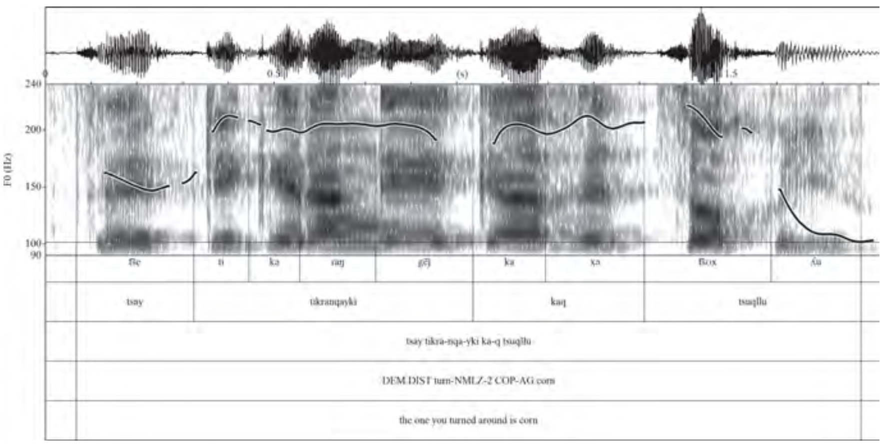


Figure 126: KP04_Conc_Q_1532[270] (declarative consisting of a rise-falling contour).

In KP04's Conc_Q_1532, given in (109) and Figure 126, there is a single large rise-falling contour encompassing the whole utterance, consisting of an initial low stretch, a rise, a high plateau, a fall, and a final low stretch. The initial low stretch is realized on the determiner *tsay*. The largest part of the following increase in pitch height is already accomplished at the beginning of the vowel segment of the initial syllable of the second word *tikranqayki*, with just the last bit of it realized as an actual

269 Note that the verb of being *ka-* is not used in Quechua in the third person present (*ka-n*) when expressing a copular relation. Instead, no verb at all is used. This only holds for the third person present; forms of *ka-* are used for all other persons and tenses. In its third person present form, *ka-* is only used to express existential meaning (*allqu kan* "there is a dog", not "x is a dog"), cf. Parker (1976: 68). The form *ka-q* used here, with the agentive suffix *–q* is not fulfilling a copular function here between *tikranqayki* "what you turn around" and *tsuqllu* "corn", instead it forms part of the complex noun phrase *tsay [[tikranqayki] kaq]*, where *tsay* is a determiner and the contribution of *kaq* has been translated in English with the construction "the one. . .". For further information on *kaq* and other aspects of Huari Quechua syntax and information structure, see Bendezú Araujo (2021).
270 https://osf.io/bgxzj/

rise. Note that the syllable *ti* is neither the penult of the entire phrase nor that of the word. Pitch then basically remains eventlessly high over the course of the two words *tikranqayki kaq* until the initial /penultimate syllable of the final word *tsuqllu*, during which it begins to fall. Again the largest part of the pitch transition takes place over voiceless segments and by the time the vowel segment of the final syllable of the word and the entire utterance is reached, it has decreased substantially and then remains low. This utterance is exemplary for an alignment pattern in which the initial rise (both of the rising-falling and the rising contour) takes place not on a particular syllable in one of the words on which the contour is realized, but at a boundary between those words. I will call this the word boundary pattern. KP04's Conc_Q_1576 ((110)/Figure 127) shows that the initial low stretch can encompass more than one word also in this pattern.

(110) KP04_Conc_Q_1576

tsay	kay	hawa-n	ka-q	wanupakush
DEM.DIST	DEM.PROX	below-3	COP-AG	burial

"that one that's below this [is the] burial"

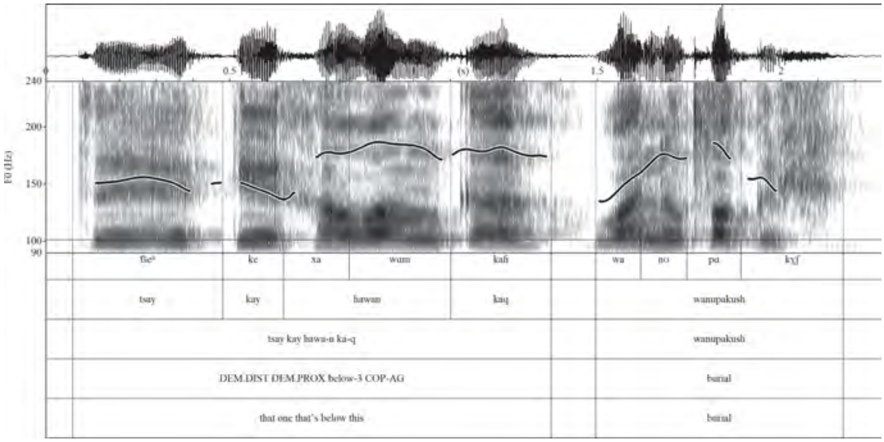


Figure 127: KP04_Conc_Q_1576[271] (declarative with a rising and a rise-falling contour).

There and in LC34's Conc_Q_0110 ((111)/Figure 128), the word boundary pattern occurs on rising contours.

271 https://osf.io/yxzuw/

(111) LC34_Conc_Q_0110

primer	qalla-na-n	ka-q-chaw	ashkash
first	begin-DISC[272]-3	COP-AG-LOC	lamb

“now the one starting the first one is the lamb”

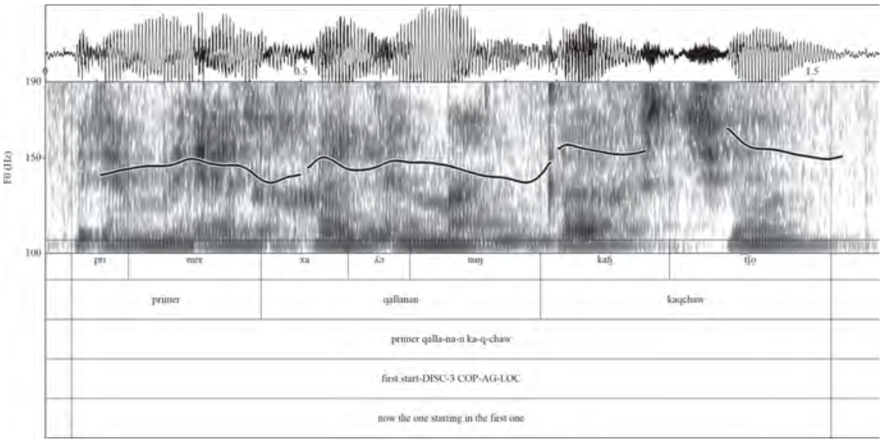


Figure 128: LC34_Conc_Q_0110[273] (part of a declarative with a rising contour).

The fall in falling contours can also be aligned with a word boundary, as Figure 125 above shows. However, often the falling pitch event stretches over both the penult of a word and a word boundary, so that it is difficult to determine which landmark it actually aligns with (cf. Figures 122 and 123 above). The alignment pattern showcased in this section leads to the conclusion that a phonological phrase in the Huari Quechua data can contain several prosodic words. The level of the phrase determines the extent of the pitch contour, and, under this alignment pattern, the boundaries of the prosodic words are the landmarks that tonal alignment refers to. However, this is only one option attested in these data.

272 The suffix *-na* glossed here as DISC(ontinuative) is sometimes used to indicate something like a change of topic, which the translation tries to approximate by the use of “now” as a sentence adverbial. In other cases its use can also be translated by “already” (cf. Parker 1976: 146).
273 https://osf.io/9mbu3/

6.1.3 Alignment with the phrase boundary and phrase-final devoicing

Regarding tonal alignment with the phrase boundary, all contours seen so far show evidence for it in an uncontroversial way: the left-peripheral tones (the first L tone in rising and rising-falling contours, the H tone in only-falling contours) must be aligned with the left edge of the phonological phrase; the right-peripheral tones (the second L tone in only-falling and rising-falling contours, the H tone in rising contours) with the right. Here I'll show that also the point of tonal transition, so to speak the "inner" alignment, refers to the right phrase edge in some of the contours. Potential cases in point are abundant: in utterances with rising contours like (97)/(Figure 114), (101)/(Figure 118), or (103)/Figure 120, the observed pitch movement of a rise taking place on or after the penult of the last word in the phrase could well result from an alignment of both the L and the H tone to the right phrase boundary and tonal crowding ensuring that the L tone only reaches into the penult because the H tone already occupies the rightmost TBU. Similarly, in falling contours, the high pitch extending until the penult of the final word, like in (109)/Figure 126, (110)/Figure 127 (rising-falling), or (112)/Figure 129 (only-falling), with the fall taking place on or after it can be analyzed by the L tone occupying the final TBU in the phrase and the H the preceding one, both seeking to align as rightwards in the phrase as possible.

(112) ZR29_Conc_Q_1661

ya	pitu	ladu-n-chaw	uusha
ok	flute	side-3-LOC	sheep

"ok beside the whistle [is] the sheep"

The obvious alternative analysis is that the H tone in both the rising and falling contours is aligned with the penult of the (final) word in these cases. Under that analysis, the fact that the (phrase-final) penult is roughly where the tonal transition takes place would be because it serves as H tone association or alignment site by virtue of being stressed (**b** in Figure 130), as opposed to because this is what happens when two tones in linear succession seek to be realized as rightmost as possible with the one to the right occupying the rightmost TBU and pushing the other TBU to the left (**a** in Figure 130).

In the next section (6.1.4), evidence will be provided that something like analysis **b** (also applied to rising and only-falling contours) must be right for some of the utterances in the data, whereas in this section I will proceed to show that for others, only **a** can hold.

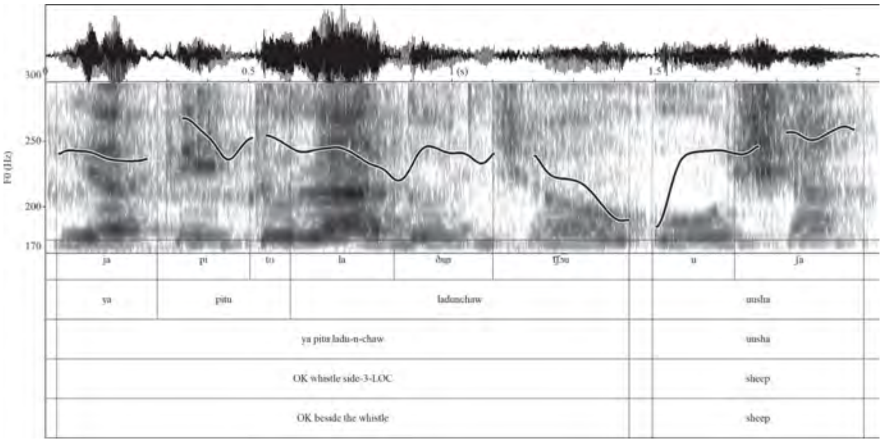


Figure 129: ZR29_Conc_Q_1661[274] (declarative with a falling and a rising contour).

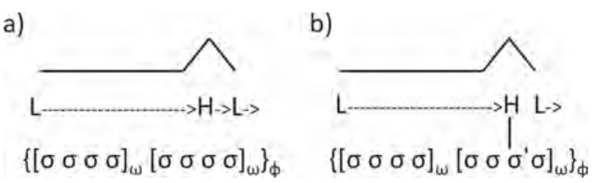


Figure 130: Two possible analyses for schematized rising-falling contours.

In this Quechua variety, vowels in phrase- and especially utterance-final syllables are sometimes devoiced, as well as high vowels following voiceless fricatives or plosives (similar to what Delforge 2011 describes for Cuzco Quechua). When domain-final devoicing takes place, the phrase-final behaviour of the tonal contour is frequently not cut short, but instead occurs earlier, further "to the left". This occurs in TP03_Cuent_Q_1663 ((113)/Figure 131) and OA32_Cuent_Q_1222 ((114)/Figure 132).

274 https://osf.io/8um7v/

(113) TP03_Cuent_Q_1663

maa	cuenta-ri-shayki	huk	cuento-ta
let's.see	tell-ITER-1.SUB>2.OBJ.FUT	one	story-OBJ

cuenta-yaa-maa-nqa-n-ta
tell-PL-1.OBJ-NMLZ-3-OBJ
"let's see, I'm going to tell you a story they told[275] me"

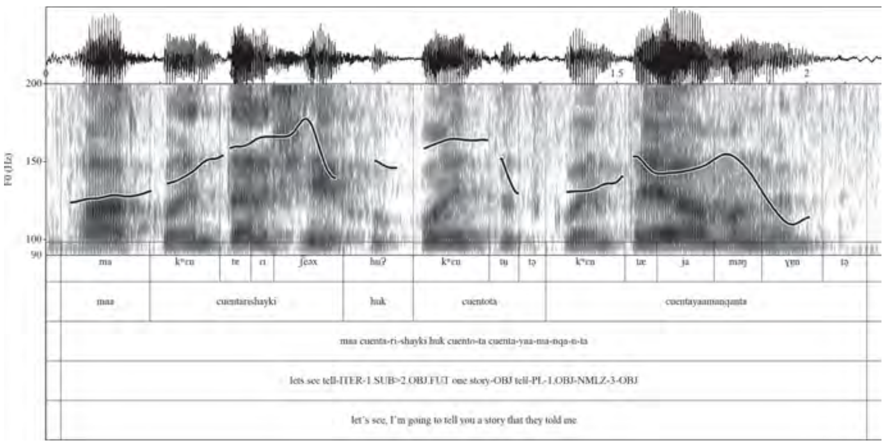


Figure 131: TP03_Cuent_Q_1663[276] (declarative consisting of three falling contours).

(114) OA32_Cuent_Q_1222

tsay	mosca-ta	miku-ski-r-ni-n-qa	imanuy-taa
DEM.DIST	fly-OBJ	eat-ITER-SUBID-FON-3-TOP	how-DETVAR

siente-ku-rqa-n-qa
feel-MID-PST-3-TOP
"when [he] ate the flies, how did [he] feel?"

275 What has been translated as a relative clause here is a tenseless nominalization with –*nqa* (homophonous, but not identical, to the marker of the third person future –*nqa*), usable in both perfective and imperfect contexts. The verb form nominalized with –*nqa* usually does not bear tense markers (cf. Parker 1976: 172).
276 https://osf.io/ka8zq/

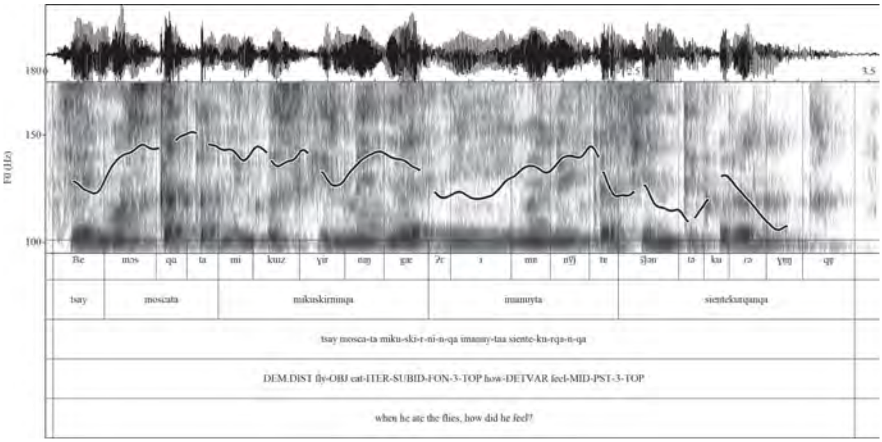


Figure 132: OA32_Cuent_Q_1222[277] (wh-question with two rising and two falling contours).

In TP03_Cuent_Q_1663's *cuentayaamanqanta*, realized as [kʷɛn.tæ.ja.məŋ.ɣḛn.tə̥], the last vowel (of the object suffix *–ta*) is produced without voicing. Yet there is no pitch peak or end of the plateau on the penult [ɣḛn][278] of the word, with the following fall cut short, as would be expected under an account where a H tone is associated or aligned with a supposedly prominent word penult. Instead, the peak is realized one syllable earlier, "displaced", as it were, onto the antepenult [məŋ], and followed by a full realization of the fall with a low elbow reached on the penult. Similarly, in OA32_Cuent_Q_1222, *sientekurqanqa*, realized as [sʲən.tə.kʉ.ɾə.ɣḛŋ.qɐ̥], with an epenthetic schwa inserted in the cluster /rq/, the peak occurs not on the word penult [ɣḛŋ], but on either [kʉ] or [ɾə], the preceding syllable(s), while it is the elbow of the following fall that is realized on the penult – followed by a (nearly) voiceless final syllable. A third example is QF16_MT_Q_1282 ((115)/Figure 133), a single-word utterance realized with an only-falling contour. The last two syllables are voiceless (the penult realized with very creaky voice, the final syllable completely devoiced), and the fall takes place on the preantepenult of the word, with the antepenult, the last voiced syllable, realized low.

277 https://osf.io/pw2ad/

278 The breathy voice diacritics [̤] here and elsewhere indicate a vowel produced audibly with increased breathy air flow and reduced energy in the formant spectra. As here, a breathy vowel often occurs in a syllable preceding a fully unvoiced one, suggesting that both are due to a gradually increasing domain-final exhalation process.

(115) QF16_MT_Q_1282
hawa-n-pa-taaku
below-3-GEN-NEG
"not below it"

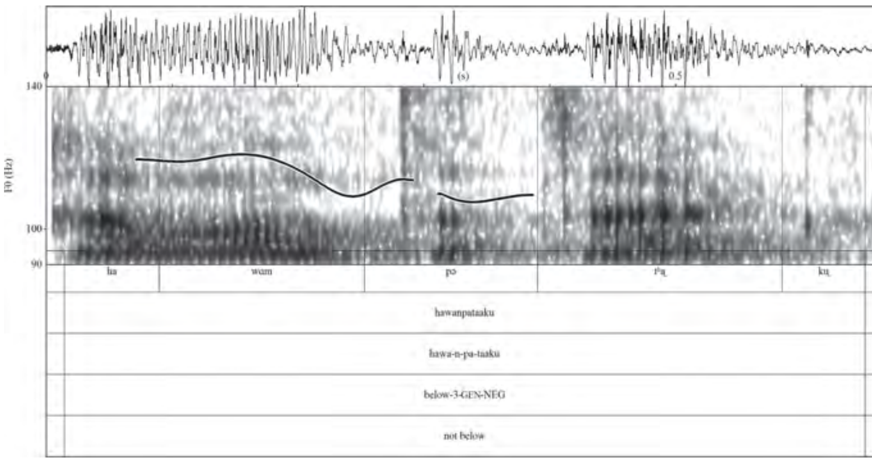


Figure 133: QF16_MT_Q_1282[279] (declarative consisting of a falling contour).

Such "displacement" is unexpected under an account where the penult is stressed at the level of the prosodic word, with the H tone as a post-lexical pitch accent anchored to it. In such a scenario, the final syllable being unvoiced should not affect the anchoring of the H tone to the supposedly prominent penult because the necessary condition for tonal association, namely the status of the penult as stressed, is determined previously to, or at least independently of, the unvoicing. Furthermore, the penult itself as stressed syllable should not be able to be devoiced at all, since deletion and other weakening processes in languages with stress usually target all other material instead of the stressed position. In Huari Spanish, a similar devoicing process occurs in the speech of some speakers, but there, affected pitch movement unambiguously shows the behaviour of pitch accents associated with stressed syllables, i.e. it is either less affected than on unstressed material or the associated pitch movement is deleted instead of displaced. See e.g. the utterance-final part of ZZ24's ELQUD_ES_26 (Figure 134). There, the utterance-final syllable [kḁ] of *roca* "rock" is also devoiced, but the preceding pitch peak due to the LH* pitch accent on the stressed syllable [ɾɔ] takes place on that syllable and is not "displaced" to

279 https://osf.io/3c2xa/

the left. On *encima*, the stressed vowel in [si̥] is devoiced due to the presence of the voiceless fricative.[280] Here it is clearly visible from the remaining pitch contour that the rising movement due to the LH* pitch accent on that syllable is removed, instead of displaced.

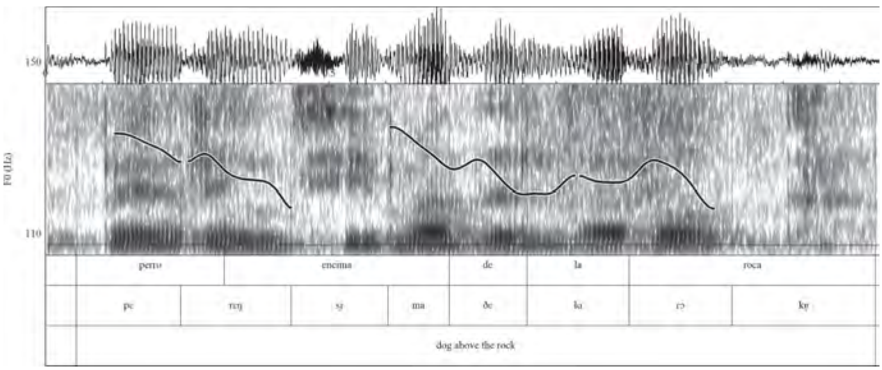


Figure 134: (part of) ZZ24_ELQUD_ES_26[281] (*perro encima de la roca* ‘dog above the rock’).

The displacement is best explained via an account which makes no reference to a prominent position determined at the word level (**a** in Figure 130), and where instead alignment of tones with TBUs happens concurrently with devoicing: phrasal tones seek alignment on the rightmost available TBUs. Those seem to be vowel nuclei with crowding disallowed. Then, the processes of final devoicing and tone alignment can go in tandem: unvoicing makes a syllable unavailable as a TBU. Consequently, rightmost alignment will then align the two right tones with the penult and the antepenult, respectively. What looks like a “displaced accent” thus simply falls out from that. Whether the unvoicing or the tone assignment happens “first” is hard to determine. However, there are also occasionally some examples like HA30_MT_Q_2044 ((116)/Figure 135), where the peak/end of the plateau and fall take place earlier even though no devoicing happens on the last syllable(s).

280 Deletion of high stressed vowels due to voiceless context does not occur frequently in the Huari Spanish data. In languages without word stress, deletion processes more often affect positions anchoring high tones: in Japanese e.g., syllables with lexical pitch accent (cf. section 3.3.2) are regularly devoiced if they have a high vowel and occur in the appropriate phonological context (Venditti et al. 2008: 480–481).
281 https://osf.io/5sfym/

(116) HA30_MT_Q_2044

y	tsay	pitu-chaw	pitu	hita-raa-ykaa-nqa-n-pita
and	DEM.DIST	whistle-LOC	whistle	throw-ITER-PROG-NMLZ-ABL

"and in that whistle, from where the whistle is lying"

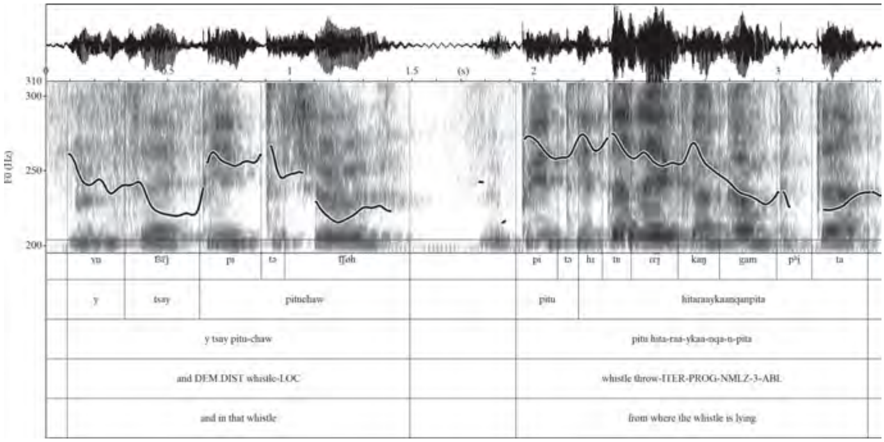


Figure 135: HA30_MT_Q_2044[282] (parts of a declarative with two falling contours).

The second phrase *pitu hitaraaykaanqanpita*, realized as [pi.tə.hɪ.tɐ.rɛ͡j.kaŋ.gam.pʰi̤.ta], is realized with an only-falling contour beginning with a high plateau that continues evenly (disregarding consonantal microprosody) until the syllable [kaŋ], on which a fall begins that reaches its low elbow in the next syllable [gam]. Here, [kaŋ] and [gam] are the preantepenultimate and the antepenultimate syllable in the phrase, respectively. A possible reason for this even earlier realization of the fall might be found in the penult, which is almost completely devoiced. It might be that the voicelessness of the penult causes the alignment of the two phrase-final tones to be realized on the two rightmost contiguous TBUs, even though the final syllable is again voiced and produced with considerable energy. An account using the word penult as the prominent position to which the high tone is associated would be at a total loss here.[283] For utterances like the ones shown in this section,

282 https://osf.io/wj9cv/

283 One theoretical alternative would be to assume fixed but varying stress positions for different lexical items, perhaps in a syllable window of some size, as in Spanish. However, this cannot be the case, since on the one hand, variation in final peak alignment occurs even in the same lexical item across different utterances. On the other, as we have seen in these examples, Quechua words can be quite long and morphologically complex, with the amount of information conveyed by one of the longer ones perhaps closer to that of a sentence in a European language than that of a single word.

analysis **a** is clearly the only option. Further evidence against analysis **b** consists in cases where no final vowel devoicing occurs, but where instead a medial vowel is entirely eliminated, leading to resyllabification with the penult itself sometimes reduced. This happens in OA32's MT_Q_2915 ((117)/Figure 136), where *pinkullutam*, produced with a falling contour, is realized as [piŋ.kʊl.tam], with the peak before the fall occurring on [kʊl], which is the penult *after* reduction, but at the cost of completely eliminating the vowel of what would have been the penult without reduction, the syllable /ʎu/ of /pin.ku.ʎu.tam/.

(117) OA32_MT_Q_2915
pinkullu-ta-m rikaa-:
flute-OBJ-ASS see-1
"I see [a] flute"

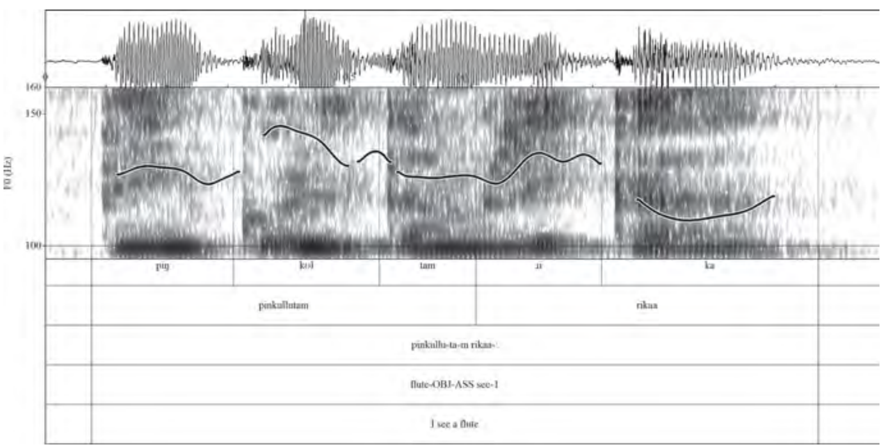


Figure 136: OA32_MT_Q_2915[284] (declarative consisting of two falling contours).

This syllable in this lexical item is not regularly reduced (cf. e.g. the realizations given in Figure 142), but there is no trace of the high back vowel following the

To suggest that one particular instantiation of a given lexical root plus a certain chain of suffixes has a lexically fixed stress on the penult, another on the antepenult and yet another on the preantepenult seems exceedingly implausible, especially given that these word forms are the result of productive formation, and the rules for combining roots and suffixes allow a vast array of possible combinations, so that each individual combination of such length will have quite a low frequency of occurrence, making a lexicalized stress position for each of them improbable. Such a hypothesis has also never been put forth for any Quechua variety.

284 https://osf.io/dj8us/

lateral here at all. This is incompatible with an account of word stress on the penult tweaking all acoustic parameters to cue this prominence, reducing atonic and preserving tonic syllables. Clearly a better hypothesis is that the reduction process is independent of word stress, with the penult just as available to be reduced as other syllables. Tonal alignment also does not make reference to stress and is affected by the segmental chain merely with regards to what qualifies as a TBU.

Applying the difference between the two analyses to rising contours yields an interesting prediction. As Figure 137 shows, under the analysis which only assumes phrase edge-alignment (**a**), the penult of the final word is expected to be low and the rise to occur on or after it, because the H tone only occupies the rightmost TBU. Under an analysis assuming alignment of the H tone with the stressed penult there, it is expected to be high and the rise to begin earlier.

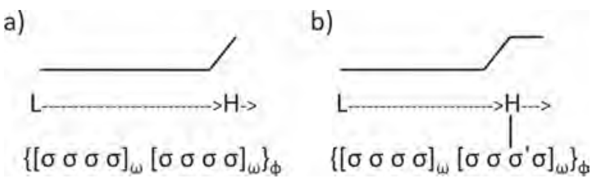


Figure 137: Two possible analyses for schematized rising contours.

This difference is borne out by observation: utterances like (101)/Figure 118 realize a contour clearly more compatible with **b**, with the penult of the phrase-final word *hawanchaw* mostly high already. On the other hand, in utterances like QF16_Conc_Q_1205 ((118)/Figure 138), the penult on the phrase-final word *wayinchaw* is clearly realized with a low stretch, and only the final syllable is high.

(118) QF16_Conc_Q_1205
runa wayi-n-chaw
person house-3-LOC
"a person in their house"

For utterances like Figure 138, analysis **a** is clearly the more adequate option, demonstrating that a subset of the rising contours here exhibits a purely edge-based alignment. Yet as just seen, analysis **b** must also hold for a different subset, in which the (final) word penult is a relevant anchor for tonal alignment. The distinction between these two variants is also upheld when considering phrases realized on individual words (see section 6.1.5). In the next section we will first consider further contours for which analysis **b** is more appropriate.

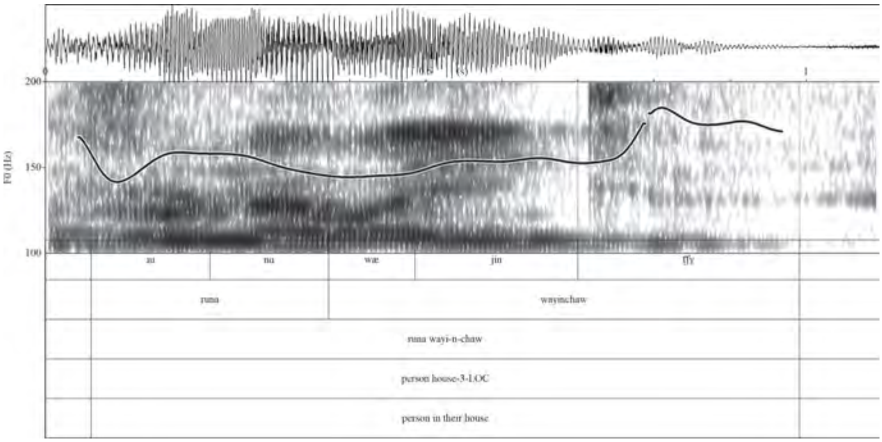


Figure 138: QF16_Conc_Q_1205[285] (two rising contours realized on part of a declarative).

6.1.4 Alignment with the word penult

The utterances discussed in this section, together with the arguments made just above, will present further evidence for the relevance of the penult of an individual word as a position with which the rise or the fall in the rising or falling contours is aligned. This solidifies the observation that in the Huari Quechua data overall, tonal alignment is possible with the edge of the phonological phrase, the edge of the prosodic word, and a regular word stress position. At least two of these alignment patterns are clearly in competition with each other, resulting in variant realiations.

In some utterances the rise in the rising contour occurs on the penult of a pre-final word in a phrase, with the following high plateau extending until the end of the phrase, cf. *aptasha kaykaan* in XU31_Conc_Q_0615 ((119)/Figure 139), or *akakakuna niyaashqa* in XQ33_Cuent_Q_1302, ((120)/Figure 140).

(119) XU31_Conc_Q_0615

runa	bolsa-n	apta-sha	ka-ykaa-n	escoba-pa	ladu-n-chaw
person	bag-3	carry-PRTCP	COP-PROG-3	broom-GEN	side-3-LOC

"a person is carrying a bag, besides the broom"

285 https://osf.io/z7x3s/

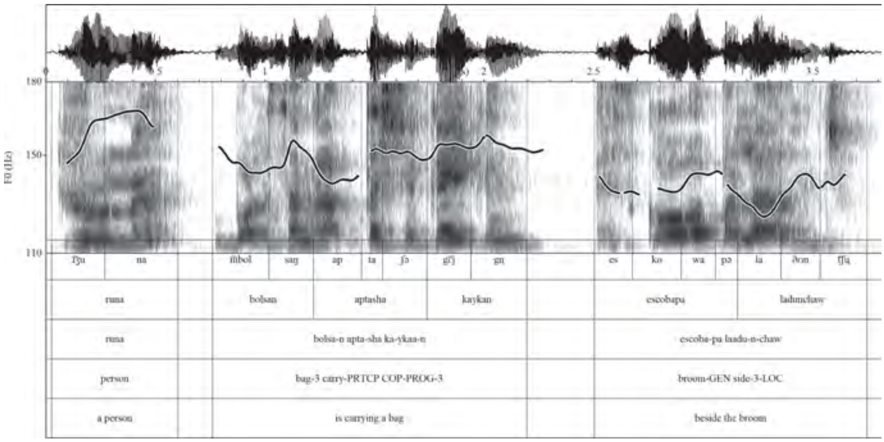


Figure 139: XU31_Conc_Q_0615[286] (declarative consisting of 5 rising contours).

(120) XQ33_Cuent_Q_1302

hampi-ku-q-mi	aywa-ykaa-:	akaka-kuna	ni-yaa-shqa
heal-MID-AG-ASS	go-PROG-1	woodpecker-PL	say-PL-PST

"'I am going to get myself healed'; [and] the woodpeckers said"

These contours are less frequent than their counterparts where the low stretch extends until the penult of the final word in a phrase.[287] Yet they do occur, and they unambiguously require an analysis in which the word penult is a relevant position for tonal alignment in a phrase. The same can be applied to falling contours with the fall occurring on pre-final words. Some of them, like (108)/Figure 125, realize the fall between two words, using the word boundary as alignment landmark, while others like (107)/ Figure 124 realize it after the penult on a pre-final word. This latter realization is an instance of alignment with the word penult in falling contours, counterparts to the rises just discussed. Note that both phrases realizing the tonal transition on a prefinal word and those realizing it on a final word share the property that only a single tonal transition takes place in the phrase. That is to say, it is only the word penult of one word in the phrase that serves as tonal alignment anchor. It seems logical to assume that this is the word that is most prominent, i.e. only σ' in ω' can serve as alignment anchor, while all other σ's of mere ωs in ϕ are ignored by tone alignment.

286 https://osf.io/q83es/

287 In section 6.2.3.3, they will be argued to cue a marked metrical structure with prominence on a prefinal word in the phrase.

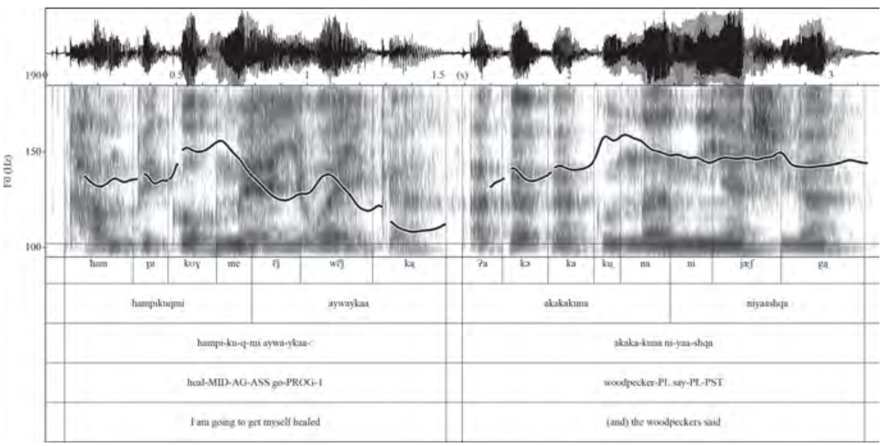


Figure 140: XQ33_Cuent_Q_1302[288] (two declaratives with one rising and one falling, and a rising contour, respectively).

In this context it should be noted that the only observable effect of this word-level prominence is precisely the possibility for anchoring the tonal transition, and nothing else. As Buchholz & Reich (2018) showed, the word penult does not consistently attract higher pitch or longer duration, and several already seen examples demonstrate that vowel quality is also not more peripheral, nor spectral energy necessarily higher in that position either: e.g. Figure 140, where the syllable [kṳ] on which the rise takes place is shorter than the surrounding ones and has less spectral energy, or Figure 118 where the vowel of [wən] on which the rise takes place is more central than on surrounding syllables. Thus even in this alignment pattern where the tonal grammar makes reference to the metrical representation, the prominence on the penult is quite abstract. Its effects are certainly quite different from what can be observed in "typical" word stress languages like English, where many more phonological processes cannot be adequately described without making reference to it (cf. Hyman 2014, section 3.3.2). In this respect, Quechua certainly bears more resemblance to Japanese, where vowels bearing a lexical pitch accent are regularly devoiced when they are high (cf. Venditti et al. 2008: 480–481; Beckman 1986). But following Hyman (2014), it would be misleading to typologize Quechua as categorically different, e.g. by labeling it a "pitch-accent", a "stressless" or an "edge-prominent" language. By adopting a "properties-based typology" (Hyman 2018) interested in whether more fine-grained individual properties are shared or not, the more informative assessment can be made that Huari Quechua

288 https://osf.io/g3x84/

utterances with this alignment pattern share only one out of 8 features with prototypical "stress-accent languages", namely that "[l]exical stress provides the designated terminal elements for the assignment of intonational tones ('pitch-accents')" (Hyman 2014: 58, cf. (10) in 3.3.2). Even this does not hold for utterances where tones seek alignment only with phrase or word boundaries. There is thus evidence for different variants in the intonational grammar available to the speakers providing the Huari Quechua data here. They differ precisely with regards to whether the information about a prominent location at the word level is relevant to their description or not. Just as Huari Spanish, the Huari Quechua data does not behave uniformly with respect to this typology.

We have now seen evidence that in a multi-word phrase, there can be three different kinds of landmarks serving as anchoring points for tonal transitions: the phrase-peripheral domain boundaries (in all variants), a phrase-internal boundary of the next lower domain, the prosodic word (in one alignment variant), and a regular prominent position, the penult of the strongest prosodic word in the phrase (in another alignment variant). Between the alignment variants, there is no reason to assume a differing tonal sequence or inventory going beyond LH for rising, HL for only-falling, and LHL for rising-falling contours. What differs between them is only where the transitions take place.

6.1.5 Contours on single-word phrases

In this section, the different intonational patterns as attested on individually phrased words will be discussed. Individually phrased here means that they were surrounded by (short) silent breaks on both sides and exhibited a recognizable complete tonal movement. That is not to say that individually phrased words are always surrounded by breaks, quite to the contrary. In section 6.2.3, only the presence of a recognizable tonal contour will be taken as sufficient evidence of phrasing, but in order to firmly establish this, the break criterion was used here to avoid ambiguities (in section 6.1.6 it is also used for operationalization).

Figures 141–146 provide examples of individually phrased Quechua words of increasing length, from two syllables (Figure 141), over three syllables (Figure 142) and four syllables (Figure 143 for rises and Figure 144 falls) to five or more syllables (Figure 145 for rises and Figure 146 for falls). The selection is somewhat shaped by necessity especially in the range of five syllables and more: those words are rarer than shorter ones, and individually phrased words by themselves are infrequent. Nevertheless, I believe they represent the range of attested intonational variation in the corpora, with the exception of the optional patterns particular to Spanish loanwords (see section 6.1.8). The following observations can be made.

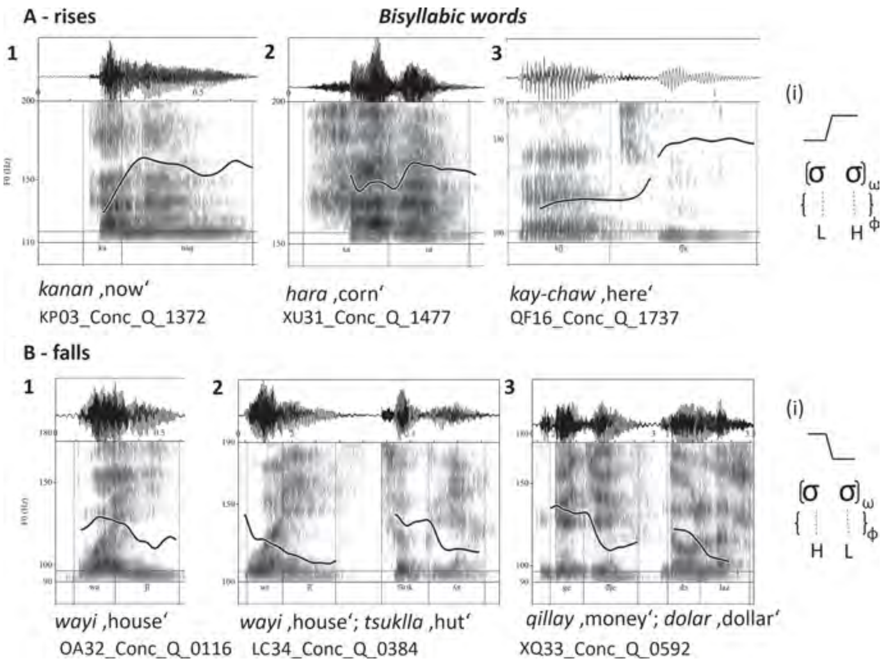


Figure 141: Individually phrased bisyllabic Quechua words. Row A 1–3:[289] words with a rising contour. Row B 1–3:[290] words with a falling contour. A (i) and B (i) are schematized tonal representations of bisyllabic rising and falling phrases, respectively.

Variability increases with syllable count: in phrases consisting of bisyllabic words, all rising contours consist of an L tone aligned with the first syllable and an H tone aligned with the second one and all falls are their mirror image in this respect, with the penult high and the final syllable low. This fact is reflected in there being only two schematic tonal representations, one for rises (A(i)) and one for falls (B(i)), for words of this size in Figure 141. This suggests that tonal alignment with phrase edges trumps alignment with the word penult under pressure of crowding; otherwise, we could expect to see high penults/initial syllables in some of the rising contours. It also suggests that rising-falling contours are not separately identifiable in phrases of this length. More variation is introduced beginning with phrases consisting of trisyllabic words. There are now two patterns for rises and falls each: in rising contours, the difference consists in whether the rise takes place before or after the penult, with

289 A1 https://osf.io/yuztx/; A2 https://osf.io/jb68t/; A3 https://osf.io/zx4dp/
290 B1 https://osf.io/uks6e/; B2 https://osf.io/f4ycp/; B3 https://osf.io/2z5wq/

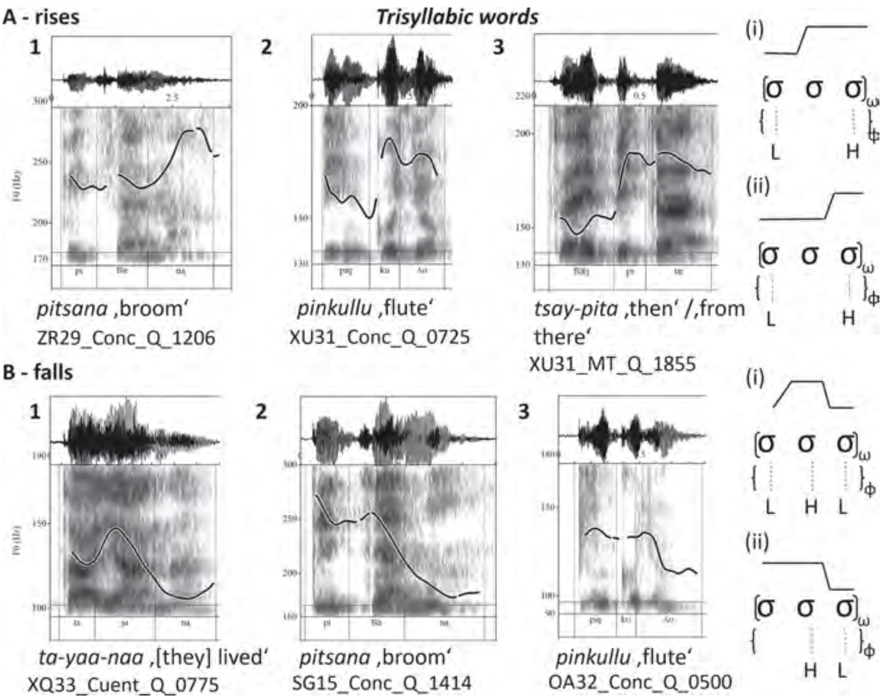


Figure 142: Individually phrased trisyllabic Quechua words. Row A 1–3: words with a rising contour, A1[291] with the rise only on the final syllable (schematic representation in A(ii)), A2–3 with the rise on the penult and maintained height on final syllable (schematic representation in A(i)). A1–2 are monomorphemic, A3 is bimorphemic. Row B 1–3: words with a falling contour, B1[292] with the initial syllable low and a rise to the penult (schematic representation in B (i)), B2–3 with both initial syllable and penult at a high level (schematic representation in B (ii)). B1 consists of three morphemes, B2–3 are monomorphemic.

the final syllable being always high and the penult thus varying between high or low, in parallel to what was observed in multi-word phrases. Pattern A(i) in Figure 142, with the penult high, is the more frequent one in the corpora considered here. In falling contours, the difference consists in that between rising-falling contours, where the contours start low, realizing a rise before the fall (B(i)), and only-falling contours, which already begin high (B(ii)). In both patterns, the penult is high and the final syllable low. The only-falling contour seems to be somewhat more frequent than the rise-falling contour for individually phrased trisyllabic words.[293] With an

291 A1 https://osf.io/cqr9m/; A2 https://osf.io/u9ntz/; A3 https://osf.io/f98ph/
292 B1 https://osf.io/fsd43/; B2 https://osf.io/atdur/; B3 https://osf.io/cjf3s/
293 This relative frequency is reversed in phrases with more syllables, cf. section 6.2.3.2.

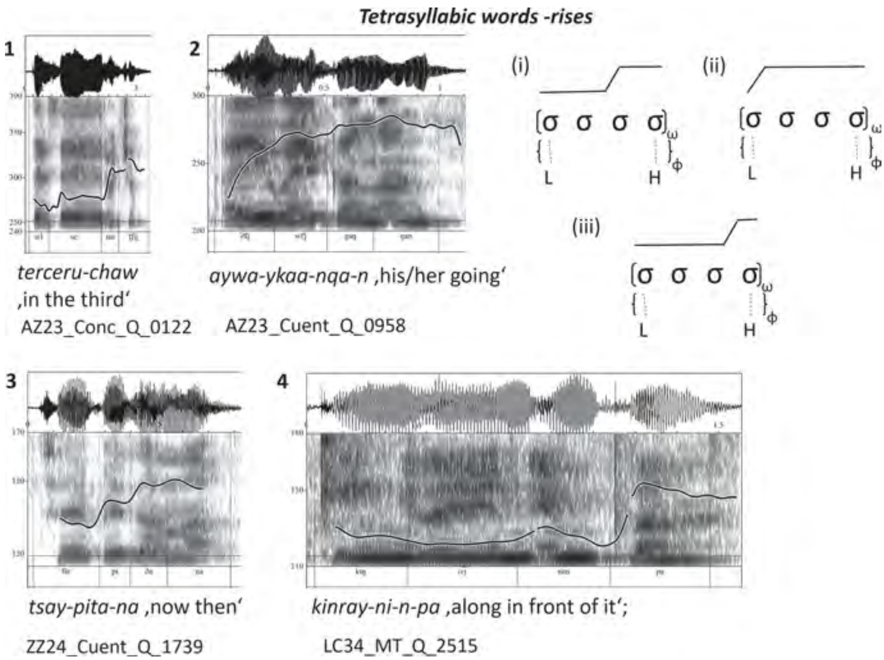


Figure 143: Individually phrased tetrasyllabic Quechua words realized with a rising contour.[294] 1: word beginning with a low stretch, with a rise and peak in the penult followed by high plateau maintained in the final syllable (schematic representation in (i)); 2: word beginning low, with most of the rise taking place already in the initial syllable, high plateau maintained until the end of the word/phrase (schematic representation in (ii)); 4: word beginning with a low stretch, with a rise and peak only in the final syllable (schematic representation in (iii)); 3: word where the rise begins already in the antepenult and continues into the final syllable (intermediate between all three representations). All words are multimorphemic.[295]

additional syllable, yet a further pattern each for the rises and the falls is attested. In individually phrased words of 4 syllables with a rising contour, there is now also a pattern in which (most of) the rise already takes place in the first syllable, and the rest of the word is realized with a high plateau ((ii) in Figure 143). There also seem to be intermediate realizations, with the rise taking place over several syllables. It was harder to find good examples of this early-rising pattern than for the other patterns. For the falling contours, there is also a third pattern, splitting the rise-falls in two: in the first, the initial low stretch extends over the first two syllables, so that only the penult is high ((i) in Figure 144), while in the second, the rise occurs already

294 1 https://osf.io/cn9aj/; 2 https://osf.io/4bkp8/; 3 https://osf.io/cnw9x/; 4 https://osf.io/c4d2b/
295 There are only very few if any Quechua words with roots longer than three syllables.

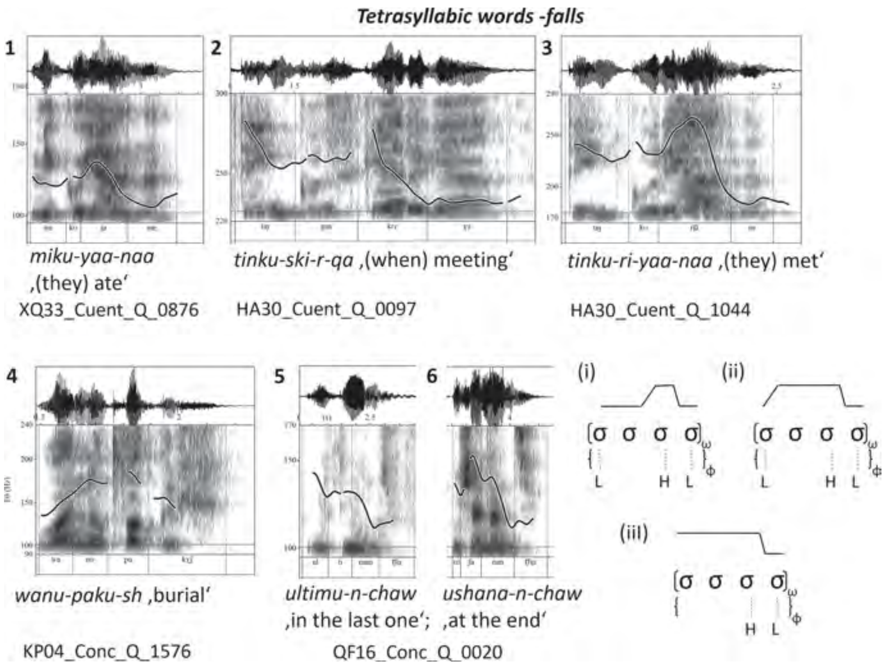


Figure 144: Individually phrased tetrasyllabic Quechua words realized with a falling contour with time-aligned syllabic transcription.[296] 1 and 3:[297] words beginning with a low stretch, with a rise and peak in the penult followed by a fall to the final syllable (schematic representation in (i)); 4 and possibly 6: words beginning low, with a rise reaching its peak already in the antepenult, plateau maintained until the end of the penult, fall to the final syllable (schematic representation in (ii)); 2 and 5: words beginning with a high plateau, fall beginning in penult and falling to final syllable (schematic representation in (iii)). All words are multimorphemic.[298]

296 1 https://osf.io/8kjng/; 2 https://osf.io/h3rjy/; 3 https://osf.io/ugahr/; 4 https://osf.io/s6mtg/; 5 & 6 https://osf.io/v9epg/

297 Analyzing /ri.ja:/ as monosyllabic [ɾ͡ja] with a complex onset in *tinkuriyaanaa*, which is clearly most faithful to the actual realization, goes against Quechua phonotactics described in Parker (1976: 55–56). If analyzed faithfully to the phonotactics as described, the word would be pentasyllabic.

298 With differing degrees of transparency. *Wanupakush* (also *wanupakusha*) is used in our data to refer to an image depicting a coffin surrounded by mourners, for which the same speakers use *entierro* 'burial' or *funeral* 'funeral' in Spanish. It is probably best analyzed as *wanu-paku-sh* (die-UNSPEC-PRTCP), containing the poorly understood –*paku*, itself possibly derived from –pU-kU, a combination which "is often understood as indicative of a commercial or professional activity, that is to say that the subject is beneficiary of an action directed at other persons" (Parker 1976: 119, my translation). It also possibly indicates an aspect of involuntariness in an action, with Carranza Romero (2003: 149) giving the meaning of the infitival verb form *wanupakuy* as "morírsele sin que pueda hacerse nada" [to die on someone without being able

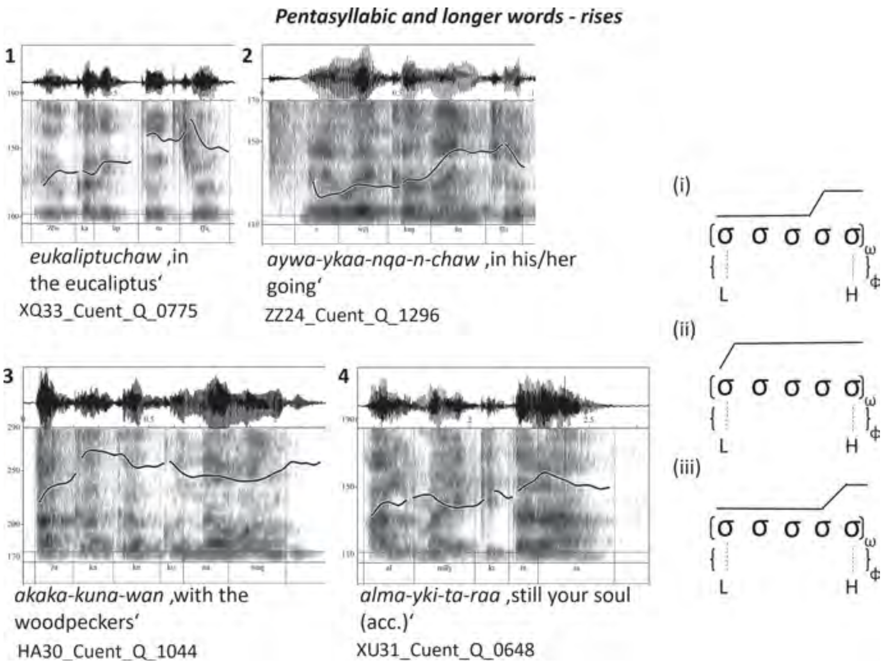


Figure 145: Individually phrased pentasyllabic and longer Quechua words realized with a rising contour.[299] 1 and 2: words beginning with a low stretch, with a rise starting just before and peaking in the penult followed by a high plateau maintained in the final syllable (schematic representation in (i)); 3: word beginning low, with most of the rise taking place already in the initial syllable, high plateau maintained until the end of the word/phrase (schematic representation in (ii)); 4: word beginning with a low stretch, with a rise in the penult and peak only in the final syllable (schematic representation in (iii)). All words are multimorphemic.

to do something] and "morírsele, ser responsable de la muerte de un animal o persona" [to die on someone, to be responsible for the death of an animal or person], but that of *ishpa-paku-y* (urinate-UNSPEC-INF) as "to urinate on oneself", *tushu-paku-y* (dance-UNSPEC-INF) as "to dance expecting thanks or payment" and *miku-paku-y* (eat-UNSPEC-INF) as "to eat at someone's disadvantage", i.e. to scrounge (all my translations). All of this points at a complex valency change meaning, whose precise workings remain unclear. The change from base verb meaning to derived verb meaning does not seem to be transparently the same between different verbs, indicating historical changes leading to lexicalization. It is unclear how productive this morpheme is. There is a further semantic stretch from the given meanings of the derived verb *wanu-paku-* to the attested meaning 'burial' after the addition of the participle *–sh* (also *–sha* or *–shqa*). In comparison, all the other words in Figure 144 are transparent results of productive morphology. I have not found any indication that morphological composition has an influence on intonation in the data here.

299 1 https://osf.io/fe78n/; 2 https://osf.io/rqj7e/; 3 https://osf.io/t98v2/; 4 https://osf.io/8d4jt/

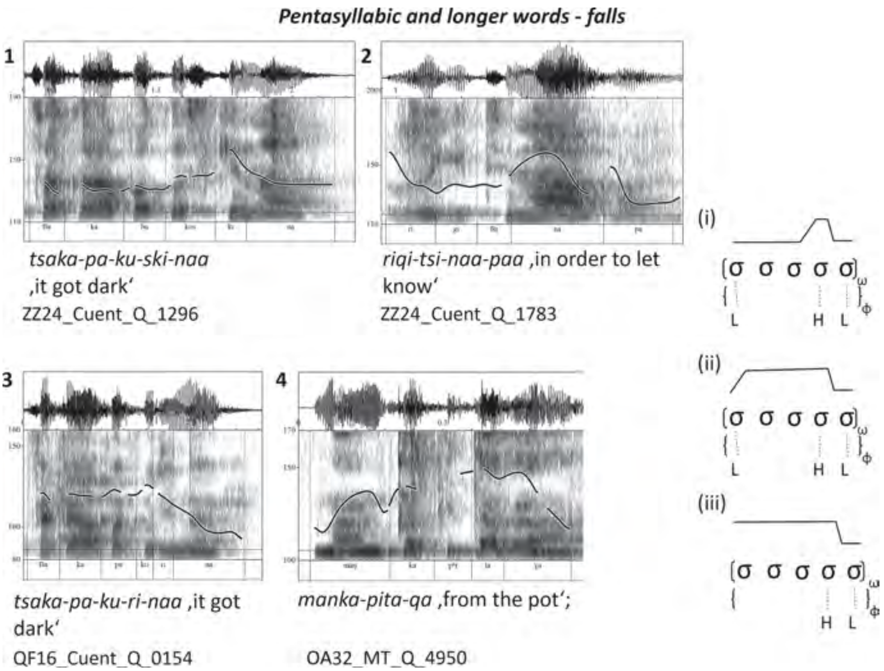


Figure 146: Individually phrased pentasyllabic and longer Quechua words realized with a falling contour.[300] 1 and 2: words beginning with a low stretch, with a rise in or before and a peak in the penult followed by a fall to the final syllable (schematic representation in (i)); 4: word beginning low, with a substantial part of the rise taking place already in the initial syllable, high plateau maintained until the end of or after the penult, fall to the final syllable (schematic representation in (ii)); 3: word beginning with a high plateau, fall beginning in penult and falling to final syllable (schematic representation in (iii)). All words are multimorphemic.

on the initial syllable, creating a high plateau-like realization on both penult and antepenult before the final fall ((ii) in Figure 144). Both patterns can be said to "continue" the trisyllabic rise-fall to equal degree, each preserving only parts of it. The plateau-forming rise-fall pattern seems to be less frequent than the other two. These three patterns persist also in individually phrased words of five or more syllables (Figures 145, 146), where some intermediate realizations are also found. The up to three variant patterns can be separated according to whether they need to make reference to the word penult as prominent position or not: of the three rising patterns, only the one which starts low and then rises so that both penult and final syllable are high ((i) in Figures 143 and 145) needs to make reference to such a position. Of the three falling patterns, none actually do in single-word phrases, because

300 1 https://osf.io/qxm84/; 2 https://osf.io/asqex/; 3 https://osf.io/ga634/; 4 https://osf.io/efqnp/

rightmost alignment of the two right tones will also always generate the attested contours. Since alignment with the word penult is clearly an active pattern in the data overall (cf. previous section), it can be assumed that it underlies some of the contours on single-word phrases as well, leaving them ambiguous.

Despite the variability, all attested patterns can be modeled using only three tonal configurations: LH for the rising contour, and LHL and HL for two falling contours, the rise-falling and only-falling contour, respectively. Comparing phrases of individual words with those composed of several ones reveals that there are no pitch events that pertain specifically to the level of individual words such as pitch accents – effectively, phrases consisting of several words behave as if they consisted of very long single words, except that they have the additional option of anchoring the tonal event of a rise or fall with one of the boundaries between the individual words they are composed of. Increasing the number of words per phrase does not increase the number of tones or pitch events, just the number of possible locations at which the tonal transitions may take place. This confirms the more specific pitch-based criterion for phrasing: only sequences on which either of these contours is fully realized should be taken to realize complete phonological phrases.

Schematic tonal representations for a few utterances that were previously discussed (see (121)-(125)) are here provided to illustrate the insights gained so far. They use square brackets for the delimitation of phonological phrases (PhPs), which are taken to be the units at which tones are assigned, just as in Huari Spanish. The assumption here is that rising contours consist of an LH tone sequence, only-falling contours of an HL tone sequence and rising-falling contours of an LHL sequence. The leftmost tone always minimally occupies the leftmost TBU. There is also a phrase-final H tone, which minimally occupies the rightmost TBU. The rightmost tone always minimally occupies the rightmost TBU. In rising-falling contours, this leaves the H tone in between these two peripheral tones, exhibiting varying alignment according to the patterns described. All tones seek to satisfy multiple alignment constraints in opposing directions, symbolized via dashed arrows. This creates the observed low and high stretches. Their interaction, discussed in section 6.3, creates the attested differences in the location of the tonal transition. Only where a word boundary occurs medially within a phonological phrase, a subscripted right bracket $_{]\omega}$ is inserted. A penult is marked with an accent mark (ˈ) and a line connects it with the H tone in the tonal tier only in the case that the information of that syllable being prominent is necessary to determine the shape of the contour. In accordance with the argumentation in section 6.1.3, this is the case when analysis **b** from Figures 130 or 137 is more appropriate than analysis **a**, also e.g. when the penult is already high in rising contours realized on more than two syllables. At this point, the analysis can remain agnostic regarding whether the H tone associates in these positions, since only alignment is necessary to generate the observed contour shapes. When the vowel of a syllable is unvoiced it is shaded in grey and cannot serve as TBU (cf. (122)). The details of constraint interaction generating these contours will be dealt with in the OT-analysis (section 6.3). The main point here is to show that using these contour variants produced entirely from tones assigned at the phonological phrase level, a sufficient tonal representation of the types of utterances encountered so far can be produced, leaving issues of information structural- or syntax-dependent phrasing aside. In particular, nothing indicates the need for potential higher-level (=IP) boundary tones: both falling and rising contours appear utterance-finally as well as medially without any difference in their shape apart from tone height. Utterance-final tonal movements, whether high or low, often have greater local excursion than medial ones. IP-final tones thus seem to only ever be copies of the final tone in the final PhP, likely obviating the need for such higher-level tones and possibly also non-recursive instantiations of higher prosodic levels here.

(121) XU31_MT_Q_1856 (cf. (104) and Figure 121)
[t͡sæ͡j ˈpi tæ]$_{\text{PhP}}$ [ul tʊ tɐ]$_{\text{PhP}}$ [ɾi ˈkæn ki]$_{\text{PhP}}$ [jæ kʊ ˈpɪ tæ]$_{\text{PhP}}$ [jaɹ ga mɯ ˈʃɐ ta]$_{\text{PhP}}$
L <-----H L H L L <----H L---> <-----H L-----------> <--- H

(122) TP03_Cuent_Q_1663 (cf. (113) and Figure 131)
[ma]$_{\omega}$ kwɛn tɐ ɾɪ ʃe͡əx]$_{\text{PhP}}$ [huʔ]$_{\omega}$ kwɛn tʉ tə̥]$_{\text{PhP}}$ [kwɛn tæ ja məŋ ɣɐn tə̥]$_{\text{PhP}}$
L-------------> H L L H L L------------> H L

(123) XQ33_Cuent_Q_1302 (cf. (120) and Figure 140)
[ħam pɪ ˈkʊɣ me]$_{\text{PhP}}$ [ɛ͡j wɛ͡j ka]$_{\text{PhP}}$ [ʔa kə ka ˈku na]$_{\omega}$ ni jæʃ ga]$_{\text{PhP}}$
L----> <-----H L H L L --------> <------------------H

(124) KP04_Conc_Q_1576 (cf. (110) and Figure 127)
[t͡seʰ]$_{\omega}$ ke]$_{\omega}$ xa wɑm]$_{\omega}$ kaɦ]$_{\text{PhP}}$ [wa nʊ pɑ kʏʃ]$_{\text{PhP}}$
L-------> <----------------H L <---H L

(125) HA30_Conc_Q_0211 (cf. (108) and Figure 125)
[wa ʃə]$_{\text{PhP}}$ [gɐn ˈtʉ t͡ʃu]$_{\text{PhP}}$ [ʔʊ kuʃ mi]$_{\omega}$ kæ͡j kn̩]$_{\text{PhP}}$
L H L <-----H L <-----H <--------L

6.1.6 Peak alignment in rise-falling contours in Huari Quechua and Spanish

In this section,[301] I provide quantitative results on peak alignment in similar-looking contours, the rise-falling contour of Quechua and pitch accents on paroxytone words followed by a low boundary tone in Spanish. In Huari Quechua, when the peak occurs phrase-finally, it is sometimes realized within the penult of the last word (cf. Figure 148), and sometimes later, in the final syllable (cf. Figure 147),[302] even on the same lexical item. In contrast, in Huari Spanish, pitch accent peaks nearly always align within the stressed syllable (cf. section 5.1.1.2, also Figure 149). We saw that some Quechua contours are ambiguous between two analyses, one in which tones are only aligned with prosodic edges, and another where the H tone is aligned with the word penult. The latter analysis is apt for cases in which the peak here is aligned within the penult, but it can hardly account for those where it is aligned in the final syllable. I made the assumption that in the Quechua data overall, (at least) these two variant alignment patterns are both active. The main question of this section is whether this holds up under quantified measurements via a comparison to the Spanish data. If the results show that it does, this confirms the suggestion that the H tone in the Quechua contours, unlike the H of the Spanish LH* pitch accent, is not always aligned or associated with the word penult, but instead sometimes aligned with the right edge of the phonological phrase (cf. the discussions in sections 3.3.3, 6.1.3).

The data considered here come from the *Quien* corpus, which was not used in the preceding sections mainly based on data from the *Conc, Cuento,* and *Maptask* corpora. For all utterances from Quechua and Spanish *Quien*, word and syllable boundaries were annotated and labelled manually in *praat*. A script (cf. Appendix C) was written to extract labels and durational measurements from these annotations, and to create pitch objects from which f0 measurements were extracted. Each pitch object created was inspected manually and perturbances due to consonantal microprosody were removed so that they would not influence the results. Only polysyllabic words were included in the analysis. Words were excluded when they were fragmentary, insufficiently voiced, or when the overall pitch range within the last two syllables did not exceed 7 Hz.[303] Table 20 gives the number of words remaining after this first process of elimination.

301 Parts of the analysis in this section was presented as a poster at Pape 21 (Buchholz 2021b).

302 Occasionally it is even realized in the antepenultimate or preantepenultimate, as seen in Figure 131 and Figure 132, but these cases are not considered here. Their existence is however in itself an argument for the lack of a strict H tone alignment with the putatively stressed word penult in the Quechua data.

303 See note 98 in section 5.1.1.2 for why this threshold was chosen.

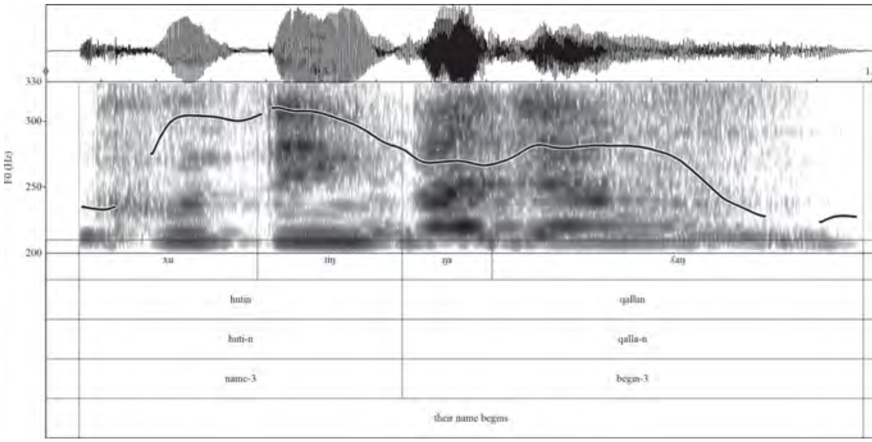


Figure 147: MS27_Quien_Q_0477[304] (part of a declarative).

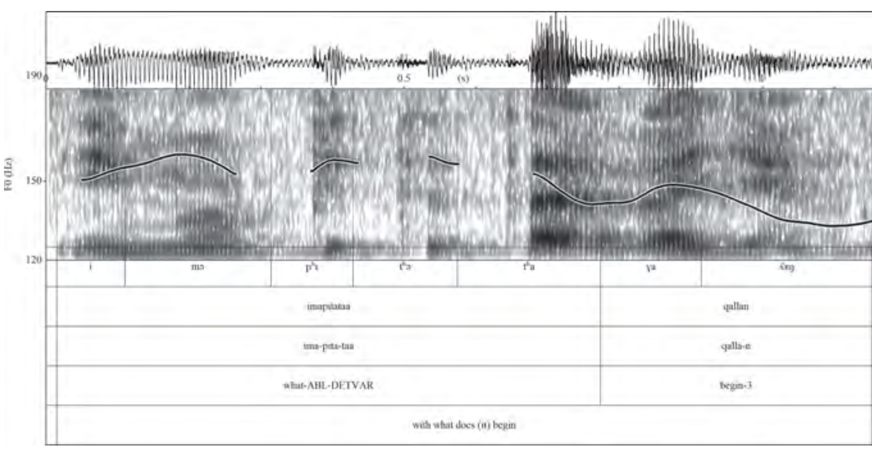


Figure 148: CJ35_Quien_Q_1351[305] (wh-question).

To operationalize only rising-falling contours for the analysis here, only words were considered whose f0 maximum position within the final two syllables preceded the f0 minimum position, i.e. which had an overall negative slope within the last two words. This was considered appropriate because after inspection of the data, complex pitch movements with more than one relevant peak within the last two syllables and final low pitch were not found, as in the majority of the Huari Quechua

304 https://osf.io/mxd5q/
305 https://osf.io/veh3q/

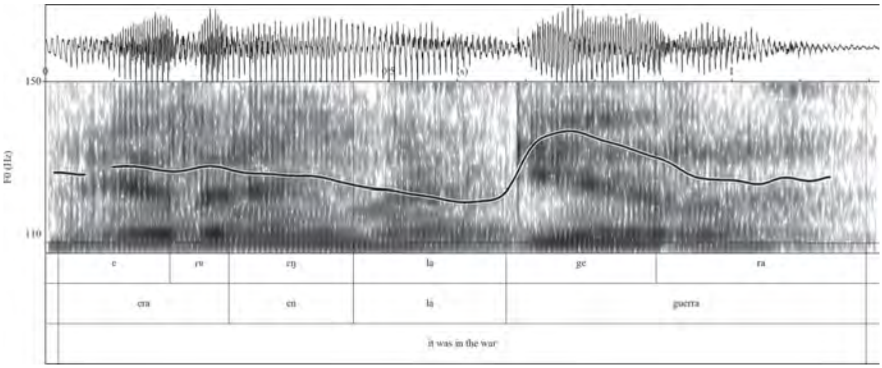


Figure 149: ZZ24_Quien_ES_0851[306] (declarative).

and Spanish data. This leaves 254 Quechua words, 143 paroxytone Spanish words, and 35 oxytone Spanish words. Proparoxytones were excluded because of their low number.

Table 20: Words from Quechua and Spanish *Quien* sorted according to length.

word length	***Quechua***		***Spanish***	
		proparoxytone	***paroxytone***	***oxytone***
2 syllables	124 (34.7%)	–	143 (56.8%)	63 (85.1%)
3 syllables	122 (34.2%)	2 (100%)	77 (30.6%)	9 (12.2%)
4 syllables	80 (22.4%)	–	29 (11.5%)	2 (2.7%)
5 syllables	21 (5.9%)	–	3 (1.2%)	–
6 syllables	9 (2.52%)	–	–	–
7 syllables	1 (0.3%)	–	–	–
Total	**357**	**2**	**252**	**74**
number of syllables	1100	6	630	161

Figure 150 visualizes the distribution of the f0 maximum position values in the final two syllables for Quechua words and Spanish paroxytone and oxytone words. The combined violin- and boxplot shows that in Spanish paroxytones, the great majority of peaks is located in the word penult, while a substantially larger portion of them is located in the final syllable in Spanish oxytones. For Quechua words, the distribution is broadest, with the majority of the values within the penult, but a number of them reaching also far into the final syllable.

306 https://osf.io/r56wf/

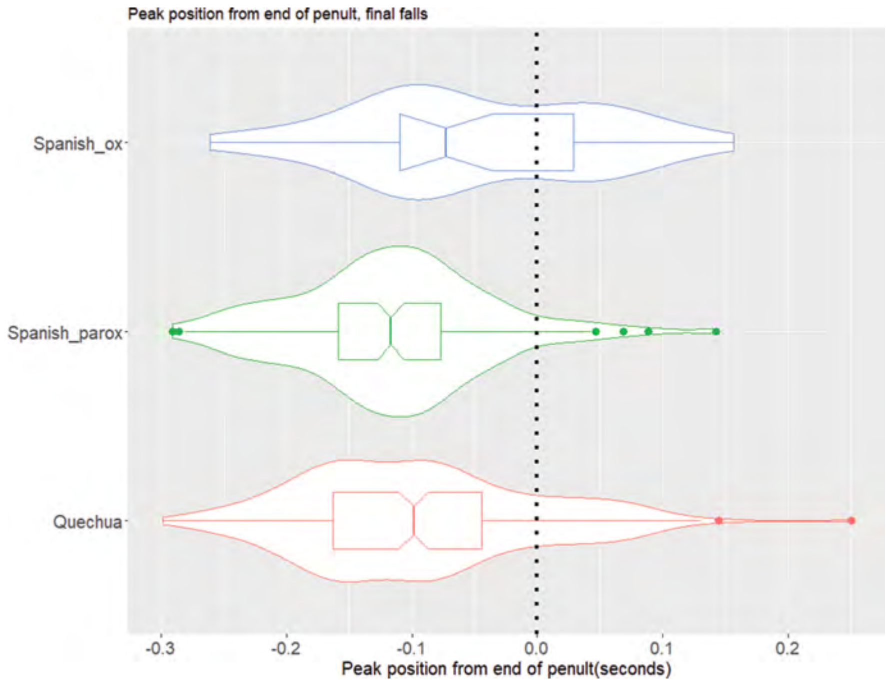


Figure 150: Final peak position in the Quechua and Spanish *Quien* words with a fall in the final two syllables. Combined violin- and boxplot for Spanish paroxytone, Spanish oxytone, and Quechua words.

The barplot in Figure 151 gives the share of peaks that are located within the penult (green) and the final syllable (red) for the three groups. For Spanish paroxytones, in 9 out of 143 words (6.3%) the f0 maximum is located in the final syllable, for Spanish oxytones it is 13 out of 35 (37.1%), and for Quechua words it is 41 out of 254 (16.5%). That only a minority of the peaks in Spanish oxytones occurs in the final syllable is perhaps surprising given the observations on peak placement in the Spanish chapter.

Most relevant here is the difference between Quechua and Spanish paroxytones given that Quechua words are also putatively stressed on the penult. A X^2-Test on the Quechua and Spanish paroxytone data yields a significant result ($X^2(1)=7.19$, $p=0.007$), suggesting that the observed difference in peak location is indeed associated with the difference between Spanish paroxytone and Quechua words. However, a further differentiation of the data is necessary. Since the aim is to especially to compare the similar-appearing phrase-final contours between Quechua and paroxytone Spanish, the words were differentiated further accord-

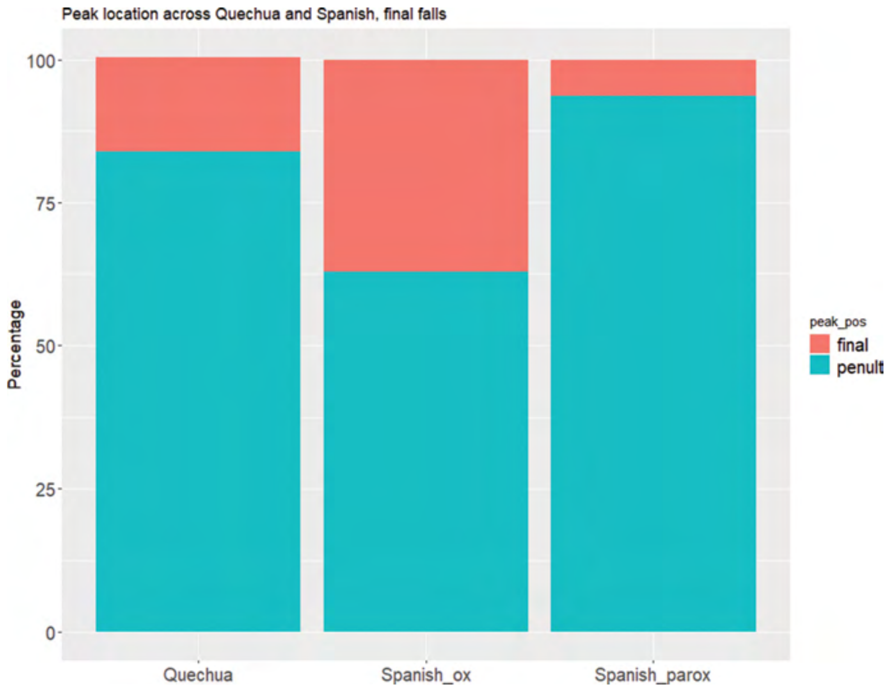


Figure 151: Final peak position in the Quechua and Spanish *Quien* words with a fall in the final two syllables. Barplot showing location of peak in either penult of final syllable in Spanish paroxytone, Spanish oxytone, and Quechua words.

ing to whether they were phrase-final or not, operationalized by whether they were followed by a pause or not.[307]

The left combined violin- and boxplot in Figure 152 gives the peak position values for Quechua and paroxytone Spanish for phrase-final and non-phrase-final words separately. The distributions appear to be somewhat different, with the medians located further towards the final syllable in the nonfinal words, but with the final Quechua words still showing the broadest distributional range. Linear regression models were used to further explore this. The goal was to find a model which best predicts the position of the f0 maximum in the final two syllables in phrase-final

307 Pauses as indicators of phrase boundaries are somewhat problematic, as is pointed out occasionally throughout this work. They were chosen because they are easy to detect automatically and in order not to base the phrase boundary criterion on final pitch movement when it is the aim of the analysis to say something about final pitch movement. Separated into phrase-final and non-phrase-final words, the counts for Quechua are 113 (phrase-final), 142 (non-phrase-final); for paroxytone Spanish 56 (phrase-final), 87 (non-phrase-final).

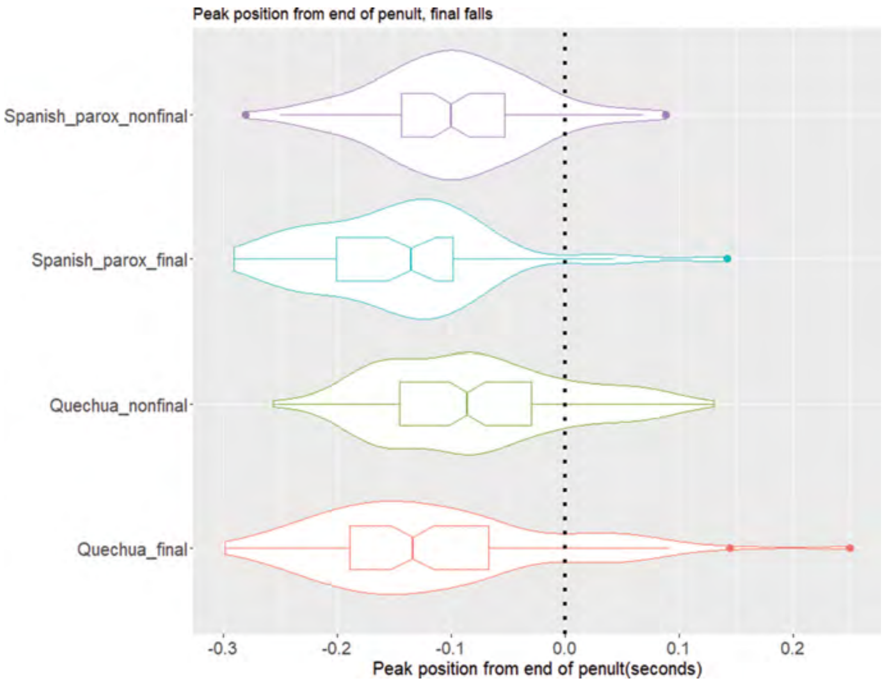


Figure 152: Final peak position in the Quechua and Spanish *Quien* words with a fall in the final two syllables. Combined violin- and boxplot for Spanish paroxytone and Quechua words in final vs. non-final position in the phrase.

words with a final fall. The predictor variables were *length of the penult*, *length of the final syllable*, and *position of the minimum* f0 value within the final two syllables. The dependent variable of the f0 maximum position could in itself be measured in three different ways: from the beginning of the penult, from its end, and from the end of the final syllable. Combinations of predictor variables and measurement reference points represent potential alignment scenarios. For the Spanish paroxytone words, the best model (R^2=0.76) turned out to be one in which the f0 maximum position was measured from the end of the final syllable with the predictor variable of the *length of the final syllable*.[308] The model results effectively mean that the final peak is on average located within the penult, at a fixed position relative to its end. This holds to a similar extent for words in phrase-final and non-phrase-final position, explaining 69–76% of the variance in the data. Following the discussion in section 3.5, this can be taken as evidence that the tone responsible for it, the H of the LH*

308 These are the model results for paroxytone Spanish words with a final fall in phrase-final position measuring from the end of the final syllable:

pitch accent, is associated with the stressed penult in paroxytones, as expected for Spanish. For Quechua, the same model applied to phrase-final words does not yield good results, explaining only about 30% of the variance (R^2=0.29). Instead, the most promising models use either all three predictor variables to predict peak position from the end of the final syllable,[309] or the *length of the penult* and the *position of*

---

	estimate β	st. error β	t-value	p-value
intercept	0.16	0.02	7.56	<0.001 ***
length of the final syllable	0.95	0.07	13.1	<0.001 ***
F(1, 54)=171.6, p <0.001				
R^2= 0.76				

Another model using both the length of the final syllable and that of the penult as predictors reached a slightly better adjusted R^2 of 0.77, but with a diminished F(2, 53) of 94.16. Because model fit always slightly increases with more predictors and because the results of the simple model are easier to interpret, I discuss only those.

The results of the same simple model applied to non-phrase-final words are as follows:

	estimate β	st. error β	t-value	p-value
intercept	0.11	0.01	8.02	<0.001 ***
length of the final syllable	0.93	0.06	13.7	<0.001 ***
F(1, 85)=189.8, p <0.001				
R^2= 0.69				

309 The model results for Quechua words with a final fall in phrase-final position using all three predictors and measuring from the end of the final syllable are:

	estimate β	st. error β	t-value	p-value
intercept	0.01	0.03	0.3	0.76
length of the penult	0.77	0.11	7.14	<0.001 ***
length of the final syllable	0.67	0.07	9.4	<0.001 ***
position of the minimum	0.29	0.1	2.8	0.007 **
F(3, 109)=43.28, p < 0.001				
Adj. R^2=0.53				

Because there are some outliers in the Quechua data that cannot be excluded on reasonable grounds, a robust regression model using *rlm* was also fitted. It slightly changes the standard errors compared to the standard regression but leaves all predictor variables significant. Because my main aim here is to compare the Quechua data with that of the Spanish paroxytones, I'm sticking with the normal regression results for better comparison. The position of the minimum in the model is measured from the same position as the maximum. The minimum position values themselves are represented moderately well with models that either take the length of the penult and of the final syllable as predictors (adj. R^2=0.58), measuring from the beginning of the penult, or only the length of the final syllable, measuring from the end of the penult (R^2=0.57), compatible with the idea that the minimum is the realization of an L tone seeking proximity to the phrase boundary.

the f0 minimum in the final two syllables to predict it from the end of the penult.[310] The amount of variance in the data they explain only amounts to 53%, more than 20% less than for the Spanish paroxytones. Their results are also more difficult to interpret. What they broadly amount to is the statement that the peak position in the final two syllables in Quechua is somehow influenced by their duration, but even that still leaves a lot of variation. The same models also fare worse when applied to words not in phrase-final position.[311] Crucially, the characteristic of the Spanish paroxytones that they realize the peak at a relative distance to some landmark within the penult, which is well captured by the model for those data, does not seem to hold well for the Quechua data that show a much broader range of variation, confirming the impression from the individual inspection of examples. This absence of an identifiable landmark within the final two syllables that the peaks in Quechua are on the whole orientated relative to suggests that there is more than one tonal alignment pattern underlying the data here, at least one of which does not refer to the putatively stressed penult (or that tonal alignment is in general much more variable).

310 The model results for Quechua words with a final fall in phrase-final position using *length of the penult* and *f0 minimum position*, measuring from the end of the penult, are:

	estimate β	**st. error β**	**t-value**	**p-value**
Intercept	0.02	0.03	0.9	0.35
length of penult	−0.86	0.1	−8.6	<0.001 ***
position of the minimum	0.27	0.06	4.74	<0.001 ***
$F(2, 110)=64.36$, $p < 0.001$				
Adj. $R^2=0.53$				

311 These are the results for the same models applied to Quechua words in non-phrase-final position:

	estimate β	**st. error β**	**t-value**	**p-value**
from the end of the final syllable				
intercept	0.01	0.03	0.3	0.77
length of the penult	0.57	0.11	5.01	<0.001 ***
length of the final syllable	0.72	0.08	8.49	<0.001 ***
position of the minimum	0.38	0.11	3.55	<0.001 ***
$F(3, 138)=43.6$, $p < 0.001$				
Adj. $R^2=0.48$				
from the end of the penult				
Intercept	−0.03	0.02	−1.27	0.21
length of penult	−0.52	0.11	−4.83	<0.001 ***
position of the minimum	0.33	0.07	4.43	<0.001 ***
$F(2, 139)=22.55$, $p < 0.001$				
Adj. $R^2=0.23$				

We explore one further possible caveat. Huari Quechua has contrastive vowel length. This raises the possibility that differences in syllable structure might influence peak placement in the Quechua data.[312] Table 21 gives the structure of the final syllable (light or heavy) compared to the location of the f0 maximum in the final two syllables (in the penult or the final syllable) in all Quechua words with a final fall and only in those that are phrase-final. There are two options for which syllables could be counted as heavy: only those with a long vowel, or also those with a consonantal coda (both have been suggested in the literature). For none of the options did X^2-tests produce significant results,[313] suggesting that heavy final syllables, whether only long or also closed, are not simply associated with a peak in the final syllable instead of the penult.

Table 21: Distribution of peak location in the final two syllables against syllable structure of the final syllable in Quechua words with a final fall.

only syllables with long vowels counted as heavy			***closed syllables and those with long vowels counted as heavy***		
Words with a final fall in both final and non-final position					
Final syllable	Peak location		Final syllable	Peak location	
	penult	final		penult	final
light (CV, CVC)	188	34	light (CV)	132	23
heavy **(CV:)**	26	7	heavy **(CV: & CVC)**	82	18
Words with a final fall in phrase-final position					
	penult	final		penult	final
light (CV, CVC)	94	15	light (CV)	69	10
heavy **(CV:)**	2	2	heavy **(CV: & CVC)**	27	7

However, if the syllable structure of both final syllables (penult and final syllable) is considered in tandem, an observable effect emerges. Figure 153 shows that the distribution of final peaks in phrase-final Quechua words with a final fall reaches far further into the final syllable in words with a “right-heavy” final syllable structure, i.e. where the penult is light, but the final syllable is heavy ((C)VC or (C)V:), than in

312 This would be expected under Stewart (1984)’s account of Conchucos Quechua stress, and under Parker (1976)’s account of the stress system of some Ancash Quechua varieties. In contrast, Hintz (2006) claims that South Conchucos Quechua stress is not sensitive to quantity. Cf. the discussion in section 3.3.3.

313 In the case of phrase-final words with only syllables with long vowels counted as heavy, the expected counts in two cells were <5, rendering the X^2-result unreliable. To remedy this, a Fisher’s exact test, which does not have that problem (Field et al. 2012: 816), was also executed on the counts. It also did not produce a significant result.

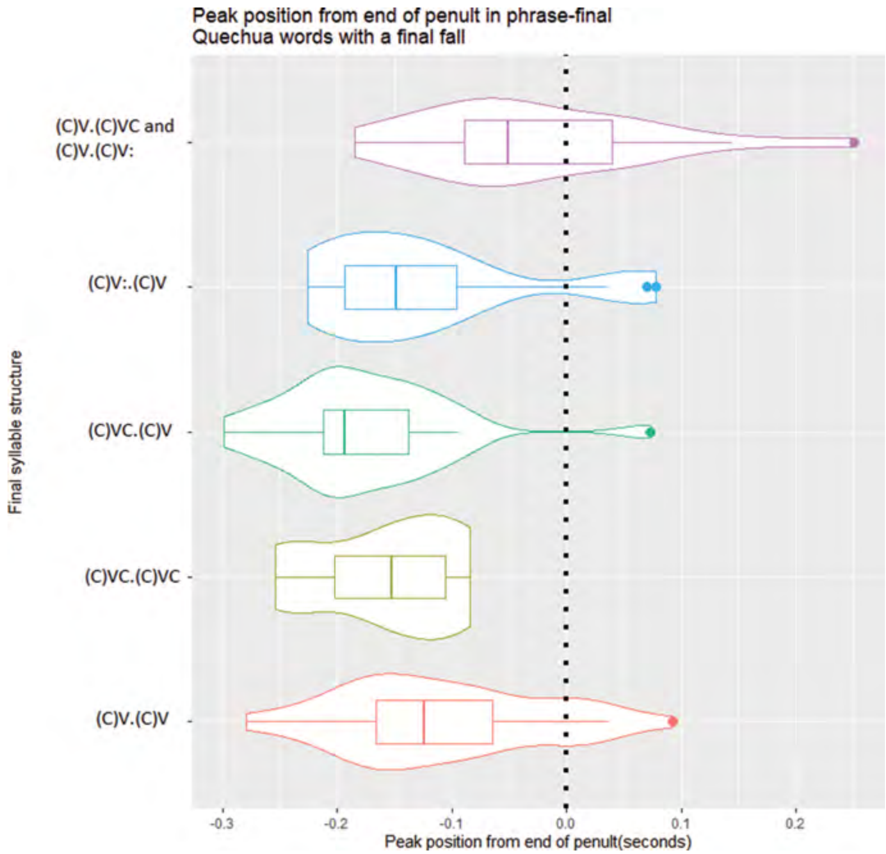


Figure 153: Final peak position in phrase-final Quechua *Quien* words with a fall in the final two syllables, sorted according to syllable structure of the final two syllables.

words with other final syllable structures. In a regression model, this effect is significant,[314] but it does not obtain with words that are not phrase-final. This is some

314 There are the model results:

	estimate β	st. error β	t-value	p-value
intercept (=(C)V.(C)V)	−0.11	0.01	−7.39	<0.001 ***
heavy.heavy	−0.05	0.03	−1.86	0.06
(C)VC.(C)V	−0.07	0.02	−2.87	0.005 **
(C)V:.(C)V	−0.01	0.02	−0.54	0.59
light.heavy	0.08	0.02	3.4	<0.001 ***
$F(4, 108)=9.35$, $p<0.001$				
Adj. $R^2= 0.23$				

indication that the *relative* syllable weight between the two final syllables at the end of a phrase has an effect on peak position. The model only explains about 23% of the variation (adj. R^2=0.23), however, so that this can only be part of the story, perhaps describing only one alignment variant active in the sample. If the categorical predictor variable of final syllable structure is replaced with the continuous predictor variable of the difference in duration between penult and final syllable, a similar effect is observed and such a model actually explains a greater share of the variation (adj. R^2=0.38).[315] This means that peaks move further rightwards relative to the boundary between the final two syllables the more the length difference between them tilts in favour of the final syllable. Plausibly, some of this is due to a complex effect of syllable structure,[316] but more than that it suggests that the peak simply seeks to be not too far away from the right phrase boundary. Still, more than half of the variation remains unexplained in this way.

For some languages, it has been found that early and late peak alignment in very similar-looking rising-falling contours differentiates between declaratives and

They indicate that right-heavy final syllable structures move the final peak significantly to the right.

The counts for the groups are: 35 ((C)V.(C)V); 11 ((C)VC.(C)VC); 25 ((C)VC.(C)V); 19 ((C)V:.(C)V); 19 ((C)V.(C)VC); 3 ((C)V.(C)V:); 1 ((C)VC.(C)V:). The categories (C)VC.(C)V: and (C)VC.(C)VC were grouped together as "heavy.heavy", and (C)V.(C)V: together with (C)V.(C)VC as "light.heavy" because of the low counts. A robust model using *rlm* was also fitted, with the same groups making significant contributions. Applying the model to non-phrase-final words did not yield significant results.

315 These are the model results:

	estimate β	st. error β	t-value	p-value
intercept	−0.14	0.01	−17.54	<0.001 ***
length difference between penult and final syllable	−0.46	0.06	−8.23	<0.001 ***
F(1, 111)=67.69, p <0.001				
R^2= 0.38				

The predictor variable was calculated by subtracting the length of the final syllable from the length of the penult, i.e. positive values indicate a relatively longer penult, and negative values a relatively longer final syllable. The same model applied to non-phrase-final words only has an R^2 of 0.14 (F(1, 140)=23.29, p<0.001).

316 The effect of syllable structure should not simply be dismissed because it also has traces in Huari Spanish. In the audio stimuli for Spanish *Cuento*, the oxytone *colibrí* "hummingbird" occurred. For some speakers, this was a new lexical item, because they use *picaflor* instead. When they reproduced it in the retelling, they produced it variably as [koliˈbrin] or [kuleˈbriŋ], with a closed final syllable. This could suggest an unmarked preference for oxytones with a heavy final syllable. The effect is subtle, however, since it does not occur with oxytones with light final syllables the speakers have lexicalized, such as *maní* "peanut".

polar questions, with the peak in polar questions aligning later than in declaratives e.g. in Neapolitan Italian (D'Imperio & House 1997) and Tashlhiyt Berber (Roettger 2017, cf. Roettger 2017: 68 for a discussion of further cases). Because the data here come from both declaratives and questions, the difference in peak alignment could be related to that. However, no significant difference could be found in peak alignment between declaratives and polar questions,[317] excluding the possibility that this difference in utterance type is responsible for the observed variation in the Quechua data.

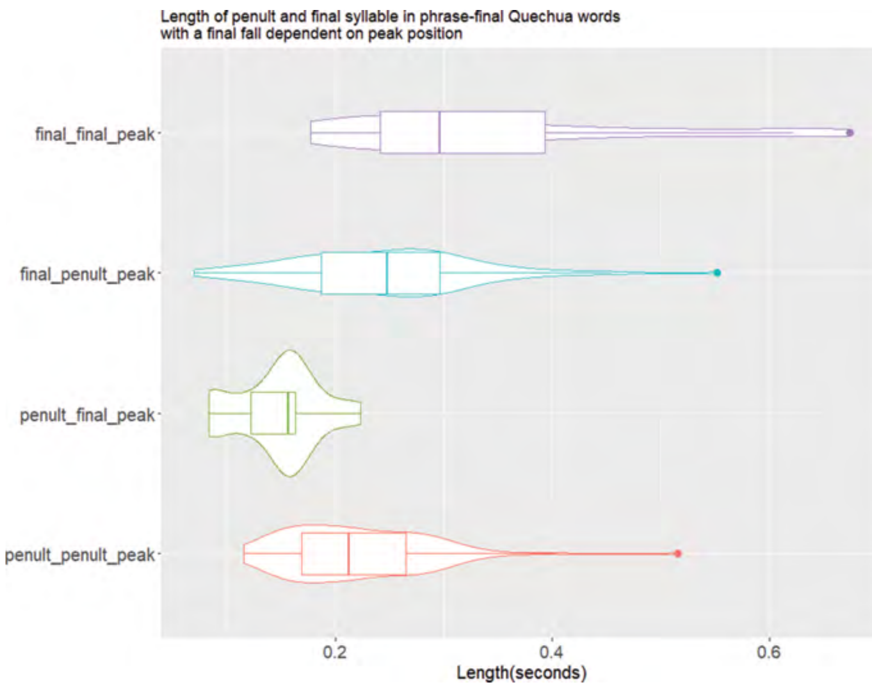


Figure 154: Length of the final two syllables in Quechua phrase-final words with a final fall sorted according to whether the peak is realized on the penult or the final syllable. "final_final_peak" means "final syllables in words with the peak on the final syllable", "final_penult_peak" means "final syllables in words with the peak on the penult", and so on. The peak is located on the penult in 96 words, and in 17 on the final syllable.

317 Wilcoxon rank-sum tests were executed on the difference in peak position between phrase-final words in declaratives and all other utterance types (neutral polar questions, biased polar questions, wh-questions) and on that between all types of polar questions vs other utterance types (declaratives and wh-questions), measuring peak position from the beginning and end of the penult, and from the end of the final syllable. No results were significant.

Finally, Roettger (2017: 47, 144) suggests that it is an indication for the association of a H tone if the syllable on which it is realized is longer than one on which it isn't realized, a process called "phonetic enhancement" (cf. section 3.5.3). Figure 154 shows the length of the penult and the final syllable, respectively, in phrase-final Quechua words with a final fall, sorted according to whether the f0 maximum in the last two syllables is located in the penult or the final syllable. It shows clearly that syllable length increases on average when the peak is located on that syllable. This holds for both the penult and the final syllable, and is independent of the fact that the final syllable is longer than the penult on average (final lengthening). The effect obtains for all Quechua words, not just phrase-final ones, and it is statistically significant.[318] The observed enhancement effect for Quechua words refers to syllable length depending on peak position, which is not the same as stress-induced lengthening.[319] For the putatively stressed word penult in Huari Quechua, Buchholz & Reich (2018) showed that it is not longer on average than surrounding syllables. Word stress position should not differ between individual instantiations of the same lexical item in the same prosodic context. Yet peak alignment does differ between individual instantiations of the same lexical item in final position here, as the comparison between examples like Figures 147 and 148 shows. Thus, the effect of lengthening according to peak position here is independent of stress. Most of the results in this section do not in fact particularly support an account whereby the H tone associates with the penult because it is (putatively) stressed. However, following Roettger (2017), the observed lengthening effect means that the H tone responsible for the peak does associate with the TBU on which it is realized. This would suggest that in the Quechua rising-falling contours observed here, at least some of the H tones associate, but not with a specific landmark within either of the two final syllables, and the property of stress is not relevant to this association. The phonetic enhancement effect together with the earlier observed effect that relatively longer or heavier syllables in the final two-syllable window seem to attract peak placement could also explain a puzzling finding in Buchholz & Reich (2018: 153–155): there it was found (based on other data than here) that CVC penults are *longer* than surrounding syllables to a higher degree than CVC syllables in other positions, while CV penults are *shorter* than surrounding syllables to a higher degree than CV syllables in other positions. This might be, at least partially, because the CVC penults

318 A Wilcoxon rank-sum test on the length of the penult depending on final peak positions for phrase-final words with a final fall yielded a highly significant result ($W=1350$, $p<0.001$); for the length of the final syllable the result is significant ($W=520$, $p=0.017$).

319 This is observed in the Spanish data: penults in phrase-final and non-phrase-final paroxytones are significantly longer than in oxytones, according to a Wilcoxon rank-sum test ($W=4206$, $p<0.001$); final syllables in oxytones longer than in paroxytones ($W=11628$, $p=0.001$).

are slightly biased[320] for bearing the H tone and realizing the peak, making them correspondingly even longer than other CVC syllables, while CV penults are slightly biased against bearing the H tone and realizing the peak, resulting in them being somewhat shorter on average than other CV syllables.

Overall, the results confirm the proposal that in part of the Quechua data, an alignment pattern is active in which tones do not associate with a putatively stressed syllable, but instead seek alignment with the edge of the phrase. In particular, peak alignment behaviour seems markedly different from that of Spanish words with a superficially similar rising-falling pitch contour, where it is uncontroversially due to association with a stressed syllable and alignment to a landmark relative to that syllable. It also seems likely that peak position in Quechua is affected by variable temporal constraints within the last two syllables, possibly influenced by syllable structure. The H tone responsible for the Quechua contour likely associates with the TBU on which it is realized at least in a subset of the cases. That so much variability has been found confirms the suggestion that there are several alignment patterns active in the data, each contributing to the varied overall picture. In section 6.3.1, I will show that the Quechua rising-falling contour under discussion is indeed generated from more than one of the alignment constraint rankings which are independently motivated there.

6.1.7 Extent of the phonological phrase and mapping to other categories

Before moving on to the discussion of loanwords from Spanish, I want to make a short excursion exploring the extent of the phonological phrase and its mapping to the categories of word and utterance via two singular examples. First, KP04_MT_Q_0393 (Figure 155), with context (126). At 36.5, KP04 gives the information that the path in the map leads above the tree. After a pause, TP03 answers with an acknowledgement token (38.5). The delay in answering seems to be interpreted by KP04 as indicating less than full understanding,[321] because he repeats his previous

320 Assuming equal distribution of syllable types in all positions, a CVC penult is equally likely ($p = 1/3$) to be followed by a final syllable that is CV, CV:, or CVC. In 1/3 of those cases, when it is followed by a CV final syllable, it is even more likely to attract the peak than the penult is anyway, according to the results in Figure 153. For CV penults, it is the other way around. Therefore, statistically, CVC penults are slightly biased to attract the peak, while CV penults are relatively biased against it.

321 See Kendrick & Torreira (2015); Bögels et al. (2015) on how delays in responding to various conversational actions are associated in with an increased likelihood for the response being dispreferred from the perspective of the recipient, i.e. in this case, incomplete understanding or agreement.

utterance at 39.3 as a request for confirmation.[322] That it is understood as such is clear from TP03's rapid response of three repeated acknowledgement tokens. KP04's repetition, the confirmation-seeking request (Figure 155), is realized as a rise-falling contour with word boundary alignment. A low stretch on the first word *hacha* is followed by a high tone on the initial syllable of *hananpa*. The high tone is maintained into *nan*, but then falls to reach a low target on the final syllable, from which it then rises again towards the end of the phrase. Clearly, there must be at least an additional H tone after the LHL of the rise-falling contour, aligned with the right edge of the phrase. This is the only example in the Huari Quechua data in which such an additional tone clearly occurs,[323] an intonational *hapax legomenon* to my knowledge.

(126) TP03 & KP04_MT_Q_0365 – 0409[324] (context for Figure 155)

time (seconds)	KP04 (*with the path on the map*)	TP03 (*without the path on the map*)
36.5	(L H-------L) hacha hana-n-pa tree above-3-GEN *above the tree*	
37.2	*(1.2)*	
38.5		(H L) mhm ya
39.3	**(L H L H)** **hacha hana-n-pa** tree above-3-GEN *above the tree*	
40.4		(H L) ya ya ya

That this additional tone is there is not only unexpected from the findings about Huari Quechua phrasal contours so far, but also from the point of view of tone distribution to TBUs. Above we saw that the rising-falling contour is never realized on bisyllabic single-word phrases, and that the "displacement" of the H tone to the penultimate voiced syllable of the phrase is also explained well by tonal crowding.

322 The confirmation KP04 seeks is that TP03 has understood so that they can move on to the next landmark; not a confirmation that the information is correct (KP04 provides the path information here, not TP03).

323 Another potential case occurs in the same *Maptask* later at 278.9, but contains several voiceless stretches and is therefore difficult to assess.

324 https://osf.io/etny4/

Yet here, at least three tones are aligned on the two syllables *nan.pa*. Morphologically, *pa* is not long (i.e. possibly bimoraic), according to Parker (1976: 84). Aside from this tonal crowding problem, three explanations for the exceptional realization with four tones are possible here. The first is that this is the only token in the corpus of a rise-fall-rise contour, which is simply so rare that it's not encountered elsewhere. That is certainly not impossible, if this contour is associated with a particularly marked pragmatic context condition. However, there are plenty of confirmation-seeking questions occurring under comparable context conditions in the Quechua data, and they do not realize this contour.

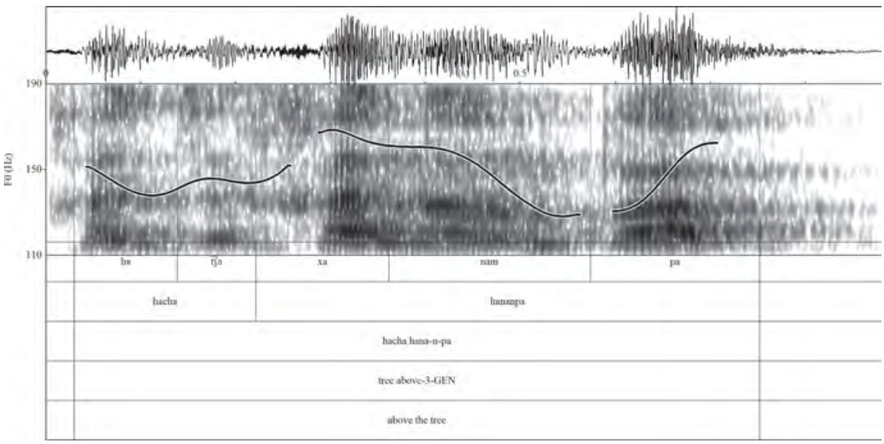


Figure 155: KP04_MT_Q_0393[325] (confirmation-seeking question consisting of a falling contour plus final rise).

The second is that this is a case of adding boundary tone movement from Spanish to an otherwise Quechua utterance. If this were a readily available option to the speakers, we might expect to see it more frequently, but it is equally not impossible. The third explanation is that the contour normally realized on the utterance-final particle *aw*, is here realized on *–pa*, with the segmental material of *aw* fully assimilated to it. This is appealing because *aw* often occurs in confirmation-seeking and other biased questions. Utterance-initially, it means "yes" (Parker 1976: 35), but utterance-finally it is used like a question tag in other languages, e.g. Spanish *no* or *verdad*. From the observation of other examples it appears that this utterance-final *aw*, if realized in its own phrase, always realizes that phrase with a rising contour. This explanation entails that this utterance, consisting of a single word, is realized with two phonological phrases,

325 https://osf.io/zmbpk/

one with a rise-falling, and an additional one with a rising contour. This might seem outlandish, but Igarashi (2014: 466–467) makes the argument that this also occasionally occurs in Japanese, when more than one Accentual Phrase (AP) is realized on a single *morphological* word.[326] This is not tantamount to saying that the *prosodic* word is larger than the phonological phrase here, just that a phrase is not always as large or larger than a *morphological* word (or a word plus segmentally assimilated particle). That remarkable gaps can exist between the extent of morphological word(s) and of a phrase is also evident from another example, AZ23_Cuent_Q_0717 ((127)/Figure 156). AZ23 is telling about a speaker change between the two characters in the *Cuento* story.

(127) AZ23_Cuent_Q_0717

hampi-pa	ni-n	ni-pti-n-qa	runa-qa	ni-n
heal-BEN	say-3	say-SUBDIFF-3-TOP	person-TOP	say-3

"'for healing', he said / when this is said, the man says"

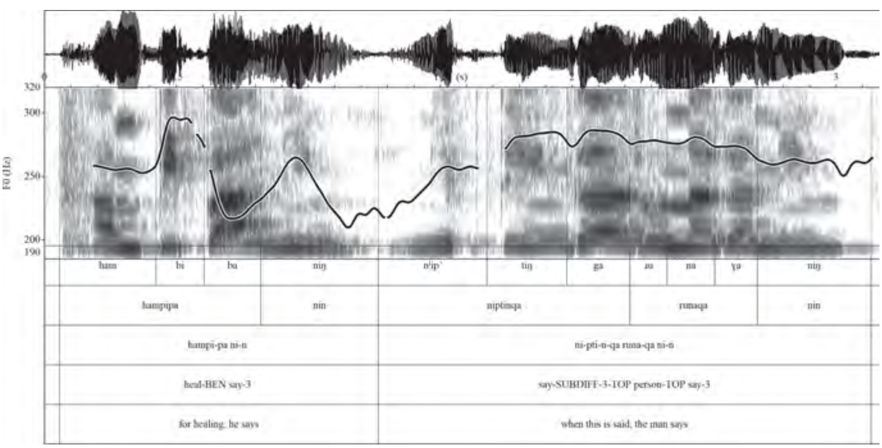


Figure 156: AZ23_Cuent_Q_0717[327] (two declaratives, consisting of two rise-falling and one rising contour, respectively).

326 For French, where the non-isomorphy between lexical words and APs is uncontroversial, Jun & Fougeron (2002: 159–160) report that more than the expected number (one initial, one final) of accentual rises can occur on a single very long word (they use *anticonstitutionellement* „unconstitutionally"). They conclude that APs can exceptionally have more rises than the expected initial and final one, but arguably a more principled analysis might be to assume, given that non-isomorphy between morphological words and APs is established in the other direction (1 AP containing >1 word), that here more than one AP maps onto a single word. Since the additional rises show a tendency to occur near morphological boundaries, such an analysis would also preserve the delimitative character of the AP pitch movements.

327 https://osf.io/z4re9/

At the beginning, *hampipa* is part of the direct speech of one of the protagonists. It is phrased separately with a rise-falling contour, followed by the verb of saying in the third person *ni-n*, which is also realized with a rise-falling contour. It realizes at least two (HL) or possibly three (LHL) tones on an at most bimoraic syllable. Phrasing here separates the content of the direct speech from the quotative verb, and that again from the subsequent utterance *niptinqa runaqa nin*, which is realized as a single rising phrase on three morphological words. It announces a speaker switch (achieved by the subordinating marker under subject change *–pti*, cf. Parker 1976: 143–144), and establishes the referential frame under which the following utterance, giving the content of the other protagonist's speech, is to be understood. The entire second utterance is thus contextually an introduction to the following one (note the topic marker *–qa* on both the verb *niptin-qa* and the noun *runa-qa*) and the proposition asserted by it is suspended and incomplete.[328]. That is likely why it is realized with a rise and without any of its components individually phrased (cf. the findings in sections 6.2.2 and 6.4). Here a single phonological phrase thus realizes one monosyllabic word, *nin* (possibly trespassing against otherwise active constraints on tonal crowding), on the one hand, and an entire utterance consisting of three polysyllabic words, on the other. The mapping between morphological words and phonological phrases is thus clearly far from predetermined and certainly malleable in order to conform to constraints imposed by contextual meaning. The case represented by the first example, with more than one phrase possibly realized on one word, is less frequent than that represented by the second, where the non-isomorphism goes in the other direction. But given that the latter possibility is a fact, the former should not be dismissed out of hand. Mapping more than one phonological phrase to a single long morphological word might also go some way in explaining the very heterogeneously reported secondary "stresses" in Quechua varieties (e.g. Parker 1976: 57–60; Adelaar 1977: 43–46; Stewart 1984; Hintz 2006; O'Rourke & Swanson 2013; cf. also the overview in Hintz 2006: 482–483). In the following, we will not fully resolve the questions posed by these two examples. I will at times return to the discussion of tonal crowding and whether it can be explained by assuming a moraic structure. In the next section, we will move on to marked intonational patterns that are restricted to loanwords from Spanish.

328 Weber (1989: 15) takes the verb of saying *ni-* to take sentential complements.

6.1.8 Marked tonal alignment on Spanish loanwords

This section presents two exceptional intonational patterns. They only occur on lexical items that are etymologically of Spanish origin,[329] and they crucially deviate from the characterizations given for the previous tonal alignment patterns. The frequency of their usage also varies considerably between the speakers in our corpora. We have already encountered several examples containting words of Spanish origin, where they do not behave any differently from words of Quechua origin. In contrast, the first of the loanword patterns covered in this section is characterized by the tonal transition, i.e. the rise or the fall, not occuring in one of the positions already established (phrase boundary, word boundary, or word penult), but instead as determined by the location of the stressed syllable in the corresponding Spanish word. That is to say, while the alignment pattern with the word penult makes reference to a fully regularly determined prominent position, this pattern requires a lexically specified word stress position, but otherwise exhibits the same contours. In the second loanword pattern discussed here, the tonal transition familiar from the Quechua contours occurs where expected from one of the three established patterns, but in addition, a separate local pitch event takes place at the location of the stressed syllable in the Spanish word. I will call the first of these the "inherited" loanword pattern, because it seems somewhat exotic when viewed from the perspective of the rest of the prosodic system, formed by the demands of a quite different system (paying much more attention to a structure like stress), like an oddity inherited from a traveling grandparent. The second pattern with the additional pitch event I call the "grafted" pattern, because the lonely localized pitch event appears not quite to fit in with the rest of the intonation, like a branch from a eucalyptus grafted on to a fir. Spanish-origin words in a Quechua utterance do not obligatorily require one of the marked loanword prosodic patterns. Using these patterns seems at best one of several options available to the Huari speakers. In the last part of this section (6.1.8.3), the distribution of the intonational patterns for Spanish-origin words in some of the corpora will be investigated. In the next section I will first give a more detailed description.

6.1.8.1 Description of the patterns

In AZ23's MT_Q_1158 ((128)/Figure 157), *abejakunapa* is realized on the one hand with a rising contour finally, with pitch reaching a high plateau in the penult [na]

329 I use "Spanish loanword" as a shorthand to refer to this group of words, but this section will show that their behaviour is far from homogenous.

and extending to the final syllable [pa].[330] Yet the syllable [be], stressed in the Spanish *abeja*, realizes a substantial first rise, followed by a gentle fall. Note that it is also realized with more duration and energy than most other open syllables in this utterance. This is an instance of the mnemonically labelled "grafted" pattern, where a local pitch event takes place on the location of the Spanish stressed syllable in addition to the "normal" tonal make-up of the phrase.

(128) AZ23_MT_Q_1158

tsay-pita	*abeja*[331]-kuna-pa	hana-n-pa	*escoba*-man	chaa-nki
DEM.DIST-ABL	bee-PL-GEN	above-3-GEN	broom-DEST	arrive-2

"from there along above the bees you get to the broom"

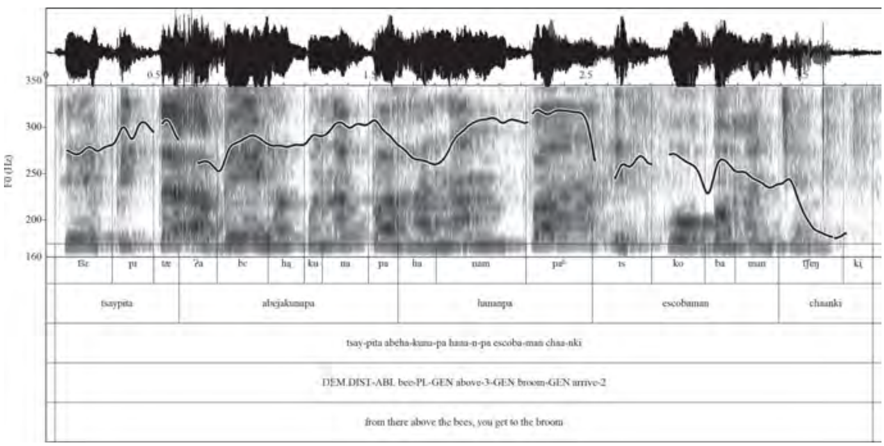


Figure 157: AZ23_MT_Q_1158[332] (declarative consisting of three rising and two falling contours).

330 I analyse the contour on *abejakunapa* as rising because I assume that the incipient fall on the final [pa] is due to the upcoming initial L of the phrase realized on *hananpa* and not because of a final L of the phrase realized on *abejakunapa*, which would make it a falling contour. This is because pitch only reaches the low elbow in [ha] instead of [pa]. However, the transition between *hananpa* and *escobaman* shows that supposedly the same underlying tones can also result in a much less ambiguous phonetic realization. To a certain extent, such analyses will always leave some space for interpretation and must also take recourse to external considerations like the hypothesis that rising contours signal continuation and falling ones finality (cf. section 6.2.2), making this here more likely to be a rising contour. The circularity in this argumentation can probably not be entirely avoided.

331 Lexical items or stems originating historically from Spanish in the examples in this section are italicized.

332 https://osf.io/9u4s6/

Compare this on the one hand with *escobaman* in the same utterance, which does not seem to pay much attention to the Spanish stress position at all, as far as can be made out, and on the other with the realizations of *terceruchaw* in OZ14_Conc_Q_0331 (see (129)/Figure 158) or *primeruchaw* in QZ13_Conc_Q_1516 ((130)/Figure 159): both of these are produced with a rising contour, with the rise taking place not on the phrase penult or the initial syllable, but at the location of the stress in the Spanish word, i.e. the antepenultimate, [se] and [me] respectively.

(129) OZ14_Conc_Q_0331

segundu	*fila*-chaw	*terceru*-chaw	chuqllu
second	row-LOC	third-LOC	corncob

"in the second row, the third one is the corncob"

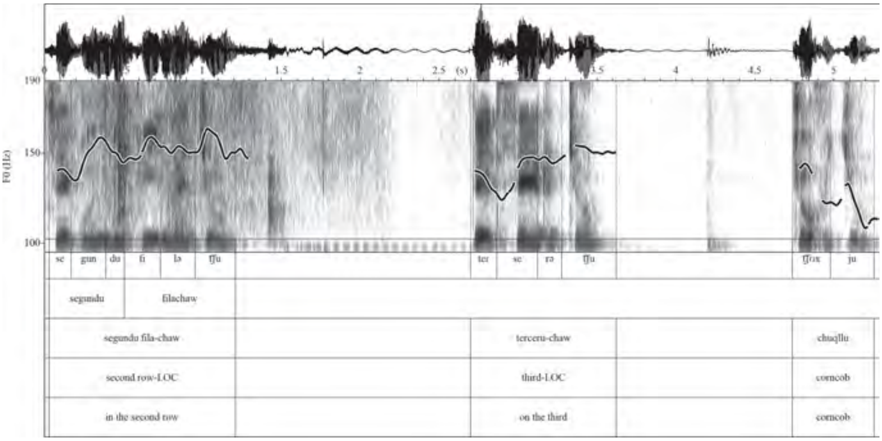


Figure 158: OZ14_Conc_Q_0331[333] (declarative with three rising and one falling contours).

(130) QZ13_Conc_Q_1516

tercera	*fila*	*primeru*-chaw	*riku*	runa
third	row	first-LOC	rich	person

"in the first one the third row [is] the rich person"

333 https://osf.io/e2b4y/

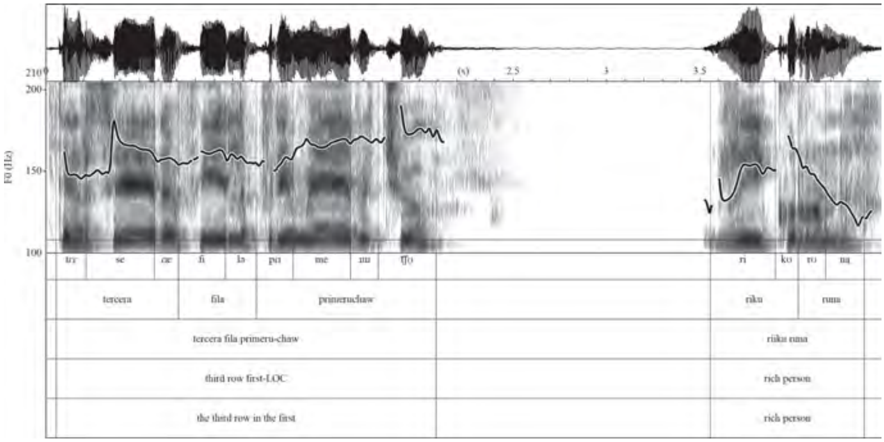


Figure 159: QZ13_Conc_Q_1516[334] (declarative with two rising and one falling contours).

These are instances of the “inherited” pattern, where the Spanish stress determines the location of the tonal transition in the phrase contour, but does not produce an additional tonal event. The “inherited” pattern isn’t always identifiable unambiguously: in both utterances, the first word, *segundu* and *tercera*, respectively, looks like it might also follow this pattern, with elevated pitch and more duration on the Spanish stressed syllable, followed by a fall. However, this might also be an instance of the rise-falling contour with one of the previously seen alignment patterns.[335] Similar considerations pertain to many bi- and trisyllabic words of Spanish origin in the data. It is noteworthy that only very few cases were found where the “inherited” pattern is unambiguously realized in a falling contour. This might suggest that the rise on Spanish accented syllables is part of what is emulated in the inherited pattern, not just the abstract property of accentedness.

The following two utterances, OZ14_Conc_Q_0013 and XU31_MT_Q_2837 ((131)/Figure 160 and (132)/Figure 161, respectively), are further examples for the forms words with an “inherited” or “grafted” accent can take.

(131) OZ14_Conc_Q_0013

segundu	*fila*-chaw	*primeru*-chaw
second	row-LOC	first-LOC

“in the second row, in the first one”

334 https://osf.io/3ejw9/

335 It is an open question whether in a bilingual community such superficially similar forms are distinguishable at the level of phonological representation in all cases. Their convergence is discussed in section 7.4.

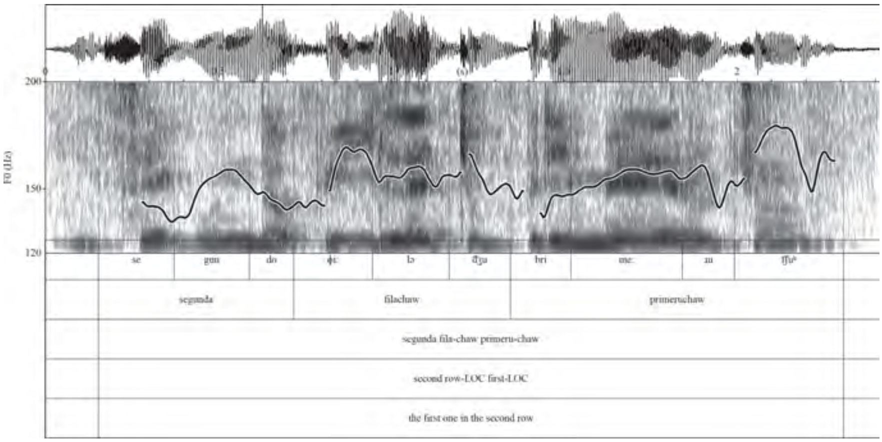


Figure 160: OZ14_Conc_Q_0013[336] (part of a declarative consisting of three phrases).

(132) XU31_MT_Q_2837

manka	*ladu*-n-chaw	atuq	tsay	*frenti*-n-chaw-raa-mi
pot	side-3-LOC	fox	DEM.DIST	front-3-LOC-CONT-ASS

ka-ykaa-n	*murcielagu*
COP-PROG-3	bat

"next to the pot, still before the fox, there is the bat"

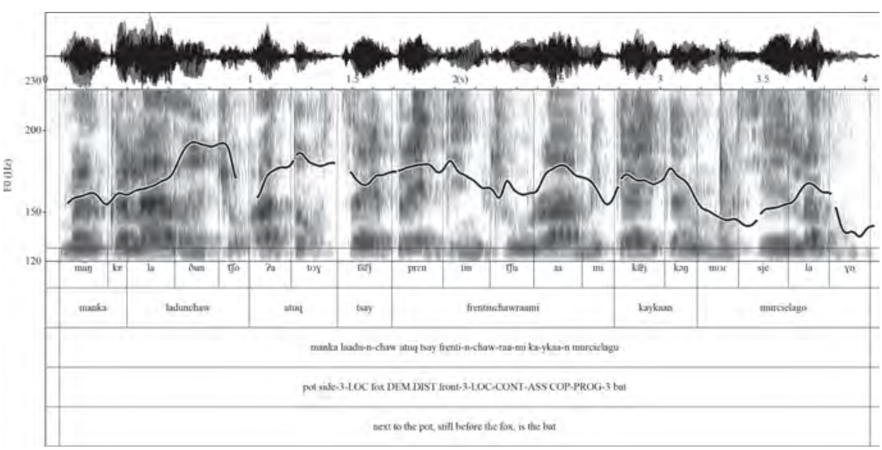


Figure 161: XU31_MT_Q_2837[337] (declarative with 4 rising and one falling contours).

336 https://osf.io/judpy/
337 https://osf.io/qbuk3/

In OZ14_Conc_Q_0013 (Figure 160), the pitch movement on *filachaw* takes place almost completely on the first syllable that would be stressed in the Spanish *fila*, while in *primeruchaw*, the rise on the Spanish stressed syllable [me] is less steep, but a further phrase-final rise can be identified much more unambiguously. In XU31_MT_Q_2837 (Figure 161), different lexical items, even though all of Spanish origin, can be seen to display a different prosodic behaviour within the same utterance: on *manka ladunchaw*, the rise ignores the Spanish stressed position [la] and realizes the rise instead on the phrase penult, [ðʉn]. On *tsay frentinchawraami*, instead, the Spanish stressed syllable [prɛn] is elevated, and pitch falls after it again to then produce a second rise-fall pertaining to the phrase-final tones, in an instance of the "grafted" pattern. On *murcielago*, again, it is the phrase or word penult that attracts the rise, while the Spanish stressed position [sje] is ignored. In section 6.1.8.3 we will see that forms derived from Spanish *lado* are almost never realized with an "inherited" or "grafted" pattern, while for *murcielago*, there is much more variation in the speech of speaker XU31. This indicates that the use of the marked loanword patterns is both lexicalized and speaker-dependent. In most of the words seen here, the Spanish stressed syllable, if it is open, is produced longer relative to other open syllables in its surroundings, no matter whether it is realized with the marked loanword patterns or as part of a phrase realized with one of the previously discussed patterns. This stands in contrast to what was observed about the penult in Quechua-origin words, which is not usually longer independent of peak position, but is coherent with similar duration behaviour in the Huari Spanish data (cf. section 6.1.6). This observation is relevant because it might be related to the following analytical challenge: in rising contours on bisyllabic words of Quechua origin (cf. Figure 141), the first syllable was seen to be usually almost fully low and only the second high, with the transition between them taking place "between" the syllables, especially when voiceless segments intervene between the two nuclei. This has been assumed to be due to constraints against tonal crowding overruling any which align the H tone with the penult (but cf. section 6.1.7). In several phrase-initial bisyllabic words of Spanish origin, on the other hand, a substantial part of the rise takes place clearly already on the first syllable, suggesting that this might be a "grafted" pitch accent and that its faithful realization is unimpeded even by the constraint against tonal crowding. One example for this was the contour realized on *riku* in (130)/Figure 159. Further ones occur on *nubi* in OZ14_MT_Q_1964 ((133)/Figure 162) and *bolsan* in XU31_MT_Q_1081 ((134)/Figure 163).

(133) OZ14_MT_Q_1964

nubi	*trenu*-kuna-wan	ka-ykaa-n
cloud	thunder-PL-INST	COP-PROG-3

"there is a cloud with thunder"

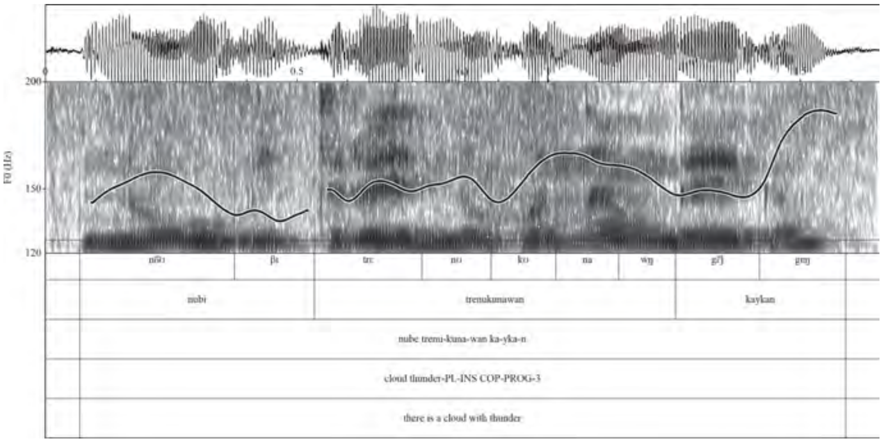


Figure 162: OZ14_MT_Q_1964[338] (declarative with a falling and two rising contours).

(134) XU31_MT_Q_1081

bolsa-n apta-ku-shqa

bag-3 grab-MED-PRTCP

"grabbing a bag"

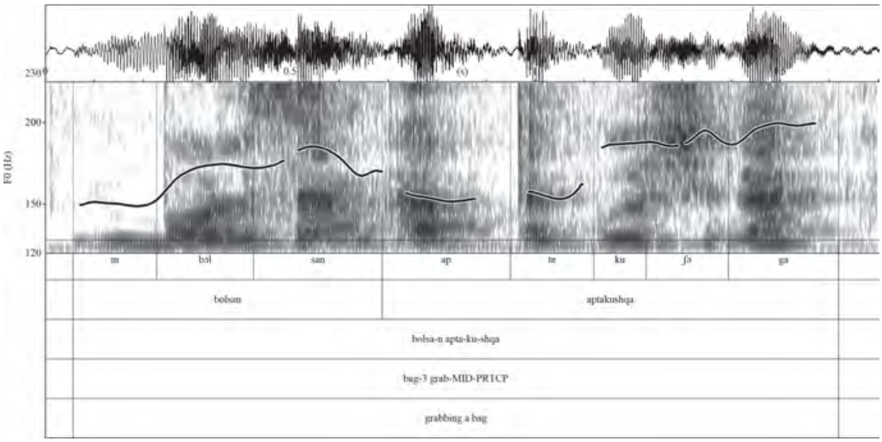


Figure 163: XU31_MT_Q_1081[339] (part of declarative consisting of two rising contours).

338 https://osf.io/rcqxa/

339 https://osf.io/fr2py/

Nubi in Figure 162, in addition to the pronounced rise fully taking place on the initial syllable, also lengthens that syllable considerably, and the vowel is in fact realized as a diphthong [o͡ʊ]. In Figure 163, the vowel in the first syllable of *bolsan* is not lengthened like that and the syllable is not actually longer than the others. This is possibly because a CV:C syllable would go against Quechua syllable structure, with CV: or CVC being maximal, and potential CV:C syllables usually reduced to CVC (Parker 1976: 56). However, the syllable here is preceded by a substantial period of nasalized phonation (segmented separately in the syllable tier in Figure 163 and transcribed with <m>), resulting in the auditive impression that the initial voiced obstruent is strongly prenasalized, yielding [mbɔlsan]. Such phonetic prenasalization, called "nasal leak", occurs frequently in voiced obstruents in phrase-initial position in languages where voiced obstruents are characterized by having negative voice onset time. It is one possible solution to the articulatory problem of having to build up subglottal pressure to maintain voicing long enough for it to be recognizable while at the same time being forced to release the labial closure due to increasing supraglottal pressure, the other being devoicing of initial voiced obstruents (cf. Solé 2012, 2014). It might be considerable that here, this articulatory phenomenon is actually exploited to enable the realization encountered in Figure 163, where the nasal segment serves as carrier for the initial low tone, like an auxiliary TBU, so that the high tone can be realized on the syllable bearing the accent in Spanish and the constraint against crowding is not violated. The nasal would then be epenthetic and tone-bearing. More controlled studies will have to be done to confirm or disconfirm such a hypothesis, but assuming the account is adequate, it points towards moras, not syllables being TBUs here. The stressed syllables in Spanish loanwords could then be bimoraic (cf. Buchholz & Reich 2017 for a similar account of a rhythmic phenomenon in Huari Spanish), and this would be reflected in their preferential realization as lengthened, diphthongized, or, where that isn't possible, with the insertion of an epenthetic sonorant segment like the prenasal. The prohibition of tonal crowding should then refer to the mora as TBU, and we should expect to see similar effects in Quechua-origin words with phonologically long vowels. However, this prediction does not always hold: e.g. in Figure 136, the word *rikaa*, with a phonologically and phonetically long final vowel, realizes the falling contour with the penult high, and all of the final syllable low. On the other hand, in Figure 129, the rise on *uusha* is realized on the first syllable containing the long vowel, parallel to the examples with Spanish loanwords above, although that syllable is not actually longer than others in the same utterance. Recall also from section 6.1.6 that in some of the utterances with falling contours, the longer the final syllable relative to the penult, the further to the right the peak is realized. Taken together, this is at least some evidence for tonal placement being sensitive to the length of the sonorant segments involved. Note that the evidence also suggests that phonological or morphological length by itself often seems not enough for a syllable to exhibit behaviour indicative of it containing two TBUs. It must instead be actually phonetically long relative to other syllables, or diphthongized, or otherwise enhanced. It is quite likely that the observed variability is again indicative of several underlying patterns. Further research is needed to come to a more definitive conclusion, especially since the tonal crowding account with syllables as TBUs has proven very useful elsewhere in the analysis.

6.1.8.2 A typological comparison

Loanword prosody is still a somewhat underresearched branch of loanword phonology. Nevertheless, several studies have compared various cases and produced initial typologies (e.g. Kubozono 2006; Kang 2010; Davis et al. 2012). Davis et al. (2012) document case studies varying both source and target language along the traditional classification of “stress”, “tone”, and “pitch accent” language and present an initial taxonomy sorting them along three binary parametres: 1), whether features of the source language play a role for the assignment of loanword prosody in the target language ([+/-SL] in their shorthand); 2), whether loanword prosody makes use of rules or constraints specific to loanwords and not occuring elsewhere in the target language ([+/-sp-loan]); and 3), whether the prosodic pattern in the adapted words is determined by segmental or suprasegmental features ([+/-pros])[340] (cf. Davis et al. 2012: 15). They also note further differences in loanword adaptation strategies: loanwords may recreate prosodic properties of the source word using the means of the

340 At first sight it might look as if there must be overlap between the parametres, especially between 1) and 3); however, they need not be dependent on each other: loanwords in Japanese are nearly universally accented following a rule first formulated by McCawley (1968): accent falls on the antepenultimate mora. According to Kubozono (2006), the same rule, reformulated in terms of a trochaic moraic foot and final syllable extrametricality (and thus of course effectively identical to the Latin Stress Rule, cf. Hayes 1995) is also valid for the majority of accented Japanese words and compounds. Because loanwords are assigned accent uniformly according to this rule, independent of source language and where stress might lie in the source language, Davis et al. (2012: 25–26) class loanword prosody in Japanese as being independent from source language prosody with regard to parametre 1), and because the rule assigning accent is quantity-sensitive, they class it as being determined by suprasegmental rather than segmental features with regard to parametre 3). An objection might be that if accent placement is determined by quantity, a feature of the source language, namely the syllable structure of the loanword, does play a role. However, this turns out to not be the case: taking *sandwich* → *san.do.ítt.chi* (μμ.μ.μ́μ.μ) as an illustration (cf. Kubozono 2006: 1141), because Japanese has such a restricted syllable structure, source language syllable structure is completely broken up, and only this segmentally refitted version of the loanword then undergoes accentuation, counting syllables not existing at all in the source language (the final syllable here) in the assignment process.

target language on a syllable-by-syllable basis (as happens with English loanwords in Hong Kong Cantonese, cf. Davis et al. 2012: 20–21), or they may recreate an overall tonal melody for words that is frequent in the source language but absent in the target language and generalize it to all loanwords from that source language, thus leveling differences between accent positions present in the source language (Japanese loanwords in Taiwanese Southern Min, cf. Davis et al. 2012: 22–23). They also state that a language may base its loanword prosody on phonetic properties of the source words, as in the case of English loanwords in Lhasa Tibetan: English words beginning with a voiced obstruents are realized with an L tone on the first syllable of the loanword (*bottle* → *po to ra* [L.H.H]), but words beginning with a voiceless obstruent receive an H tone on the first syllable (*police* → *pu li su* [H.H.H]; the initial syllable of the word being the only place where there is a possible specification of tone being either H or L, all other syllables always being H), ignoring the stress position in the source word and also leveling the voicing distinction segmentally, which does not exist in Tibetan. Their interpretation is that this is due to the well-known articulatory property of voiceless obstruents of raising pitch in the following vowel (cf. Hombert et al. 1979; Hanson 2009; Kirby & Ladd 2016; Ladd & Schmid 2018), so that the initial high tone preserves as much of the phonetic properties of the source word as possible given the system of the target language (cf. Davis et al. 2012: 19–20).

Applying this taxonomy to the Spanish loanwords in Quechua, with regards to 1), both the "inherited" and the "grafted" pattern clearly pay attention to prosodic features of the source language, namely the acoustic prominence of the Spanish stressed syllable as realized by tonal movement and duration. They treat it differently though: in the "inherited" pattern, what happens on the Spanish stressed syllable is somehow equated to the rise taking place in the rising contour, with the rise then produced on a syllable where it would not normally appear. In the "grafted" pattern, what happens on the Spanish stressed syllable is treated as something separate from the phrasal intonation of Quechua, but it is still emulated prosodically, as an additional pitch event localized to that syllable. In cases where the Spanish loanword is not realized with either the "inherited" or "grafted" pattern but treated like any other word, the Spanish stressed syllable is at least still produced with longer duration if it is open. Thus, the source language prosody always plays a role, but it can surface in two different ways tonally or just via duration ([+SL]). With regards to 2), the "inherited" and "grafted" patterns clearly constitute prosody specific to loanwords, since both require an irregular lexical specification of a prominent position. This specification is lexicalized, since it can cause the pitch event on the Spanish stressed syllable to appear on syllable positions in Quechua words that are outside the Spanish three-syllable window. The "grafted" pattern additionally has a tonal specification that is unique to these words. In those cases where words of Spanish origin are not realized with the "inherited" or "grafted" pattern, no loanword-specific prosody is employed, because long vowels are a regular feature in Quechua. However, the option for a loanword-specific prosody exists and is frequently employed, so overall we may classify our case as [+ sp-loan]. With regards to 3), it is clear that it is suprasegmental and not segmental features which determine the resulting prosody of these loanwords, thus [+pros]. The property of accentedness itself may be taken as that which effects the special prosody of these loanword, since the speakers are all bilingual, we rarely find any "wrong" stress placement when they speak Spanish, and they can therefore be safely assumed to have robust knowledge about stress positions in their Spanish lexicon. Yet the fact that the "inherited" pattern seems to come up mostly in rising contours, and that pitch movement and duration are not equally distributed in the production of these words as loanwords in Quechua seems to indicate that the speakers take the phonetic correlate of this accentedness not to be an inevitable property of the lexical word form – which is also reflected in their usage when speaking Spanish.

In sum, it is certainly possible to apply Davis et al's (2012) taxonomy to our case. However, in the data here, several realizational varieties for these loanwords can be found, some of which would require variant classifications according to the taxonomy.

The loanword intonational patterns can be characterized in relation to the ones seen before. The tonal alignment patterns make reference to different prosodic landmarks, to the phrase-peripheral domain boundaries, to a phrase-internal boundary of the next lower domain, the prosodic word, and to the word penult of the strongest word in the phrase, a regularly determined prominent position independent of the lexical identity and morphological makeup of the words involved. To this, only in words of Spanish origin, the loanwords patterns add a lexicalized and thus not fully regular prominent position, the stress position "inherited" from the Spanish word (e.g. (135), with the Spanish stressed syllable marked with an asterisk).

(135) "Inherited" accent on a Spanish-origin word (from OZ14_Conc_Q_0331, cf. (129)/Figure 158)

[tɛr se* rə t͡ʃu]$_{\text{PhP}}$

|

L$_\phi$ H$_\phi$ --------->

The "grafted" pattern even produces a pitch accent at this position that is additional to the phrase tones. In other words, while the "inherited" pattern affects the alignment of the phonological phrase tones of Quechua in providing a unique lexically specified position the H tone is attracted to, the "grafted" version causes pitch accent tones to be additionally assigned for that position. In this sense, the "grafted" pattern is in theory not an alternative to any of the patterns described before, but an addition, as it can possibly occur together with any of them. The additional pitch accent of the "grafted" pattern can here be analyzed as LH* in ToBI-notation, because it seems to be characterized by a rise with a peak on the accented syllable. This is plausible in light of the findings from chapter 5 on Spanish, where this pitch accent was found to be the only regular pitch accent on stressed syllables in a variety of contexts. Because many of the words with a "grafted" accent (e.g. *abejakunapa* in Figure 157 or *frentinchawraami* in Figure 161) clearly exhibit a fall or return to low after this LH* accent on the Spanish stressed syllable, which then extends until the rise pertaining to the end of the phrase, another L tone should be assumed there. The easiest way to account for this is by assuming that alignment rankings for the LH* tones are as in the Spanish "main" variant (cf. section 5.3.2.1), and that their alignment ranking dominates that of all the Quechua phrasal tones, so that they are all realized sequentially following the pitch accent tones. Then, the dip following the LH* peak is due to the presence of the L tone that is otherwise initial in Quechua phrases. For this, the LH* tones must be specified as different from the phrasal tones so that different constraints can apply to them. Since the tones in both Spanish and Quechua are assigned at the level of the PhP, the differential specification might have to make reference to the two different languages. That would in fact capture the notion that these accents are somehow intrusive.[341]

(136) "Grafted" pattern on a Spanish-origin word (from XU31_MT_Q_2837, cf. (132)/Figure 161) with LH* pitch accent and two different sets of tones
[t͡se͡j]$_{]\omega}$ prɛn* tin t͡ʃu ɹa mi]$_{PhP}$
|
L_S H_S L_ϕ-----> H_ϕ L_ϕ

Using biological metaphors like "inherited" and "grafted" evokes images of crossbreeding, and of the admixture of elements from different pedigrees that do not really fit well together because they belong to different "systems". At first glance, this seems apt in the situation at hand, where two "language systems" are so obviously in contact with each other and where we as linguists can so easily discern the "foreign" provenance of loanwords in the speech of these speakers because we are well-acquainted with the historiography of the two languages in question that treats them as separate for the majority of their documented and projected existence. However,

341 Considering that intonational contours can be lexicalized on individual lexical items when occurring sufficiently often under similar context conditions (Calhoun & Schweitzer 2012; Schweitzer 2012; Schweitzer et al. 2015), another possibility might be that the Spanish pitch accent tones have become part of the lexicalized representation. Then the alignment could refer to these "lexicalized" tones vs those assigned by the PhP.

it is worthwhile to switch perspectives here and to consider the phenomenon from a typological viewpoint. From the view of languages like (Northern varieties of) Basque (Elordieta 1998; Hualde 1999; Hualde et al. 2002), Tokyo Japanese (Pierrehumbert & Beckman 1988; Kubozono 2008; Kawahara 2015), or Turkish (Levi 2005; Kamali 2011), having a subset of the lexicon, so-called accented words, with prosodic specifications *in addition* to the ones valid for all other, unaccented, words or phrases, is par for the course. To demonstrate: in (Tokyo) Japanese, accent manifests as a H tone on the accented mora followed by an L tone on the following one, i.e. a fall that has been analyzed as H*L in Japanese versions of ToBI (J-ToBI, cf. Venditti 2005). AP-initially,[342] there is a rise with an L tone on the initial and a H tone on the following mora. If the lexical accent is initial, it supersedes the AP-initial rise (cf. (137)a). A mora left unspecified for tone by either of these mechanisms receives the tone from the rightmost specified mora (variously analyzed as tonal spreading, copying or interpolation, cf. Kawahara 2015: 449–450 and the works cited there). These mechanisms fully determine the tonal shape of all Japanese words at the word- and AP-level (cf. for trisyllabic words + an unaccented suffix in one AP[343] (137)a-d). Accented words need an additional lexical specification (location of the accent) that unaccented words are lacking, but behave the same otherwise. Accent position is almost exclusively cued by pitch; duration is not a good cue for it and also signals lexical contrasts non-culminatively (cf. Beckman 1986; Kubozono 2015 amongst others).

(137) Tokyo Japanese (adapted from Kawahara 2015: 448–452)

a. ínochi-ga (city-NOM) (initial accent)

i no chi -ga

H L

342 In all the works on Japanese, Basque and Turkish cited in this section, an Accentual Phrase (AP) or Phonological Phrase (PhP) is assumed to be the minimal prosodic unit at which tones are assigned and where accent is culminative. This manifests e.g. in the analysis for Japanese in that only one accented word can occur in an AP, optionally with additional unaccented words. It should be kept in mind that the phrase-initial (and/or phrase-final, in Turkish and Quechua) tones properly belong to the phrase, and not to the word, as the pitch accent does. So in an AP with two unaccented words, like *yamada-ga uta-tte-ru* (Yamada-NOM sing-PROG-COP “Yamada is singing”), there is only one phrase-initial rise on the first word, and the second word is realized just as a high plateau, without further pitch events (cf. Venditti et al. 2008: 459). Adjusting for their specific prosody, this holds also for Basque and Turkish. Since APs frequently map to a single word, the regular phrase-final or –initial tones have often been interpreted as a regular lexical accent, e.g. in traditional descriptions of Turkish (cf. Féry 2017).

343 Higher-level prosodic units can add further tones leading to interactions with those given in an actual utterance.

b. kokóro-ga (heart-NOM) (penultimate accent)
ko ko ro -ga
L_ϕ H L

c. atamá-ga (head-NOM) (final accent)
a ta ma -ga
L_ϕ H_ϕ H L

d. miyako-ga (city-NOM) (unaccented)
mi ya ko -ga
L_ϕ H_ϕ

Northern Biscayan Basque has a surprisingly similar system:[344] words are divided into accented and unaccented ones, the accent always takes the form of a fall after the accented syllable (H*L) and is free to occur on any syllable in the word except for the final one in most cases, there is a word-initial rise, and tones spread across unspecified syllables (leftwards from the accented syllable, according to Hualde 1999: 950). In a position that is not directly preverbal (the focal position according to Hualde et al. 2002, in which unaccented words also receive a falling H*L accent in final position), accented words boast both a surface pitch event and a word-level tonal specification (the position of the accent) in addition to the ones found on unaccented ones (cf. (138)a-c; as in Japanese, higher-level tones will interact with the AP-level ones). Duration is also not a good cue for accent position, according to Hualde & Beristain (2018), although it is not used to mark lexical contrasts.

(138) Northern Biscayan Basque (adapted from Hualde 1999; Hualde et al. 2002)

a. gixon-an-ari (man-GEN.SG-DAT.SG „to the one of the man“) (unaccented)
gi xo na na ri
L_ϕ H_ϕ

344 Basque prosody is described as varying enormously despite the geographically limited area where it is spoken. Hualde et al. (2002: 548–549) profess a belief that prosodic differences can be delineated and spatially contained by establishing isoglosses, but also hint gently at the insight that the dialectological enterprise of pinning down the prosodic features of Basque both geographically and typologically must needs make use of idealization.

b. gixon-án-ari (man-GEN.SG-DAT.PL „to the ones of the man“) (antepenultimate accent[345])

gi xo na na ri

L_ϕ H_ϕ H L

c. gixón-an-ari (man-GEN.PL-DAT.SG „to the one of the men”) (preantepenultimate accent)

gi xo na na ri

L_ϕ H L

According to Levi (2005); Kamali (2011), Turkish also shares the distinction between accented and unaccented words and the accent in accented words is a fall with an H tone on the accented syllable followed by an L tone (analyzable as H*L). AP-initially it does not have a rise but an L tone, and APs in prenuclear position[346] have a final H tone. This has traditionally been taken to be a regular oxytone accent on those words that are classed as unaccented in Levi (2005) and Kamali (2011), cf. Göksel & Kerslake 2005). Thus in prenuclear position, accented words also require an additional lexical specification in comparison to unaccented ones and add a tonal specification and an additional pitch event to the phrase in comparison to a phrase made up only of unaccented words (cf. (139)a-d). In nuclear position, APs are realized with an initial and final low tone. This manifests as a low plateau followed by a final fall in APs with an unaccented word and accented words are differentiated from unaccented ones in that they realize this fall earlier, after the accented syllable which is sometimes preceded by a compressed rise (cf. Kamali 2011: 74–77). Duration is not a reliable cue for accent position and is also not lexically contrastive (cf. Levi 2005).

345 This variety of Basque has “preaccenting” suffixes (Hualde 1999): suffixes that cause the preceding syllable to be accented. Preaccenting often distinguishes plural suffixes from segmentally identical singular ones.

346 The literature on Turkish splits IPs up into a prenuclear, nuclear, and postnuclear part, based on an interplay of syntactic, prosodic and information structural criteria (cf. Féry 2017: 251–253). Here it is only relevant that the right edge tone of APs differs with position: in prenuclear position, it is H_ϕ, in (post)nuclear position it is L_ϕ.

(139) a. Turkish (adapted from Göksel & Kerslake 2005; Kamali 2011)
okuldá-ymış-lar (school-EVID-PL, "apparently they are/were at school", preantepenultimate accent on an AP in prenuclear position[347])

o	kul	da	y	mış	lar
|		|	|		|
L_ϕ		H	L		H_ϕ

b. limónlu-ya (Limonlu-DAT "to Limonlu", antepenultimate accent on an AP in prenuclear position)

li	mon	lu	ya
|	|	|	|
L_ϕ	H	L	H_ϕ

c. bunal-an-lar-ı (get.overheated-REL-PL-ACC "those that get overheated", no accent on an AP in prenuclear position)

bu	na	lan	la	rı
|				|
L_ϕ				H_ϕ

d. yöndlendir-íyor (forward-IMPERF "s/he forwards", penultimate accent on an AP in nuclear[348] position)

yön	len	dir	i	yor
|			|	|
L_ϕ			H	L_ϕ

These languages as well as others[349] make a regular distinction between unaccented and accented words, with accented words a minority in the lexicon. The

347 Turkish is also usually analyzed as having a class of preaccenting suffixes. The evidential copular suffix *–(y)mIş* (capitalized vowels indicate a vowel subject to harmony) is such a suffix. The imperfective *–(Í)yor* causes the accent to fall on its first syllable, as indicated by the orthographic accent. In a chain of suffixes, the leftmost preaccenting one determines accent position (cf. Göksel & Kerslake 2005: 29–33).

348 I have not found any information on what happens when a word in prenuclear position is accented on the penult: the L of the H*L pitch accent and the phrasal H_ϕ would then possibly compete to occupy the only remaining rightmost TBU. Because of this lack of information, use a penult-accented word in nuclear position as example here. In that position, the right-edge phrase tone of the AP is L_ϕ, and no such potential conflict ensues.

349 For example, Swedish and some of the Germanic varieties in the South Low Franconian and Central Franconian tone accent area, following the privative analyses by Riad (1998) for Swedish, Gussenhoven (2000a, 2000b) for Roermund and Gussenhoven & van der Vliet (1999) for Venlo Dutch, Peters (2008) for Hasselt Dutch. Cf. also Féry 2017: 194–200 for a discussion. Modern Hebrew also differentiates three noun classes according to stress position: words with final stress that always shifts to the final syllable when suffixes are added, words with penultimate stress that shifts in this way, and words with a lexically determined stress position that does not shift under

difference in type frequency between the groups is also reflected in terms of structural markedness: "accented" words are more marked in that they require further specifications than "unaccented" ones. That is also true in Huari Quechua, when comparing the loanword patterns with the others: the "inherited" and "grafted" patterns require a further tonal specification at the word level (a lexicalized accent position), and in the "grafted" pattern this also translates to additional tones and an additional pitch event. In this light, despite knowing about the historical origins of the "Spanish" words, we should not be averse to considering this particular variation as simply one prosodic option (or two closely related options) available for a subset of the lexical items the speakers have at their disposal. As the discussion on loanword prosody has shown, despite variation in their realization it is clear that the loanwords are not produced just as if the speakers were speaking Spanish. Phonological phrases that contain them form a particular prosodic subtype in Quechua sharing attributes both with phrases with only "native" words and with Spanish stressed words. It makes sense to call these words an "accented" class in Quechua.

The distinction between accented and unaccented words also seems to interact in a non-trivial way with loanword phonology in other languages: according to Kubozono (2008: 167), only 29% of native Japanese words are accented, but 93% of loanwords[350] are. In a similar vein, Davis et al. (2012: 14) attest that the marked or accented class of Modern Hebrew words, which is smallest among native words but to which most loanwords belong, is a modern innovation. They suspect, based on Zuckermann (2003), that it might have come into existence through the large share of vocabulary from Germanic and Slavic languages the modern language has adopted. In Turkish, unaccented words form the majority of native word stems, but most loanwords are accented (Göksel & Kerslake 2005: 26–27). In the Basque case, Hualde (1993, 1999: 984) even argues that the present Western Basque accentuation system (including Northern Biscayan) arose from a historical development in which a system that originally had no word accent changed to the present distinction between accented and unaccented words due to the large-scale adoption of accented loanwords from Latin and Romance languages. If that is correct, the

suffixation. The last type is seen as marked, because it requires further constraints in addition to those acting on the first two classes (cf. Bat-El 2005; Davis et al. 2012). Becker (2003: 45–46) calls the first two classes 'unaccented' and the third 'accented'. Loanwords from languages with lexical stress such as Germanic or Slavic languages or Arabic usually preserve their stress position and are of the last class, which has by far the fewest members among native words and might even have only been formed under the influence of those languages (Davis et al. 2012: 17; Zuckermann 2003). Hebrew is not usually analysed as a "pitch accent" language.

350 From languages other than various historical stages and varieties of Chinese, the so-called Sino-Japanese words, that form a large and integral part of the modern Japanese lexicon. In them, the share of accented words is 49% (Kubozono (2008: 167)).

beginning stages of this process could probably be imagined as a scenario quite similar to the Quechua situation described here. Basque also shows a remarkable prosodic variety across the different regions where it is spoken, despite occupying a relatively small geographical area. Besides the pitch accenting system described for Northern Biscayan, there are also varieties without a distinction between accented and unaccented words, some with a fixed penultimate accent position and some where there is more than one accent shape (a rising one besides the falling one, cf. Hualde & Beristain 2018). This is also not wholly unfamiliar from our ongoing discussion on Quechua, except that here we find a similar breadth of variation in a corpus made up only of the speech of speakers hailing from and residing in basically the same locality, the town of Huari and its environs. This suggests that the space of possibilities in both cases has similar dimensions. Other interesting parallels to the three languages here briefly discussed are that in Huari Quechua as well as in Japanese, Basque and Turkish (to a lesser degree), tones tend to copy (or spread) across tonally unspecified syllables, leading to plateaux and low stretches, and that duration is not a good cue for marking accent position (apart from in the words of Spanish origin in our data). This set of languages is however also separated by some properties (cf. also Ito 2002 for a prosodic comparison of Basque and Japanese): in Huari Quechua and Japanese, duration is employed in lexical contrasts, but not in Basque and Turkish. In Quechua and Turkish, phrase-initial L tones occur, while in Japanese and Basque there are initial rises (LH_{ϕ}). While Japanese, Basque and Turkish all have a lexical accent that has been analyzed as H*L, the pitch accent identified in the "grafted" pattern is LH*, like the most frequent pitch accent in Huari Spanish. The Quechua "grafted" pattern is similar to phrases with accented words in Turkish in that it forms a pitch accent in addition to the tone(s) at the right edge of the phonological phrase, while the "inherited" pattern is more similar to the Basque case, where it is the same phrase tones that appear at the right edge of phrases (only if they are in focus position in Basque), and not additional ones, that are employed to mark the accented position in phrases with accented words. Turkish and Huari Quechua also share the property of having two options for the rightmost phrase-level tone, H_{ϕ} and L_{ϕ}, differentiated by whether the AP occurs in prenuclear of nuclear position in Turkish and distinguishing the rising from the falling contour in Quechua.

So far, we have only compared AP/PhP- and word-level prosody here. At higher levels of prosodic structure, further points of commonality and separation can be added: e.g., while Kamali (2011: 86–87) argues that in Turkish there is only a single higher-level boundary tone, H%, which is optional (but cf. Göksel & Kerslake 2005: 35–39 for a differing account), Japanese has an array of complex higher-level tones (L%, H%, HL%, HLH%) called "boundary prosodic movements" (BPM; Igarashi et al. 2013; Igarashi 2015, cf. also Venditti et al. 2008: 471–476). They encode pragmatic contrasts. In Huari Quechua so far, there is no evidence for additional higher-level tones. However, this should be taken as a preliminary finding: as Venditti et al. (2008: 472–473) point out, the full inventory of BPMs in Japanese has until recently been overlooked (including by Pierrehumbert & Beckman 1988), and this is not least because some of them are exceedingly rare: in a corpus of more than 178,000 phrases, the two least frequent BPMs, HL% and HLH%, only occurred 419 (0.2%) and 14 (0.008%) times. Their distribution is also extremely dependent on genre and context. That makes it not unlikely that they aren't present, or haven't been recognized, in our Huari Quechua sample (cf. the singular example of KP04_MT_Q_0393 in section 6.1.7). Northern Biscayan Basque and Turkish assign phrase accents based on a mixture of prosodic and syntactic conditions: in Turkish, all APs in prenuclear position receive a final rise (H- in Kamali 2011); in Northern Biscayan Basque, even unaccented words receive a final pitch movement that is identical in form to the falling accent when they are the last word in a phrase that is in focus position, i.e. directly preceding the verb, according to Hualde et al. (2002: 549–550). A purely prosodic version of this condition, where only the strongest word in the phonological phrase has what could be called an accent in a completely regular position, is quite a close-fitting description of the Quechua pattern seeking alignment with the word penult.

Discussions of the prosody of Japanese, Basque, and Turkish often group them together under the label of "pitch-accent languages" because of their several shared properties (cf. Féry 2017). Should Huari Quechua now be added to this group as a further member, based on the number of intriguing similarities we have encountered? Despite these similarities, it should be kept in mind that in our data, different prosodic variants of both Spanish and Quechua assume prosodic configurations that each on their own resemble what has been described as quite diverse prosodic "types" in the literature. In the next section we will see that the same speaker may produce the same lexical item of Spanish origin in the same conversation one time with an "inherited" accent, another with a "grafted" one, and yet another with one of the three patterns that would also be used on phrases composed only of words of Quechua origin. To the extent that Quechua prosody is similar to that of the three other "pitch accent" languages, this would then be more valid for the first two instances than for the third one, while phrases produced with an exclusively edge-oriented alignment pattern would perhaps more closely resemble French or Korean. Thus I would hesitate to say that simply because we can analyze the prosodic behaviour of these Spanish-origin words in parallel to the distinction between unaccented and accented words in other languages, it is reasonable to classify Quechua as belonging to a prosodic "prototype" like the T-type pitch accent languages proposed in Hualde et al. (2002). In addition, we simply do not know very much about the tonal properties of very many languages, so that creating prototypes because two or three languages whose prosody we have studied a little more seem to show parallels in some analyses might be premature. The theoretical cast of the analysis itself is also crucial: Igarashi (2015: 561) points out that French has been described both as having an accent-based prosody (e.g. by Post 2002) and an edge-based prosody (e.g. by Jun & Fougeron 2000, 2002), which are elsewhere seen as representing two different prosodic types. On the other, the idea of prosodic types themselves seems difficult to maintain: Hyman (2006) argues that the exponents of "pitch accent languages" are too heterogeneous for it to be a useful type, and Beckman & Venditti (2010) expand this assessment also to "tone languages" and "stress languages" after detailed case studies, voicing the suspicion "that the appearance of prototypes comes from looking too closely at just one or two of the functions in which tone participates, as well as from being thoroughly immersed in the consensus assumptions of specialists in just one or two Sprachbund regions" (Beckman & Venditti 2010: 642). Hyman (2014) essentially concurs by proposing property-based (instead of type-based) typological comparison.

Perhaps the problem in assigning types lies more in assuming that they are valid for whole "languages" rather than in the observation itself that certain properties seem to cluster and form recurring patterns. It would then be a much more interesting question to ask what it is about these properties that makes it likely or preferable for them to occur together in some system, independent of whether this system is a historical language or just one possible variant besides others in the repertoire of an individual. In the following we will look at the loanword patterns from a quantitative perspective, in order to see how they vary by speaker, token and type of discourse.

6.1.8.3 Distribution of the loanword patterns across speakers and lexical items

To conclude the section on loanword patterns we'll look at the occurrence of pitch movements determined by the Spanish stress position in a subset of the corpora, the seven *Conc*s and *Maptask*s. The purpose here is not to provide a comprehensive quantification of their frequency of occurrence but to demonstrate that pitch movements determined by the Spanish stress position vary in our corpus at least along the following dimensions: according to speaker and according to lexical item; possibly also according to conversation type. In the following two tables, the occurrence of Spanish-origin word tokens is given and their frequency relative to all word tokens. Their realization is classified as either without pitch movement due to the Spanish stress position (i.e., like any other Quechua word), with pitch movement due to the Spanish stress position (i.e. grouping the "inherited" and "grafted" patterns together), or as indeterminable. The counts are separated by speaker for the two sets of corpora of *Conc* (Table 22) and *Maptask* (Table 23). The definition of "word" here used is that of the Quechua morphological word: any stem together with its suffixes. Words were classed as indeterminable if either there was no or not enough pitch movement in the phrase overall to classify it, or if what was found could be interpreted to represent one of the Quechua alignment patterns just as much as tonal alignment relative to the Spanish stressed syllable with no way of telling them apart (e.g. a rising contour on an individually phrased trisyllabic word where the tonal transition takes place on the penult and the penult also corresponds to the stressed syllable in the Spanish word). The counts here are thus necessarily the result of some interpretation, but they broadly indicate differing strategies between speakers.

Table 22: Prosodic patterns on Spanish-origin word tokens in the seven Quechua *Conc* corpora. The counts in the last column exclude word tokens classed as "indeterminable".

Speaker	**Corpus (length in seconds)**	**Total number of words by that speaker**					**Ratio number of Spanish-origin words /all words**	**Ratio loanword pattern words / Spanish-origin words (without "indeterminable")**
		Spanish-origin words				**Total**		
		***No* pitch movement due to Spanish stressed syllable**	**Pitch movement due to Spanish stressed syllable**	**Indeterminable**	**Total**			
OZ14	Conc	2	9	5	16	31	0.52	0.56
QZ13	(180.7)	12	9	11	32	48	0.67	0.28
QF16	Conc	4	0	3	7	48	0.15	0
SG15	(218.8)	1	7	8	16	79	0.20	0.44
TP03	Conc	2	0	1	3	39	0.08	0
KP04	(172)	4	0	3	7	70	0.10	0
AZ23	Conc	33	2	11	46	70	0.66	0.04
ZZ24	(222)	8	1	2	11	20	0.55	0.09
XU31	Conc	6	1	1	8	40	0.2	0.13
OA32	(192.3)	8	1	0	9	64	0.14	0.11
ZR29	Conc	6	0	5	11	36	0.31	0
HA30	(300.2)	7	3	0	10	117	0.09	0.3
XQ33	Conc	3	0	3	6	29	0.21	0
LC34	(193.6)	3	0	2	5	36	0.14	0

Table 23: Prosodic patterns on Spanish-origin word tokens in the seven Quechua *Maptask* corpora. The counts in the last column exclude word tokens classed as "indeterminable".

Speaker	Corpus (length in seconds)	Total number of words by that speaker					Ratio number of Spanish-origin words /all words	Ratio "inherited" accents / Spanish-origin words (without "indeterminable")
		Spanish-origin words				Total		
		No pitch movement due to Spanish stressed syllable	Pitch movement due to Spanish stressed syllable	Indeterminable	Total			
OZ14	MT	7	10	8	25	151	0.17	0.4
QZ13	(234.2)	2	3	5	10	80	0.13	0.3
QF16	MT	7	2	2	11	103	0.11	0.18
SG15	(148.4)	1	0	0	1	11	0.09	0
TP03	MT	5	1	8	14	198	0.07	0.07
KP04	(303)	17	8	14	39	301	0.13	0.21
AZ23	MT	11	2	2	15	95	0.16	0.13
ZZ24	(176)	2	2	1	5	50	0.10	0.4
XU31	MT	57	16	4	77	622	0.12	0.21
OA32	(777.6)	12	5	5	22	181	0.12	0.23
ZR29	MT	2	0	1	3	34	0.09	0
HA30	(262.8)	16	4	17	37	205	0.18	0.11
XQ33	MT	1	0	1	2	103	0.02	0
LC34	(328)	11	3	12	26	266	0.10	0.12

Some speakers (OZ14, QZ13, SG15) tend to realize a substantially higher number of the Spanish-origin word tokens they produce with pitch movement on the Spanish stressed syllable than others (AZ23, TP03, ZR29, XQ33), while the rest seems to be in-between. However, we also see that the counts differ according to the corpus type (*Conc* or *Maptask*). Two factors must be noted related to the nature of the experimental tasks: firstly, the proportion of Spanish-origin word tokens that are produced also varies by speaker and corpus type, thus varying the number of opportunities at which either alignment pattern could be produced. Although the central objects in both tasks were the same and controlled for by the experimenters, speakers were still free in their choice of name for these objects and of all other lexical items they used in the task. Secondly, speakers speak more or less, not only by individual preference, but also because of the constraints of the task: in the

Maptasks, it is usually the speaker with the path already drawn on the map who will speak more. In *Conc*, it depends on the course of the game: if a speaker begins the game who happens to memorize all or nearly all of the locations of the objects correctly, they will speak much more than the other speaker, yet if their game is more balanced, so their speaking portions will tend to be. Table 24 gives the counts for each speaker individually, but pools the data from both corpora together, somewhat correcting for these two factors.

Table 24: Ratios (in %) of Spanish-origin words against all words, and of pitch movement classified as due or not due to the Spanish stressed syllable, or indeterminable, on Spanish-origin word tokens, across both *Conc* and *Maptask* taken together.

Speaker	**Corpus**	**Ratio Spanish-origin words / all words (%)**	**Pitch movement due to Spanish stressed syllable / Spanish-origin words (%)**	**No Pitch movement due to Spanish stressed syllable / Spanish-origin words (%)**	**Indet. (%)**
OZ14	MT+ Conc	22.5	46.3	22	31.7
QZ13	MT+ Conc	32.8	28.6	33.3	38.1
QF16	MT+ Conc	11.9	11.1	61.1	27.8
SG15	MT+ Conc	18.9	41.2	11.8	47.1
TP03	MT+ Conc	7.2	5.9	41.2	52.9
KP04	MT+ Conc	12.4	17.4	45.7	37
AZ23	MT+ Conc	37	6.6	72.1	21.3
ZZ24	MT+ Conc	22.9	18.8	62.5	18.8
XU31	MT+ Conc	12.8	20	74.1	5.9
OA32	MT+ Conc	12.7	19.4	64.5	16.1
ZR29	MT+ Conc	20	0	57.1	42.9
HA30	MT+ Conc	14.6	14.9	48.9	36.2

Table 24 (continued)

Speaker	Corpus	Ratio Spanish-origin words / all words (%)	Pitch movement due to Spanish stressed syllable / Spanish-origin words (%)	No Pitch movement due to Spanish stressed syllable / Spanish-origin words (%)	Indet. (%)
XQ33	MT+ Conc	6.1	0	50	50
LC34	MT+ Conc	10.3	9.7	45.2	45.2
Avg.	MT+ Conc	17.2	17	49.2	33.8

Table 24 allows us to make a better judgment about differences between comparable[351] speakers. The most general conclusion is that no speaker produces all or even more than half of their Spanish-origin words identifiably with one of the loanword patterns (but keep in mind the sizeable number of tokens classed as indeterminable), but probably also no speaker (with sufficient data) is entirely incapable of it. Apart from that, reliable conclusions about genuine individual differences with regards to their loanword-prosodic preference can only be drawn for some speakers: e.g. speaker AZ23 has a ratio of Spanish-origin word tokens that is comparatively high and also roughly approximate to that of speakers OZ14 and QZ13. All three of them also produce a decent absolute number of tokens. Yet, her ratio of loanword pattern realizations is the lowest of all speakers with sufficient data, while OZ14 and QZ13's are among the highest, with almost 40% difference between AZ23 and OZ14. This is also not due to differences in lexical choice. The pairs of OZ14 and QZ13, and AZ23 and ZZ24, respectively, follow very similar strategies in their *Conc* tasks: they describe the locations of the object cards in the game with reference to the rows and columns that are laid out on on the table. Other speaker pairs instead prefer to describe the locations via relative reference to other cards (e.g. *tsaypa hanan kaq* "the one above that one", KP04_Conc_Q_0225), or to the border of the grid (e.g. *washa kantu-chaw* "at the edge over there", HA30_Conc_Q_0212). OZ14, QZ13, AZ23 and ZZ24 all refer to the rows and columns by numbers using

351 Comparable are those that produce enough words overall, with an approximately similar ratio of Spanish-origin words. The extremely low ratios of loanword pattern realizations by XQ33 and TP03 should e.g. be taken with some caution, because they use Spanish-origin words themselves quite infrequently, and the low and high ratios by ZR29 and SG15, respectively, because they both produce comparatively few words overall.

the Spanish ordinal numerals (*primer/o/a, segund/o/a, tercer/o/a*) as lexical stems.[352] Table 25 lists the Spanish-origin word types these speakers use in *Conc* together with their token occurrence. The majority of Spanish-origin word tokens in their two *Conc* corpora belongs to four lexical item types, *fila* 'row', and the three first ordinal numbers, *primer(@)*[353] 'first', *segund@* 'second', and *tercer(@)* 'third' (cf. rows with grey background in the table).

Table 25: Spanish-origin lexical item types with their word forms and token number of occurrence in the Quechua *Conc* corpora of QZ13 & OZ14 and AZ23 & ZZ24, respectively, according to speaker.[354]

lexical item	word form	QZ13	OZ14	AZ23	ZZ24	all
ahí 'there'	ahí	–	–	1	–	1
faltar 'to lack'	falta *lack.<u>3S.PRES.IND</u>*[352]	–	–	1	–	1
	falta-n *lack-3*	–	–	1	–	1
fila 'row'	fila	2	–	–	–	2
	fila-chaw *row-LOC*	8	5	13	3	29
cajón 'box'	kahun-chaw *box-LOC*	–	–	1	–	1
lado 'side'	ladu-n *side-3*	1	–	–	–	1
	laa-ni-n-chaw *side-FON-3-LOC*	–	1	–	–	1
primer-(@) 'first'	primer	2	–	3	1	6
	primer@	1	1	2	1	5
	primer@-chaw *first-LOC*	1	2	5	2	10
ric@ 'rich'	riiku	1	–	–	–	1
segund@ 'second'	segund@	5	2	3	1	11
	segund@-chaw *second-LOC*	2	1	4	1	8
tercer-(@) 'third'	tercer	2	–	1	–	3
	tercer@	1	2	5	1	9
	tercer@-chaw *third-LOC*	5	2	5	1	13

352 Some of the other speaker pairs also refer to the rows and columns, but use relative location terms for differentiation (*punta* 'front', *waqta* 'back') or the Quechua numerals (*huk* 'one', *ishkay* 'two', *kimsa* 'three') that do not seem to have special ordinal forms (e.g. *kimsa fila-chaw* 'in the third row' and *kimsa allqu* 'three dogs').
353 The "@" here stands in for a vowel segment that could be either *–a* or *–o*, the Spanish gender desinences. When the speakers produce these in a Quechua utterance, they do not pay much attention to gender agreement, if applicable. There are also occasional cases of gender mismatch in Spanish, e.g. *segund**o** fil**a*** (OZ14_Conc_ES_1785). This has been noted in the speech of Quechua-Spanish bilinguals before (Escobar 2000, 2011). These segments often have a centralized vowel quality, also when the tonal transition takes place on them.
354 Spanish-origin morphological glosses (verb endings) are given in italics and underlined, Quechua-origin morphology as well as all stem glosses are given in italics only.

Table 25 (continued)

lexical item	word form	QZ13	OZ14	AZ23	ZZ24	all
últim@ 'last'	ultim@-chaw *last-LOC*	1	–	–	–	1
ver 'to see'	vistes *see.2S.INDEF.IND*	–	–	1	–	1
all		32	16	46	11	105

(140) AZ23_Conc_0122
primer *fila*-chaw *terceru*-chaw *primer* *fila*-chaw *terceru*-chaw pitu
first row-LOC third-LOC first row-LOC third-LOC whistle
"in the third one in the first row, in the third one in the first row [is] the whistle"

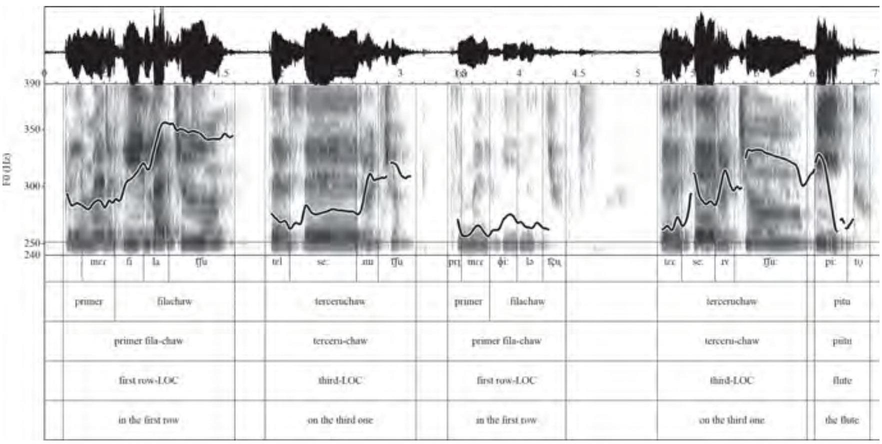


Figure 164: AZ23_Conc_Q_0122[355] (declarative with four rising and one falling contours).

This section has already seen a number of examples from QZ13 and OZ14's *Conc* with loanword pattern realizations on these words. For comparison, see AZ23_Conc_0122 ((140)/Figure 164), where almost all of these lexical items appear in phrases without the Spanish stressed position as locus of pitch movement. Instead, they exhibit Quechua phrasal accentuation. Just like words of Quechua origin, Spanish-origin words like *primer* can be realized as part of the initial low stretch without any pitch events, frequently if they are in a modifying position. Thus even on nearly identical lexical items in the same task, there are individual speaker differences (between AZ23 on the one side and QZ13 and OZ14 on the other) with

355 https://osf.io/g6qux/

regards to the use of loanword prosody. Conversely, the importance of lexical identity can also be shown fairly unequivocally, in that the same speakers under the same context conditions can treat different lexical items of Spanish origin very differently prosodically. This is seen when comparing the realizations of forms of *lado* 'side' with those of *murciélago* 'bat' in the Quechua *Maptask* and *Conc* corpora by speakers XU31 and OA32. Fortunately, those two speakers, upon encountering the conflicting section in their maps, patiently began the game anew several times, determined to iron out the kinks in their interaction by simply redoing everything again and retracing their steps. This has resulted in a *Maptask* of nearly 13 minutes of continuous conversation, with similar but varying utterances containing always the same set of lexical items. One of the objects causing a conflict in their maps was the bat, so it is quite often referred to. In addition, both speakers seem to have incidentally forgotten the Quechua word for 'bat', *tsiqtsi* (or only one of them forgot and the other followed him in that usage), so they use forms of *murciélago* throughout this long map task. In most other corpora by other speakers, *tsiqtsi* is used to refer to a bat, and its occurrence in their *Cuento* shows that *tsiqtsi* is also in XU31's and OA32's lexicon. *Murciélago* in their map task thus is probably something of an ad hoc-substitute for *tsiqtsi* via access to the multilingual lexicon available to the speakers. Morphosyntactically, it is used in the corpus like any other noun, suffixed with Quechua nominal morphology. *Murciélago* is a fortunate case also in that it is proparoxytonic in Spanish, so that even if it is phrased as a an individual word without suffixes in Quechua, accentuation on the Spanish stressed syllable /sje/ can be differentiated from a phrase-final rise beginning on the penult. On the other side, forms of *lado* are used by nearly every speaker at least once in their Quechua data. Etymologically "native" words for 'side' are found in one dictionary, like *awi* or *niq* (cf. Carranza Romero 2003: 40, 140), but in another, actually older one, *laadu* is the only entry with the meaning 'side' (Parker & Chavez Reyes 1976: 260). In our data as well, only forms of *lado* are used. There is also a shortened form, *laa*, but speakers use both with the same meaning[356] in the same task, and both forms can be and are suffixed with Quechua nominal morphology. Note that even though *lado*

356 Though perhaps not with the same morphosyntax: in this very small corpus, XU31 uses *laa* exclusively following either the demonstrative *kay* or the locative nouns *hana* 'above' and *hawa* 'below', while *lado* is used everywhere else (see Table 26). This might suggest that *kay laa, hana laa* and *hawa laa*, as well as perhaps others involving basic directions, have become fused lexical expressions for him with the meaning 'this side', 'upper side', and 'lower side', respectively. However, OA32, while using it only twice overall, produces *laa* once following *manka* 'pot' in *manka laa-n-chaw* (pot side-3-LOC) 'beside the pot', which is far less plausible as a lexicalized expression. For both it is the case that while a noun preceding a form of *lado* (as well as other location nouns) may optionally be suffixed with the genitive –*pa* in a construction marking relative reference / possession of the form *X(-pa) ladu-n(-suffix)* (X-(GEN) side-3(-suffix)) '()X's side', such as *manka-pa* seems to have a much more comprehensive history as a loanword in Quechua than *murciélago*, it is of course also an active item in the Spanish lexicon of these bilingual speakers, and stressed 'correctly' when used in Spanish by them. However, it can be assumed that *lado*, both in their Spanish and in their Quechua, will have a much higher frequency in the daily usage of these speakers than either *murciélago* or *tsiqtsi*.

Table 26: Token counts for word form types of *murciélago*, *lado, escoba,* and *abeja* in the Quechua *Conc* and *Maptask* corpora of the speaker pair XU31 and OA32 according to prosodic pattern.

speaker	token word form	"inherited" pattern	"grafted" pattern	no pitch movement due to Spanish stress	indeterminable	all
murciélago (in Conc *and* Maptask*)*						
XU31	murcielagu	1	6	1	–	8
	murcielagu-ta *bat-OBJ*	–	1	–	–	1
	murcielagu-qa *bat-TOP*	–	–	1	–	1
	murcielagu-pa *bat-GEN*	–	1	–	–	1
	all	*1*	*8*	*2*	–	*11*
OA32	murcielagu	1	–	2	1	4
	murcielagu-ta *bat-OBJ*	–	1	1	–	2
	murcielagu-m *bat-ASS*	–	1	–	1	2
	murcielagu-man *bat-DEST*	–	1	–	–	1
	all	*1*	*3*	*3*	*2*	*9*
both		***2***	***11***	***5***	***2***	***20***
lado (in Conc *and* Maptask*)*						
XU31	ladu-n *side-3*	–	–	1	–	1
	ladu-n-chaw *side-3-LOC*	1	–	7	–	7
	ladu-n-chaw-mi *side-3-LOC-ASS*	–	–	5	–	5
	ladu-n-pa *side-3-GEN*	–	–	13	–	13
	ladu-n-pa-m *side-3-GEN-ASS*	–	–	2	–	2
	ladu-n-pita *side-3-ABL*	–	–	1	–	1
	laa-ni-n-chaw *side-FON-3-LOC*	–	–	5	–	5
	laa-ni-n-chaw-mi *side-FON-3-LOC-ASS*	–	–	1	–	1

ladu-n 'the pot's side', the dependent-marking with *–pa* does not seem to occur when the possessor is followed by *laa* instead of *lado*, which might suggest affix status for *laa*.

Table 26 (continued)

speaker	token word form	"inherited" pattern	"grafted" pattern	no pitch movement due to Spanish stress	indeterminable	all
	laa-ni-n-chaw-raa-mi *side-FON-3-LOC-CONT-ASS*	–	–	1	–	1
	laa-ni-n-pa *side-FON-3-GEN*	–	–	3	–	3
	laa-ni-n-pa-mi *side-FON-3-GEN-ASS*	–	–	1	–	1
	li-n-pa-mi *side-3-GEN-ASS*	–	–	1	–	1
	all	*1*	–	*41*	–	*42*
OA32	ladu	1	–	–	–	1
	ladu-n *side-3*	–	–	1	–	1
	ladu-n-chaw *side-3-LOC*	–	–	8	–	8
	laa-n-chaw *side-3-LOC*	–	–	1	–	1
	laa-ni-n-chaw *side-FON-3-LOC*	–	–	1	–	1
	all	*1*	–	*11*	–	*12*
both		***2***	–	***62***	–	***64***
escoba (in Conc *and* Maptask*)*						
XU31	escoba-man *broom-DEST*	–	–	3	–	3
	escoba-pa *broom-GEN*	–	–	1	–	1
	all	–	–	*4*	–	*4*
OA32	escoba-pa *broom-GEN*	–	–	*1*	–	*1*
both		–	–	***5***	–	***5***
abeja (in Conc *and* Maptask*)*						
XU31	abeja-kuna *bee-PL*	–	2	–	–	2
	abeja-ta *bee-OBJ*	–	–	1	–	1
	all	–	*2*	*1*	–	*3*
OA32	abeja-kuna *bee-PL*	*1*	–	–	–	*1*
both		***1***	***2***	***1***	–	***4***

Table 26 gives the counts for word form types of *murciélago* and *lado* in the Quechua *Conc* and *Maptask* corpora by XU31 and OA32, sorted according to the prosodic pattern realized on them. While overall counts differ of course, there is a very clear trend: phrases containing word forms of *murciélago* (usually produced with a high back vowel in the last segment) are realized with one of the patterns sensitive to the Spanish stress position more than half of the time (13/20 times = 65%) here. Phrases containing forms of *lado* almost never are (2/64 times = 0.03%). This is pretty solid evidence that even with the same speakers in the same task, lexical identity is a relevant factor for the prosody of Spanish-origin words in Quechua. Relating this to their history sketched above, the obvious hypothesis seems to be that with increasing integration into the Quechua lexicon, the marked prosodic patterns sensitive to Spanish stress become less frequent and the prosody thus more regular. The table also includes counts for realizations of forms of *escoba* and *abeja*, with the former never being produced with a pattern sensitive to Spanish stress, and the latter three out of four times. While the absolute numbers here are far too low to draw any definite conclusions, this suggests that the difference in prosodic realization is not specific to *murciélago* and *lado*, but is instead more generally influenced by lexical identity. It is perhaps noteworthy that most of the other speakers use *pitsana* in Quechua for 'broom', but no speaker in the data refers to bees with any other word than *abeja*. Other speakers, including even AZ23 with her overall low preference for the loanword pattern, also produce the "grafted" pattern on it ((128)/Figure 157). This seems to contradict the hypothesis of increasing prosodic regularization concomitant with lexical integration of loanwords, but perhaps *abeja* is simply infrequent overall. More subtle factors are clearly at play here than just the two we have been able to nail down, speaker preference and lexical identity. I leave a more in-depth exploration of them for further research.

In conclusion, the tonal alignment patterns showing sensitivity to the "inherited" lexical stress position in Spanish-origin words is a marked but not infrequent and presumably stable prosodic option for bilingual Huari Quechua speakers. Its frequency of realization has been shown to vary between speakers and individual lexical items. Typologically, it seems to occupy a kind of niche in that the lexical alternation between words specified for an irregular accent position and those that lack this specification in a language otherwise employing tones mostly for the optimization of phonological/accentual phrases finds its counterpart in several unrelated languages across the world, whose prosodic systems in other aspects also display similarities to what we have found for Quechua to greater or lesser extent. Any language that shows this phenomenon straddles a typological feature boundary: it both has and hasn't a lexically specified accent position for words which attracts tones realizing a pitch accent.

6.2 Phrasing and relation to meaning-based categories

Until now, attested Huari Quechua prosodic patterns corpus have been categorized only according to characteristics necessary for their prosodic definition. An important further aspect is the relation to meaning-based categories and the signalling of differential information structure. This section will explore this issue based on a sub-corpus of 415 utterances from 1 *Maptask* (TP03&KP04), 1 *Cuento* (AZ23&ZZ24), and 7

Conc (TP03&KP04, QZ13&OZ14, SG15&QF16, AZ23&ZZ24, ZR29&HA30, XU31&OA32, XQ33&LC34) tasks. These constitute nearly all the utterances produced in these corpora; excluded were only a small number because of very quiet (whisper-like) speech, too much background noise, or an abundance of hesitations and false starts. The utterances were analyzed and broadly annotated by hand according to their morphosyntax, information structure (explicit or likely implicit QUD, focus-background division and information status of referents (given/new)), and intonation. In terms of intonation, this meant annotating tonal contours according to the patterns established in the previous sections and how they mapped onto words. All together, this provides insight into what information structural and syntactic configurations pattern frequently with which contours, and thus what factors might contribute to contour choice whether a contour is realized across one or several words (phrasing). In turn, this will allow us to form hypotheses about prosodic and metrical structure and its mapping to words in Quechua. The sub-corpus is quite heterogeneous, comprising data from 14 speakers engaging in three different communicative tasks in spontaneous speech. To give one example, overall only 120 of these 415 utterances contain verbs,[357] while the other 295 are made up only of nouns (and a few adverbs and particles). However, of those 120, 98 stem from the two *Maptask* and *Cuento* corpora, leaving only 22 for the seven *Conc* corpora. This is because largely, conver-

357 A variety of Quechua suffixes allow conversion between the two main word classes, verbs and nouns, so that the resulting word form can then take suffixes exclusive to the resulting word class. Accordingly, a word form having undergone conversion to a noun from a verb (eligible to take nominal suffixes and behave syntactically like a noun, e.g. being argument to some predicate) is counted as a noun, and vice versa. This is not a central issue here, but there are quite complex cases where verbs with filled argument slots are nominalized and serve then again as arguments in higher-order predicates, such as in (i).

(i) ZZ24_Cuent_Q_1825

huqta	chuspi-ta	wañu-tsi-nqa-n-pa	alma-n-ta	qara-sha
six	fly-OBJ	die-CAUS-NMLZ-3-GEN	soul-3-OBJ	give-PRTCP

"[he] gave [him] the souls of the six flies he had killed"

Here, *huqta chuspita* is the direct object to the predicate *wañutsinqanpa*, which is nominalized by –*nqa* and the possessive modifier (indicated by the genitive *-pa*) to *almanta*, in turn the direct object of the main verb *qarasha*, whose subject and indirect object are both null and only provided by context. Translating this as a relative clause with the six flies as the head noun in English seems natural, but it is the nominalized verb form which is the head in the Quechua sentence (evidenced by the genitive which indicates the relation between the souls and those they belonged to). This structure cannot be faithfully rendered in English. An approximation would be something like "he gave the souls of those he had killed (the six flies)", with the object of the subordinate verb in parentheses and an additional pronominal relative head. What is nominalized here is not just the verb *wañutsi-*, but the entire clause *huqta chuspita wañutsi-*. Such cases make it clear that "nominalization" in Quechua is perhaps a misnomer, or at least a rather complex issue. Cf. also Lefebvre & Muysken (1988).

sational moves in *Conc* are of the broad form "[noun phrase denoting a location] is [noun phrase denoting an object]", with the copula in the third person present tense not realized. This heterogeneity, in itself interesting, makes it difficult to draw certain generalizations from the entire dataset. In the following we will look at correlational patterns taken from subsets of the data that are within themselves comparable to a degree, and also explore further sources of variation existing between the corpora.

6.2.1 Brief excursion: The question of (polar) questions

In this section, I will briefly discuss and exclude the possibility that the rising contour is used exclusively or even mainly to signal questions, in particular polar questions. Wh-questions are not any more often realized with a fall than declaratives, as far as could be determined. They do not show any noticeable differences to them intonationally. Note that the wh-word in most cases is phrased on its own with a (rising-) falling contour (cf. Figures 132, 148, 165).

(141) CJ35_Quien_Q_0163

imanir-taa	gerra-qa	ka-sh	imanir-taa	pelya-ykaa-yaa-nqa
why-DETVAR	war-TOP	COP-PRTCP	why-DETVAR	fight-PROG-PL-3.FUT

"why was there a war? why would they be fighting?"

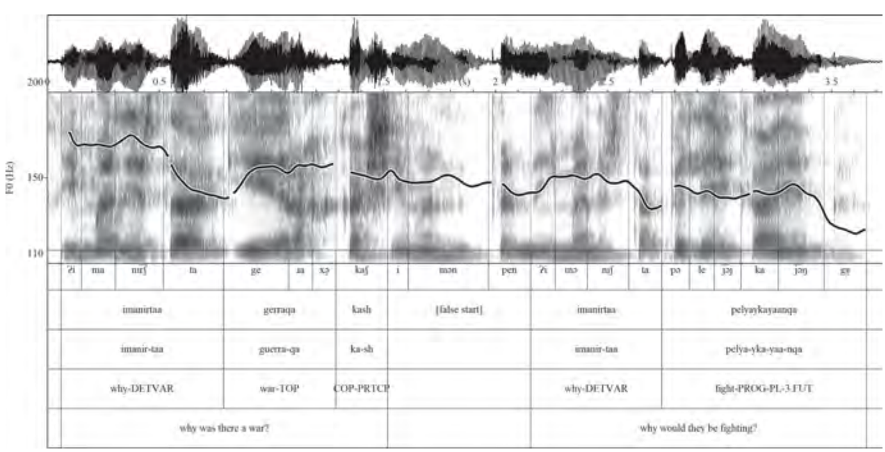


Figure 165: CJ35_Quien_Q_0163[358] (two wh-questions).

358 https://osf.io/wq9s5/

Regarding rising contours in polar questions, I investigate their occurrence using data from the *Quien* task, which was also used in section 6.1.6. It was originally intended to elicit questions in particular. Morphology and particles play a role for polar questions in Huari Quechua: the morpheme *–ku* is often associated with polar questions and can attach to any word in a sentence; if *–ku* is present in a sentence, the evidentials *–m(i)/-sh(i)/-ch(i)* cannot be (Parker 1976: 148–149). It also occurs in alternative questions (cf. Figure 166): both alternatives are here marked with *–ku*, optionally the Spanish *o* 'or' is also inserted between them. The first alternative in them is realized with a rising contour, and the second with a falling one that often seems downstepped in pitch level.

(142) CJ35_Quien_Q_0349

ruku-na-ku	hoobin-lla-raq-ku	tsay	runa-qa
old-DISC-Q	young-LIM-CONT-Q	DEM.DIST	person-TOP

"is the person old already or still young?"

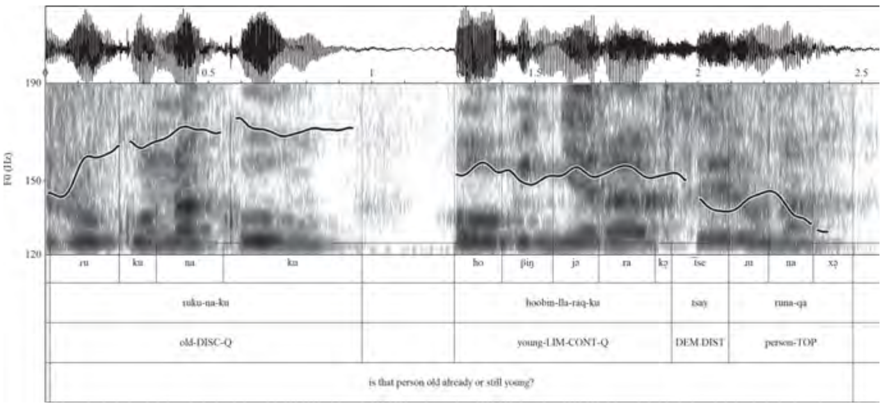


Figure 166: CJ35_Quien_Q_0349[359] (alternative question).

(143) XQ33_Quien_Q_0059

aqtsa-sapa-ku	ka-shqa
hair-AUGM-Q	COP-PRTCP

"did s/he have a lot of hair?"

359 https://osf.io/kh29b/

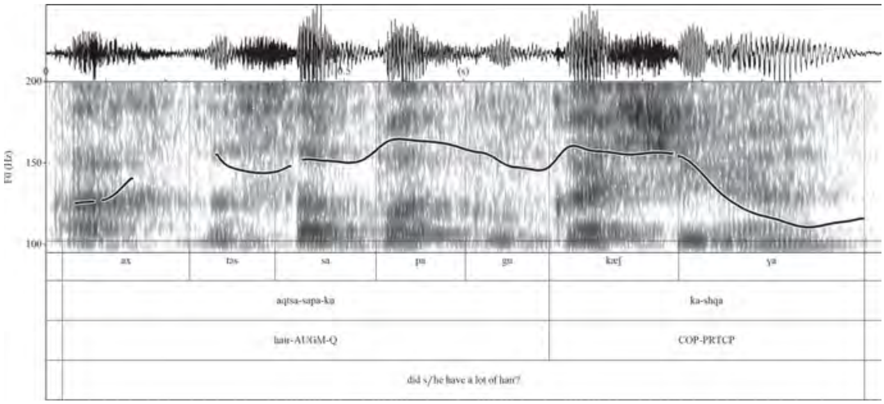


Figure 167: XQ33_Quien_Q_0059[360] (neutral polar question with *-ku*).

Also relevant for questions is the particle *aw*, which when used sentence-initially means "yes", but in our corpus is more often used utterance-finally as a question tag (cf. also section 6.1.7). Table 27 gives counts for occurrences of polar questions in *Quien*, separated by type of question. Figure 167 gives an example for a neutral polar question. "Neutral" polar questions mean those where the speaker is inquiring for a truth value on a proposition introduced by the root sentence, when no prior knowledge regarding either outcome can be assumed for the speaker or to be present in the common ground (here based on discourse context observation, cf. Farkas & Bruce 2010: 94–95; Dayal 2016: 4). "Biased" are all manner of polar questions with an epistemic bias in the context, accessible to the speaker either because of a previous utterance or via inference from the common ground, e.g. types of clarification questions and confirmation-seeking questions (cf. Farkas & Bruce 2010; Łupkowski & Ginzburg 2016). The category was here chosen to be fairly broad. The table compares both types of question against the presence of –*ku* and *aw*, and the realization with a rising contour on the main utterance and separately on *aw*. The counts in the table broadly confirm findings reported by Cole (1982) and O'Rourke (2005) on other varieties of Quechua that polar questions are not in general realized as rises. However, they also offer interesting qualifications of that general result. It seems that in particular, neutral polar questions without –*ku* are prone to being realized with a rising contour, while both –*ku* and a biased question make a rising realization less likely (cf. Muntendam 2017 for similar findings on this association in Southern varieties of Quechua).

360 https://osf.io/unpge/

Table 27: Occurrences of polar questions in the *Quien* corpus, ordered by type of question, coocurrence with the suffix *–ku*, utterance-final use of particle *aw*, and utterance-final realization with a rise or fall.

Type of polar question	***-ku***		**with *aw***	**final rise on main utterance**		**rise on *aw***
				yes	**no**	
Neutral	+	13	0	3	10	--
	–	10	0	10	0	--
Biased	+	4	3	1	3	0
	–	13	3	2	11	2
Both	+	17	3	4	13	0
	–	23	3	12	11	2
Total	40		6	16	24	2

Regarding *aw*, its use seems restricted to biased questions, supporting the hypothesis based also on observing it throughout the Huari Quechua data that it serves as a question tag for confirmation-seeking questions, with a similar meaning to Spanish tags *no*, *verdad*, or *eh*, German *oder*, or English *innit* or *huh*. Inspection of individual examples reveals that when *aw* is integrated into the preceding phrase, it is usually not realized as a rise (but recall Figure 155 from section 6.1.7), but if it is produced in a separate phrase, that phrase is always realized with a rising contour. Individual examples like (144) suggest that *aw* and the realization with a rise further differentiate within biased questions.

(144) SG15_QF16_Cuento_Q_1233–1382 (context for Figures 168 and 169)[361]

time (seconds)	SG15	QF16
123.3		ima-taq tsay runa ka-rqa-n what-DETVAR DEM.DIST person COP-PST-3 *what was that man*
124.6		**hampikuq aw** healer yes a healer, right
125.4	hampikuq healer *a healer*	

361 https://osf.io/tay7p/

126.1		hampikuq runa
		healer person
		a healer
127.0		y naani-qa
		and path-TOP
		and what about the path
128.6		karu-tsura ka-ra-n o
		far-CONJ COP-PST-3 or
		was it rather far, or
129.7	karu-sh ka-naa	
	far-REP COP-PST.REP	
	it was far (it is said)	
130.6		karu
		far
		far
131.0	mh-[mh]	
131.2		[y] mirkapa-ta ima-ta-taa apa-naa
		and provisions-OBJ what-OBJ-DETVAR
		bring-PST.REP
		and what did he bring as food
133.0	(3.3)	
136.3	**mirkapa-ta apa-sha**	
	provisions-OBJ	
	bring-PRTCP	
	he brought food	
137.6		aha
		yes

(144) is an excerpt from *Cuento* by SG15 and QF16. QF16, a school teacher, has just told the story to SG15 for the first time. Instead of waiting for her to re-tell it by herself, he asks her a number of questions to elicit the main points of the story. At 123.3, he asks her what kind of man the protagonist was, and at 124.6 suggests an answer himself, *hampikuq aw* "a healer, right?" (Figure 168). This is clearly a question with the highest possible confirmation bias, as he himself just told her the story about which he is asking her now, i.e. he is both asking the question and has the epistemic authority to decide whether the proposition is true or not. This is produced with a falling contour on the main part of the question, and a sharp rise on *aw*, typical for this use of *aw* when phrased separately.

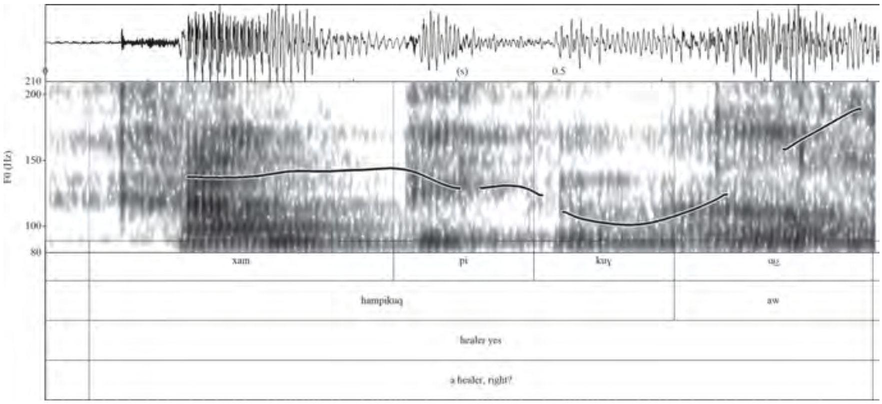


Figure 168: QF16_Cuento_Q_1246[362] (confirmation-seeking polar question with *aw*).

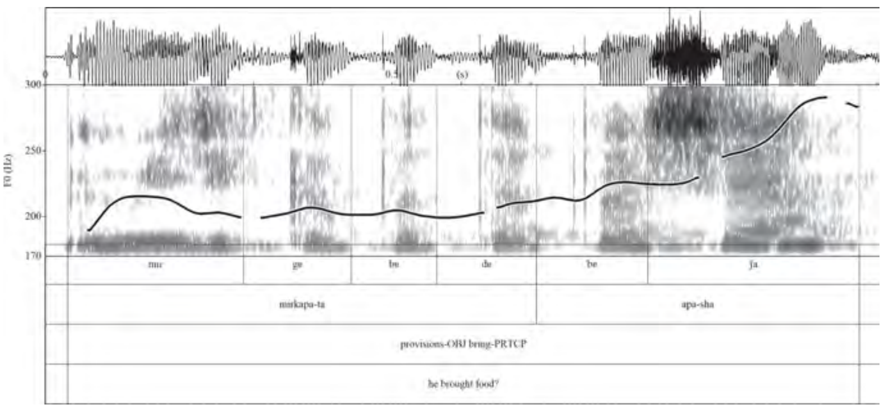


Figure 169: SG15_Cuento_Q_1363[363] (clarification-seeking polar question).

Later, at 131.2, QF16 asks what kind of food the healer had taken as provision on his journey. SG15 responds at 136.3 with the polar question *mirkapata apasha* “he brought food?” (Figure 169). This is also a question biased towards confirmation (that is clear from QF16’s response via confirmation token at 137.6), but it is also very different: in terms of epistemics, SG15 has no authority to decide whether the proposition she has asked about is true. The information she asks for she has just received in the form of a presupposition in QF16’s preceding wh-question and she asks the person who is its source about the presupposition, whether she is correct

362 https://osf.io/9fw5u/
363 https://osf.io/hrxyq/

in assuming that the presupposed proposition should be entered into CG. The two utterances are also different in terms of their relative positions in the discourse (cf. Farkas & Bruce 2010): QF16's question at 124.6 is a provocation (an elaboration on the directly preceding turn, which is also a provocation), while SG15's question at 136.3 is a response to another question. More specifically, it is a clarification request for intended content, one of several types of such query responses according to the typologies proposed in Ginzburg (2012: 148–155); Łupkowski & Ginzburg (2016). Formally, it is also different, in that it is realized without *aw*, and with a rise on the question itself not strictly aligned to just one syllable. From the Huari Quechua corpus known to me, it seems that *aw* as a tag never occurs in contexts such as the one of SG15's question, with a strong epistemic imbalance and as part of a clarification request asking about a proposition formulated as a presupposition but received as new information.[364] Such clarification requests are also rarely realized with *–ku* or without a rise, in contrast to neutral polar questions, which show *–ku* with falling intonation in nearly half the cases.

The connection between epistemic bias, the position of a turn in the discourse as provocation or response, whether a question targets at-issue or non-at-issue material of a preceding move, the presence of *–ku* and *aw*, and the prosodic realization of polar questions should be a promising field for further research to uncover whether any more restrictive correspondence actually exists and which semantic factors are really relevant. Here however, the findings from Table 27 must suffice to indicate that polar questions are not always associated with rises. The converse is

364 If that is true it would suggest that questions with *aw* as tag are similar in meaning to some English/German/Spanish tag questions which would also not be felicitous in such a context: the a-responses with tag in (i) that seem infelicitous when the proposition they ask about has been entered (as a presupposition) into CG by the provocation, presumably because they express a kind of inference based on available unconflicting information and are thus odd when the information they represent as inferred has in fact just been given. In contrast, other forms of biased questions are felicitous in such contexts: the b- and c-responses, that seem to express some kind of inconsistency of the proposition presupposed by the provocation with previous information, with the degree of this "incredulity"/ "counterexpectation" arguably even stronger in the c-responses.

(i) **A:** What did she bring as food for the journey? / Was hat sie für Proviant mitgenommen? / ¿Qué llevó de fiambre?
  a. **B:** She brought food, #right/#didn't she? Sie hat Proviant mitgenommen, #oder? / ¿Llevó (algo de) fiambre, #no/#verdad?
  b. **B:** So she brought food? / Also hat sie Proviant mitgenommen? / ¿Entonces llevó fiambre?
  c. **B:** She brought food, did she now? / Hat sie jetzt Proviant mitgenommen? /¿Así que llevó fiambre?

also not true, as both previously seen examples and the results in the next sections show that rising contours on phrases and utterances in the *Conc, Maptask,* and *Cuento* corpora – overwhelmingly declaratives – are very frequent. Their distribution is instead governed by other factors.

6.2.2 Moves with rising and falling contours in Conc

This section investigates how rising and falling contours are distributed across the *Conc* corpora (cf. the description in section 2.4). Following the entire discourse in *Conc* allows for keeping track of the information status and level of activation of the individual referents represented by the pictures and their locations quite well. As mentioned above in 6.2, conversational moves realized in *Conc* in Quechua very often consist of a specification of a location at which speakers believe that a depicted item is, or a specification about the item that they believe to be at that location. Logically, and permitted by the relatively free word order of Quechua, we would expect the sequential realization of such conversational moves to form a coherent guessing sequence to follow one of two patterns: a) *Location – Item*, or b) *Item – Location*, where Item and Location stand for simple or complex expressions encoding the item and the location, respectively. Sometimes both of these components together are produced as one utterance, but often a move, especially one specifying the location, will consist of several parts that are all individual utterances, and sometimes there are simply longer silent breaks between the utterances realizing the component moves, so that there is no one-to-one correspondence between moves and utterances.[365]

365 As working definition for "utterance", I take a very simple temporally and phonetically oriented one: a chunk of speech which a speaker produces in one go, without larger breaks or hesitations in between. Thus an utterance defined this way may or may not perfectly overlap with a conversational move, defined via criteria of discourse meaning, or a sentence, defined via syntactic criteria. It also does not necessarily have to be coterminous with an IP or other larger phonological unit, e.g. if it is an aborted or interrupted utterance, although I assume an utterance that is succesfully produced as planned by the speaker to contain one or integer multiples of one IP (by definition of what an IP is). It is also separable into smaller chunks – phrases – according to prosodic criteria.

(145) KP04_Conc_Q_0634–0658[366] (example conversational move sequence pattern a)

time					
63.4	(L	H ------------------------------->)		*location-specifying move*	*one conversational move sequence matching an item to a location*
	tsay	punta-yki	ka-q-lla-chaw		
	DEM.DIST	front-2	COP-AG-LIM-LOC		
	the one in front of you				
64.5	(L H)			*(confirm)*	
	aha				
65.0	(H)	(L H L)		*item-specifying move*	
	tsay	wanupakush			
	DEM.DIST	burial			
	that's the burial				

(145) is an example for such a conversational move sequence following pattern a), specifying first the location and then the item found at that location (the image on the card when turned around). At 63.4, the speaker begins the sequence by naming the location in one utterance realized in a single rising phrase. He then confirms at 64.5, after the game master has moved his hand towards the correct card, and at 65.0 he names the item he believes to be shown on the card at that location using a rise-falling contour on the phrase realizing the item and concluding the sequence. Here it is still counted as one sequence if a speaker specifies the location using several utterances, or if they follow the first utterance naming the item by additional ones giving a further description of it. For the purposes of counting, I also take additional utterances specifying the location after the item has been named (if the game master has not yet reached the intended card) but before the card has been turned around as belonging to the same sequence following pattern a).

366 https://osf.io/rnq5y/

(146) OA32_Conc_Q_1198–1213 (example conversational move sequence pattern b)[367]

time

119.8	(L H L) pinkullu flute *the flute*		*item-specifying move*	*one conversational move sequence matching an object to its location*
121.3	((L) H) tsuku-pa hut-GEN *beside the hut*	(L H L) ladu-n-chaw side-3-LOC	*location-specifying move*	

(146) is an example for a conversational move sequence following pattern b). The speaker first names the item with a rise-falling contour (in 119.8), and then specifies its location with an utterance consisting of first a rising and then a rise-falling phrase (in 121.3). For this pattern, further utterances referring to the same item, occurring after the move specifying the location but before the card has been turned around, are also counted as belonging to the same move sequence.

Table 28 shows the number of move sequences in the seven Quechua *Conc* corpora separated by speaker pairs and according to the sequential pattern they follow. 12 item cards were used in each game. The diverging numbers of move sequences between speaker pairs reflects the varying number of false guesses.

Table 28: Number of conversational move sequences in seven Quechua *Conc* corpora.

	move sequences overall[368]	**order Loc-Obj (pattern a)**	**order Obj-Loc (pattern b)**	**not applicable**[369]
TP03 & KP04	17	16	1	0
SG15 & QF16	17	12	5	0
QZ13 & OZ14	15	15	0	0
AZ23 & ZZ24	16	16	0	0
XU31 & OA32	18	0	17	1

367 https://osf.io/5j46n/

368 Counted here are only move sequences with which a guess about a location and an item at that location was made, not moves peripheral to the game, such as questions, banter between the speakers, etc. Those only make up a very small part of the spoken content of any of the *Conc* corpora and are ignored here.

369 One move consisted of the speaker only saying the name of the item, at the moment when all other cards had been solved already, so that no sequence exists.

Table 28 (continued)

	move sequences overall	order Loc-Obj (pattern a)	order Obj-Loc (pattern b)	not applicable
ZR29 & HA30	21	20	1	0
XQ33 & LC34	12	12	0	0
All	116	91	24	1

As seen from Table 28, overall, sequential pattern a) occurs with much greater frequency than pattern b). However, this is also a matter of speaker (pair) preference: only the speaker pairs SG15 & QF16 and XU31 & OA32 use pattern b) more than once, XU31 & OA32 never using pattern a), which is preferred by all other speaker pairs.[370] That the patterning correlates so well with speaker pairs might point to these patterns reflecting at least partially two different available *strategies* for reducing the context set (cf. Roberts 2012b). That is to say, if the task of the game is modeled as attempting to answer the super-QUD "What item is in which location?", then one strategy is to find answers to sub-QUDs that are formed using the elements of the set of possible locations as fixed variables imposing a referential restriction (= topics, Kuno 1982; Roberts 2011), and to ask for the corresponding items as open variables; and the other works the other way around, forming sub-QUDs with the items as fixed variables and asking for their corresponding locations. A reason for the preference of pattern a) over b) by most of the pairs might be that the setup of the game actually induces a bias for this, because with the cards lying face down on the table, the positions themselves (such as "the first card in the third row") are already established and unchangeable, while it is the identities of the cards *at* these positions which are in doubt. Furthermore, the locations are continuously present in the extralinguistic context, while the items' names have to be actively recalled by the speakers. In other words, the locations can be seen as more active relative to the cards' identities, and thus would lend themselves to be linguistically realized as given and topical material, i.e. signaling a QUD structure which asks for the identities of the items using the elements of the set of locations as anchors (= topics; assuming that information is packaged in the somewhat iconic order old-new / theme-rheme in Quechua (cf. Weber 1989: 427, who argues this for the related Huallaga variety of Quechua I). Additionally, naming an item for the first time before the card has been turned around can be seen as making a separate additional discourse commitment (in the sense of Farkas & Bruce 2010) from when

370 In Spanish *Conc* the speakers also broadly follow these preference patterns, with XU31 & OA32 preferring pattern b) over pattern a), while the other speaker pairs mostly follow pattern a).

a location is first named: because the set of items is not constantly available for inspection (the cards are turned upside down on the game table), naming an item makes a commitment to that item actually being a member of the set (i.e., remembering its presence correctly from before). A speaker can be just as wrong about this as about the separate discourse commitment that is made when relating an item to a location. Following pattern b) means making this discourse commitment initially as an existential presupposition, separately from the assertion that relates the item to the location. In pattern a) on the other hand, whatever cues might be used to mark this assertion (e.g. falling intonation, morphological marking, or word order) can conceivably be marked once on the last element in the sequence (= the expression denoting the item) and be understood to apply to both discourse commitments made. Note however that also in pattern a) the location is normally at least specified in each utterance anew; there is usually no topic continuity from a previous utterance. This can have a reflex in the utterance form, as argued in section 6.4.3. In this view, it would probably make sense for speaker pairs to collaboratively follow one or the other strategy, because this allows for more economic signaling, and perhaps also for exploiting deviations from the established pattern in order to signal a marked information structure. Hypotheses along these lines well be explored further below, but for now note that the only speaker pair where the majority pattern is not the one established by the first speaker's first move is SG15 & QF16, who are probably the two speakers least acquainted with each other from outside of the experiments.[371]

Table 29 relates conversational moves to intonation by counting all utterances belonging to a move sequence matching a location to an item, separated according to the corpus, whether they are part of a location- or item-specifying move and whether they are rising or falling. For the purposes of this table, utterances consisting of several phrases (with an identifiable pitch movement on each) were counted as rising or falling if the last phrase was rising or falling. If a single utterance realized both a location-specifying and an item-specifying move, the decision before the analysis was to evaluate the two moves separately if there was a clear phrase boundary separating the expressions in the utterance along that line. As it turned out, this was nearly always unambiguously the case, i.e. there was always an identifiable phrase boundary at the boundary between an item-specifying and a location-specifying move, evidence in itself that the division between these types of moves is reflected in the prosodic realization. Thus this table counts rising or

371 While the other pairs are all either siblings or friends of a similar age who volunteered as pairs for our experiments, QF16 and SG15 are a school teacher and a cook who did not know each other well before paired for the experiment, each having volunteered individually.

falling *moves*, under the assumption that the final phrase in an utterance signaling such a move has priority. Further below, we will also look at the composition of utterances into phrases and their prosodic contours.[372]

Table 29: All moves that are part of a guessing sequence in the seven Quechua *Conc* corpora, separated according to whether they specify a location or an item, and whether they are realized finally rising or falling.

	loc rise	**loc fall**	**loc indet.**[373]	**obj rise**	**obj fall**	**obj indet.**	***all***	**total loc**	**total obj**	**% rises loc**	**% falls loc**	**% rises Obj**	**% falls Obj**
TP03 & KP04	14	3	2	2	17	0	38	19	19	73.7	15.8	10.3	89.5
SG15 & QF16	13	12	2	2	13	1	43	27	16	48.2	44.4	12.5	81.3
QZ13 & OZ14	22	4	0	1	15	0	42	26	16	84.6	15.4	6.3	93.8
AZ23 & ZZ24	31	1	2	2	16	1	53	34	19	91.2	2.9	10.5	84.2
XU31 & OA32	3	23	5	8	11	2	52	31	21	9.7	74.2	38.1	52.4
ZR29 & HA30	6	19	2	12	12	1	52	27	25	22.2	70.4	48	48
XQ33 & LC34	13	0	0	0	12	2	27	13	14	100	0	0	85.7
All	102	62	13	27	96	7	307	177	130	57.6	35	20.8	73.9

Broadly speaking, location-specifying moves have a majority tendency to be rising (57.6 %), while item-specifying moves have a stronger tendency to be falling (73.9 %). Because this tendency is so much stronger for item-specifying moves, falls are slightly more frequent than rises overall (158 vs 129, or 51.5% vs 42%), even though location-specifying moves are more frequent than item-specifying ones (177 vs 130, or 57.7% vs 42.4%). One reason why there is no stronger correlation between rises and location moves on the one hand, and falls and item moves on the other, could be that there are speaker pairs who preferentially produce pattern b), and that actually rises occur at a non-final position in the sequence and falls at a final one, independent of the move type. That would mean that rising contours are effectively

372 See section 6.4.3 for a more in-depth look at the *Conc* moves by one speaker pair, ZR29 & HA30.
373 Moves were classed as „indeterminable" when they were realized either totally flat, with no discernible pitch movement that could be related to the contours established, or when large parts were voiceless. The indeterminables make up the missing percent to 100 in the last four columns in the table.

instances of what has been called “continuation rises” for many languages (cf. Chen 2007: 108; cf. Beckman et al. 2002; Aguilar et al. 2009; Estebas Vilaplana & Prieto 2010 for the use with regards to Spanish), i.e. final rises or high pitch signalling that the current topic is to be continued or that the speaker intends to maintain the turn. The cross-linguistic tendency for this has been hypothesized to result from the fact that pitch will tend to be higher at the beginning of an utterance due to heightened subglottal pressure because air is typically inhaled before speaking. Thus the expected values initial high pitch and final low pitch are supposed to signal a new topic and finality, respectively, while conversely, initial low and final high pitch, are cues signaling continuation (cf. the ‘production code’ hypothesis in Gussenhoven 2002a: 51–52, 2004: 89–90). The prediction would be that rising contours occur on location moves and falls on item moves for speakers following pattern a), and vice versa for speakers following pattern b). Four of the seven speaker pairs following pattern a), TP03 & KP04, QZ13 & OZ14, AZ23 & ZZ24, and XQ33 & LC34, bear out this prediction, suggesting that the association of rising contours with continuation and falling contours with finality is generally correct. However, the data from the other pairs suggests a more complex picture. On the one hand, the speaker pair ZR29 & HA30, who also nearly exclusively follow pattern a) (cf. Table 28), nonetheless realize 70.4 % of their location moves with falls, and about half (48%) of their item moves with rises. With these two speakers, circumstantial knowledge suggests that the high number of rising item moves might be due to uncertainty or perhaps more accurately the overt signaling of uncertainty due to the adoption of a social stance for whose outward construction that is seen as desirable.[374] XU31 & OA32 present another case. These two exclusively follow pattern b), and as expected from that, the great majority (74.2 %) of their location moves are falling, but the item moves

374 The two speakers were both 19 years old at the time of recording and seemed to the experimenters to present an outwards social identity that aims at attributes such as “cuteness”, “non-threatening femininity”, “fun”, “making an effort to not seem overly scholarly ambitious”. They frequently giggled during the course of the experiment, palatalized their speech somewhat, and gave statements to the effect that they didn’t know how to proceed but then showed themselves to do very well after only little encouragement. This perception is of course filtered through the lens of the experimenters’ own expectation with regards to presentation of gender as shaped by their experiences of persons with similar stylistic behavioural choices and their own sociocultural upbringing, which is not the same as that of the speakers. Thus there might be misperceptions inherent in this assessment due to these divergent cultural expectations, and any evaluation of another’s internal motives for outwards behaviour is of course necessarily conjecture. With these caveats, it seemed to the experimenters that what was projected by these two speakers was uncertainty. In Spanish, they also regularly realize the item move with a final rise in their *Conc* (again despite usually guessing right), supporting the hypothesis that this rather expresses a speaker-specific stylistic choice or stance than an attitude towards each individual proposition.

also fall more than half of the time (52.4 %). At the beginning of the task, both speakers introduce the items with falling contours, but after the sixth completed guessing sequence, XU31 begins to introduce them with a rising contour, while OA32 keeps realizing falling item moves until the end of the task. This can be considered in different ways: realizing a sequence only with falling moves (as OA32 does) can be interpreted as simply not signaling the coherence of the sequence as a whole. With an initial item move realized with a fall, OA32 signals only that the utterance or the move ends, but remains equivocal about its position in a larger sequence. In QUD terms, he thus does not signal that his assertion is a partial answer to a super-QUD like WHICH ITEM IS WHERE?, but merely that it answers WHICH ITEM? (specifying which is a separate commitment, as argued above). Given that it is known to all participants that realizing only an item move is not a sufficient communication in the terms of the game, OA32 does not run a great risk of having the turn taken from him or being misunderstood. Realizing the item moves with falls is therefore cost-reducing (in terms of articulatory effort and lack of commitment to forward planning) for the speaker and only a little cost-increasing for the hearer. Another not necessarily rivaling explanation is that falling contours are more likely to be understood as signaling an assertion. Because of the game's induced bias with regards to information status of the two sets of referents, as proposed above, there is some incentive to signal the additional discourse commitment about the identity of one of the items pictured on the cards as an assertion: this is the more strongly contested set of referents and specifying a member of it has higher informational value. As mentioned above, naming one of them requires correct memory retrieval both about their identity (WHICH ITEMS ARE IN THE SET?) and their position (THE ITEM IS WHERE?), while specifying the location only requires the latter; realizing falling intonation on the initial item move would then signal this discourse commitment clearly. Somewhat paradoxically, this might also go some length to explain the "uncertain" rises on item moves by ZR29&HA30, and why they do not occur on the location moves: the speakers might not want to commit to making the inherently stronger assertion about both item identity and location with confidence, but risk less by signaling certainty about just the location at which they think an item is. OA32 and initially XU31 as well then could be seen to follow one of two strategies which pattern b) forces a speaker to choose between: signaling both discourse commitments separately, while XU31 from sequence six onwards chooses instead to forgo this in exchange for highlighting the coherent nature of the entire move sequence. A third attempt at explanation could be a purely phonological one: OA32 uses "continuation rises" only *within*, but not *between* utterances, thus optimizing the prosodic unit "utterance". I do not propose that either of these explanations is true at the exclusion of the others. In any case, such a claim could not reasonably be made on the basis of the evidence available here. Instead, I would suggest that these are constraints (from different linguistic domains) that all act on the prosodic realization of a move, and that the heterogeneity in observable outcomes between speakers indicates both their violability and that they can be differently weighted.

In sum, the evidence considered here suggests that rising contours are most broadly associated with something describable as “openness” or “incompleteness”. This includes intention to continue a turn or a topic, an indication that a coherent unit has not yet reached its end, or the absence of a discourse commitment (interpretable as uncertainty, which presumably also lends itself to the use in some question types). Falling contours, conversely, are associated broadly with “finality” or “completeness”, which encompasses closing a turn or a topic, or the presence of a discourse commitment (also interpretable as certainty). The contour forms are clearly not very strictly tied to any of these individual meanings, with only distributional tendencies discernible and context used for specification. Specifically, as seen, despite lacking commitment to the truth of a proposition, questions are regularly formed with falling contours, and speaker attitude also seems to have some leeway in selecting which meaning a contour can be used to cue in a given context.

6.2.3 Distribution of contours in nominal sequences

The preceding section looked at the distribution of rising and falling contours across larger units, i.e. conversational moves and the utterances they are realized by. This is evidence for phrasing and a functional differentiation between rises and falls at a high level of prosodic structure. In the following, we will focus on phrasing within single utterances, specifically in sequences composed of nominal elements.

6.2.3.1 Noun phrases and other nominal sequences

This section takes a closer look at how contours are distributed within utterances and the conclusions this leads to about (default) prominence differences between several words in a sequence. The subset of data used here is that of all noun sequences in the nine sub-corpora (seven *Concs*, one *Maptask* and one *Cuento*) consisting of at least two nominal forms realized together without breaks and that could form a single nominal constituent (noun phrase) together syntactically. Whether they do form a noun phrase is only in doubt for a smaller subset, those consisting of a demonstrative (*kay* / *tsay*) or a bare noun, plus another noun. The demonstratives can be used both adnominally/attributively, and pronominally. Their form does not distinguish between these uses, just as with the English or Spanish demonstrative forms *this/that* or *este/ese*. If the demonstrative is adnominal and attributive, it forms a constituent with the noun as head, modifying the noun ((147)a). The same sequence of words can be a copular phrase, with the copula omitted in the 3rd person singular present, if the demonstrative is used as a pronoun. The demonstrative and the noun each form their own noun phrase then ((147)b). The same ambiguity exists for combinations of a bare (unsuffixed) noun plus another noun either unmarked for case or for locative case (-*chaw*), since a bare noun can either modify the following noun ((148)a, cf. (149)a) or also stand with it in a copular relation ((148)b).

(147) ambiguous use of demonstratives

a. adnominal/attributive use (cf. (114)/Figure 132)

[kay/tsay	wanupakush]$_{NP}$
DEM.PROX/DIST	burial

"this/that burial"

b. pronominal use (cf. (145)/Figure 170)

[kay/tsay]$_{NP}$	[wanupakush]$_{NP}$
DEM.PROX/DIST	burial

"this/that is [the/a] burial"

(148) ambiguous N_1-N_2 combinations

a. N_1 as modifier of N_2

[manka	chawpi-chaw-mi]$_{NP}$
pot	middle-LOC-ASS

"in the middle of the pot"

b. N_1 as copular subject

[manka]$_{NP}$	[chawpi-chaw-mi]$_{NP}$
pot	middle-LOC-ASS

"the pot [is] in the middle"

In the data considered here, context nearly always distinguishes between these uses and the annotation made use of that information. If used attributively, the demonstratives are ususally realized as part of the initial low stretch of a contour, phrased together with the noun they modify (cf. (110)/Figure 127, (106)/Figure 123, (114)/Figure 132). Yet if a demonstrative has the function of pronominal subject in a copular sentence, it can also be phrased separately, perhaps because it is then often also a topic, e.g. in (145)/Figure 170. The need to cue the syntactic or information structure interacts with the constraint against tonal crowding here: there is only a high tonal target realized on *tsay*, but the subsequent low target makes it clear that a new phrase begins with *wanupakush*. Thus it seems that phrasing can disambiguate structure. All other instances of nominal sequences considered here I take to form a nominal constituent, of varying length and complexity. Some examples (not just from *Conc*) are given in (149):

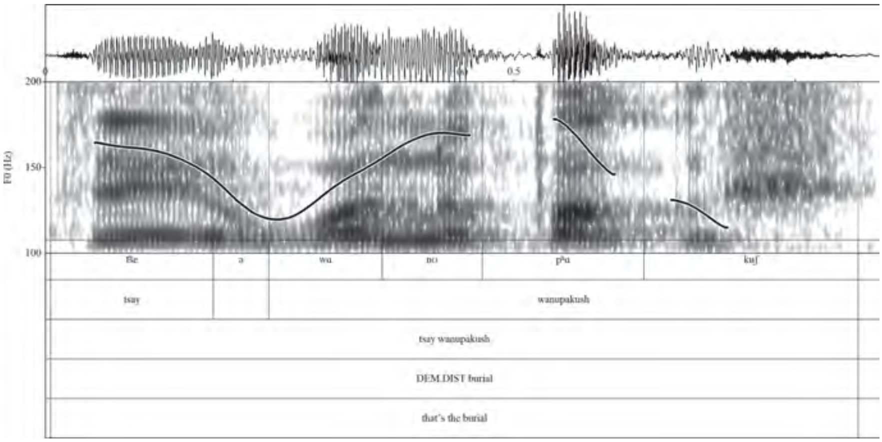


Figure 170: KP04_Conc_Q_0650[375] (declarative; copular sentence with pronominal subject realized as two phrases).

(149) Examples of nominal constituents in the Quechua corpora

a. ZZ24_Cuent_Q_1783
hacha hampi-ta
plant medicine-OBJ
"plant medicine"

b. XQ33_Conc_Q_0543
chuqllu-pa aqtsallku-n
corn-GEN corn.hair-3
"corn hair"

c. QZ13_Conc_Q_1358
runa wahi
person house
"person [and] house"

d. OA32_Conc_Q_1832
runa-wan wayi
person-INST house
"person with/and house"

375 https://osf.io/wsb8y/

e. LC34_Conc_Q_1249

runa	hawa-n-chaw
person	below-3-LOC

"below the person"

f. KP04_MT_Q_0703

pinkullu-pa	hana-n-pa
flute-GEN	above-3-GEN

"above the flute"

g. SG15_Conc_Q_0510

kay	laa	ka-q-chaw
DEM.PROX	side	COP-AG-LOC

"the one on this side"

h. XQ33_Conc_Q_1427

runa	wayi-n-man	aywa-ykaa-q	ladu-n-chaw
person	house-3-DEST	go-PROG-AG	side-3-LOC

"beside the person going into the house"

i. ZZ24_Cuent_Q_1825

huqta	chuspi-ta	wañu-tsi-nqa-n-pa	alma-n-ta
six	fly-OBJ	die-CAUS-NMLZ-3-GEN	soul-3-OBJ

"the souls of the six flies he had killed"

The examples demonstrate that what are counted here as a nominal constituents are rather varied: (149)a, c and e are examples of binominal phrases denoting a single entity where N1 modifies what kind of N2 it is, a conjunct of N1 and N2 in a coordination, and a location where N1 specifies relative to what object N2 is to be interpreted, respectively; all without internal morphological marking but with external (head-)marking in the cases of (149)a and e. (149)b, d, and f denote the same kind of semantic relations, but with internal marking (on N1) to specify the relation between N1 and N2. (149)g is an example for what might perhaps best be translated via a headless relative copular clause, or one using the pronominal dummy "one" in English: the demonstrative *kay* modifies *laa* 'side', which acts as subject of the copula *ka-*, marked with the agentive nominalizer *–q* and the locative to yield a single NP with the meaning "the one that is on this side".[376] (149)h and i showcase particularly complex examples involving relativization (cf. note 355 for a discussion of ZZ24_Cuent_Q_1825). Since this is not a work on syntax, this short exposition must suffice to point out the variety of nominal constituents here treated more or less uniformly. In what follows, I will occasionally make reference to the internal

376 For an in-depth discussion of the uses of the nominalized copula, see Bendezú Araujo (2021).

variety of these noun phrases when it seems to affect prosody, but it can't be done full justice here. One unifying characteristic is that they are all right-headed, both syntactically and semantically, or can be analyzed thus without problems ((149)c and d are actually ambiguous in this regard but could well be right-headed). In fact, only very few constructions could not securely be analyzed as right-headed. They always involve either the instrumental/comitative marker –*wan*, such as (150), the morpheme –*yuq*, marking the possession of an entity denoted by a noun which –*yuq* is attached to as an attribute of another entity (151), and in one case a postnominal locative modification (152), in all the nominal constituents treated here. This general tendency for right-headedness will be relevant for the following analysis.

(150) ZZ24_MT_Q_1635 (cf. Figure 171)
runa ushqu ñawi-wan
person cat eye-INST
"person with cat eyes"

(151) AZ23_Conc_Q_0057
achillku chuqllu-n-yuq
corn.hair corn-3-POSS
"corn hair with its corn"

(152) AZ23_Conc_Q_1734
aya kahun-chaw
corpse box-LOC
"corpse in a box"

6.2.3.2 Distribution of tonal patterns and alignment variant preferences

This section only treats sequences consisting of nominal elements. This is done to somewhat reduce the heterogeneity of the data and improve its comparability when quantified. As will be seen in section 6.4, the broad results of this section can be extended to phrases with verbs, but verbs also introduce further interesting complications that would be confounding factors here. The subset of nominal sequences considered consists of 487 words. Morphologically, their word forms are made up of 1225 syllables, for an average syllable count of 2.52 syllables per word. However, in their actual realizations, there were several (especially but not always word-final) cases of vowel devoicing and/or deletion and also some cases of consonant assimilation or deletion, sometimes severely altering the resulting word forms so that a different syllabification seemed more appropriate. A revised syllable count aiming to account for these reduction phenomena comes to 1179 syllables, or an average word length of 2.42 syllables per word. In the following, syllable counts will be given with a slash between the morphological syllables first and the actually realized ones second. Of the 487 words, 30 were produced either almost completely flat or with such substantial voiceless stretches that no confident assessment of their tonal behaviour could be made. The remaining 457 words were produced in what was analyzed as 292 separate phrases (see below for criteria), yielding a ratio of 1.57 words and 3.95 / 3.82 syllables per phrase. This compares well with similar ratios given for French and Seoul Korean in Jun & Fougeron (2000: 217, 2002: 150), where 2.3–2.6 words / 1.2 content words and 3.5–3.9 syllables per Accentual Phrase were counted for French, and 1.2 content words and 3.2 syllables for Seoul Korean, based on reading tasks. Compared with French, Quechua probably has fewer function word types, since a lot of the functions they are used for in French are fulfilled by suffixes. However, those words that can be counted as separate function words do occur with high token frequency. Refining the comparison, all instances of the demonstratives *kay* and *tsay*, nominalized forms of the copula *ka-* (mostly *ka-q* and *ka-ykaa-q*), and the numeral *huk* 'one', occasionally used similarly to an indefinite marker, were counted as function words, even if suffixed.[377] All other occurring words are content words. Nominalized tokens of *ka-* occur 36 times, *kay* 19, *tsay* 30 and *huk* 5 times, for 90 function words out of 487 in total. Out of those, 5 tokens were realized in phrases that were discounted because of flat or otherwise indeterminable realization. Applying these changes yields a content word ratio per phrase of (457−85)/292 = 1.27, rather similar to those given for French or Korean.[378] This would suggest that what has here been counted as a phonological phrase is similar in its extent to what has been identified as an Accentual Phrase (AP) in those languages. However, Igarashi (2014) makes the case for Japanese that dialects differ in whether an AP is allowed to include more than one word or not. Thus, APs would not have to include the same average number of words across different languages.

377 These are function words semantically, because they have more of a grammatical function than a lexical meaning. However, morphosyntactically they are nouns and verbs like any word belonging to those classes: the demonstratives *kay* and *tsay* and the numeral *huk* can be followed by all nominal suffixes, and the possibility that they modify another noun by being placed in front of it without any suffixes is also available for semantically richer, more content-like nouns as well (cf. (149)a), although there are probably restrictions on such a use. The verb of being *ka-* likewise takes the same verbal suffixes and has the same morphosyntactic behaviour as any other verb, except that it is normally unrealized in the third person present when serving a copular function (not when it functions as a verb of existence or location). Looking at actual usage, the counts here suggest that these "function words" take up a sizeable share of token frequencies, and my impression is also that the five word forms given here will occur without suffixes, as monosyllables, more often than "content" words.

378 Cf. also the ratio of 1.12 content words per pitch accent for Huari Spanish in section 5.1.1.2.

The criteria used to group a stretch of Quechua speech into phrases were the following: in general, a phrase had to conform to one of the tonal contour shapes established in section 6.1. That is to say, with the exception of phrases including localized pitch accents on stress positions in words of Spanish origin (the "grafted" pattern), a phrase consists of exactly one high tone, possibly forming a plateau-like realization but not discontinuously so, either preceded by a low tone (rising pattern), or followed by it (only-falling pattern), or both followed and preceded by one (rise-falling pattern). Such a pattern had to be produced in a continuous, fluent fashion in order to be counted. Word sequences were excluded when they were completely flat or the pitch signal was too discontinuous, e.g. because of longer hesitations, or because of devoicing of phonologically voiced segments, but not just because of the presence of unvoiced consonants, if the overall pattern could still be recognized with some confidence. Since the data are semi-spontaneous, and Huari Quechua has few voiced consonants, stronger exclusion criteria would have been too restrictive. Within the three main categories of rises, only-falls and rise-falls, phrases were further distinguished according to the position of the tonal transition(s): in the case of the rises and only-falls, whether it (the rise or the fall, respectively) took place at the final internal word boundary in the phrase, a prefinal word boundary, the penult of the final, or a prefinal word, or another syllabic position in the final word or a prefinal word, for six subcategories each. The defining criterion for the location of the transition was the boundary of the high tonal target, i.e. the left boundary in the case of rises, and the right boundary in the case of falls. Thus the position of a tonal transition was classified according to where the local pitch maximum was determined to be located within the context of a rise or fall. This was done by visual and auditory inspection of the annotated utterance in *praat*. In the case of the rise-falls, both the rise and the fall could in principle take place at either a final or prefinal word boundary, or at the penult or another syllabic position in either a prefinal or the final word. Here I refer only to phrase-internal boundaries as word boundaries, so "the final word boundary" is that between the penultimate and the final word, and a "prefinal word boundary" is a boundary between any two words inside the phrase not involving the final word. Out of all the possible combinations, the following eleven subcategories were initially formed to class the phrases:

(153) Alignment subcategories of the rise-falling contours
  a) with the rise taking place at the final word boundary and the fall in the penult of the final word,
  b) with the rise taking place at a prefinal word boundary . . . and the fall at the final word boundary,
  c) . . .and the fall in the penult of the final word,

d) . . .and the fall later in the penult of a prefinal word,
e) with the rise taking place in the penult of a prefinal word . . .and the fall at the final word boundary,
f) . . .and the fall in the penult of the final word
g) . . .and the fall in the penult of a later prefinal word,
h) both the rise and the fall on the last two syllables of the final word,
i) both the rise and the fall on the last two syllables of a prefinal word,
j) another combination in which the high stretch covers most of the final word,
k) another combination in which the high stretch does not cover most of the final word.

It turned out that combinations (153) c), d), f), and j) do not occur in the corpus, which is why they are not listed in Table 32 below. Combination e) possibly occurred once, but that token is ambiguous between the rise taking place on the penult and at the initial word boundary of the prefinal word, because that word is bisyllabic. It has therefore been counted as an instance of combination b), and e) also does not occur in the table. Combinations h) and i) lump the syllabic positions of penult and final syllable together. This was done in order to reduce the number of combinations, and because we have seen that final peak alignment varies between these two positions in this contour (cf. section 6.1.6).

In this manner, using breaks only as additional boundary cues, it was possible to manually delimit and classify nearly all phrases unambiguously. Occasionally, it was difficult to decide whether several words in succession formed one or more phrases, with pitch movement possibly attributable to phrase tones on some or all of the words, but scaled much larger on the final one. Those cases had to be decided between constituting phrasal tonal movement or just random fluctuation. They were individually decided based on whether the prefinal movement was audibly percebtible, whether each identifiable tonal target plausibly corresponded to one of the assumed underlying tones of a phrase, how the scale of the prefinal movement compared relative to the final movement given the speaker's range, and whether it surpassed a threshold of 7 Hz difference. An example of such a case is Figure 171 (cf. (150)): on the two prefinal words *runa* and *ushqu*, rising pitch movements have a range from local minimum in the first syllable to local maximum in the second at 136–146 Hz and 133–142 Hz, respectively. These ranges are clearly smaller than that on the final word, with 134–159 Hz from minimum to maximum. Here the the prefinal rises could be audibly perceived (although weakly), and not only three peaks, but also three low tonal targets, could be made out, corresponding to a sequence LH LH LH of three rising contours. Therefore, this example was counted as consisting of three rising phrases, even though the pitch difference on the first two words is only just above the 7 Hz threshold.

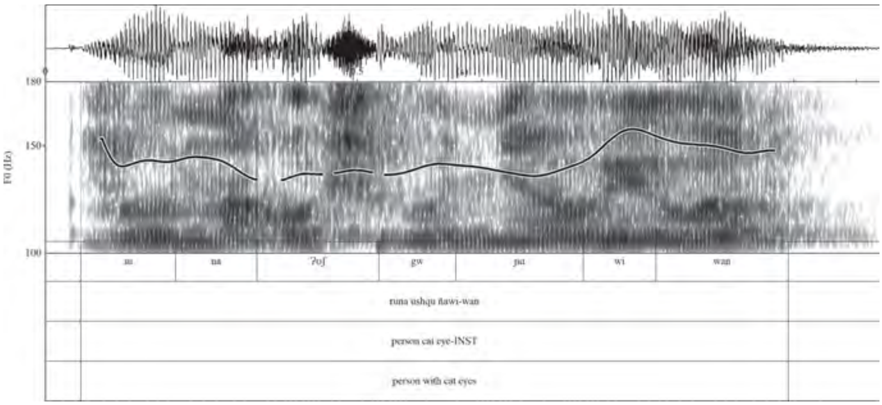


Figure 171: ZZ24_Cuent_Q_1635.[379]

In doubt, the tendency in such cases was to decide them in favour of a grouping involving more phrases rather than fewer.[380] Another ambiguous situation for analysis arose whenever a word with a rising contour on a syllabic position was directly followed by another word with an only-falling contour on a syllabic position. This could possibly also be counted as a single phrase with a rise-falling contour. Recall section 6.2.2 where in similar cases, when counting those as two separate units, location moves could nearly always be separated from item moves. Here, unless the pitch transition between the two words was completely smooth and there was absolutely no interruption whatsoever, this was counted as constituting a phrase break, instead of one rise-fall phrase, which contributes to why subcategory (153) f) has no counts.

Using these analytical criteria, the dataset was found to contain the following distribution of phrases, in rising (Table 30), only-falling (Table 31), and rising-falling (Table 32) contours.

379 https://osf.io/4aqbd/

380 In section 6.4, it will be argued that such phrases with systematic scaling differences form larger prosodic units together.

Table 30: Rising contour patterns in the Quechua nominal sequences subset, in all phrases (rows 2–4) and in phrases containing more than one word (rows 5–7).

	tonal transition location	**final word boundary**	**prefinal word boundary**	**final penult**	**prefinal penult**	**final other syllable**	**prefinal other syllable**	**total**
all phrases	**counts**	**26**	**8**	**63**	**12**	**40**	**4**	**153**
	% of all rises	17	5.2	41.2	7.8	26.1	2.6	100
	% of all phrases	8.9	2.7	21.6	4.1	13.7	1.4	52.4
phrases of >1 words	**counts**	**26**	**8**	**28**	**12**	**12**	**4**	**90**
	% of all >1-word rises	28.9	8.9	31.1	13.3	13.3	4.4	100
	% of all >1-word phrases	17.2	5.3	18.5	8	8	2.7	59.6

Table 31: Only-falling contour patterns in the Quechua nominal sequences subset, in all phrases (rows 2–4) and in phrases containing more than one word (rows 5–7).

	tonal transition location	**final word boundary**	**prefinal word boundary**	**final penult**	**prefinal penult**	**final other syllable**	**prefinal other syllable**	**total**
all phrases	**counts**	**4**	**1**	**33**	**0**	**12**	**1**[381]	**51**
	% of all only-falls	7.8	2	64.7	0	23.5	2	100
	% of all phrases	1.4	0.3	11.3	0	4.1	0.3	17.5
phrases of >1 words	**counts**	**4**	**1**	**9**	**0**	**5**	**1**	**20**
	% of all >1-word only-falls	20	5	45	0	25	5	100
	% of all >1-word phrases	2.7	0.7	6	0	3.3	0.7	13.3

381 This single count comes from a case with "inherited" pattern on the Spanish word *último* "the last".

In the rising contours, three of the four tokens in the least frequent category, with the rise taking place on a syllable other than the penult in a prefinal word (column 8 in Table 30), are cases of “inherited” patterns on oxytonic Spanish loanwords, with the Spanish stress position causing the rise to take place on the final syllable of the prefinal word. In the fourth token, the rise takes place gradually over the whole word, which seems to be an infrequent phonetic variant. Hence a contour with the rise in a prefinal word taking place on a syllable other than the penult seems unavailable without a lexically specified prominent syllable like in the Spanish loanwords (cf. the OT-analysis in section 6.3.2). In the rise-falling contours, no tokens were found where the rise took place at a word boundary, but the fall at a syllabic position, or vice versa, except for combination a), where the rise takes place at the final phrase-internal word boundary and the fall on the penult of the final word. This latter combination is a special case deriving from phrase-final tonal crowding (cf. the OT-analysis in section 6.3.1).

Table 32: Rising-falling contour patterns in the Quechua nominal sequences subset, in all phrases (rows 2–4) and in phrases containing more than one word (rows 5–7).

	tonal transition locations (cf. (153))	**a)**	**b)**	**g)**	**h)**	**i)**	**k)**	**total**
all phrases	**counts**	**12**	**6**	**2**	**56**	**8**	**4**	**88**
	percent of all rise-falls	13.6	6.8	2.3	63.6	9.1	4.6	100
	percent of all phrases	4.1	2.1	0.7	19.2	2.7	1.4	30.1
phrases of >1 words	**counts**	**12**	**6**	**2**	**10**	**8**	**3**	**41**
	percent of all >1-word rise-falls	29.3	14.6	4.9	24.4	19.5	7.3	100
	percent of all >1-word phrases	8	4	1.3	6.6	5.3	2	27.2

A first result from all three tables taken together is that rises are slightly more frequent (n = 153; 52.4%) than only-falls and rise-falls taken together (n = 139; 47.6%), but rises and falls broadly each make up about half of all tokens. Recall that in section 6.2.2, falling *moves* turned out to be in the majority, i.e. final phrases in individual moves. That rising *phrases* are relatively more frequent at this lower level of observation here is compatible with the hypothesis that they cue incompleteness, with moves in which the final phrase is falling also preferentially containing rising phrases prefinally. Secondly, patterns making use of a word boundary as tonal transition points are in a minority (34 tokens/22.22% of rises, 5 tokens/9.8% of only-falls, 18 tokens/20.45% (categories a) and b)) of rise-falls) compared to those using a syllabic position, overall. However, these counts include many phrases consisting of only a single word, where word boundaries cannot be the tonal transition locus by definition. In multi-word phrases (lines 5–7 each of Tables 30–32), patterns making reference to a word boundary make up a more sizable share, namely 37.8% of rises, 25% of only-falls and 43.9% of rise-falls. Yet while they are certainly a relevant variant for multi-word phrases, word boundary-patterns are also clearly not in complementary distribution with patterns making reference to a syllabic position, with the former used in multi-word phrases, and the latter only for single-word phrases. Instead, patterns making reference to a syllabic position are the majority variant in all phrases.

Table 33: Distribution of phrase patterns across the nominal sequences subset Quechua, ordered according to subcorpus, phrase type (rising, only-falling, rise-falling), tonal transition location (all (=syllabic position + word boundary) vs word boundary), and phrase length (all vs. only multi-word phrases).

		Rises	*of which* **wb**	*Only-falls*	*of which* **wb**	*Rise-falls*	*of which* **wb**	**Total**	**% wb** *rises*	**% wb** *only-falls*	**% wb** *rise-falls*	**% wb all**
all phrases	*TP03&KP04 MT*	30	4	15	1	20	8	65	13.3	6.7	40	20
	TP03&KP04 Conc	17	10	1	0	7	4	25	58.8	0	57.1	56
	SG15&QF16 Conc	9	1	7	1	7	0	23	11.1	14.3	0	8.7
	QZ13&OZ14 Conc	16	2	3	1	7	0	26	12.5	33.3	0	11.5
	XU31&OA32 Conc	13	0	4	0	12	0	29	0	0	0	0
	ZR29&HA30 Conc	12	5	14	2	8	0	34	41.7	14.3	0	20.6
	XQ33&LC34 Conc	16	5	2	0	0	0	18	31.3	0	--	27.8
	AZ23&ZZ24 Conc	18	5	2	0	7	0	27	27.8	0	0	18.5
	AZ23&ZZ24 Cuent	22	2	3	0	20	1	45	9.1	0	5	6.7

Table 33 (continued)

		Rises	*of which* **wb**	*Only-falls*	*of which* **wb**	*Rise-falls*	*of which* **wb**	**Total**	**% wb** *rises*	**% wb** *only-falls*	**% wb** *rise-falls*	**% wb all**
only multi-word phrases (>1 word)	*TP03&KP04 MT*	11	4	7	1	15	8	33	36.4	14.3	53.3	39.4
	TP03&KP04 Conc	13	10	0	0	6	4	19	76.9	--	66.7	73.7
	SG15&QF16 Conc	7	1	2	1	5	0	14	14.3	50	0	14.3
	QZ13&OZ14 Conc	9	2	2	1	1	0	12	22.2	50	0	25
	XU31&OA32 Conc	4	0	3	0	3	0	10	0	0	0	0
	ZR29&HA30 Conc	8	5	5	2	6	0	19	62.5	40	0	36.8
	XQ33&LC34 Conc	12	5	0	0	0	0	12	41.7	--	--	41.7
	AZ23&ZZ24 Conc	16	5	1	0	0	0	17	31.3	0	--	29.4
	AZ23&ZZ24 Cuent	10	2	0	0	5	1	15	20	--	20	20

Individual speaker preference also seems to play a role in the variation between word boundary and syllabic position as tonal alignment anchor: speaker pair XU31 & OA32 do not produce word-boundary-aligned phrases at all in the 26 nominal phrases of their *Conc* corpus, and speaker pair AZ23 & ZZ24 produce only 7 word boundary-aligned phrases in the 72 nominal phrases of their *Conc* and *Cuento* corpora together, or 9.7%. On the other end, TP03 & KP04 produce 29 (including 11 of the 12 tokens of pattern a) of the rise-fall phrases overall) of them in the 90 nominal phrases (52 of them multi-word) of their *Conc* and *MT* corpora together, amounting to 32.2% of all their phrases, and 55.8% of their multi-word phrases. Table 33 shows that this difference persists also if only the *Conc* corpora are compared, suggesting that an explanation in terms of different functions is less likely. I also could not find evidence that the distinction between locating the tonal transition at a word boundary or a syllabic position is functional in a careful qualitative investigation of all the Huari Quechua data in their contexts. The most likely hypothesis is therefore that the difference is one of individual speaker preference, i.e. a prosodic feature that is socioindexically meaningful, part of a speaker's style.

6.2.3.3 Phrasal prominence and relation to information structure and information status

This section investigates a feature of multi-word phrases that does seem to serve a functional purpose. The argument here is based on the assumption from metrical phonology that metrically assigned relative prominence extends to all levels of the prosodic structure, at each level (n) marking exactly one (the head) of the subordinate metrical constituents contained within it at the level below (n-1) as strong (s), and all others as weak (w, cf. section 3.3.1). I argue that this metrical structure also exists in Huari Quechua, and that the location of the high tone and the stretch (plateau) extended from it cue prominence at the phrase level.[382] I follow Ladd (2008: 223, 252), Calhoun (2010b), Jun (2014b), Féry (2017) in assuming that there is a default or unmarked prominence pattern, specific to languages and constructions,[383] which determines the location of the strong position in a phrase. This default prominence is compatible with a range of contexts, specifically at least those contexts where the focus domain as determined from the QUD is as broad as the phrase itself. Such an unmarked default is argued to be crucial for assigning a prominence structure based on expectation even in the (relative) absence of phonetic cues (Ladd 2008: 257–259, 271; Calhoun 2010b). My hypothesis is that in the subset of the Quechua data investigated here, on nominal sequences also syntactically right-headed in the majority, default prominence in a phrase is right-

382 It is usually assumed that high pitch, as the result of high tones, is more prominent than low pitch resulting from low tones, ceteris paribus. This has been shown in perception experiments and is part of the predictions of the Effort Code (Gussenhoven 2002a: 50, 2004: 85). Note that pitch height here is relational, never absolute, and relative pitch excursion size is therefore the most important cue to prominence. The Effort Code is based on the association of pitch movement with articulatory effort, so that passages with more pitch variation are perceived as more effortful, and in turn as more prominent. That is why I assume that in the rising and only-falling contours it is the word on which the tonal transition takes place that is most prominent (at the same time, the Effort Code also predicts rising contours to be more prominent than falling ones). Note however that language-specific assocations of pitch and information structure can sometimes go against this general trend: In Akan, Kügler & Genzel (2012) found that corrective focus is associated with relatively *lower* pitch.

383 It seems that languages can differ both in where main prominence falls within the same type of construction, and in whether it changes across constructions in the same language. In polar questions, English places prominence on the final lexical noun and on the verb only if no noun is available, while Russian always places it on the verb. In Russian, statements differ from polar questions in that main prominence goes instead to the final lexical noun (cf. Ladd 2008: 224–225).

most (as in Spanish and many other languages, cf. Ladd 2008: 252),[384] the last word being strong (s), and the others weak (w). The findings in this section will also shed light on whether for the data here, a similarly indirect and probabilistic relationship between prosodic cues, metrical structure and information structure can be assumed as the discussion in section 3.7.3 has suggested holds for different varieties of Spanish as well as Cuzco Quechua.

The assumption of phrasal prominence and of the relation between high tones and prominence is used here to generalize across the various identified phrasal contours. For the nominal sequences subset of the data, I assume that the relation between contours and metrical structures is the following, with phrasal prominence either final (on the final word in the phrase) or prefinal (on a prefinal word):

- *In rising contours*: those contours with the rising tonal transition either at the final boundary between words, or on a syllabic position in the final word cue final prominence. All others cue prefinal prominence.
- *In rising-falling contours*: those contours with the rising tonal transition taking place either at the final word boundary or at a syllabic position within the final word, i.e. where the high tone or plateau is entirely within the final word, cue final prominence. Contours where the final word is entirely excluded from the high plateau cue prefinal prominence. Those where a plateau begins before the final word but extends into it are counted as indetermined.
- *In only-falling contours*: only those with the falling tonal transition in the penult or the final syllable of the final word cue final prominence, all others cue prefinal prominence.

384 But not all, such as Japanese (cf. Jun 1993, 2005b; Venditti et al. 1996; Venditti et al. 2008). Venditti et al. (2008: 466, 477, 480–481) are at pains to emphasize that there is no structural or surface equivalent to a „nuclear“ or „sentence accent“ in Japanese; pitch accent tones are only assigned lexically and never postlexically. Prominence relationships or focus are signaled by pitch range expansion and compression and phrasing only. Mostly the same applies to (Seoul) Korean, except that no lexical accents exist either. However, it is still the case that culminative tonal marking at the default phrase edge is compatible both with a constituent positioned at that edge being narrowly at-issue and a QUD as broad as the phrase (Venditti et al. 2008: 481), just as in Spanish or English, except that the default phrase edge is to the left in Japanese, instead of the right. It seems plausible to take this edge-inversion as an effect of the inversion in the default metrical relation (s-w instead of w-s) between constituents, and thus to say that the differences only lie in the cues and this basic direction in the metrical structure, which nonetheless underlies all four languages, as Ladd (2008: 278–279) does. Note that this account of Japanese focus marking has been put in question by Ishihara (2011, 2016, 2017), who suggests that metrical structure in Japanese does not observe culminativity, and that phrasing is instead determined by syntax.

In addition, when a noun sequence is realized on several phrases, the entire sequence is counted as cueing final prominence only if the phrases are all rising or rise-falling and the final one clearly has the largest pitch span (meaning that sequences like (150), with a pitch contour as seen in Figure 171, will be counted as having final prominence). This implies that I take such phrases (PhPs) to form a larger unit together (either an IP or a recursively iterated larger PhP), in which the same default prominence relation (w-s) holds. If the phrases are of the same type but the final one does not have the largest pitch expansion, the sequence is counted as cueing prefinal prominence. All other combinations of phrases in a sequence are counted as indetermined.

Classing the nominal sequences dataset in this way means taking as basic unit not an individual phrase but an individual sequence (often but not always corresponding to a syntactic noun phrase). This was done in order to be able to say something about the relationship of words in a sequence where words are individually phrased. The absolute counts are thus different from the ones in Tables 30–33.

Table 34 gives counts for the prominence patterns (final, prefinal, indetermined) per nominal sequence classed according to information structure within the noun sequence. The information structural annotation the classification is based on was done on the entire Quechua dataset. That is to say, even though only the nominal sequences are considered here, for each corpus the entire discourse was analyzed and the annotation for the nominal sequences results from this. The nominal sequences were categorised according to which information structural function they most likely played based on an analysis of the discourse context, using the QUD model of discourse as well as the context model of Farkas & Bruce (2010) as analytical guidance, as laid out in section 3.7. The categories are thus explicitly not based on formal criteria like morphosyntactic or prosodic form. Such a contextual analysis has its limits, because an implicit QUD cannot always be determined exactly in actual conversation. When context did not allow for a precise determination of the implicit QUD in order e.g. to decide between a more or less narrow focus domain, the implicit QUD making less contextual assumptions (i.e. in general a broader focus) was chosen. The categories in the table are to be understood in the following: **Broad focus** means that the context was judged to most likely imply an (implicit, rarely explicit) QUD that does not impose a focus-background division within the nominal sequence itself, while the sequence as a whole is at-issue with respect to the current QUD. **Topic (broad)** also means that the context was judged to most likely imply an (implicit, rarely explicit) QUD that does not impose a focus-background division within the nominal sequence. Here however, the sequence as a whole is not at-issue with respect to the current QUD (i.e., backgrounded), but instead topical, i.e. it serves as a referential or predicational anchor or restriction for the current at-issue material relating it to the discourse progression via relevance. **Topic-comm** means that the context made it plausible to infer a division within the nominal sequence in terms of at-issueness, such that the earlier part[385] is a topic (backgrounded with respect to the current QUD but establishing a relation of relevance with regards to a preceding QUD) and the later at-issue with respect to the current QUD. Those cases where context allowed to decide ambiguous noun sequences as copular (cf. (147), (148)) make up the majority of this category, with the reasoning that the possibility to freely omit subjects in Quechua if they are already active in the discourse (pro-drop) conversely makes it likely for them to be topical if not omitted (cf. Lambrecht 1994: 137). **Prefinal focus (narrow)** also means that context made it plausible to infer a division within the nominal sequence in terms of at-issueness, such that a prefinal constituent could be taken to be at-issue, with the domain of at-issueness minimally excluding the final word. **Final focus (narrow)** is the complement to prefinal focus (narrow), in that the domain of at-issueness here could be understood to cover minimally the final word and to exclude at least one prefinal word. **Prefinal focus (corr)** and **Final focus (corr)** cover the same context-inferred separations of the nominal sequence as prefinal (narrow) and final (narrow), respectively, except that it was also possible to make out a salient alternative, uttered in the preceding context, to which the focal constituent is a correction (i.e. as a divergent answer to the same QUD).

Table 34: Prominence patterns in the noun sequences dataset of Quechua according to information structure relation within the noun sequences.

		information structure within the noun sequence							
		broad focus	topic (broad)	topic-comm	prefinal focus (narrow)	final focus (narrow)	prefinal focus (corr)	final focus (corr)	total
prominence	final	50	39	1	5	9	0	1	105
	prefinal	19	7	1	16	0	1	0	44
	indeterminate	40	22	8	2	1	0	0	73
	total	109	68	10	23	10	1	1	222

Table 35 classes the same data according to the information status profile of the nouns in the sequence. The categories used in the table are informed mainly by the classification in Baumann & Riester (2012), who themselves incorporate insights

385 Categories do not specify the location between which words in a sequence the division takes place, only the order. The 487 words occurred in 222 nominal sequences consisting of at least 2 and on average 2.19 words.

Table 35: Prominence patterns in the noun sequences dataset of Quechua according to information status (givenness/newness) within the noun sequences.

		Information status within the noun sequence				
		all new	all given	given-new	new-given	total
prominence	final	40	33	20	12	105
	prefinal	28	6	3	7	44
	indeterminate	41	10	19	3	73
	total	109	49	42	22	222

from several preceding works, such as Chafe (1976, 1994); Gundel et al. (1993); Prince (1981, 1992). However, several distinctions made in those works have here been classed on either side of a binary division, given or new. This was done as an adaptation to the type of data as well as in order to reduce the number of categories to a manageable size, so that the main question regarding a relationship between prominence and information status within the nominal sequence could be investigated. The categories were created using the following annotation schemes: if a referent occurred for the first time in the discourse (the respective task corpus), the referring expression it was denoted by was annotated as new; if it occurred repeatedly, as given. This corresponds mostly to the concept of referential givenness of Baumann & Riester (2012), but disregards the several more fine-grained distinctions they make within the broader category of referential givenness. In *Conc* and *MT*, the "new" here corresponds most closely to their *r-unused-known* ("discourse-new item which is generally known", Baumann & Riester 2012: 138), or *r-environment* (discourse-new items that refer to "visible objects in the communicative environment which are not available in the speech setting by default", cf. 139), because the participants were able to familiarize themselves with all the referents (the figures on the cards) before playing (*Conc*) or had their images in front of them (*MT*) during the game. In *Cuento*, what is "new" here can occasionally also be their *r-new* ("specific or existential indefinite introducing a new referent", Baumann & Riester 2012: 138), with the difference that there is no formal marker on nouns for definiteness in Quechua. Annotating only according to referential givenness means that if two expressions are not coreferential, an expression that is formally partially the same as another having occurred earlier will not be counted as given (e.g. *kantu-chaw* in first *hana kantuchaw* "at the upper border" and later *washa kantuchaw* "at the border over there"). Referential givenness is only applicable to referring expressions: often a nominal sequence in the data, if it consists of only one NP such as *hana kantuchaw*, is referentially given or new only as a whole, since only the noun together with its modification denotes a single identifiable referent in the discourse. The two components *hana* and *kantuchaw* do not refer to separate referents individually.[386] So far, this leaves out another aspect of givenness, called lexical givenness in Baumann & Riester (2012). Lexical givenness refers not to referents but only to expressions, even individual lexical items. An expression is lexically new if it occurs for the first time, and given if it, a synonym, a hypernym, or a holonym occurred previously in the discourse. There is a difference in temporal restriction from the treatment of referential givenness implemented by Baumann & Riester (2012: 144–145) and first postulated by van Deemter (1994): referring expressions are retained for a shorter while than their referents, and can thus become "new" again. An annotation according to lexical givenness allows us to annotate a givenness difference even within a nominal sequence that forms a single referring expression: thus, *kantuchaw* in *washa kantuchaw* is (lexically) given if *hana kantuchaw* occurred a short enough distance before. The distance chosen here was 5 utterances, which is similar to the 5 intonation phrases chosen by Baumann & Riester (2012), or else 25 seconds. Lexical givenness enters the data in Table 35 only as *difference* in givenness *within* a nominal sequence; i.e. some NPs that denote new referents might be partially made up of expressions that are given, like *kantuchaw* in *washa kantuchaw* in the situation just described. This is then counted as **new-given** (column 5 in Table 35). If the opposite order obtains (e.g., *washa* is lexically given because of a recent previous occurrence of *washa ladu* and then the referentially new NP *washa kantuchaw* is introduced), it is classed as **given-new** (column 4). The opposite situation, where a given referent is partially expressed via lexically new expressions, would have been treated in the same way but did not occur. In cases where lexical and referential givenness were completely opposed, a decision would have had to be made regarding which type of givenness would have been given precedence.[387] However, this did not happen on the data

386 Note that Quechua here works differently from the languages Baumann & Riester (2012) base their work on: they state that referential givenness can only apply at the level of the DP, which is formally different from an NP in languages like English or German. Quechua has no formal marking in this respect. On its own, *kantuchaw* or just *kantu* can be a full referring expression if context allows only a single interpretation (e.g. there just being one border) and no overt specifier such as *hana* is necessary. An analysis might assume that such a noun then has a covert specifier and is hence a DP, but this could only be argued from the broader context, i.e. when it allowed us to deduce the specificity of *kantu* in such a case. The DP assumption thus has to be argued somewhat circularly. The point is that languages that like Quechua lack formal markers for definiteness and related nominal categories suggest that the relation between information status and nominal syntax is less clear-cut ("only DPs can be referring expressions") than it would seem from the perspective of languages that do have them.

387 When a given referent is wholly expressed via lexically new expressions, Baumann & Riester (2012: 146, 150) hypothesize but cannot demonstrate, using their German corpus study, that referential givenness should trump lexical newness insofar as the resulting expression should still be considered here. Modifying demonstratives and nominalized forms of the copula were annotated according to the referential givenness of the whole expression (cf. Baumann & Riester 2012: 143–144).

One of the main findings in the two tables is that phrases with final prominence are overall in the majority, making up nearly half (47.3%) of all observations. This in itself is some evidence that final prominence is indeed a kind of default at the level of the phrase, at least if we accept the assumed relation between high tone position and prominence that is the basis of the prominence classification here. The prevalence of this prominence pattern is also relevant if we recall the pattern called phrase accentuation in the analysis of the Spanish data (cf. section 5.1.3.1). The phrase accentuation of Spanish has also been shown to cue final prominence. In terms of contour shape, it is very similar to a rise-fall contour of Quechua in which the high tone is aligned with a syllabic position in the final word (column h in Table 32, accounting for about a quarter of rising-falling contours on multi-word phrases). Equally, falling contours with prefinal prominence seem not only to resemble contours with postfocal deaccentuation in Spanish, but to also serve somewhat similar functions if the findings here are correct. My aim here is not to argue for one being the origin of the other, but to point out the structural and

---

deaccented. The opposite case is also unclear. Baumann & Riester (2012: 147, 150) hypothesize that new referents expressed by given lexical items should also be deaccented, but again their German data does not confirm this. An example in van Deemter (1994: 5) is interesting in this regard:

(i) Clinton visited many towns; when he finally arrived in Clinton, he was late.

Here, the first "Clinton" refers to the former US president and the second to a town of the same name. In this case, despite the lexical givenness of the expression "Clinton", its second occurrence is thought to be obligatorily accented (van Deemter 1994: 5). This stands in contrast to an example from Büring (2007: 448):

(ii) A: Why do you study Italian?
B: I'm married to an Italian.

In (ii), the second occurrence of "Italian" is thought to be deaccented because of its lexical givenness. While van Deemter (1994: 5) takes the accent on the second "Clinton" to be due to the difference in denotation between the two Clintons but also suggests that focus or contrast might play a role, Baumann & Riester (2012: 136, footnote 13) propose that the information of there being a town called Clinton has a degree of "extra newness", overriding lexical givenness here. They provide a further example:

(iii) Clinton shares his name with a town; when he finally arRIVED in Clinton, he was late.

Here, the second occurrence is thought not to be accented anymore. All of this seems to indicate that referential newness can actually overrule lexical givenness on occasion, contra the hypothesis by Baumann & Riester (2012).

realizational similarities. It seems plausible that the bilingual speakers here take recourse to a core of cueing strategies available to them independent of the language they speak as part of the repertoire of their speech community (which does not preclude their also being typologically frequent), and which they further adapt according to other more divergent structural requirements.

For the counts in both tables, χ^2-tests were done to check against the null hypothesis that rows (prominence location) and columns (information structure / status) are independent of each other. Fisher's exact tests were also done because the expected values in some cells in both tables were low enough to affect the reliability of the χ^2-test (some were below 5, none were below 1; cf. Field et al. 2012: 816). For Table 34 (information structure), the two single counts for "final (corr)" and "prefinal (corr)" were reassigned to "final (narrow)" and "prefinal (narrow)". The results of the χ^2-test (χ^2= 63.557, df = 8, p = $9.3\text{x}10^{-11}$) and Fisher's exact test (two-sided, p = $4.49\text{x}10^{-9}$) are both highly significant. For Table 35 (information status), the results of χ^2-test (χ^2 = 22.804, df = 6, p = 0.00086) and Fisher's exact test (two-sided, p = 0.0006) are also both highly significant. This suggests that columns and rows in both tables are not independent of each other, i.e. that both information structure and information status as annotated are associated with the position of prominence in the data. The adjusted standardized residuals for each cell reveal that only some counts contribute to the overall significance. For information structure, among them are the counts for final prominence under narrow final focus (10 counts including one from final (corr); z = 2.972, p < 0.01) and those for the inverse case, prefinal prominence under narrow prefinal at-issueness (17 counts including one from prefinal (corr); z = 6.638, p < 0.001, note that the expected value for this cell is just below 5 at 4.76, so the p-value is probably not fully reliable). This indicates that these respective information structural conditions both associate with the two different prominence patterns more frequently than expected under the null hypothesis of independence.

Table 36: Adjusted standardized residuals from the χ^2-test of independence for prominence position according to information structure (cf. Table 34, the "corr" counts are here integrated into the "narrow" counts). Exclamation mark indicates a cell where the expected value is <5; *, **, and *** indicate a p-value of <0.05, <0.01, and <0.001, respectively.

	focus (broad)	**topic (broad)**	**topic-comm**	**prefinal (narrow)**	**final (narrow)**
final prominence	−0.418	1.994*	−2.417! *	−2.750**	2.972**
prefinal prominence	−0.877	−2.366*	−0.797!	6.638 (!) ***	−1.691!
indeterminate	1.188	−0.112	3.246!**	−2.711**	−1.723 !

Of the complementary conditions final prominence under narrow pre-final at-issueness ($z = -2.75$, $p < 0.01$) and prefinal prominence under narrow final at-issueness ($z = -1.691$, $p > 0.05$), both with negative z-scores, only the former reaches significance, while the latter rather narrowly misses it (also with an expected value of only 2.2). Overall, this is some evidence for the position of a constituent being narrowly at-issue being reflected in the position of prominence, as expected: narrow final at-issueness associates positively with final prominence and narrow pre-final at-issueness positively associates with pre-final prominence (with the caveat of the expected values here being just below 5) as well as negatively with final prominence. The counts do not fully allow us to conclude that narrow final at-issueness also associates negatively with prefinal prominence, although they point in this direction. In the case of broad topics, but not of broad focus, there is a positive association with final prominence ($z = 1.994$, $p< 0.05$) and a negative one with prefinal prominence ($z = -2.366$, $p < 0.05$). This is at least some evidence in the expected direction, namely that a broad information structure on the noun sequence (no internal division) associates with final prominence.

Table 37: Adjusted standardized residuals from the χ^2-test of independence for prominence position according to information status (cf.Table 35). Exclamation mark indicates a cell where the expected value is <5; *, **, and *** indicate a p-value of <0.05, <0.01, and <0.001, respectively.

	all new	all given	given-new	new-given
final prominence	−3.107 **	3.184 **	0.046	0.717
prefinal prominence	2.154 *	−1.507	−2.289 *	1.487 !
indeterminate	1.474	−2.106 *	1.893	−2.025 *

In the case of information status (Table 37), an unexpected result is that nominal sequences classed as completely new are significantly negatively associated with final prominence ($z = -3.107$, $p < 0.01$) and positively associated with prefinal prominence ($z = 2.154$, $p < 0.05$). Somewhat more expected is that being completely given is significantly positively associated with having final prominence ($z = 3.184$, $p < 0.01$), and that a sequence with a given-new partition is negatively associated with prefinal prominence ($z = -2.289$, $p < 0.05$). The latter two results might be taken as some evidence that prominence is final by default, i.e. if nothing else intervenes, and that relative newness in the sequence attracts prominence. For the surprising result about the all-new sequences, recall that the annotation used a threshold of 5 utterances or 25 seconds distance (adapted from a similar threshold in Baumann & Riester 2012) between two mentions of a word for it to be once again classed as lexically new instead of given. This means that some instances of noun phrases like *primera fila-chaw* “in the first row” in *Conc* were counted as all-new because the last mention of *fila* was beyond this limit. However, it is quite plausible that in particular such frequent words that were essential to how some of the speakers played the *Conc* game were actually active for longer, possibly having a kind of "default" givenness status in this task. This suggests that the threshold for re-classing an expression as lexically new after it had previously already been given should probably be set at a further distance than the one proposed in Baumann & Riester (2012), at least for such types of rather repetitive conversational tasks. It might also be worth considering in further resarch whether different lexical items can be taken to have different thresholds of this kind, depending on how central they are to the type of conversational task at hand. The following will shortly explore further possible factors.

A partial explanation for the ambiguous findings regarding prosodic cues related to information status overall might be found when looking at individual examples. Once again, a difference seems to exist between speaker pairs in terms of their preferred strategy. Section 6.1.8.3 showed that the speaker pairs AZ23 & ZZ24, on the one hand, and QZ13 & OZ14, on the other, occupy opposite ends of a spectrum with regards to how they treat Spanish-origin words. This could be demonstrated particularly well because they both played *Conc* describing card locations by referring to them via a kind of grid coordinate, made up of an ordinal number of Spanish origin plus a form of the Spanish-origin *fila* 'row' to specify the row, and then another instance of an ordinal number plus the locative *–chaw* to specify the column. In realizing these elements they also differ regarding the prosodic treatment of information status and/or structure. Over the course of their *Conc* game, the four speakers all uttered the sequence *x fila(-chaw)* several times (OZ14 5, QZ13 10, AZ23 13, ZZ24 3 times), where x stands for one of the Spanish ordinal numbers from one to three. Forms of *fila* were mostly (but not always, see above) lexically given because of their frequent occurrence, while the preceding ordinals *primer(a)/segund@/tercer(a)* were mostly lexically new, or, if they had recently occurred already, also given. In any case, given that these four speakers always began the specification of a location at which they wanted to guess the identity of an object with such an instance of *x fila(chaw)*, in the context of how speakers played the game, the ordinal number *x* is the more informative part of the phrase, since *fila(chaw)* was highly expectable both from its position as immediate follower of the ordinal number (the only other attested option for following an ordinal number in these corpora is the locative suffix) and as part of the initial phrase for each new location-specifying move, since the specification of the column by numeral+*chaw* always followed after the row specification. In the information structural annotation, these phrases were uniformly annotated as beingly broadly at-issue, because assuming a more specific implicit QUD like IN WHICH ROW AND COLUMN IS WHICH ITEM? instead of the broader WHERE IS WHICH ITEM? depends exactly on the task-specific information status of items like *fila*, which is in question here. However, it should be clear that in terms of both information status and structure, as well as regarding plausible assumptions about how speaker and hearer here expected the discourse to evolve given the absolute regularity of these sequences,[388] the immediate contexts for these utterances in both corpora (QZ13 & OZ14 and AZ23 & ZZ24) are as similar as could be. With that in mind, it is remarkable that the two pairs realize these phrases in quite distinct fashion. QZ13 & OZ14 realize them either with two pitch peaks ("grafted" pattern) on the two Spanish stressed syllables (in the numeral and the form of *fila*), plus a rise towards the end of the final syllable, or with just a rise or rise-fall on the Spanish stressed syllable of the numeral, with *filachaw* without its own pitch event ("inherited" pattern), i.e. in the latter case with a pattern counted as cueing prefinal main prominence (see Figures 158–160 above). AZ23 & ZZ24, on the other hand, most often realize them with a rise that takes place on the penult of the second word, *filachaw*, or at the word boundary between the first and second word, i.e. with a pattern counted as cueing final prominence (see Figures 114 and 164 above). QZ13 & OZ14's realizations might be taken as evidence that they interpret the context as calling for narrow prefinal focus. Their realization with prefinal prominence could also mean that prosody in their case is sensitive to the relative givenness of *filachaw* compared to the preceding numeral. In turn that would suggest that AZ23 & ZZ24's realization are *not* sensitive to it. Thus QZ13 & OZ14, who in terms of paying heed to Spanish stress positions are more "Spanish-like" in their Quechua prosody, here behave less "Spanish-like", as relative givenness is generally thought not to be cued in Spanish (cf. Cruttenden 2006; Hualde 2002; Hualde & Colina 2014). Conversely, AZ23 & ZZ24, who pay little heed to Spanish stress positions, would here be more "Spanish-like" in their insensitivity to relative givenness. Whatever explanation actually holds, it seems clear that under nearly identical context conditions, these two speaker pairs choose two systematically different prosodic realization strategies.

Another aspect worth exploring is the treatment of forms of the nominalized copula (*ka-q-*). Wether this is also a matter individual preferences or something else is not obvious here. In modifying prenominal position, both *ka-q* and the demonstratives *kay* and *tsay* are almost never the location of the tonal transition, i.e. they are either uniformly low or high, and many word boundary rises occur at the word boundary after one of these modifiers.[389] On the other hand,

388 Cf. Calhoun 2010a; Turnbull et al. 2015 on language-specific relations of predictability and consequent informativity to prosodic prominence, and Clopper et al. 2018 for indications that "predictability" might not be a unified phenomenon.

389 The only case where a prenominal *ka-q* is not part of the initial low or high stretch is KP04_Conc_0324, *kay kaq ladu kantu-chaw* "at the border that is at this side", where *kay kaq ladu* is with *kaq* in postnominal position as the last word in a phrase (more frequent than its occurrence as prenominal modifier), some of the speakers (TP03, KP04, QF16, SG15, ZR29), do not realize the final phrasal tonal transition on forms of *ka-q*. They realize the final rise or fall on the preceding content noun, so that the resulting phrase is counted here as cueing prefinal prominence (18 instances in total, see Figures 116, 126, 127 above for examples). Of those, 14 tokens are in sequences classed as all-new, because of the referential newness of the entire sequence. Other speakers (OA32, LC34), however, do realize the final movement on forms of *kaq* or at its initial word boundary (8 instances in total, see Figure 128 for an example). Of those, only 5 are in sequences classed as all-new. This is not a clear-cut difference between speakers because KP04, SG15 and LC34 also each realize *ka-q* in the respective other way at least once. The contexts in which *ka-q* occurs are also much more heterogeneous than is the case with *x fila(-chaw)*. In all, only five *ka-q*-final sequences in the whole corpus are classed as all-given, in contrast to 24 classed as all-new (of which 14 have prefinal prominence, 5 final prominence and 5 are indeterminate). This could suggest that the function of *ka-q* is somehow related to introducing new referents, see Bendezú Araujo (2021) for more on its meaning. That *kaq*-final sequences with pre-final prominence are so much more often new rather than given, together with the described tendency by a majority of speakers to not extend the phrasal high tone to it, goes some further way in explaining why all-new sequences are associated with prefinal prominence. The behaviour of *ka-q* and the demonstratives seems to indicate that a lexical factor is also involved in determining which part of a phrase the high tone is realized on, possibly also subject to individual variation.

The foregoing discussion suggests that in the data summarized in Tables 34–37, speaker-and likely also task-specific strategies interacting with lexical factors are hidden. If the behaviour of each individual speaker were observed on more data, individually regular strategies for the cueing of prominence and its relation to information status and information structure would likely emerge. This task is left to future research.

Overall, the results in this section support the hypothesis that the relation between prosodic cues, metrical structre, and information structure in Huari Quechua is quite as indirect and distributional as discussed in section 3.7.3 for Spanish and Cuzco Quechua. They also support the hypothesis that the high portion of phrasal contours is associated with prominence, and that that promi-

realized in a rise-fall phrase, with both the tonal transitions, the initial rise and the final fall, taking place at the initial and final word boundary of *kaq*, respectively. The elevated pitch on *kaq* is also clearly audible. I have no explanation for why this is the case.

nence profile of a phrase provides cues to its information structure and the information status of the elements it is composed of. In the next section, this main finding will be used for the OT-analysis of the Huari Quechua contours. Section 6.4 will then look at individual examples in context, expanding the discussion to sequences containing verbs, and use findings from this section as guide for the analysis of how information structure relates to prosody in those examples.

The entire analysis in this section has involved several steps of somewhat subjective interpretation in the annotation process. I think that the approach is still justified, since it allows for some exploratory generalizations despite the spontaneous nature of the data. Hopefully, they can also serve as basis for the better formulation of testable hypotheses in future research.

6.3 OT-Analysis

In this section, the tonal alignment patterns will be translated into OT constraints and thereby related to each other as well as to Huari Spanish intonation, providing the second half of the answer to question (36)b, the first half of which was given in section 5.3. I will derive constraint rankings generating all attested phrasal contours and alignment variants. Beginning with the rise-falling contours (6.3.1), it will be established that alignment with the word boundary (cf. section 6.1.2) and alignment with the word penult (cf. section 6.1.4) require different rankings, thus constituting a word boundary-pattern and a word penult-pattern, respectively, while alignment with the phrase boundary (cf. section 6.1.3) occurs in both of them to different degrees. The variation in alignment of the final peak in the phrase between the penultimate and final syllable (cf. section 6.1.6) will be shown to naturally arise from the constraint rankings for the word boundary- and the word penult-patterns. The rising and the only-falling contours will be covered in sections 6.3.2 and 6.3.3, and the "inherited" and "grafted" patterns in 6.3.4. The rankings moving from the word boundary-variant via the word penult-variant to the loanword variants describe a progression from a purely edge-oriented prosody to one where prominent syllabic positions become successively more integrated, the opposite direction of the one in which we progressed in the Spanish OT analysis (section 5.3). This is one of the points of connection to the Spanish analysis, but not the only one. We already saw that certain contours exist in both Huari Spanish and Quechua. Up to a certain point, they can be interchangeably generated by several different constraint rankings. Thus the tonal grammar of the speaker community can be shown to have both language-specific peripheries and a common central space.

Starting with assumptions, I assume that a metrical relation between words in a phonological phrase exists which makes exactly one word more prominent than the others (cf. section 3.3.1). I also assume that as a default, it is the rightmost word that is strongest in the metrical representation. Default here means what is to be understood or expected in the absence of signaling to the contrary (cf. Calhoun 2010). Such a representation is ambiguous with regards to information structure, as it might indicate an answer to a broad QUD or one in which only the rightmost element is at-issue (cf. sections 3.7.3, 6.2.3.3). See section 6.2.3.3 for evidence and a more detailed argumentation. I assume that the location of the high tone in a contour cues prominence, in the following way (cf. section 6.2.3.3):

In rising contours: the most prominent word in the phrase is cued by the rising tonal transition taking place on a syllabic position on that word or at the left boundary of that word. That is to say, if the rise takes place on the initial of two words in a phrase, and the final word is realized with a high plateau, then the initial word is taken to have higher prominence. But if it occurs at the word boundary between them or anywhere on the final word, the final word has the higher prominence.

In rise-falling contours: the most prominent word in the phrase is again cued by the rising tonal transition taking place on a syllabic position on that word or at the left boundary of that word. That is to say, in a 3-word phrase where the rise takes place at the left boundary of the middle word and the fall takes place on its penult, the middle word is taken to be prominent. If the rise takes place at the left boundary of the final word and the fall on its penult, the final word it taken to be prominent. The analysis here omits phrases where the rise takes place at the beginning of or within one word and the fall on a later one. As the findings in section 6.2.3.2 indicate, they are very infrequent variants (category g) in Table 32, 2 tokens) and could also consist of a rising phrase plus a falling one.

In only-falling contours: the most prominent word is that inside or at the right word boundary of which the fall occurs.

Regarding which forms are actually attested, I use the counts given in Tables 30–33 from section 6.2.3.2. For the input forms of Quechua, I assume the penult to be marked as the prominent syllable in a word (consisting of a stem plus suffixes) underlyingly, but this only has an effect in the ranking for the word penult-pattern, showing that the Quechua tonal grammar pays varying attention to stress in the sense of the typology by Hyman (2014). Strictly speaking, if a speaker produced only utterances in the word boundary-variant, there would be no grounds to assume the presence of a prominent position at the word level in their grammar at all. However, in our data, no speaker is consistent in this (cf. section 6.2.3.2). I therefore take the penult to be prominent, but this prominence is solely expressed by serving as an alignment anchor in the word penult-variant. I assume the three contours, the rising-falling, the rising, and the only-falling one, to be different in their tonal input. The rising-falling contour is assumed to have the input tonal sequence $L_{\phi Lt}$, H, $L_{\phi Rt}$; the rising contour $L_{\phi Lt}$, H; and the only-falling contour H, $L_{\phi Rt}$.

For reasons of space, the analysis will use only phrases containing two words. It easily extends to phrases of only one word because in those cases "internal" alignment is outranked by alignment with the phrase edges, and what would have to be considered for longer phrases will be remarked on individually. The following are the alignment constraints[390] to build the varying rankings from (all either taken or slightly adapted from Gussenhoven 2000a, 2004):

(154) ALIGN ($L_{\phi Rt}$, Rt_{ϕ}) align the right L tone with the right edge of the phonological phrase

(155) ALIGN ($L_{\phi Lt}$, Lt_{ϕ}) align the left L tone with the left edge of the phonological phrase

(156) ALIGN (H, Rt_{ϕ}) align the H tone with the right edge of the phonological phrase

(157) ALIGN (H, Lt_{ϕ}) align the H tone with the left edge of the phonological phrase

(157) ALIGN (H, Lt_{ω}) align the (left edge of the) H tone with the left edge of a prosodic word[391]

(159) ALIGN (H, $Lt_{\omega'}$) align the (left edge of the) H tone with the left edge of the strongest prosodic word in PhP

(160) ALIGN (H, $Rt_{\omega'}$) align the (right edge of the) H tone with the right edge of the strongest prosodic word in PhP

(161) ALIGN (H, σ', Rt): align the right edge of H with the right edge of a stressed syllable

(162) ALIGN (H, σ', Lt): align the left edge of H with the left edge of a stressed syllable

These constraints are automatically in competition with each other. Their competition will generate both the observed long low stretches and high plateaux and the

390 The constraints used for the Quechua analysis were mostly already introduced in the Spanish OT-analysis (section 5.3). I repeat them here for convenience.
391 The complementary constraint aligning the H with the right edge of the prosodic word was not found to have an effect in any of the contours.

tonal transitions between them. In order for them to be able to do this, we furthermore need three high-ranked faithfulness constraints ranked above all of them.

(163) LINEARITY: the sequence of tones in the output is the same as in the input

(164) MAXIO(T): every tone in the input has a correspondent in the output

(165) NOCROWD: a TBU has only one tone

The TBU is assumed to be the syllable here. The puzzle about cases where more than one tone occupies a single syllable (cf. the discussions in sections 6.1.7 and 6.1.8.1) will be mostly left aside. We would need further systematic data to really decide this point. In principle, an interesting solution might be that deciding between moras and syllables as TBU is an additional variable choice for speakers. In both cases, NOCROWD would be kept high-ranked. Another observed effect of tonal organization under temporal constraints is already integrated in the ranking here: NOCROWD is ranked above all alignment constraints, necessary for the "displacement" effect observed in section 6.1.3, but underneath MAXIO(T), so that when a phonological phrase is mapped onto a monosyllabic word, the tones will be realized despite crowding. However, since there are also attested cases where tones are clearly truncated because of crowding, this is likely not the whole story. Recall that Huari Spanish also shows variability in tonal realization under temporal constraints (section 5.3.3).

Regarding association constraints, the findings on phonetic enhancement in section 6.1.6 suggest that the H tone at least in the rising-falling contours associates some of the time, but not that the stressed penult always associates with a tone (so that $\sigma' \leftarrow T$ is never active, unlike in Spanish, cf. 5.3.2). For such cases, three association constraints are needed:

(166) $H \rightarrow$ TBU: the H tone is associated with a TBU

(167) $(\sigma')_{\omega'} \leftarrow T$: associate a tone with the stressed syllable of the metrically strong prosodic word in a phonological phrase

(168) NOASSOC: TBUs are not associated with tones

Their ranking must be $H \rightarrow$ TBU >> NOASSOC >> $(\sigma')_{\omega'} \leftarrow T$ to conform to the observations. $H \rightarrow$ TBU and NOASSOC likely have overlapping distributions (Boersma & Hayes 2001), so that sometimes, the H tone fails to associate, because not every syllable on which the peak is realized is longer than adjacent syllables. In the loanword patterns, $(\sigma')_{\omega'} \leftarrow$ T is then also promoted above NoAssoc, because of the observed lengthening of Spanish stressed syllables in them. These association constraints do not interfere with the alignment constraints generating the different contour variants discussed in the following, so they are not given in the rankings.

6.3.1 Rise-falling contours

The progression through the analysis of the variants begins with the rising-falling contours. Since they include one more tone than the other two contours, their analysis establishes the basic mechanisms that can then be adapted to the analysis of the bitonal contours (sections 6.3.2 and 6.3.3).

6.3.1.1 Word boundary-variant

In all the constraint rankings here, the three faithfulness constraints are ranked above the alignment constraints. For rise-falls, the highest-ranked alignment constraints, Align ($L_{\phi Rt}$, Rt_{ϕ}) and Align ($L_{\phi Lt}$, Lt_{ϕ}), establish the initial and final L tone as initial and final phrasal boundary tones, respectively. They are followed by the constraints aligning the H tone with the boundaries of the prominent prosodic word, Align (H, $Lt_{\omega'}$) and Align (H, $Rt_{\omega'}$). Their ranking ensures that in this variant, a high plateau-like realization will form on that word. The relative ranking of Align ($L_{\phi Rt}$, Lt_{ϕ}) >> Align (H, Rt_{ϕ}) below them ensures that in phrases in which a prefinal word is prominent, a low stretch will form on all words following the prominent word. Align (H, Lt_{ω}) >> Align ($L_{\phi Lt}$, Rt_{ϕ}), the next constraints in the ranking, are relevant for phrases with more than two words, which are not treated in the tables here: their ranking results in the plateau-like realization extending further to the left of the prominent word, either up to the first internal word boundary, as attested e.g. in Figure 126, or to another prefinal one, as e.g. in Figure 127.[392] The three constraints below have no influence in this ranking, their relative ranking results from how they are ranked in the other variants. That I leave them ranked relative to each other at all is because I want to change as little as possible between the variant rankings.

392 How the difference between these two could be captured is not covered here. Note that in both examples, the initial low stretch extends across the demonstratives, and the plateau begins with the first content word.

(169) Alignment constraint ranking for the word-boundary variant (with multiple alignment of the H tone) of Quechua (rise-falls)
LINEARITY, MAXIO(T), NOCROWD >> ALIGN ($L_{\phi Rt}$, Rt_{ϕ}) >> ALIGN ($L_{\phi Lt}$, Lt_{ϕ}) >> ALIGN (H, $Lt_{\omega'}$) >> ALIGN (H, $Rt_{\omega'}$) >> ALIGN ($L_{\phi Rt}$, Lt_{ϕ}) >> ALIGN (H, Rt_{ϕ}) >> ALIGN (H, Lt_{ω}) >> ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) >> ALIGN (H, Lt_{ϕ}) >> ALIGN (H, σ', Lt) >> ALIGN (H, σ', Rt)

The low ranking of the syllabic alignment constraints (161) and (162) means that the word-boundary pattern does not make reference to a syllabic position at all. This is because section 6.2.3.2 established that rise-fall phrases are unattested that combine a rise taking place at a syllabic position within a word with a fall taking place at a boundary between the words, or vice versa, except for the case where the rise takes place at the boundary preceding the final word, and the fall taking place on that word's penult or final syllable ((153)a), cf. Table 32). This seeming exception results from the same ranking: the fact that the fall at the right edge of the H tone in rise-falling contours often aligns with the penult of the final word, even if the rise at its left edge aligns with the preceding word boundary, simply falls out from the high-ranking constraint aligning the L tone with the right edge of the phrase, ALIGN ($L_{\phi Rt}$, Rt_{ϕ}), in combination with NOCROWD (cf. winning candidate **a** in Table 38 and the left contour in Figure 135 for an example). In individually-phrased words, any of the variant rankings ((169), (170), (173)) will result in at least the penult (plus possibly preceding syllables) being high for the same reason, as attested (cf. section 6.1.5). Bisyllabic individually-phrased words are a special case, because they don't ever seem to realize a tritonal contour. Solving this is not trivial. In principle, I would like to propose that an OCP-constraint is active and ranked above the alignment constraints, but below NOCROWD. This would ensure that the rising-falling contour on such phrases is always realized bitonally with H and $L_{\phi Rt}$, because even though the alignment of $L_{\phi Lt}$ is ranked higher than that of H, the high-ranking OCP-constraint will prevent candidates in which H is not realized, and the ranking of ALIGN ($L_{\phi Rt}$, Rt_{ϕ}) over ALIGN ($L_{\phi Lt}$, Lt_{ϕ}) ones in which $L_{\phi Rt}$ is not realized. However, MAXIO(T), which has to be ranked above NOCROWD as pointed out above because of observed bitonal contours on monosyllabic phrases, would prevent any effects of the OCP constraint. Recall from section 5.3.3 that Cho & Flemming (2015) found continuously variable tonal reduction under continuously variable time constraints (increased speech rate) in Korean, but categorical tonal deletion of the second L in APs with less than four syllables (normally with an LHLH contour). This seems comparable to our case of bisyllabic phrases: one of the tones of the tritonal contour is categorically not realized. That the deleted tone is not the H is likely truly due to an OCP-effect as described, but we have to assume that the deletion occurs prior to the step when the alignment constraints take effect (Kiparsky 2015), since for them

MAXIO(T) must be high-ranking, so that its interaction with NOCROWD produces the other observed continuous crowding and compression phenomena. While something like this must be going on, truly solving this puzzle must be a task for the future. This is yet another case showcasing the complexities of phenomena surrounding tonal crowding, truncation and deletion, and their categorical vs. continuous expression.

Returning to the main discussion, the same ranking (169) also produces attested rise-falling contour variants when the initial word is prominent in the phrase, with the parallel displacement of the H tone by one TBU away from the left word boundary, now due to the high-ranking ALIGN ($L_{\phi Lt}$, Lt_{ϕ}) ensuring that the phrase-initial TBU is occupied by the L tone (cf. winning candidate **d)** in Table 39, exemplified on individually-phrased words in Figures 144 and 146). In the tableaux, brackets mark the extent of the prosodic words, curly brackets the phrase. Accents mark the strongest element within its domain (syllable within prosodic word and prosodic word within phonological phrase). Dashed lines signify alignment of a tone with a syllable, dashed arrows indicate multiple alignment. Black dots are tonal targets.

Table 38: OT-Tableau with the constraint rankings to arrive at the correct alignment behaviour for the word-boundary variant *with* multiple H tone alignment of a *rise-falling* Quechua phonological phrase (ϕ) containing two prosodic words (ω) consisting of four syllables (σ) each, with the *final word prominent* in the phrase.

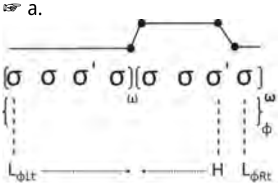
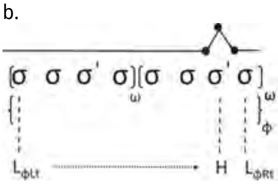
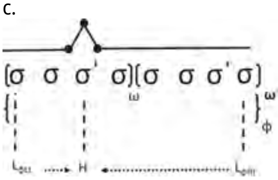
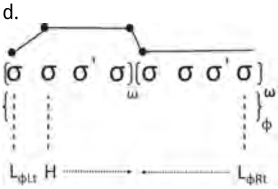
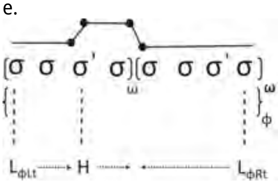
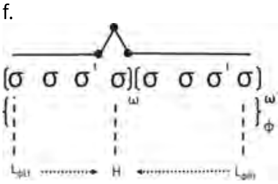
[(σσ'σ)$_ω$ (σσ'σ)$_{ω'}$]$_ϕ$ $L_{ϕLt}HL_{ϕRt}$	**ALIGN ($L_{ϕRt}$, $Rt_ϕ$)**	**ALIGN ($L_{ϕLt}$, $Lt_ϕ$)**	**ALIGN (H, $Lt_{ω'}$)**	**ALIGN (H, $Rt_{ω'}$)**	**ALIGN ($L_{ϕRt}$, $Lt_ϕ$)**	**ALIGN (H, $Rt_ϕ$)**	**ALIGN (H, $Lt_ω$)**	**ALIGN ($L_{ϕLt}$, $Rt_ϕ$)**	**ALIGN (H, $Lt_ϕ$)**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**
☞ a. 				*	**7***	*		****	****	**	
b. 			**!	*	**7***	*	**	**	**6***		

Table 38 (continued)

$[(\sigma\sigma'\sigma)_{\omega}\ (\sigma\sigma'\sigma)_{\omega'}]_{\phi}$ $L_{\phi Lt}HL_{\phi Rt}$	ALIGN ($L_{\phi Rt}$, Rt_{ϕ})	ALIGN ($L_{\phi Lt}$, Lt_{ϕ})	ALIGN (H, $Lt_{\omega'}$)	ALIGN (H, $Rt_{\omega'}$)	ALIGN ($L_{\phi Rt}$, Lt_{ϕ})	ALIGN (H, Rt_{ϕ})	ALIGN (H, Lt_{ω})	ALIGN ($L_{\phi Lt}$, Rt_{ϕ})	ALIGN (H, Lt_{ϕ})	ALIGN (H, σ', Lt)	ALIGN (H, σ', Rt)
c. 			**!	**5***	***	**5***	**	**6***	**		
d. 			***!	****	****	****	*	**7***	*	*	*
e. 			**!	****	****	****	**	**6***	**		*
f. 			*!	****	****	****	*	**5***	***	*	*

Even for the attested phrase pattern where both the rise and fall only take place at the end of the phrase (e.g. Figures 132, 175), instead of the rise ocurring already at the beginning of the final word, no reference needs to be made to any syllabic position within a word, as winning candidate **b)** in Table 40 shows. The only difference between the rankings (169) and (170) is that in (170), the constraints ALIGN (H, Lt_{ω}) and ALIGN (H, Lt_{ω}), high- and mid-ranking, respectively, in (169), have been downranked, effectively preventing the H tone from also aligning leftwards and creating a plateau. Note that the constraint ALIGN (H, $Rt_{\omega'}$), seeking to align the H tone with the right edge of the prominent word, is still active in (170).

Table 39: OT-Tableau with the constraint rankings to arrive at the correct alignment behaviour for the word-boundary variant *with* multiple H tone alignment of a *rise-falling* Quechua phonological phrase (ϕ) containing two prosodic words (ω) consisting of four syllables (σ) each, with the *prefinal word prominent* in the phrase.

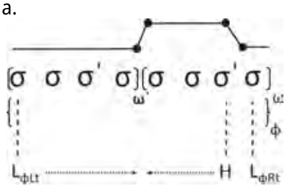
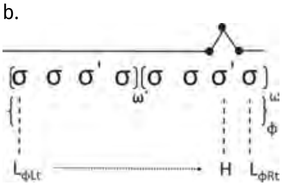
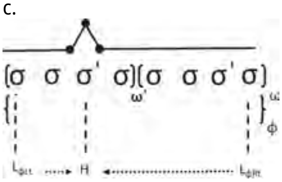
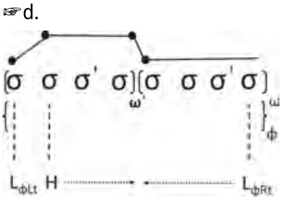
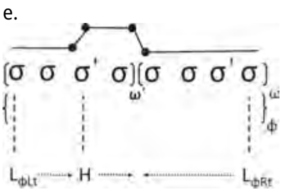
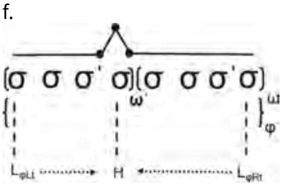
$[(\sigma\sigma'\sigma)_{\omega}{}' (\sigma\sigma'\sigma)_{\omega}]_{\phi}$ **$L_{\phi Lt}$ H $L_{\phi Rt}$**	**ALIGN ($L_{\phi Rt}$, Rt_{ϕ})**	**ALIGN ($L_{\phi Lt}$, Lt_{ϕ})**	**ALIGN (H, $Lt_{\omega'}$)**	**ALIGN (H, $Rt_{\omega'}$)**	**ALIGN ($L_{\phi Rt}$, Lt_{ϕ})**	**ALIGN (H, Rt_{ϕ})**	**ALIGN (H, Lt_{ω})**	**ALIGN ($L_{\phi Lt}$, Rt_{ϕ})**	**ALIGN (H, Ltϕ)**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**
a. 			****!	***	**7***	*		****	****	**	
b. 			**6*!**	***	**7***	*	**	**	**6***		
c. 			**!	*	***	**5***	**	**6***	**		
☞d. 			*		****	****	*	**7***	*	*	*
e. 			**!		****	****	**	**6***	**		*

Table 39 (continued)

$[(\sigma\sigma'\sigma)_{\omega'}\ (\sigma\sigma'\sigma)_{\omega}]_{\phi}$ $L_{\phi Lt}$ H $L_{\phi Rt}$	ALIGN ($L_{\phi Rt}$, Rt_{ϕ})	ALIGN ($L_{\phi Lt}$, Lt_{ϕ})	ALIGN (H, $Lt_{\omega'}$)	ALIGN (H, $Rt_{\omega'}$)	ALIGN ($L_{\phi Rt}$, Lt_{ϕ})	ALIGN (H, Rt_{ϕ})	ALIGN (H, Lt_{ω})	ALIGN ($L_{\phi Lt}$, Rt_{ϕ})	ALIGN (H, $Lt\phi$)	ALIGN (H, σ', Lt)	ALIGN (H, σ', Rt)
f. 			***!		****	****	*	5*	***	*	*

ALIGN (H, $Rt_{\omega'}$) here prevents ALIGN ($L_{\phi Rt}$, Lt_{ϕ}) from pushing further to the left, thus enabling ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) to push rightwards and create the observed low stretch, even though it is lower-ranked than ALIGN ($L_{\phi Rt}$, Lt_{ϕ}). This ranking without reference to a stressed penult must exist, because the "displaced" pitch peaks (cf. section 6.1.3) cannot be generated from a ranking in which tones align with the word penult. Thus this must be one of the rankings that create a phrase-final rise-falling contour in Quechua. A similar contour is generated with pitch accents and boundary tones in Huari Spanish (section 5.3.2), and it is also generated in the word penult-pattern (cf. section 6.3.1.2).

(170) Alignment constraint ranking for the word-boundary variant (without multiple alignment of the H tone) of Quechua, rise-falls (changes from (169) in bold)
LINEARITY, MAXIO(T), NOCROWD >> ALIGN ($L_{\phi Rt}$, Rt_{ϕ}) >> ALIGN ($L_{\phi Lt}$, Lt_{ϕ}) >> ALIGN (H, $Rt_{\omega'}$) >> ALIGN ($L_{\phi Rt}$, Lt_{ϕ}) >> ALIGN (H, Rt_{ϕ}) >> ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) >> ALIGN (H, Lt_{ϕ}) >> ALIGN (H, σ', Lt) >> ALIGN (H, σ', Rt) >> **ALIGN (H, $Lt_{\omega'}$)** >> **ALIGN (H, Lt_{ω})**

That ALIGN (H, $Rt_{\omega'}$) is still ranked so high has consequences that only emerge when a prefinal word instead of the final one is prominent in the phrase. Table 32 in section 6.2.3.2 counts 8 tokens where both the rise and the fall take place on either the penult or the final syllable of a prefinal word. Looking at these examples individually, the majority realize the high tone on either of these syllabic positions, but not both. That is to say, in one variant the rise takes place at the left edge of the penult, and the fall at its right edge, while in the other, the rise takes place at the left edge of the final syllable and the fall at its right edge, which is also the word

Table 40: OT-Tableau with the constraint rankings to arrive at the correct alignment behaviour for the word-boundary variant *without* multiple H tone alignment of a *rise-falling* Quechua phonological phrase (ϕ) containing two prosodic words (ω) consisting of four syllables (σ) each, with the *final word prominent* in the phrase.

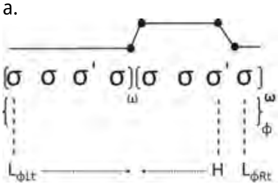
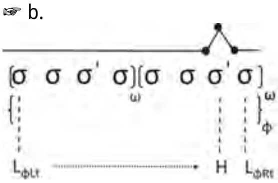
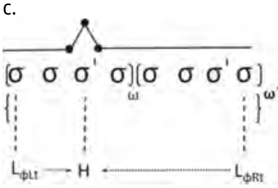
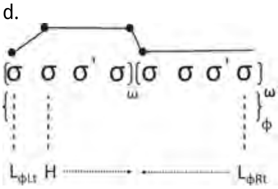
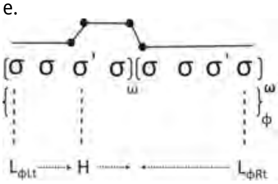
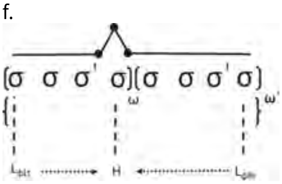
$[(\sigma\sigma'\sigma)_{\omega}\ (\sigma\sigma'\sigma)_{\omega'}]_{\phi}$ $L_{\phi Lt}HL_{\phi Rt}$	ALIGN ($L_{\phi Rt}$, Rt_{ϕ})	ALIGN ($L_{\phi Lt}$, Lt_{ϕ})	ALIGN (H, $Rt_{\omega'}$)	ALIGN ($L_{\phi Rt}$, Lt_{ϕ})	ALIGN (H, Rt_{ϕ})	ALIGN ($L_{\phi Lt}$, Rt_{ϕ})	ALIGN (H, Lt_{ϕ})	ALIGN (H, σ', Lt)	ALIGN (H, σ', Rt)	ALIGN (H, $Lt_{\omega'}$)	ALIGN (H, Lt_{ω})
a. 			*	7*	*	****!	****	**			
☞ b. 			*	7*	*	**	6*			**	**
c. 			5*!	***	5*	6*	**			**	**
d. 			****!	****	****	7*	*	*	*	*	*
e. 			****!	****	****	6*	**		*	**	**

Table 40 (continued)

$[(\sigma\sigma'\sigma)_\omega\ (\sigma\sigma'\sigma)_\omega{}']_\phi$ $L_{\phi Lt}HL_{\phi Rt}$	ALIGN ($L_{\phi Rt}$, Rt_ϕ)	ALIGN ($L_{\phi Lt}$, Lt_ϕ)	ALIGN (H, $Rt_{\omega'}$)	ALIGN ($L_{\phi Rt}$, Lt_ϕ)	ALIGN (H, Rt_ϕ)	ALIGN ($L_{\phi Lt}$, Rt_ϕ)	ALIGN (H, Lt_ϕ)	ALIGN (H, σ', Lt)	ALIGN (H, σ', Rt)	ALIGN (H, $Lt_{\omega'}$)	ALIGN (H, Lt_ω)
f. 			****!	****	****	**5***	***	*	*	*	*

boundary. As an example of the latter variant, see TP03_Conc_Q_0917 (Figure 172), as an example of the former, HA30_Conc_Q_2398 (Figure 173).

(171) TP03_Conc_Q_0917
hatun runa
big person
"big person"

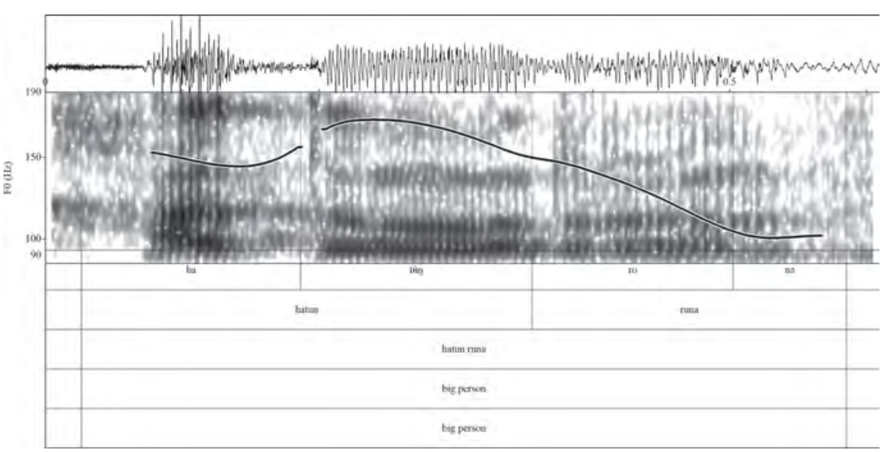


Figure 172: TP03_Conc_Q_0917[393] (part of a declarative with a rise-falling contour).

393 https://osf.io/8nvrq/

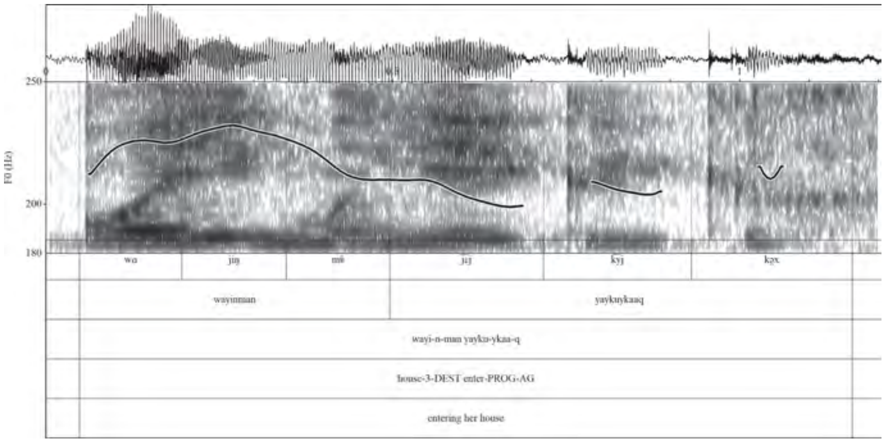


Figure 173: HA30_Conc_Q_2398[394] (part of a declarative with a rise-falling contour).

(172) HA30_Conc_Q_2398
wayi-n-man yayku-ykaa-q
house-3-DEST enter-PROG-AG
"entering her house"

The variant produced by TP03 is what the ranking (170) predicts when applied to a phrase in which a prefinal word is prominent: ALIGN (H, $Rt_{\omega'}$) makes sure that the H tone occupies the final syllable position in the prominent prefinal word, but ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) is unimpeded in its effort to push the phrase-initial L tone as far right as possible until it hits upon the H tone occupying the final syllable and causing the H tone to be restricted to only that TBU (cf. winning candidate **f)** in Table 41). In particular, the stressed penult does not attract the H tone and is instead realized as part of the low stretch produced by ALIGN ($L_{\phi Lt}$, Rt_{ϕ}). The fact that this variant is attested corroborates the validity and plausibility of ranking (170) for both finally-prominent and non-finally-prominent rise-falling phrases without multiple H alignment. The rising contour exhibits similar variation (cf. section 6.1.3). A similar analysis to represent it is made in section 6.3.2. For the variant as produced by HA30, where the H tone is restricted to the penult of the prefinal word, a ranking is needed in which alignment to the prominent syllabic position is high-ranked. This will come next. It seems plausible to take the competition between these rankings, leading to variable alignment with either the final syllable or the penult at several occasions

394 https://osf.io/uxqtc/

in the overall data, to be one of the reasons for the quantitative findings on variable peak alignment in section 6.1.6.

Table 41: OT-Tableau with the constraint rankings to arrive at the correct alignment behaviour for the word-boundary variant *without* multiple H tone alignment of a *rise-falling* Quechua phonological phrase (ϕ) containing two prosodic words (ω) consisting of four syllables (σ) each, with the *prefinal word prominent* in the phrase.

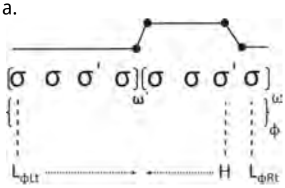
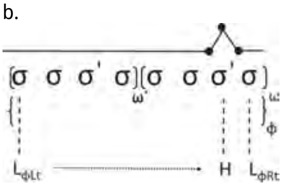
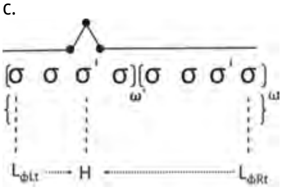
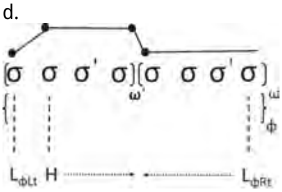
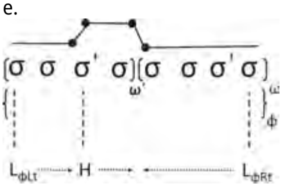
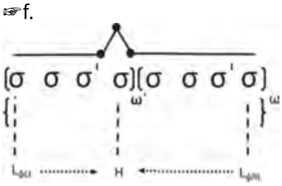
$[(\sigma\sigma'\sigma)_{\omega'}\ (\sigma\sigma'\sigma)_{\omega}]_{\phi}$ $L_{\phi Lt}$ H $L_{\phi Rt}$	**ALIGN ($L_{\phi Rt}$, Rt_{ϕ})**	**ALIGN ($L_{\phi Lt}$, Lt_{ϕ})**	**ALIGN (H, $Rt_{\omega'}$)**	**ALIGN ($L_{\phi Rt}$, Lt_{ϕ})**	**ALIGN (H, Rt_{ϕ})**	**ALIGN ($L_{\phi Lt}$, Rt_{ϕ})**	**ALIGN (H, Lt_{ϕ})**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**	**ALIGN (H, $Lt_{\omega'}$)**	**ALIGN (H, Lt_{ω})**
a. 			*****!**	**7***	*	****	****	**		****	
b. 			*****!**	**7***	*	**	**6***			**6***	**
c. 			***!**	***	**5***	**6***	**			**	**
d. 				****	****	**7*****!**	*	*	*	*	*

Table 41 (continued)

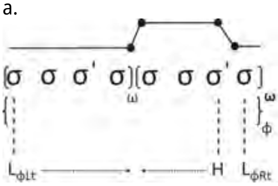
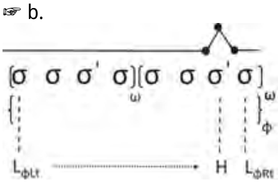
$[(\sigma\sigma'\sigma)_{\omega'}\,(\sigma\sigma'\sigma)_{\omega}]_{\phi}$ $L_{\phi Lt}$ H $L_{\phi Rt}$	ALIGN ($L_{\phi Rt}$, Rt_{ϕ})	ALIGN ($L_{\phi Lt}$, Lt_{ϕ})	ALIGN (H, $Rt_{\omega'}$)	ALIGN ($L_{\phi Rt}$, Lt_{ϕ})	ALIGN (H, Rt_{ϕ})	ALIGN ($L_{\phi Lt}$, Rt_{ϕ})	ALIGN (H, Lt_{ϕ})	ALIGN (H, σ', Lt)	ALIGN (H, σ', Rt)	ALIGN (H, $Lt_{\omega'}$)	ALIGN (H, Lt_{ω})
e. 				****	****	**6*!**	**		*	**	**
☞f. 				****	****	**5***	***	*	*	***	*

6.3.1.2 Word penult-variant

The previous rankings, in which alignment is only oriented towards prosodic edges, cannot generate examples like Figure 173. For them, a ranking is needed in which the two constraints militating for alignment between the right and left edge of the H tone with the right and left edge of the prominent syllable, Align (H, σ', Rt) and Align (H, σ', Lt), are promoted to a higher position. This is done in ranking (173). Winning candidate **b)** in Table 42 on phrases with final prominence is again the phrase-final rise-falling contour, also generated in the "phrase accentuation" variant of Huari Spanish (cf. sections 5.3.2, 5.3.3), and from the word-boundary ranking without multiple H tone alignment (170) on finally prominent phrases, candidate **b)** in Table 40. Note again that the H tone here might well associate, as it does in the Spanish contour, but this is not relevant for generating the contours (see above). Align (H, $Rt_{\omega'}$) here is kept at a relatively high position, only having moved Align (H, σ', Rt) and Align (H, σ', Lt), above it. This keeps the difference between (170) and (173) minimal and also has the desired effect of keeping the high tone within the prominent word in winning candidates, even though ALIGN (H, σ', Rt) and ALIGN (H, σ', Lt) are not specified for this.[395]

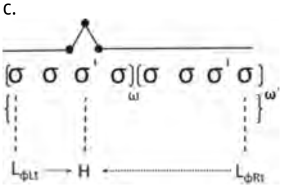
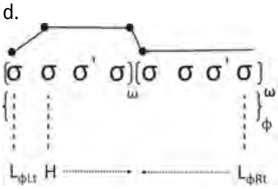
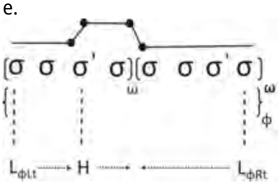
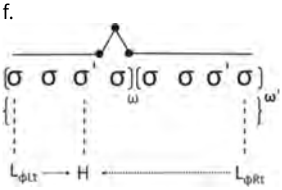
(173) Alignment constraint ranking for the word-penult variant of Quechua (rise-falls)
LINEARITY, MAXIO(T), NOCROWD >> ALIGN ($L_{\phi Rt}$, Rt_{ϕ}) >> ALIGN ($L_{\phi Lt}$, Lt_{ϕ}) >> ALIGN (H, σ', Lt) >> ALIGN (H, σ', Rt) >> ALIGN (H, $Rt_{\omega'}$) >> ALIGN ($L_{\phi Rt}$, Lt_{ϕ}) >> ALIGN (H, Rt_{ϕ}) >> ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) >> ALIGN (H, Ltϕ) >> ALIGN (H, $Lt_{\omega'}$) >> ALIGN (H, Lt_{ω})

Table 42: OT-Tableau with the constraint rankings to arrive at the correct alignment behaviour for the word-penult variant of a *rise-falling* Quechua phonological phrase (ϕ) containing two prosodic words (ω) consisting of four syllables (σ) each, with the *final word prominent* in the phrase.

[(σσ'σ)$_\omega$ (σσ'σ)$_{\omega}$']$_\phi$ $L_{\phi Lt}HL_{\phi Rt}$	**ALIGN ($L_{\phi Rt}$, Rt_{ϕ})**	**ALIGN ($L_{\phi Lt}$, Lt_{ϕ})**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**	**ALIGN (H, $Rt_{\omega'}$)**	**ALIGN ($L_{\phi Rt}$, Lt_{ϕ})**	**ALIGN (H, Rt_{ϕ})**	**ALIGN ($L_{\phi Lt}$, Rt_{ϕ})**	**ALIGN (H, Ltϕ)**	**ALIGN (H, $Lt_{\omega'}$)**	**ALIGN (H, Lt_{ω})**
a. 			****!**		*	**7***	*	****	****		
☞ b. 					*	**7***	*	**	**6***	**	**

395 With a change in relative ranking between ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) and ALIGN (H, Ltϕ), the very infrequently attested variant where the rise takes place on the penult of one word and the fall in that of a later word (combination g) in (153) and Table 32 in section 6.2.3.2, would be the winning candidate, indicating perhaps that their relative ranking distributions overlap just enough to occasionally produce the rare variant (cf. Boersma & Hayes 2001). Another option for keeping the H tone within the prominent word might be to say that it associates secondarily with that word, as Roettger (2017) does for Tashlhiyt Berber, but this is disallowed in Gussenhoven (2004).

Table 42 (continued)

$[(\sigma\sigma'\sigma)_\omega\ (\sigma\sigma'\sigma)_{\omega'}]_\phi$ $L_{\phi Lt}HL_{\phi Rt}$	**ALIGN ($L_{\phi Rt}$, Rt_ϕ)**	**ALIGN ($L_{\phi Lt}$, Lt_ϕ)**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**	**ALIGN (H, $Rt_{\omega'}$)**	**ALIGN ($L_{\phi Rt}$, Lt_ϕ)**	**ALIGN (H, Rt_ϕ)**	**ALIGN ($L_{\phi Lt}$, Rt_ϕ)**	**ALIGN (H, Ltφ)**	**ALIGN (H, $Lt_{\omega'}$)**	**ALIGN (H, Lt_ω)**
c. 					**5***!	***	**5***	**6***	**	**	**
d. 			*!	*	****	****	****	**7***	*	***	*
e. 				*!	****	****	****	**6***	**	**	**
f. 			*!	*	****	****	****	**5***	***	*	*

Applying the same constraint ranking to phrases in which the prefinal word is prominent now finally yields the winning candidate **c)** in Table 43, exemplified by HA30_Conc_Q_2398 (Figure 173), the variant realization to candidate **f)** in Table 41.

The idea here has been to keep the analysis of the word-penult pattern as close to the spirit of a tonal grammar paying more heed to phrases and edges than to words and prominent syllables. In this sense, this is an analysis of the word-penult pattern that is minimal in its concession to the penult as a prominent position in the word. That is in keeping with the characterization of Quechua prosody in section

Table 43: OT-Tableau with the constraint rankings to arrive at the correct alignment behaviour for the word-penult variant of a *rise-falling* Quechua phonological phrase (ϕ) containing two prosodic words (ω) consisting of four syllables (σ) each, with the pre*final word prominent* in the phrase.

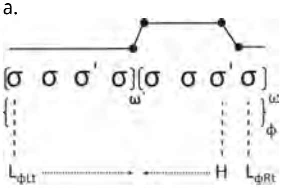
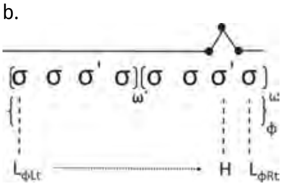
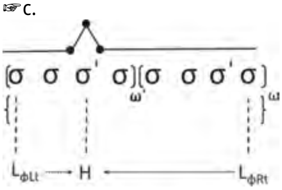
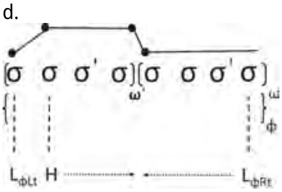
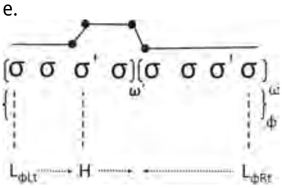
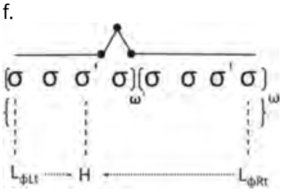
$[(\sigma\sigma'\sigma)_{\omega'}\ (\sigma\sigma'\sigma)_{\omega}]_{\phi}$ $L_{\phi Lt}$ H $L_{\phi Rt}$	**ALIGN ($L_{\phi Rt}$, Rt_{ϕ})**	**ALIGN ($L_{\phi Lt}$, Lt_{ϕ})**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**	**ALIGN (H, $Rt_{\omega'}$)**	**ALIGN ($L_{\phi Rt}$, Lt_{ϕ})**	**ALIGN (H, Rt_{ϕ})**	**ALIGN ($L_{\phi Lt}$, Rt_{ϕ})**	**ALIGN (H, Ltϕ)**	**ALIGN (H, $Lt_{\omega'}$)**	**ALIGN (H, Lt_{ω})**
a. 			****!**		***	**7***	*	****	****	****	
b. 					*****!**	**7***	*	**	**6***	**6***	**
☞c. 					*	***	**5***	**6***	**	**	**
d. 			***!**	*		****	****	**7***	*	*	*
e. 				***!**		****	****	**6***	**	**	**

Table 43 (continued)

[(σσ'σ)$_{ω}$' (σσ'σ)$_{ω}$]$_{ϕ}$ **$L_{ϕLt}$ H $L_{ϕRt}$**	**ALIGN ($L_{ϕRt}$, $Rt_{ϕ}$)**	**ALIGN ($L_{ϕLt}$, $Lt_{ϕ}$)**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**	**ALIGN (H, $Rt_{ω'}$)**	**ALIGN ($L_{ϕRt}$, $Lt_{ϕ}$)**	**ALIGN (H, $Rt_{ϕ}$)**	**ALIGN ($L_{ϕLt}$, $Rt_{ϕ}$)**	**ALIGN (H, Ltϕ)**	**ALIGN (H, $Lt_{ω'}$)**	**ALIGN (H, $Lt_{ω}$)**
f. 			*!	*		****	****	5*	***	***	*

6.1, as making some reference to a prominent syllabic position, but minimizing the effects of this. The analysis is kept here so much on this “accentless” side of things not because I think that more accent-oriented grammars are not available to the speakers, but because I want to map out the space of grammatical possibility that the Huari speakers can take recourse to by tracing its peripheral boundaries. It is clear that the speakers have access to the grammars represented by the rankings (169) and (170), which do not need to make reference to a prominent syllabic position at all. From the Spanish OT-analysis it is clear that they also have access to the other end of the spectrum, the “main variant” of Spanish, where even stressed syllables in nonprominent words in the phrase form pitch accents. In between, they have a variety of only subtly different options both at the level of the contours produced and of the constraint rankings utilized to generate them at their disposal. This middle ground is accessible from either direction and could therefore be seen less as belonging to either language but to the structural communicative resources available to the speakers of this community. In the following, the constraint rankings will be adapted to rising contours.

6.3.2 Rising contours

For the rising contours, the rankings for the rise-falling contours can be adapted. The global difference is that all constraints making reference to the right L tone here have no target and are therefore omitted. In addition, the constraint aligning the H tone with the right phrase boundary ALIGN (H, $Rt_{ϕ}$) is promoted to a position where it is undominated by all other alignment constraints militating against it.

Thus we have a phrase-initial L tone and a phrase-final H tone and the remaining constraints generate the attested contours. This ensures, together with the faithfulness constraints, that bisyllabic phrases will always realize the initial syllable low and the final high, as attested (see Figure 141 and the discussion in section 6.1.5), even under the word-penult variant ranking (176). Since the promotion of ALIGN (H, Rt_{ϕ}) means that ALIGN (H, $Rt_{\omega'}$) will never be able to differentiate between any candidates that comply with ALIGN (H, Rt_{ϕ}), i.e. all attested candidates, it has also been omitted here.

6.3.2.1 Word boundary-variant

With these changes, the ranking for the word-boundary variant of the rising contour with multiple H tone alignment, parallel to that of the rise-falls (169), is the following:

(174) Alignment constraint ranking for the word-boundary variant (with multiple H tone alignment) of Quechua (rises)
LINEARITY, MAXIO(T), NOCROWD >> ALIGN (H, Rt_{ϕ}) >> ALIGN ($L_{\phi Lt}$, Lt_{ϕ}) >> ALIGN (H, $Lt_{\omega'}$) >> ALIGN (H, Lt_{ω}) >> ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) >> ALIGN (H, Lt_{ϕ}) >> ALIGN (H, σ', Lt) >> ALIGN (H, σ', Rt)

This ranking correctly selects the winning candidate **b)**, in the case where the final word is prominent in the phrase, and **e)** when the prefinal word is prominent (see Tables 44 and 45, respectively).

In sections 6.1.3 and 6.1.5, it was already discussed that there are two variants of the rising contour when the rise takes place at the end of the phrase-final word: in one, the high tonal target is clearly realized already on the penult, in the other, the penult is realized as low and only the final syllable is high (e.g. (101)/Figure 118 vs. (118)/Figure 138, and examples for both in Figure 142, cf. also the schematic difference in Figure 137). This latter variant is generated by ranking (175), parallel to ranking (170) for the rise-falling contours. In it, ALIGN (H, $Lt_{\omega'}$) has been ranked down, allowing ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) to have a much stronger effect. This yields the desired result, generating winning candidate **c)** in Table 46 under the condition of the final word in the phrase being prominent, but it poses something of a conundrum when applied to a phrase in which the prefinal word is prominent (see below). It is also something of a deviation since it ranks ALIGN (H, $Lt_{\omega'}$) down, which comes back up again for the word-penult pattern.

Table 44: OT-Tableau with the constraint rankings to arrive at the correct alignment behaviour for the word-boundary variant *with* multiple H tone alignment of a *rising* Quechua phonological phrase (ϕ) containing two prosodic words (ω) consisting of four syllables (σ) each, with the *final word prominent* in the phrase.

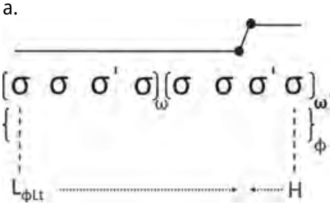
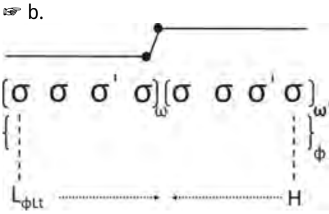
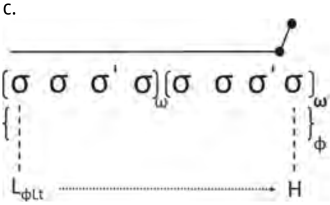
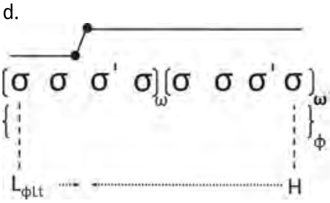
[(σσ'σ)$_{ω}$ (σσ'σ)$_{ω'}$]$_{ϕ}$ **L$_{ϕLt}$HL$_{ϕRt}$**	**ALIGN (H, Rt$_{ϕ}$)**	**ALIGN (L$_{ϕLt}$, Lt$_{ϕ}$)**	**ALIGN (H, Lt$_{ω'}$)**	**ALIGN (H, Lt$_{ω}$)**	**ALIGN (L$_{ϕLt}$, Rt$_{ϕ}$)**	**ALIGN (H, Lt$_{ϕ}$)**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**
a. 			****!**	**	**	**6***		*
☞ b. 					****	****	**	*
c. 			*****!**	***	*	**7***	*	*
d. 			****!**	**	**6***	**		*

Table 44 (continued)

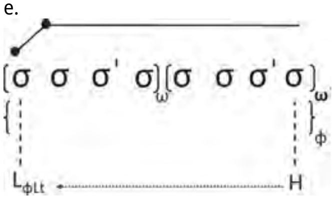
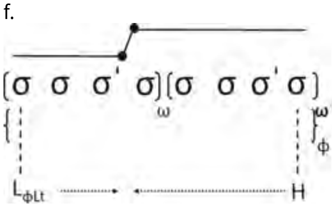
$[(\sigma\sigma'\sigma)_{\omega}\ (\sigma\sigma'\sigma)_{\omega}']_{\phi}$ $L_{\phi Lt}HL_{\phi Rt}$	ALIGN (H, Rt_{ϕ})	ALIGN ($L_{\phi Lt}$, Lt_{ϕ})	ALIGN (H, $Lt_{\omega'}$)	ALIGN (H, Lt_{ω})	ALIGN ($L_{\phi Lt}$, Rt_{ϕ})	ALIGN (H, Lt_{ϕ})	ALIGN (H, σ', Lt)	ALIGN (H, σ', Rt)
e. 			***!	*	**7***	*	*	*
f. 			*!	*	**5***	***	*	*

Table 45: OT-Tableau with the constraint rankings to arrive at the correct alignment behaviour for the word-boundary variant *with* multiple H tone alignment of a *rising* Quechua phonological phrase (ϕ) containing two prosodic words (ω) consisting of four syllables (σ) each, with the *prefinal word prominent* in the phrase.

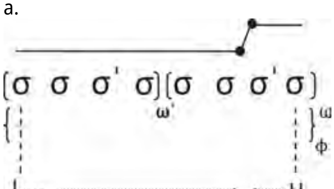
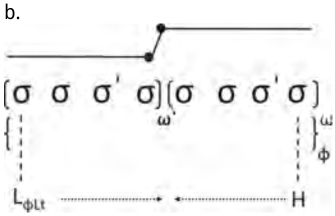
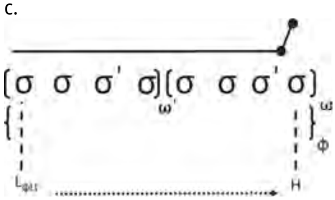
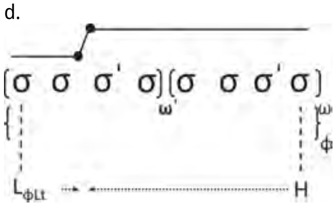
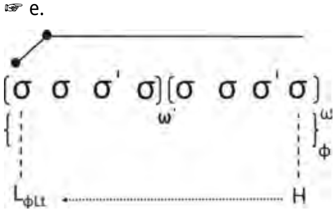
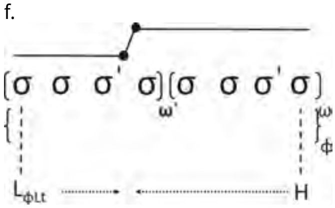
$[(\sigma\sigma'\sigma)_{\omega}'\ (\sigma\sigma'\sigma)_{\omega}]_{\phi}$ $L_{\phi Lt}HL_{\phi Rt}$	ALIGN (H, Rt_{ϕ})	ALIGN ($L_{\phi Lt}$, Lt_{ϕ})	ALIGN (H, $Lt_{\omega'}$)	ALIGN (H, Lt_{ω})	ALIGN ($L_{\phi Lt}$, Rt_{ϕ})	ALIGN (H, Lt_{ϕ})	ALIGN (H, σ', Lt)	ALIGN (H, σ', Rt)
a. 			**6*!**	**	**	**6***		*

Table 45 (continued)

$[(\sigma\sigma'\sigma)_{\omega}{}' \, (\sigma\sigma'\sigma)_{\omega}]_{\phi}$ **$L_{\phi Lt}HL_{\phi Rt}$**	**ALIGN (H, Rt_{ϕ})**	**ALIGN ($L_{\phi Lt}$, Lt_{ϕ})**	**ALIGN (H, $Lt_{\omega'}$)**	**ALIGN (H, Lt_{ω})**	**ALIGN ($L_{\phi Lt}$, Rt_{ϕ})**	**ALIGN (H, Lt_{ϕ})**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**
b. 			****!		****	****	**	*
c. 			**7*!**	***	*	**7***	*	*
d. 			**!	**	**6***	**		*
☞ e. 			*	*	**7***	*	*	*
f. 			***!	*	**5***	***	*	*

(175) Alignment constraint ranking for the word-boundary variant (without multiple H tone alignment) of Quechua (rises)
LINEARITY, MAXIO(T), NOCROWD >> ALIGN (H, Rt_{ϕ}) >> ALIGN ($L_{\phi Lt}$, Lt_{ϕ}) >> ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) >> ALIGN (H, Ltϕ) >> ALIGN (H, σ', Lt) >> ALIGN (H, σ', Rt) >> ALIGN (H, $Lt_{\omega'}$) >> ALIGN (H, Lt_{ω})

Table 46: OT-Tableau with the constraint rankings to arrive at the correct alignment behaviour for the word-boundary variant *without* Multiple H tone alignment of a *rising* Quechua phonological phrase (ϕ) containing two prosodic words (ω) consisting of four syllables (σ) each, with the *final word prominent* in the phrase.

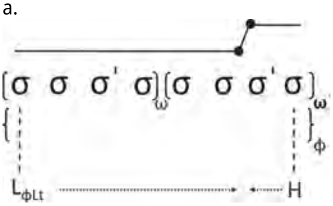
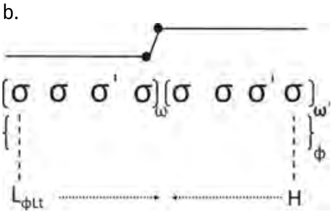
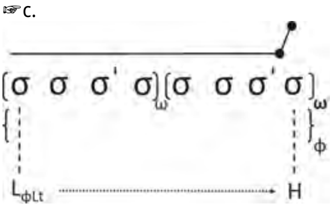
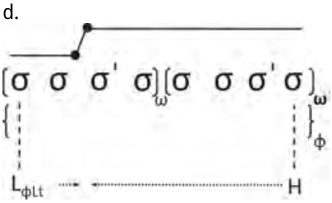
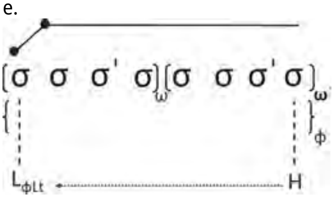
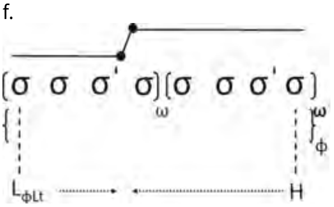
$[(\sigma\sigma'\sigma)_{\omega}\ (\sigma\sigma'\sigma)_{\omega'}]_{\phi}$ **$L_{\phi Lt}HL_{\phi Rt}$**	**ALIGN (H, Rt_{ϕ})**	**ALIGN ($L_{\phi Lt}$, Lt_{ϕ})**	**ALIGN ($L_{\phi Lt}$, Rt_{ϕ})**	**ALIGN (H, Lt_{ϕ})**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**	**ALIGN (H, $Lt_{\omega'}$)**	**ALIGN (H, Lt_{ω})**
a. 			**!	**6***		*	**	**
b. 			****!	****	**	*		
☞c. 			*	**7***	*	*	***	***

Table 46 (continued)

$[(\sigma\sigma'\sigma)_\omega\ (\sigma\sigma'\sigma)_{\omega}']_\phi$ $L_{\phi Lt}HL_{\phi Rt}$	ALIGN (H, Rt_ϕ)	ALIGN ($L_{\phi Lt}$, Lt_ϕ)	ALIGN ($L_{\phi Lt}$, Rt_ϕ)	ALIGN (H, Lt_ϕ)	ALIGN (H, σ', Lt)	ALIGN (H, σ', Rt)	ALIGN (H, $Lt_{\omega'}$)	ALIGN (H, Lt_ω)
d. 			**6*!**	**		*	**	**
e. 			**7*!**	*	*	*	***	*
f. 			**5*!**	***	*	*	*	*

In the case of a prefinal prominent word, the ranking does not yield a separate contour variant. Instead it again selects candidate **c)**, meaning that this ranking does not differentiate between final and prefinal prominence via an intonational contour[396] (the tableau has been omitted here to save space). In comparison, in the

396 No other ranking of the existing constraints could be found that would have been able to distinguish between the prominence configurations. That is because the ranking promotes ALIGN ($L_{\phi Lt}$, Rt_ϕ) to a nearly undominated position in order to produce the attested winning candidate **c)**. A theoretical possibility would have been to create a new constraint, like e.g. ALIGN ($L_{\phi Lt}$, $Rt_{\omega'}$), and ranking it above ALIGN ($L_{\phi Lt}$, Rt_ϕ). That would have produced candidate **b)** in the case of prefinal prominence and still kept **c)** as the winning candidate in the case of final prominence. However, it would have meant both introducing a new constraint purely for this case and making **b)** in turn ambiguous between a reading in which the final word is prominent and one in which the prefinal rise-falling contours, the ranking always differentiates according to prominence via separate contours. However, the distribution of rising contour variants as given in Table 30 in section 6.2.3.2 supports the analysis here, because the sixth contour that is “missing” from the comparison with the other variants, where the rise takes place not at the penult but the final syllable of a prefinal word (candidate **f**), only occurs when that word is an oxytonic Spanish loan, and the position for the rise is thus lexically specified (it is also very rare). So the gap in the paradigm of contours generated from the rankings corresponds to that found in the data. Whether this ranking occasionally really produces contours where the position of the H tone is misaligned with the prominent word, or whether some process acts to counteract this, remains unknown.

6.3.2.2 Word penult-variant

For the ranking generating contours in the word penult-variant, ALIGN (H, $Lt_{\omega'}$) is promoted back to position 5 among the aligment constraints, while ALIGN (H, σ', Lt) and ALIGN (H, σ', Rt) are ranked above it. This parallels the change from the word boundary- to the word penult-variant in the rise-falling contours, but note that here it is ALIGN (H, $Lt_{\omega'}$), and not ALIGN (H, $Rt_{\omega'}$), which is needed to ensure that the rise takes place in the prominent word. The ranking (176) produces the winning candidate **a)** in Table 47 under the condition that the final word is prominent, the direct variant to **c)**, which is derived from ranking (175). Under the condition of the prefinal word being prominent, the ranking also generates the attested winning candidate **d)** in Table 48.

(176) Alignment constraint ranking for the word-penult variant of Quechua (rises)
LINEARITY, MAXIO(T), NOCROWD >> ALIGN (H, Rt_{ϕ}) >> ALIGN ($L_{\phi Lt}$, Lt_{ϕ}) >> ALIGN (H, σ', Lt) >> ALIGN (H, σ', Rt) >> ALIGN (H, $Lt_{\omega'}$) >> ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) >> ALIGN (H, Ltϕ) >> ALIGN (H, Lt_{ω})

6.3.3 Only-falling contours

The analysis for the only-falling contours is complementary to that of the rises in terms of which constraints are used, but otherwise runs entirely in parallel. Here,

word is prominent. Since it would not have generated an attested candidate not generated otherwise, and since candidate **f)** is only attested in the “inherited” pattern and therefore cannot be the target contour here, the addition of such a constraint would be spurious and also not reduce the number of ambiguous contours.

Table 47: OT-Tableau with the constraint rankings to arrive at the correct alignment behaviour for the word-penult variant of a *rising* Quechua phonological phrase (ϕ) containing two prosodic words (ω) consisting of four syllables (σ) each, with the *final word prominent* in the phrase.

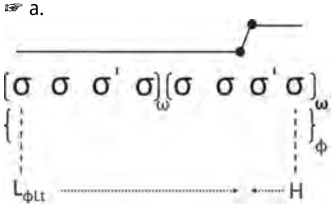
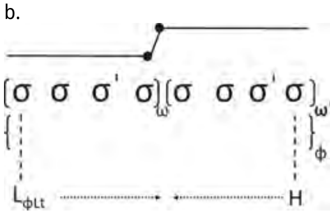
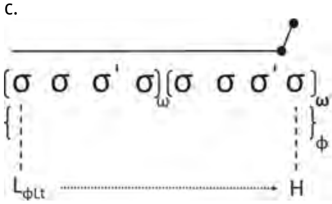
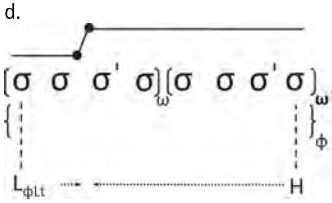
$[(\sigma\sigma'\sigma)_\omega\ (\sigma\sigma'\sigma)_{\omega'}]_\phi$ **$L_{\phi Lt}HL_{\phi Rt}$**	**ALIGN (H, Rt_ϕ)**	**ALIGN ($L_{\phi Lt}$, Lt_ϕ)**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**	**ALIGN (H, $Lt_{\omega'}$)**	**ALIGN ($L_{\phi Lt}$, Rt_ϕ)**	**ALIGN (H, Lt_ϕ)**	**ALIGN (H, Lt_ω)**
☞ a. 				*	**	**	**6***	**
b. 			****!**	*		****	****	
c. 			***!**	*	***	*	**7***	***
d. 				*	**	**6*!**	**	**

Table 47 (continued)

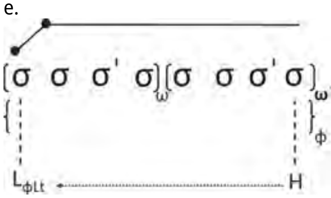
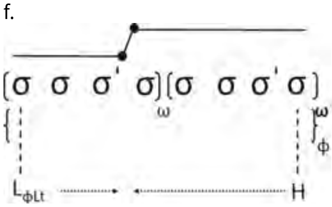
[(σσ'σ)$_ω$ (σσ'σ)$_ω$']$_ϕ$ **L$_{ϕLt}$HL$_{ϕRt}$**	**ALIGN (H, Rt$_ϕ$)**	**ALIGN (L$_{ϕLt}$, Lt$_ϕ$)**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**	**ALIGN (H, Lt$_{ω'}$)**	**ALIGN (L$_{ϕLt}$, Rt$_ϕ$)**	**ALIGN (H, Lt$_ϕ$)**	**ALIGN (H, Lt$_ω$)**
e. 			*!	*	***	**7***	*	*
f. 			*!	*	*	**5***	***	*

Table 48: OT-Tableau with the constraint rankings to arrive at the correct alignment behaviour for the word-penult variant of a *rising* Quechua phonological phrase (ϕ) containing two prosodic words (ω) consisting of four syllables (σ) each, with the *prefinal word prominent* in the phrase.

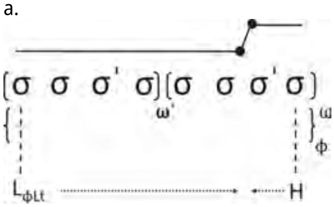
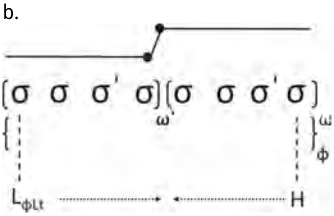
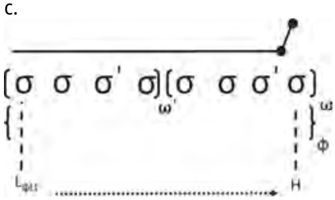
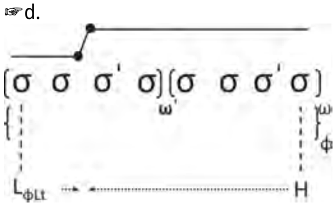
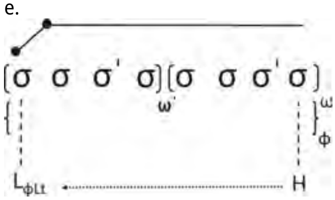
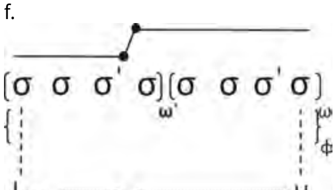
[(σσ'σ)$_ω$' (σσ'σ)$_ω$]$_ϕ$ **L$_{ϕLt}$HL$_{ϕRt}$**	**ALIGN (H, Rt$_ϕ$)**	**ALIGN (L$_{ϕLt}$, Lt$_ϕ$)**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**	**ALIGN (H, Lt$_{ω'}$)**	**ALIGN (L$_{ϕLt}$, Rt$_ϕ$)**	**ALIGN (H, Lt$_ϕ$)**	**ALIGN (H, Lt$_ω$)**
a. 				*	**6*!**	**	**6***	**

Table 48 (continued)

[(σσ'σ)$_{\omega}$' (σσ'σ)$_{\omega}$]$_{\phi}$ **L$_{\phi Lt}$HL$_{\phi Rt}$**	**ALIGN (H, Rt$_{\phi}$)**	**ALIGN (L$_{\phi Lt}$, Lt$_{\phi}$)**	**ALIGN (H, σ', Lt)**	**ALIGN (H, σ', Rt)**	**ALIGN (H, Lt$_{\omega'}$)**	**ALIGN (L$_{\phi Lt}$, Rt$_{\phi}$)**	**ALIGN (H, Lt$_{\phi}$)**	**ALIGN (H, Lt$_{\omega}$)**
b. 			****!**	*	****	****	****	
c. 			***!**	*	**7***	*	**7***	***
☞d. 				*	**	**6***	**	**
e. 			***!**	*	*	**7***	*	*
f. 			***!**	*	***	**5***	***	*

the constraints making reference to the left L tone are omitted and ALIGN (H, Lt_{ϕ}) is ranked directly underneath ALIGN ($L_{\phi Rt}$, Rt_{ϕ}), so that all winning contours will start high at the left edge and be low at the right edge, with what happens in between determined by the other constraints. Again complementary to the rises, the interaction of these constraints with NOCROWD and the other high-ranking faithfulness constraints also ensures that on bisyllabic phrases, the left syllable is always high and the right always low, and here this does not run counter to the penult being occupied by the high tone in the word-penult variant. Here it is ALIGN (H, $Lt_{\omega'}$) which does not have any distinguishing effect between winning candidates, so it will be omitted. On the other hand, ALIGN (H, $Rt_{\omega'}$) plays the crucial role of making sure that the high tone is realized within the prominent word here, just as in the rise-falling contours. The only-falling contour is the least frequently occurring contour in our data, particularly on phrases consisting of more than a single word, as Table 31 in section 6.2.3.2 shows. This makes some of the expected alignment combinations in multi-word phrases exceedingly rare. In particular, there is not a single token of an only-falling contour on a multi-word phrase in which the fall takes place on the penult of the prefinal word. Additionally, no cases were found in which the H tone did not extend at least to the right boundary or the penult of the prominent word, so that the separate ranking for a word boundary-variant without multiple H alignment is not included here. No contours are attested for the only-falling variant that specifically necessitate the word penult-variant at all, since on phrase-final words, both the word boundary-variant and the word penult-variant result in superficially the same contour. The ranking for the word penult-variant is still given here. More data is needed to see whether these gaps persist or are accidental. Its low frequency could suggest that the only-falling contour is simply an apheretic variant of the rise-falling contour, but it does not exclusively occur when there are insufficient phrase-initial TBUs to realize the initial L tone on, as a number of examples attests.[397] Because the treatment of the only-falling contours is entirely in parallel to that of the rises, I will not give tableaus here and merely state the rankings for the three alignment variants.

(177) Alignment constraint ranking for the word-boundary variant of Quechua (only-falls)
LINEARITY, MAXIO(T), NOCROWD >> ALIGN ($L_{\phi Rt}$, Rt_{ϕ}) >> ALIGN (H, Lt_{ϕ}) >> ALIGN (H, $Rt_{\omega'}$) >> ALIGN ($L_{\phi Rt}$, Lt_{ϕ}) >> ALIGN (H, Rt_{ϕ}) >> ALIGN (H, Lt_{ω}) >> ALIGN (H, σ', Lt) >> ALIGN (H, σ', Rt)

397 Cf. the examples in section 6.1.5 showing that the two falling variants collapse on individually phrased bisyllabic words, with the initial L being omitted, but from trisyllabic words onwards, both variants exist side by side.

(178) Alignment constraint ranking for the word-penult variant of Quechua (only-falls)
LINEARITY, MAXIO(T), NOCROWD >> ALIGN ($L_{\phi Rt}$, Rt_{ϕ}) >> ALIGN (H, Lt_{ϕ}) >> ALIGN (H, σ', Lt) >> ALIGN (H, σ', Rt) >> ALIGN (H, $Rt_{\omega'}$) >> ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) >> ALIGN (H, Lt_{ϕ}) >> ALIGN (H, Lt_{ω})

6.3.4 The marked loanword patterns and moving towards Spanish

Turning to the patterns only found on Spanish-origin words, the "inherited" pattern will be generated by the rankings for the word penult-variant, but with a different specification regarding which syllable is stressed in the lexical entry for the word than the default penult specification for Quechua words. In this respect, the difference between the word penult- and the "inherited" variant has more to do with differences in lexical entries than tonal alignment. However, section 6.1.8.3 showed that speakers differ non-categorically in whether they align according to the Spanish stress position or the penult-default in realizations of the same lexical word. Since all speakers are bilingual and do not stress words "wrongly" in Spanish, the difference in tonal alignment position can hardly be due to a difference in stress position in the lexical entries for the same item between different speakers. It can however be captured by introducing two new constraints that seek to align the left and right edge of the H tone with the left and right edge of the lexically accented, as opposed to the regularly stressed, syllable, respectively. This captures the different natures of the truly lexically specified stress position in Spanish loanwords vs. that of the entirely regular penult word stress of native Quechua words.

(179) ALIGN (H, σ_{LEX}, Lt): Align the left edge of the H tone with the left edge of the lexically accented syllable

(180) ALIGN (H, σ_{LEX}, Rt): Align the right edge of the H tone with the right edge of the lexically accented syllable

If these are each ranked above their counterparts ALIGN (H, σ', Lt) and ALIGN (H, σ', Rt) in a word-penult ranking, the result will be the "inherited" pattern. If they are ranked below them, the normal word penult-pattern results. The respective ranking distributions of these constraint pairs must be at different distances for each speaker here, making it more or less likely for them to produce one of the variants preferentially. However, since section 6.1.8.3 also showed a clear difference between lexical items in the production of the marked loanword patterns, the true explanation for the variation here cannot be restricted to a difference only between speakers.

Regarding the “grafted” pattern, I said in section 6.1.8.2 that the additional pitch accent tones LH* are specified separately and have the same constraint rankings as in the Spanish “main” variant which outranks that of the Quechua phrasal tones. That means also assuming that the H tone here associates. That ranking is repeated here in (181), adapted so that its tones are specified as being from Spanish (H_S and L_S) and the alignment and association constraints refer to the lexically accented syllable.

(181) Constraint ranking for the “main” variant of Spanish adapted for the “grafted” accent of Quechua
ALIGN (H_S, σ_{LEX}) >> $(\sigma_{LEX})_{\omega'} \leftarrow T_S$ >> $H_S \rightarrow$ TBU >> $\sigma_{LEX} \leftarrow$ T >> ALIGN (L_S, Rt_ϕ) >> ALIGN (L_S, Lt_ϕ) >> ALIGN (H_S, Lt_ϕ) >> NOASSOC >> ALIGN (H_S, Rt_ϕ)

These constraints can now be inserted into the ranking for the Quechua word-penult variant, above all except the faithfulness constraints.

(182) Alignment constraint ranking for the “grafted” accent pattern on Spanish-origin words in Quechua (rise-falls)
LINEARITY, MAXIO(T), NOCROWD >> **ALIGN (H_S, σ_{LEX})** >> **$(\sigma_{LEX})_{\omega'} \leftarrow T_S$** >> **$H_S \rightarrow$ TBU** >> **$\sigma_{LEX} \leftarrow$ T** >> **ALIGN (L_S, Rt_ϕ)** >> **ALIGN (L_S, Lt_ϕ)** >> **ALIGN (H_S, Lt_ϕ)** >> **NOASSOC** >> **ALIGN (H_S, Rt_ϕ)** >> ALIGN ($L_{\phi Rt}$, Rt_ϕ) >> ALIGN ($L_{\phi Lt}$, Lt_ϕ) >> ALIGN (H, σ', Lt) >> ALIGN (H, σ', Rt) >> ALIGN (H, $Rt_{\omega'}$) >> ALIGN ($L_{\phi Rt}$, Lt_ϕ) >> ALIGN (H, Rt_ϕ) >> ALIGN ($L_{\phi Lt}$, Rt_ϕ) >> ALIGN (H, Lt_ϕ) >> ALIGN (H, $Lt_{\omega'}$) >> ALIGN (H, Lt_ω)

In (182), the constraints from the Spanish main variant are in bold. This combined ranking will generate the attested contour of the “grafted” pattern, because the alignment of the differentially specified Spanish pitch accent tones outranks that of the Quechua phrasal tones. There is no need to assume that LINEARITY applies to the whole sequence $L_S H_S\ L_\phi H_\phi L_\phi$, but just to those tones belonging to the same source, in keeping with the analysis for IP-level boundary tones for Spanish (section 5.3.3). The analysis further supports the relational underspecification of tones at a relevant level of representation, since the L that is phrase-initial in all other Quechua variants here is phrase-medial.[398] That the Spanish-origin stressed syllable is differentially specified as a lexical accent position captures the insight gained from section 6.1.8.2 on how the loanword patterns resemble

398 Alternative analyses would involve either crossed association lines or the stipulation of an otherwise unattested H*L pitch accent, and a more complex integration of the different constraints into a single ranking.

lexical accent systems in other languages. The analysis of the “grafted” pattern further demonstrates how aspects of the Quechua and Spanish tonal grammars can be integrated, in the middle ground of the prosodic possibility space the speakers can take recourse to. Unlike the phrase-final rise-fall contour, which is a convergent intonational option accessible via at least three different constraint rankings for the speakers (two from Quechua, one from Spanish), the “grafted” pattern represents another option available to the speakers. It is a nonconvergent exploitation of the possibilities of two different tonal grammars: a lexically specified Spanish accent position uncoupled from the three-syllable window together with a peripherally determined Quechua default prominence position; Spanish pitch accents together with Quechua phrasal tones. It results in something that is best characterized as belonging to the repertoire of these speakers, rather than either language.

6.4 Context-based analysis of prosodic and other cues to information structure

Section 6.2.3.3 presented findings that default prominence in nominal sequences is phrase-final, compatible both with broad focus contexts and other information structural configurations, and that narrow final and narrow prefinal focus associate significantly with final and prefinal prominence, respectively. The analysis was restricted to sequences consisting of nominal elements for a more homogeneous dataset. This section (6.4) will qualitatively explore further cases that extend the analysis to entire utterances also involving verbs, and on units larger than individual phrases. I will begin by showing that final main prominence is also the default in verbal utterances from *Cuento* (section 6.4.1). That section will establish the prosody of “broad focus” in the sense of asserting larger chunks of information with no internal IS-partition in longer utterances. Describing the intonation appropriate to such contexts is also important because it establishes a default and contrast foil for turns which do contain internal informational structuring because of discourse contexts involving more negotiation, as in the subsequent sections on utterances from *Maptask* (section 6.4.2) and *Conc* (section 6.4.3). The examples investigated here corroborate the hypothesis that the right boundary of (narrowly or broadly) focused constituents tendentially seeks to align with the right boundary of a phrase (a broad cross-linguistic tendency, cf. Féry 2013 and the discussion in section 3.7.3). Coherence of individual PhPs in a group and their prominence profile is argued to be signaled via tonal scaling. In such groups of PhPs for which context suggests no internal IS-partition (broad focus), local pitch span on prefinal PhPs is reduced compared to that on the final PhP, thus cueing rightmost prominence (section

6.4.1). In groups for which context does suggest an internal IS-partition (prefinal narrow focus), peak height and pitch level are reduced (downstepped) after the prefinal PhP whose right boundary corresponds to the right boundary of the focus domain (section 6.4.2). This cueing of prominence across individual PhPs constitutes grounds for assuming a prosodic grouping at a higher level, but this grouping is not cued via a separate set of boundary tones, and instead only via differences in scaling. This suggests a prosodic structure in which the PhP is recursive. Besides investigating the prosodic cues for prominence above the level of individual PhPs, this entire section (6.4) is also concerned with the interaction of prosodic cues with those from other domains (word order and morphology) for information structure. We will consider this throughout the entire section, but it will come particularly to the fore in in section 6.4.3. Having established prosodic cues for final and prefinal prominence on units larger than individual PhPs as a contrast foil in sections 6.4.1 and 6.4.2 on data from *Cuento* and *Maptask* respectively, utterances from *Conc* will be used in section 6.4.3. to demonstrate how cues from prosody, word order, morphology and context interact in complex negotiations of information. It will be shown that they can align or misalign in doing so. I will conclude by arguing that neither of the different formal cues directly "mean" categories from information structure, each obeying formal constraints specific to their domain, but that their interaction is exploited to navigate subtle information structural differences in complex contexts.

6.4.1 "Broad focus" in Cuento

This section will consider data from *Cuento*, the task where speakers hear a recorded story and then re-tell it to each other. The *Cuento* corpora are the most narrative corpus type covered in this study. The active speaker in this task has relatively great liberty for telling the story they heard from the recording or the other speaker in whichever way they choose, they can pace themselves and rarely need to anticipate interruptions or challenges to how they build up the common ground. The discourse progresses at the speed of their choice. Because in a narrative format like *Cuento* there is very little risk of being misunderstood or contested, and speakers have comparatively much time to plan ahead, they can effectively unilaterally determine what content should be added to the common ground and in what units. They can minimize elements that are not at-issue and intended to help specify and negotiate which material should enter CG in what way (i.e. elements intended for common ground *management* instead of *content*, cf. Krifka 2007). Therefore they can to a certain extent forego IS-related internal prosodic structuring of utterances and instead increase the use of prosody for demarcating maximal coherent units of material intended to enter CG in one go (matching large prosodic units to conversational moves). Because of this, the intonation observed in *Cuento* exemplifies how new information is asserted in large chunks without the anticipation of problems or the need to negotiate as reflected in internal IS-partition, i.e. what is often called "broad focus". We will see how default rightmost prominence is cued via scaling across units of more than one individual PhP to demarcate those larger chunks corresponding to the broad focus domains. This constitutes prosodic phrasing at a level above (minimal) PhPs. In the absence of a need for marking internal IS-partitions, phrasing can also sometimes come to be employed to cue high-level syntactic structure like relative constructions or subordinations.[399] We'll see how this plays out in the first sequence from AZ23 & ZZ24's *Cuento*.

6.4.1.1 First Cuento sequence: Establishing a default for broad focus

(183) AZ23_Cuent_Q_0104–0274[400] (context for Figures 174 and 175)[401]

time (seconds)	AZ23
10.4	(L H-----) !(H----------)
	unay-shi ka-naa huk
	long.ago-REP COP-PST.REP one
	they say a long time ago there was a
12.1	L
	uhm

399 This claim is related yet different to that by Féry & Ishihara (2010: 37) that in an "all-new sentence" (one with broad focus), "the formation of prosodic phrases as well as the tonal pattern and scaling depend entirely on the morpho-syntactic structure". Here we will see instead that scaling in particular can cue broad focus on an utterance irrespective of syntactic factors, e.g. whether a sentence is verb-final or object-final. The effect of syntactic structure is instead observable sometimes in the boundary location of individual PhPs that form a larger unit whose coherence is signaled by scaling and which is the domain of broad focus.

400 The first line for each turn gives a basic intonational analysis, with low (L) and high (H) tones aligned at the position in the line below where their targets are first realized. H tones form plateaux rightwards only if indicated by a dashed line, L tones form low stretches by default. Round brackets () indicate phrase boundaries of PhPs/APs, square brackets [] indicate the boundaries of larger units. Where it isn't quite clear whether individual phrases should really be analyzed as such, because of severely reduced scaling (see text), the boundaries and tones in doubt are in grey. The exclamation mark (!) before a phrase indicates downstep of the entire phrase (pitch level), the upside-down exclamation mark (¡) indicates upstep.

401 https://osf.io/m397j/

12.6	(L H-------------)
	hampikuq runa
	healer person
	healer
13.6	[(L H)(L H)(L H L)]
	allaapa yacha-naa hampi-ku-y-ta
	much know-PST.REP heal-MID-INF-OBJ
	[who] knew a lot [about] healing
16.2	(L H-----)(LH------)
	tsay-qa huk dia-qa aywa-n
	DEM.DIST-TOP one day-TOP go-3
	then one day he went
18.3	(H----------------L)
	hampi-y ashi-q
	heal-INF search-AG
	looking for medicine
19.6	(L H)
	karu-pa
	far-GEN
	far away
21.2	(L H-----------)
	tsay-chaw-shi
	DEM.DIST-LOC-REP
	then
22.9	L H L
	tsaka-y
	get.dark-INF
	getting dark
24.0	H
	tar-
24.3	(L H---------)
	tsaka-pa-ku-ski-n
	get.dark-ITER-MID-ITER-3
	it is getting dark
25.7	![(L H)(L H)(LH L)]
	mana-raq hampi-y-man chaa-r-nin
	no-CONT heal-INF-DEST arrive-SUBID-3
	[with him] not yet having come upon the medicine

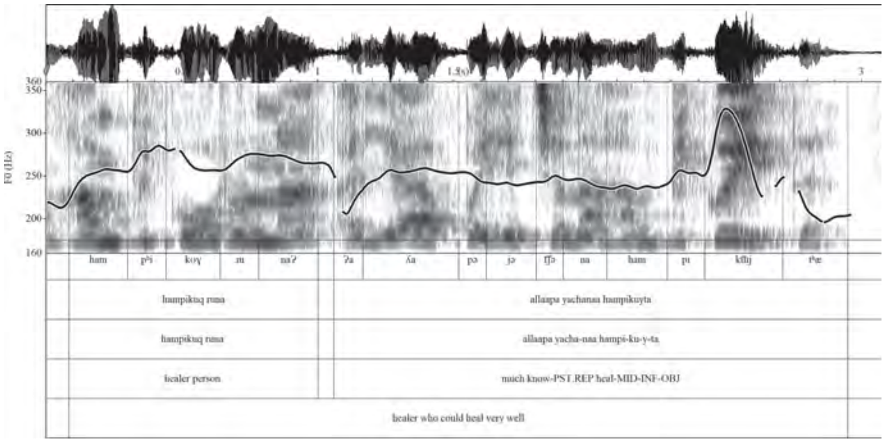


Figure 174: AZ23_Cuent_Q_0126.[402] Cf. (183) for the context.

In (183), nearly each turn corresponds to the assertion of a new proposition. The propositions are nearly all predicated about the protagonist, who, once introduced (at 12.6), is implicitly understood as topical subject and as most prominent discourse referent referred to without a (pro)nominal expression. In terms of the model of implicit QUDs by Riester and colleagues (cf. section 3.7.2), the utterances in (83) all answer a QUD like WHAT HAPPENED (NEXT)? or WHAT DID THE PROTAGONIST DO?, with almost no material repeated from one utterance to the next (i.e. all material is new). Thus these utterances almost all have the broadest possible focus domain.

Figure 174 (at 12.6 and 13.6) and Figure 175 (at 24.3 and 25.7) are examples in point for how high-level prosodic phrasing is cued in such broad focus utterances. The boundaries of high-level syntactic units are demarcated via a similarly high level of prosodic phrasing: the separation between the head noun *hampikuq runa* and its (quasi-) relative phrase *allaapa yachanaa hampikuyta* in Figure 174, and that between the two verb phrases, *tsakapakuskin* and *manaraq hampiman chaarnin* in Figure 175 are both cued by short breaks in the flow of speech. In both cases the first and shorter unit is realized with a rising contour signaling continuation; the second and longer one with a rise-fall signaling finality. Scaling at the level of local pitch span acts to cue coherence across lower-level phrases: each individual word in both of the larger phrases could be assumed to form a PhP with a rising contour. This is evidenced by low tonal targets at the beginning of each of the prefinal words, followed by small rises and then a return to low towards

402 https://osf.io/dmf6u/

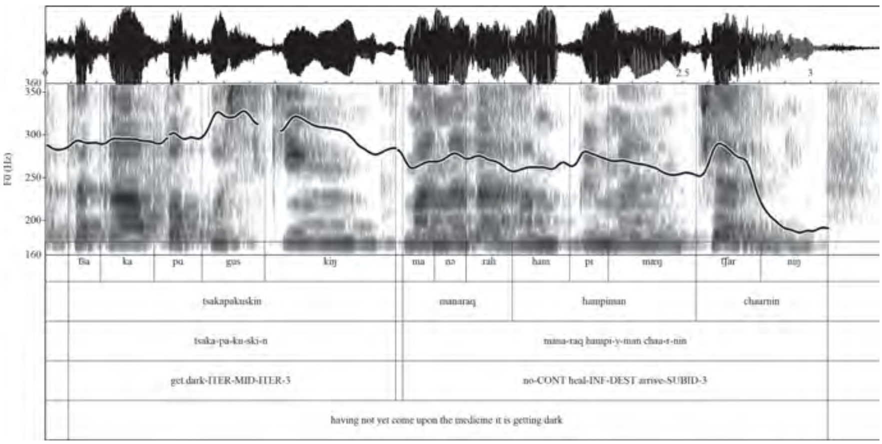


Figure 175: AZ23_Cuent_Q_0243[403] Cf. (183) for the context.

the beginning of the next (cf. also Figure 171 in section 6.2.3 and the discussion there – the analytical uncertainty due to the reduced scaling is marked in (183) by setting the corresponding tones and phrase boundaries in grey). On the final word in both cases, *hampikuyta* and *chaarnin*, pitch events are unambigously identifiable. Whether we analyze individual PhPs on each word or not, prosodically the two prefinal words in the larger phrases are clearly cued to be subordinate to the final one via local pitch scaling. Thus in both cases all three words are part of a larger unit with final prominence (parallel to the findings in section 6.2.3). The pattern of a pronounced rise-fall on the final word, dwarfing any preceding pitch movement in terms of local scaling, is found in the majority of longer utterances in this *Cuento* corpus.[404] Often, as here, the prefinal phrases in the larger unit are rising, if identifiable, and the final one is rising-falling. Since the context clearly suggests broad focus, it is assumed that such a larger unit with a single prominence profile can in fact be made up of PhPs with different tonal contours, as here. Such sequences were classed as indeterminable with regards to their prominence profile in section 6.2.3.3.

This global contour shape cueing rightmost prominence across a large unit seems unaffected by syntactic differences in the sentences corresponding to it, but some of these differences are nonetheless reflected in the prosody. In Figure 175 the first verb is syntactically the main verb, the second, *chaa-r-ni-n*, is marked

403 https://osf.io/9h8u3/

404 Cf. section 5.1.3.1, where very similar context conditions were argued to favour the Spanish phrase accentuation, which is also formally similar to the broad focus contours discussed here.

morphologically as subordinate by the morpheme *–r*. They are phrased separately, and their subordinating relationship is likely reflected in the downstep (in terms of pitch level) between the first and second verb phrase in Figure 175, which does not occur in Figure 174. Yet the syntactic structure of the sentence in Figure 175 does not result in a global contour signaling prefinal prominence (on the main verb): the rising contour shape on the main verb phrase suggests continuation, and even though it is scaled higher in terms of pitch level than the final subordinate verb phrase, it is the latter which is realized with the final rise-fall, also scaled larger in terms of local pitch span. Highest prominence is therefore still final in the entire utterance in Figure 175. This suggests that different scaling cues (pitch level compared between phrases vs. pitch span on local pitch events) can signal different structures: some phrasing matching relevant internal syntactic partitions while at the same time signaling no internal IS-partition (=broad focus). We will see this again in the next examples. For global contour shape, word order also does not seem to matter. In Figure 174 the final word is the direct object, while in Figure 175 it is the verb, yet the contour is the same and context suggests broad focus in both cases. In Figure 174, the utterance does not map onto one sentence, but on the continuation of one, plus another. (183) shows that its first part is the completion of a sentence begun at 10.4: *unayshi kanaa huk . . . uhm . . .* ***hampikuq runa*** "a long time ago there was a . . . uhm . . . **healer person**", with a filled pause occurring between the numeral *huk* and the noun it modifies. The second part, *allaapa yachanaa hampikuyta*, could be a full sentence with a null pronominal subject: "[s/he] knew a lot about healing". Prosodically however, the two parts of Figure 174, *hampikuq runa* and *allaapa yachanaa hampikuyta* are more closely fitted together than its first part is with the preceding *unayshi kanaa huk*, from which it is separated by a mostly filled pause of more than 500 miliseconds (likely signaling continuation, as does the rising contour on *unayshi kanaa huk*). *Hampikuq runa* is realized with a rising contour, also indicating continuation, and separated by only a very short break (90 miliseconds) from *allaapa yachanaa hampikuyta*. The realization of Figure 174 as a single utterance suggests a reading of its second part as a relative clause with the first part as head noun. Syntax and prosody thus seem to be able to contribute independently to discourse cohesion.

6.4.1.2 Second Cuento sequence: Syntactic and prosodic subordination

(184) AZ23_Cuent_Q_0274–0525[405] (context for Figures 176 and 177, continuing right after (183))

time (seconds) AZ23

27.4 *pause*

28.2 (L H)
tsay-chaw *(0.7)*
DEM.DIST-LOC
then

29.8 (L H--------)(L H--------)
ala-tsi-ku-n mantsa-ku-n
be.cold-CAUS-MID-3 be.afraid-MID-3
he is cold, he is afraid

32.0 L
y
and
and

32.5 (H---------------------------------L)
tsiqtsi-kuna-si yuri-pu-ski-n
bat-PL-ADD appear-DIR-ITER-3
even bats appear

35.2 (L H L)
ala-tsi-ku-yka-pti-n
be.cold-CAUS-MID-PROG-SUBDIFF-3
[while] he is getting cold

36.4 [(L H)¡(L H--------)]
mallaqa-y-ni-n tsari-ski-n
be.hungry-INF-FON-3 grasp-ITER-3
hunger grasps him [= he gets hungry]

38.8 (L H--------------------------------)(L H L)
tsay apa-naa chuqllu-ta-wan aytsa-ta
DEM.DIST bring-PST.REP corn-OBJ-INST meat-OBJ
he had brought corn and meat

405 https://osf.io/s9kqr/

41.6 (L HL)(LHL)(L H L)
almuersu-n na-n mirienda-n-ta
lunch-3 PSSP-3 snack-3-OBJ
his lunch, thing, his snack

44.4 (L H-------------)(L H----)(L H------------)
tsay-chaw-na-m allaapa mallaqa-ski-n-lla-raa
DEM.DIST-LOC-DISC-ASS much be.hungry-ITER-3-LIM-CONT
then he is still very hungry

47.2 L
y
and

47.7 [(H L) (H L) (L H L)]
hoqta chuspi-ta tsari-ski-r miku-ski-n
six fly-OBJ grasp-ITER-SUBID eat-ITER-3
he catches and eats six flies / catching six flies he eats [them]

51.1 (L H L)
pacha-n hunta-na-n-paq
stomach-3 fill-NMLZ-3-BEN
in order to fill his stomach

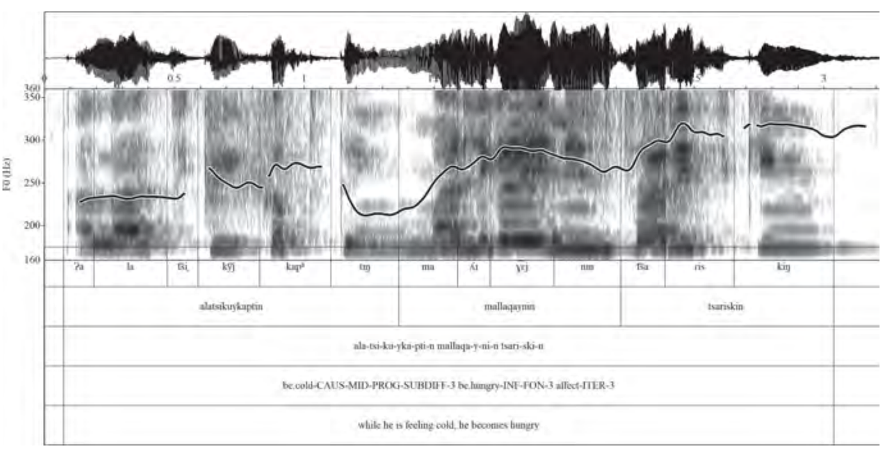


Figure 176: AZ23_Cuent_Q_0352.[406] Cf. (184) for the context.

406 https://osf.io/4as7m/

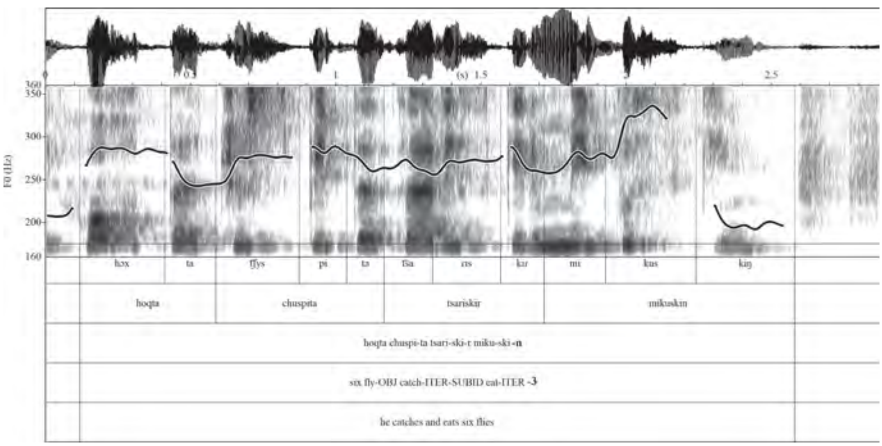


Figure 177: AZ23_Cuent_Q_0477.[407] Cf. (184) for the context.

This section further discusses utterances that exhibit some internal phrasing at the level of individual PhPs, while still also cueing broad focus overall. Some of this internal phrasing corresponds to boundaries between high-level syntactic divisions. The utterances in both Figures 176 and 177 are of sentences that contain two verb phrases, a main verb and a subordinated one each. In Figure 176 the subordinated verb phrase is *alatsikuykaptin* „[while][408] he is getting cold“, and the main one is *mallaqaynin tsariskin* “he gets hungry” (lit. “his hunger grasps [him]”). Here, *alatsikuykaptin* clearly forms its own rise-falling phrase, with the final low tone reaching a low target early on the final syllable [tɪŋ] (cf. the rising contour on *tsakapakuskin* in Figure 175 where the low target is only reached with the onset of the next word). *Mallaqaynin tsariskin* is realized as two rising phrases, with the second, on *tsariskin*, upstepped: the H tone in the second phrase reaches far higher than the one in the first, and the initial L of the second phrase is also just a little lower than the H on the first and far higher than that phrase’s initial L. It is clear that the boundary between *alatsikuykaptin* and *mallaqaynin tsariskin* is stronger than that between the phrases on *mallaqaynin* and *tsariskin*. This suggests the latter are part of one larger prosodic unit which contains both PhPs formed on

407 https://osf.io/wur83/

408 The semantic relation between the action conveyed by the main verb and that conveyed by the verb marked as subordinated by *–pti* is underspecified: it can be interpreted as temporal (consecutive or contemporaneous), causal, conditional etc. depending on the context (Parker 1976: 143–144, 160). Parker (1976: 160) states that what they express is that the action or state denoted by the subordinated clause is a “prerequisite” for that denoted by the main clause. I adopt that formulation in the main text.

the two words, but not that on *alatsikuykaptin*, like A(BC) or (A(BC)). The high-level separation between *alatsikuykaptin*, on the one hand, and *mallaqaynin tsariskin*, on the other, reflects the fact that these form two separate propositions, "he got cold", and "he got hungry", both partial answers to the superordinate QUD WHAT HAPPENED ONCE IT GOT DARK? (answered by the entire sequence (184)), with "he got cold" already having been asserted at 29.8 and now serving as a prerequisite for "he got hungry". *Alatsikuykaptin* here establishes a reference back to the scene set with the previous utterances (at 29.8–32.5). The separation is also reflected in the syntax: *-pti* in *alatsikuykaptin* marks a subordinated verb with a different subject from that of the main clause (Parker 1976: 143). Its subject is not the same as that of the main verb *tsariskin* "s/he grasps", whose subject is the preceding *mallaqaynin* "his hunger", with only the possesive marker *–n* (3^{rd} persons singular) coreferential with the protagonist. In the following utterance at 38.8, the healer is reactivated as subject with the demonstrative *tsay* instead of a null pronoun, confirming that the subject change took place here.

The final rise on the utterance indicates thematic continuation in the sense that the superordinate QUD WHAT HAPPENED ONCE IT GOT DARK? is not yet fully answered: the issues of hunger and lack of food are continued. The utterances at 38.8 and 41.6 provide answers to the subordinate QUD WHY WAS HE HUNGRY?, with 38.8 answering another subordinate QUD about how the healer had brought corn and meat (HAD HE BROUGHT SOMETHING TO EAT?), but that this was for lunch (41.6, answering subordinate QUDs along the lines of DID HE HAVE ANY LEFT? or WAS IT ENOUGH FOR DINNER, TOO? by implicature[409]). These two utterances each completely answer their respective QUDs as well as together answering the QUD WHY WAS HE HUNGRY?, and are produced with rise-falling contours. The fact that they provide a background to the superordinate QUD of WHAT HAPPENED ONCE IT GOT DARK AND HE GOT COLD AND HUNGRY is then brought back to the fore in 44.4, which repeats that he was hungry.

409 How to best analyze implicit QUDs in this sequence is a tricky question which cannot be fully treated here. To give an example, strictly speaking, according to the principles laid out in Riester et al. (2018); Riester & Shiohara (2018); Riester (2019), at 38.8 the implicit QUD would have to be WHAT DID HE DO?, because neither the verb *apanaa* 'he had brought' nor its objects *chuqllutawan aytsata* 'corn and meat' are previously mentioned in the context. However, the previous assertion that the healer got hungry arguably evokes the question in a rational interlocutor whether a competent healer setting out on a long journey through uninhabitated country did not bring any provisions with him. In such a context, the action denoted by the verb *apanaa* would be contextually given, making the implicit QUD something like WHAT HAD HE BROUGHT? or even HAD HE BROUGHT ANY FOOD?. Possibly, this is also cued by the objects being postverbal, because then word order is compatible with a prosodic realization cueing final prominence (on the food items), unlike a verb-final one. Future research will have to show how it is possible to include such deliberations into the generation of implicit QUDs in a principled way.

The morpheme *–na* in *tsaychawnam mallaqaskinllaraa* indicates a kind of topic change (Leonel Menacho and Gabriel Barreto, p.c.), and the utterance is realized with rising contours, suggesting that a further partial answer to the superordinate QUD is still to be given. This happens at 47.7/Figure 177. There, *huqta* and *chuspita* likely form only-falling phrases each (more clearly on *huqta*). The falling contours possibly serve to highlight the six flies as new referents that are important in that they will play a crucial role in the story later, while at the same time keeping the phrases formed on *huqta chuspita* prosodically subordinate to the larger unit as signaled by the reduced local pitch span on them compared to that on the final word *mikuskin*. The subordinated verb *tsariskir* and the main verb *mikuskin* directly following it are realized with a single rise-falling phrase with the subordinated verb realized only on the initial low stretch. In terms of local pitch span, the final pitch movement on *mikuskin* is larger than all the preceding ones in the utterance, serving to cue that not *huqta chuspita* alone, although realized with falling contours, but the complex proposition conveyed by the entire utterance, is the answer to the QUD (HE GOT HUNGRY AND?). In Figure 177, the subordinated verb and the main verb are phrased together, while in Figure 176, they were phrased apart. This fits the tighter syntactic relationship the two verbs have here: the subordinating morpheme *–r* in *tsariskir* indicates subordination with the same subject as the main verb *mikuskin*, in contrast to *–pti* used before. The subordinate verb and the main verb here not only share the same subject, the null pronoun coreferential with the protagonist, but also the object, the six flies (*huqta chuspita*). Word order here is the unmarked OV.[410] The Quechua utterance arguably achieves an effect of cueing the object as new while at the same time signaling a tight relationship between the two verbal actions (so that they can be interpreted as conveying a single complex proposition) and broad focus overall, something which no single English translation can do. The two options given in (184) each only achieve some of these properties, either unmarked word order or the subordinating relation between the two verbs. Morphosyntax and prosodic phrasing here align to cue that the utterances in Figures 176 and 177 have different information structures: in Figure 176, the two verb phrases denote two separate propositions, with the one denoted by the main verb phrase *mallaqaynin tsariskin* asserted and the one denoted by *alatsikuykaptin* serving as a prerequisite (having been previously asserted). In Figure 177, in contrast, a single proposition is asserted composed of two verbal actions (grasping and eating).

These individual examples support the hypothesis from section 6.2.3.3 that larger prosodic groupings exists which can include several PhPs as identifiable from their tonal make-up. The extension of this unit is signaled via local pitch span

410 Cf. sections 6.4.2.3 and 6.4.3 for discussions on which word orders are unmarked.

differences: relatively smaller or reduced pitch span was observed on prefinal PhPs in such units, while the final PhP had the greatest pitch span. Such groups of PhPs with greatest pitch span in final position all occurred in contexts with comparatively broad focus, supporting the hypothesis that they form larger units on which prominence is final by default. The groupings created by prosodic phrasing, both at this higher level and at that of individual PhPs, interact with groupings created by morphosyntax, either to mutually reinforce when aligning, or to convey additional grouping information when not aligning. Tonal contour shape of individual PhPs (rising/falling) seems also able to cue distinctions aiding discourse coherence independent of focus, suggesting that global phrasing cued by relative differences in local pitch scaling (at a level above individual PhPs), local contour shape, and local phrasing (at the level of PhPs) each may contribute different types of information.

6.4.2 Complex contexts and marked prosody in Maptask

In the following, context examples from *Maptask* will be examined, with quite different interactional dynamics and resulting prosody. Instead of long, uninterrupted narrative passages produced by one speaker, there is much more back-and-forth between the speakers, negotiating CG increments via clarification requests and specifications of how a proposition relates to elements already in CG. We will see how pitch scaling plays a role in cueing IS-partitions within larger prosodic units.

6.4.2.1 First sequence: Initial and final backgrounded constituents

In the first sequence, KP04 (with the path on the map) and TP03 (without it) have already been playing the maptask for about three and a half minutes. They proceeded without problems until coming upon the spot where the maps differ (in the upper part of the map, the skunk and the lightning have swapped places, cf. Figures 188 and 189 in Appendix B). Confusion ensues and they realize that their maps are different. They change strategies: instead of KP04 leading the conversation by giving descriptions of the path for TP03 to follow, he now asks TP03 a series of questions about the layout of this part of his map, trying to bring the landmarks, mostly known by now, into their correct relative positions. Arguably, the superordinate QUD that best represents this strategy has three open variables, two location variables and one path variable, as in FROM WHICH LANDMARK TO WHICH LANDMARK DOES THE PATH GO WHICH WAY?. This is itself a subquestion to the more general WHAT IS THE CORRECT PATH? guiding the game overall. The speakers begin pursuing this new strategy starting from the lightning, which they identified as one of the landmarks that has a divergent location. This is where (185) takes off:

(185) TP03&KP04_MT_Q_2010–2216[411] (context for Figures 178 and 179)

time (seconds) | KP04 (*with the path on the map*) | TP03 (*without the path on the map*)

201.0
(L H--------)
y tsiqtsi-qa
and bat-TOP
and the bat

202.4
(L H----) (LH L)(L H---------------------L) (L H--------L)
tsiqtsi-qa manka este tsay na-lla-chaw-mi[412] este tsay manka-pa
bat-TOP pot uhm DEM.DIST PSSP-LIM-LOC-ASS uhm DEM.DIST pot-GEN
the bat is just to the thing of the pot, of that pot

205.8 — *(2.2)*

208.0 — (TP03)
(L H-----------L) !(H L)
huk costado-chaw ka-yka-n
one side-LOC COP-PROG-3
it is on the other side

209.2
(L H---) !(L H-----)(H----------L)!(H-----L)
tsawra-qa entonces kay-naw-mi rura-shun
then-TOP then DEM.PROX-SIMIL-ASS do-1.PL.FUT
so then this is what we'll do

211.5
(L H-------)
tsay tillakuq
DEM.DIST lightning
that lightning

213.5
(L H -----------------L)(LH)
tillaku-yaq-cha aywa-nki aw
lightning-TERM-ASS go-2 right
you go to the lightning right

214.5 — (TP03)
(LH)
ajá

215.1 — (TP03)
(L H) (L H L)!(H--------L)
ti[llakuq] hawa-n-pa-m (?aywa-yka-n)
lightning below-3-GEN-ASS (?ir-PROG-3)
(?it's going) below the lightning

215.2
(L H)
[tillaku]

---

411 https://osf.io/pcu9b/

412 The morpheme *–mi* is here attached to a semantically empty predicate: *na-lla-chaw-mi*, where *na-* is a placeholder morpheme speakers use when for whatever reason (lexical retrieval problems etc.) they cannot say the actual word they mean (Leonel Menacho, p.c.). Thus *na-lla-chaw-mi* is an assertion that something (*manka* 'the pot') is just at the location that is specified by something the speaker cannot at the moment name.

216.5 (L H---------)
tillaku-pa uh
lightning-GEN
the lightning uh

217.1 (L----H---------------) (L H L) !(LH L)
tilla punku-pita may-chaw-tan manka
lightning front-ABL where-LOC-DET.VAR pot
from in front of the lightning, where is the pot

219.5 (L H) (L H----------------------------------L)(LH L)
tillaku-pa kay-naw hawa-n-man-pa-chaw-chi manka
lightning-GEN DEM.PROX-SIMIL below-3-DEST-GEN-LOC-CONJ pot
from the lightning, it's like this in towards below it, the pot

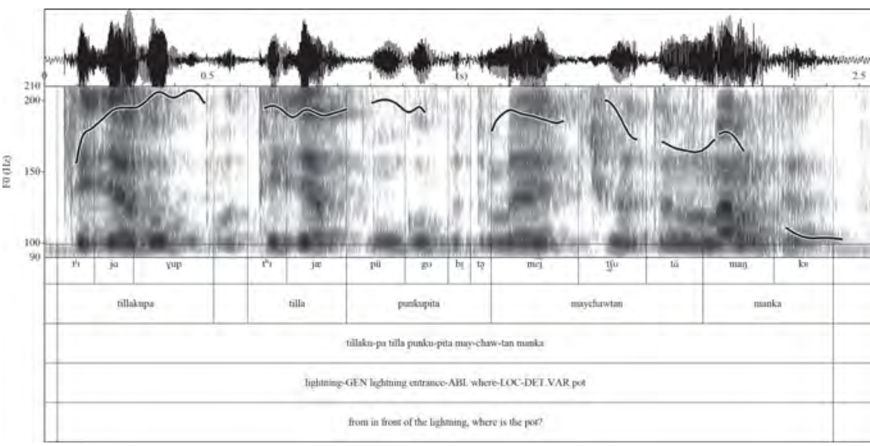


Figure 178: KP04_MT_Q_2165.[413] Cf. (185) for the context.

At 217.1, KP04 asks where the pot (*manka*) is from in front of the lightning (*tilla(kuq)*). Both objects are given referentially from the preceding discussion about their locations. However, while *tillakuq* is active from the preceding turn, *manka* was last mentioned at 202.4, nine turns and 13 seconds previously. At 217.1, *tilla punkupita* is a topic in the sense that it specifies the first location variable in the superordinate QUD FROM WHICH LANDMARK TO WHICH LANDMARK DOES THE PATH GO WHICH WAY?. It referentially restricts the open proposition denoted by the wh-question. Although the lightning is active, *tilla punkupita* is here realized fully probably because it changes roles: in KP04's turn at 213.5 it was the goal, but is now the origin. This topic is the leftmost constituent in the utterance, realized with a single rising phrase. The final constituent *manka* is fully realized because the pot is not

413 https://osf.io/ptjnu/

active enough in the directly preceding context: had the previous utterances been about it as subject, *maychawtan* would have been contextually specific enough as a question. However, as it is, *tsiqtsi* "the bat" might have been another possible candidate in the context. Thus *manka* here specifies the second locational variable in the super-QUD, but it is also referentially given. The wh-constituent *may-chaw-tan* is in focus, asking for the specification of the path variable relative to lightning and pot in the super-QUD. Context leaves ambiguous whether the pot is only part of the question background, or separately specified, i.e. a second topic.[414] Prosody separates the wh-word *maychawtan* from the final *manka*, realizing both with a falling contour each. *Manka* is downstepped in terms of pitch level from *maychawtan*, but still clearly realizes its own contour. The initial topic is realized with a separated rising phrase. In response, TP03 produces a declarative (Figure 179) with similar syntax and prosody at 219.5.

In his answer, the lightning, here *tillakupa*, is again realized first on a separate phrase with a rise. As in the question, it functions as a topic here, restricting the reference of the at-issue part of the proposition. The at-issue part, corresponding to the wh-constituent in the question, is expressed by the locative noun phrase *kaynaw hawanmanpachawchi*[415] "like this in towards below", realized as one rise-falling

414 This difference could perhaps best be described by refering to the kind of hierarchical QUD-structure the question implies. I suggest that it is the difference between two structures with differing depths. In the first with three question nodes, the super-QUD FROM WHICH LANDMARK TO WHICH LANDMARK DOES THE PATH GO WHICH WAY? leads on to a first subordinate question, FROM WHICH LANDMARK TO THE POT DOES THE PATH GO WHICH WAY?, and only then to the one actually asked, FROM THE LIGHTNING TO THE POT DOES THE PATH GO WHICH WAY?. In the second with only two question nodes, both landmark variables are filled at the same time in only one subordinate question. The former results in a question in which the pot is only backgrounded, while it is a second topic in the latter. In English, this difference would arguably be cued by the position of the strongest prominence (in capitals) in the corresponding question:

(i) A. From the lightning, WHERE is the pot? ['pot' = backgrounded]
B. From the lightning, where is the POT? ['pot' = second topic]

In both questions, what is at-issue/in focus is only the part denoted by the wh-element.

415 The location-denoting noun *hawa* is here marked with not just one, but three, locative suffixes: -*man* (glossed as DEST, marking a destination for a path), -*pa* (glossed as GEN, 'genitive', often also marking a location along which a path passes), and –*chaw* (glossed as LOC, marking a static location). Such multiple marking is not frequent in our corpora, but similar cases are also reported in Parker (1976: 86–87). The translation tries to reproduce this with the sequence of prepositions. It's possible that this proliferation of locative descriptors is due to TP03 struggling a bit to find the right formulation for the actual spatial layout (note the use of *kaynaw* 'like this' and his continued efforts to describe the location of the objects in subsequent turns): on his map, lightning, bat and pot form an isosceles upside-down triangle, with lightning and bat at the left and right top corners and the pot at the bottom corner. Thus, the pot is both to the right and below the lightning.

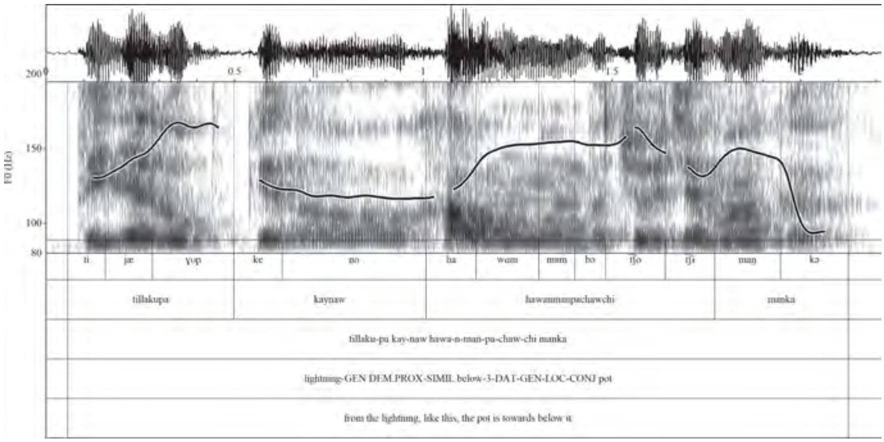


Figure 179: TP03_MT_Q_2195.[416] Cf. (185) for the context.

phrase. This constituent is marked with the conjectural evidential *–chi*, one of the evidentials said to attach to focused constituents in Southern varieties of Quechua (cf. Muysken 1995; Sánchez 2010; Muntendam 2015). Morphosyntactic marking here aligns with prosody and context to mark the focus domain, but we will see in subsequent examples that this is not always the case. The final constituent is again *manka*. It is also realized via its own phrase, this time not downstepped. It is a resumptive topic here, specifying the second referent about which the at-issue predication has been made even though this relation is already evident from the context. Even though *manka* is not downstepped as in the preceding wh-question, it is notable that the overall contour of these two utterances is clearly distinct from the broad focus utterances in *Cuento*. Here, context and morphosyntax indicate an IS-partition with narrow focus on the locational constituent, and this is cued prosodically. On the one hand, via the placement of a right phrase boundary of a falling phrase after the focal constituent, and on the other via a global contour whose scaling does not cue final main prominence and thus broad focus: the peak on the locational constituents is not scaled lower than that on the final constituent, *manka*.

6.4.2.2 Second sequence: Differential interpretations for initial elements

This section demonstrates that the same initial element can receive different interpretations in two utterances, even though local prosody and word order are very similar. It also demonstrates how verbfinal sentences do not exhibit the rise-falling

416 https://osf.io/8v7rn/

contour cueing rightmost prominence if the context is not compatible with broad focus, as in *Cuento*.

(186) TP03 & KP04_MT_Q_2221–2329[417] (context for Figure 180; directly continued after (185))

time (seconds) | KP04 (*with the path on the map*) | TP03 (*without the path on the map*)

222.1
L H
tsay-pa-
DEM.DIST-GEN
there

223.7
(L H--------------L) !(H------------L)
tsay-pita may-pa-tan tsiqtsi ka-yka-n
DEM.DIST-ABL where-GEN-DETVAR bat COP-PROG-3
from there where is the bat

226.0
(H------L)
tsiqtsi-qa-
bat-TOP
the bat

227.1
(0.5)

227.6
(L H----------------L)(L H)
tsay na ka-q-chaw este tillakuq
DEM.DIST PSSP COP-AG-LOC that lightning
where that thing is, the lightning

229.8
(HL)
ya

230.5
(L H-----) !(L H L)!(H L)
ya manka y tsiqtsi-m ka-yka-n
ok pot and bat-ASS COP-PROG-3
yes the pot is there with the bat

231.8
(L H)(H-------------------L) !(H----L) (L H----------L)
manka chawpi-chaw-mi ka-yka-n [y hawa-n-kuna-chaw
pot middle-LOC-ASS COP-PROG-3 below-3P-PL-LOC
the pot is in their middle and below them

KP04 continues to ask about the location of the other objects. At 223.7, he asks where the bat is. In response, TP03 produces two turns (226.0 – 227.6) separated by a pause,

417 https://osf.io/bg78v/

asserting that the bat is where the lightning is. The overall sequence is here again one of topic followed by at-issue/focal content, which in turn is followed by a postposed NP. The first turn consisting only of the constituent *tsiqtsi-qa* is marked morphologically as topic but not produced with a rise. The focused constituent *tsay na kaqchaw* "where that thing is" is right-aligned with a rise-falling phrase. The postposed replacement for *na*, *tillakuq* "the lightning", is produced with a rise. KP04's response here is only via an acknowledgement token (229.8). TP03 seems to agree that the assertion he has just made is not sufficient as answer to KP04's question given the purposes of the conversation, because he continues with an elaboration (230.5–231.8, cf. Figure 180).

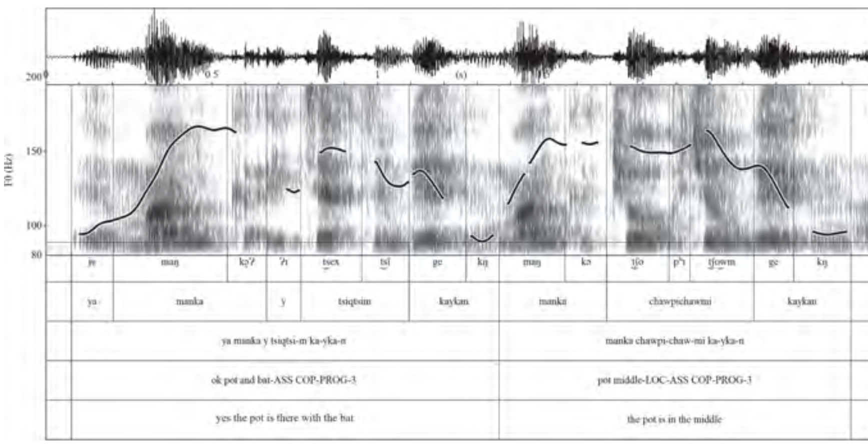


Figure 180: TP03_MT_Q_2305.[418] Cf. (186) for the context.

These two utterances both initially realize *manka* with a rising phrase, followed by a further constituent morphologically marked with *–m(i)*, and realized with a falling phrase. Then follows the existential verb *kaykan*, realized with a downstepped falling phrase. Both of these are best understood as answering the QUD WHAT IS THERE?, with "there" the location the discourse has progressed to previously,[419] i.e. where the lightning (*tillakuq*) is and where the bat (*tsiqtsi*) has been asserted to be. That this is the appropriate implicit QUD is signaled by the presence of the verbs of being. The verb of being *ka-* is omitted in the 3rd person present when it is used

418 https://osf.io/ajm2n/

419 As also noted in section 5.1.3 on Spanish, in the *Maptask*, the progression from landmark to landmark along the path on the map in the conversation of the participants is parallel to the progression the speakers make in the discourse via the negotiation of how the context set can be reduced. The landmarks and the path the speakers have successfully agreed on are, quite literally, their *common ground*.

as a copula, but not when used existentially, or when it serves as a locative copula, similar to Spanish *hay* (cf. Parker 1976: 68). This use is frequently expressed via the progressive form, as here. As part of the background of the QUD, the verbs are also not part of the at-issue content of these two utterances that answer it. Since the previous context has established that at the location under discussion, the bat (*tsiqtsi*) and the lightning (*tillakuq*) are there, what is new in the first utterance (230.5) is the information that the pot (*manka*) is also there, and it is asserted that it is there together with the bat, while the bat being there on its own is presupposed. This difference in information status is likely responsible for the prosodic realization of these two elements. *Ya manka* is realized in a separate rising phrase, like an initial topic, although it quite clearly isn't one here – the pot is the one referent that hasn't already been asserted to be in the location they are discussing. *Y tsiqtsim* is realized in a separate rise-falling phrase, but downstepped relative to the rising phrase on *ya manka*. This realization, instead of one e.g. in which the entire sequence is realized with a finally prominent rise-falling phrase (as seen in *Cuento*, appropriate for a context in which none of the referents are given or presupposed to be at the location), reflects the context conditions outlined above, which the translation also attempts to recreate: "yes the pot is there with the bat", instead of e.g. "yes the pot and the bat are there". *Tsiqtsim* and *kaykan* are both downstepped, but there are additional cues that their information structure differs: in both cases, the downstep in pitch level can be taken to cue that the element to the left of the downstep is more prominent. In the case of *ya manka* and *y tsiqtsim*, in addition, the phrase on the first element is rising, and the second is marked with *–m(i)*.[420] Assuming leftwards scope for *–m(i)* (in accordance with Muysken 1995; Sánchez 2010), this would support the interpretation that the new information that the pot is there with the bat is asserted. Yet the referents forming this proposition differ in status: that the pot is there with the bat is at-issue, but the entailed proposition that the pot is there is new, while that of the bat being there is presupposed.

420 Marking a constituent with one of the evidentials *–m(i)/-ch(i)/-sh(i)/-cha* is thought to indicate focus on that constituent (Muysken 1995; Sánchez 2010). Yet see note 84 in section 3.7.3.2 as well as the analysis in section 6.4.3 for problems with this account. Beyond this syntagmatic function, this work does not discuss the paradigmatic semantic contribution of the different evidentials besides the very broad characterization that *–m(i)* and *–cha* mark information as either asserted or based on "best possible grounds" Faller (2002, 2003), *-ch(i)* marks information as based on conjecture and *–sh(i)* as based on hearsay. In work on southern Quechua varieties, a consensus seems to have been reached that the meaning contributed by the evidentials falls under the broad category of non-at-issue meaning (similar to that of discourse particles and intonation, cf. Zimmermann 2011; Truckenbrodt 2012), but whether it is a conventional implicature (Roberts 2017) or a speech act operator (Faller 2014) seems undecided. Hintz & Hintz (2017) pursue a somewhat different route but are specifically concerned with the evidentials of South Conchucos Quechua.

Between *tsiqtsim* and *kaykan*, on the other hand, the metrical relation cued is also strong-weak, but the left element *tsiqtsim* here is realized with a falling phrase and marked with *–m(i)*, both used to cue the right edge of focal material. This aids the contextual interpretation that *kaykan* is only backgrounded.

In the second utterance (231.8), *manka chawpichawmi kaykan*, *manka* is also realized with a rise, but the following *–mi*-marked locative noun is not down-stepped. Here, *manka* being realized with a separate rising phrase aids in its interpretation as topical subject. Having been copular subject in the preceding utterance, the pot is a prime topic candidate already, so that in a comparable sequence, *manka* could probably be omitted and the utterance still be understood to be about the pot. However, here there are three highly salient referents in the discourse: the pot, the bat and the lightning, and the guiding QUD for the entire section of the conversation is how they are located relative to each other. Under these circumstances, it is more cooperative of the speaker to disambiguate the topic referent instead of omitting it. If *manka chawpichawmi* were here realized as a single rise-fall with the rise only occurring on the final element, like in the broad focus utterances in *Cuento*, the sequence would be ambiguous. As illustrated in (148) in section 6.2.3, it allows two interpretations, one with *manka* as modifier of *chawpi* ("x is the middle of the pot") and another with it as copular subject ("the pot is in the middle"). I argue that a realization as a single rise-falling phrase with final prominence like in *Cuento* would favour the former reading, or at least not disfavour it, whereas a realization as here, with two separate phrases and the first not metrically subordinate to the second, is more compatible with the latter reading. The latter reading is undoubtedly correct in the context, because the third utterance here (see (185)), *y hawankunachaw* "and below them", with *-n-kuna* marking possession by a plural possessor, clarifies that *chawpi* "the middle" is also to be related to this plural referent, the two other salient entities in the context, the bat and the lightning. This is further confirmed by looking at TP03's map, on which no objects are inside the pot, but where the three objects are located exactly as he describes: pot, bat and lightning form an upside-down triangle, with the pot at the bottom, so that it is both between and below the other two.

Comparing the two utterances produced at 230.5 and 231.8, we have seen that despite initial similarities, they are quite different in their information structure, especially regarding the role of the initial *manka* "the pot". In the first utterance, its presence at the location is part of what is asserted, while in the second, it is topical, forming part of the background but selecting between elements competing in the context. Although this element by itself is realized in both cases with a rising phrase, I argue that the differences in tonal scaling between the elements in the two utterances act together with those in morphosyntax and the presence/absence of the conjunction to signal these differential interpretations.

6.4.2.3 A positional template for information structural roles?

The observations here and on *Cuento* uncontroversially support the hypothesis of right-alignment of a (narrowly) focused constituent with the right boundary of a phrase, specifically a falling one. However, in order to better understand how prosody interacts with word order and especially the role of the separately phrased initial and final constituents I've been calling topics here, I will discuss and build on two proposals in the literature on Quechua about the relative position of elements in a sentence and their information structural roles.

For Huallaga Quechua (also Quechua I), Weber (1989: 427–435) proposes that an information packaging pattern for sentences exists in which zero, one, or several -*qa*-marked constituents are initial, followed by a number (zero also possible) of preverbal constituents plus the verb, of which one is evidential-marked, and then optionally followed by further -*qa*-marked constituents. The function of –*qa* is to mark elements that relate a sentence to its context, and compared to the notion of topic (cf. Weber 1989: chapter 20). The preverbal material including the evidential-marked constituent is broadly linked to focus (Weber 1989: 419, 428–431). Sánchez (2010) builds upon this proposal for southern varieties of Quechua II in a generative framework. In her account, fronted constituents in the "left periphery" of a sentence can be either topical or focal and then require marking with –*qa* or the evidentials, respectively. Postverbal constituents in the "right periphery" can only be topical, never focal, are not always marked with –*qa*, and are sometimes realized with "low pitch and voiceless vowels typically associated with breathy voice" (Sánchez 2010: 38–39). Elements in the left periphery can be contrastive (Sánchez 2010: 29), but not the postverbal topics (Sánchez 2010: 51). However, the elements at the right edge are also said to "disambiguate between potentially competing topics" (Sánchez 2010: 175), relativizing the claim about an absence of contrast. I take the two proposals to capture an important observation about a relation between relative positions and information structural roles, with elements in a central position conveying the "main" or at-issue content, and those in the periphery conveying instructions about how to relate that content to the ongoing discourse, very broadly speaking, and in agreement with other observations that forms encoding more discourse-related meanings seem to occupy peripheral positions in larger units universally (e.g. Zimmermann 2011: 2034 with respect to the clause). Yet our observations suggest that for our data, some of the generalizations do not hold. Sánchez (2010: 12–13) states that canonical word order corresponding to broad focus in Cuzco Quechua is word-final both in transitive and intransitive sentences, yet examples like AZ23_Cuent_Q_0126 ((183), Figure 174) show that in our data, object-final utterances can also be compatible with a broad focus reading and be realized with a contour cueing final prominence. Examples like Figures 178 and 179 also show that initial constituents (fronted according to

Sánchez 2010: 29–31) not marked with *-qa* can still be inferred to be topical from the context. It seems unclear how best to define the units that determine centrality and periphery here.

I would like to reframe the proposals by Weber (1989) and Sánchez (2010) somewhat, as a kind of template that serves as a heuristic default, presumably based on the frequent experience of pairings of constituent order and context, and involved in the generation of expectations, like metrical structure (see sections 3.3.1, 3.7.3). While Weber (1989) and Sánchez (2010) mainly frame their proposals as being about morphological marking and word order within sentences, I suggest that it can apply to utterances or even sequences of turns, instead of individual sentences, and that prosody also contributes to cueing the position of some element relative to this template. That is to say, I propose that this is a structure based on the typical sequencing of information when it is negotiated in interaction (cf. Wiltschko 2021). I do not mean to say at all that the proposals in terms of morphosysntax are wrong, but that prosodic and morphosyntactic cues to discourse and information structure can be complementary or even in conflict with each other and that they all need to be taken into account to really arrive at an apt IS-interpretation. This is based on the broader idea that cues to IS are not categorical, but distributional or probabilistic.

Taking what the two proposals have in common, I add suggestions on how observations about prosody and information structure can be formulated with reference to the template, assumed to have (at least) three positions: two peripheral positions for elements that could be said to be somehow topical (given or active in the context and filling a variable in the referential specification connecting the proposition being asserted to its discourse context), an initial and a final one. Between them is a position on which material is realized that is at-issue. We could also go on and describe another circumferential position, even more initial or final, in which response particles or tags such as *aw* are located, but I will leave this to the side here.[421] I suggest that prosody cues the position of material relative to the template

421 Cf. the proposal of an "interactional spine" by Wiltschko (2021). She proposes that languages have a grammaticalized structure derived from generalizations about how conversations normally proceed, which contains two layers concerned with how information is negotiated in interaction, the responding and the grounding layer, and one layer concerned with the actual informational content, the propositional layer (Wiltschko 2021: 72–92). Her grounding layer can be roughly equated with the two peripheral non-at-issue positions in the template I propose, and the propositional layer with the central at-issue position. The utmost peripheral positions of e.g. response particles and tags would correspond to the interactional layer. Importantly, she also does not assume that the maximal unit for these domains of grammar is the sentence, but that it can extend to turn sequences. The general idea that 'preposed' and 'postposed', i.e. peripheral, elements somehow fulfill more discourse-related functions than their more syntactically integrated counterparts at a in three ways. Firstly, via phrasing: a boundary between two positions in the template is at least a PhP-boundary (but not the other way around). Secondly, via pitch contour choice: material in the initial position is preferentially realized with rising contours, material in the central and final positions with falling contours. Thirdly, via metrical structure as cued by tonal scaling: the middle position has the highest prominence and is scaled highest. That would mean that a reduction in tonal scaling/ downstep on an element cues that it is not in the central position. Applying this template to the utterances analyzed so far predicts that those we have seen from *Cuento* mostly consist of material in the central position, which squares well with them having broad/final focus. The template allows to unify observations about the interaction between word order and syntagmatic uses of prosody, which is absolutely necessary in my opinion given that they are intimately connected via their essential use of sequentiality. An object in syntactically marked postverbal position can nonetheless be part of an assertion with broad focus, as in Figure 174, or narrow final focus, as in the utterance at 38.8 in (184), if it is realized as part of a rising-falling contour with rightmost prominence. At the same time, we can say that constituents like *manka* in Figures 178 and 179 are in a relevant sense postposed, because they are phrased separately and not scaled higher than preceding elements, and that this correlates with them having a function that connects the at-issue proposition with the preceding context, even though there is no verb in these sentences relative to which they could be postverbal. Regarding Figure 176 from *Cuento*, we can now say that *alatsikuykaptin* is realized separately in the initial position, which fits its function of establishing a relation to the preceding discourse context for the proposition that is asserted, similar to the initial topics in the *Maptask* utterances.[422] I also suggest that tonal scaling cues relative prominence also *within* each of these three positions: as seen in *Cuento*, a sequence of phrases in which the last one is scaled highest in both pitch level and span is compatible with rightmost prominence. In contrast, I propose that a sequence of phrases that are all downstepped relative to each other from left to right cues leftmost prominence. Thus the copular verbs in

more central position in an utterance is of course at least as old as Givón (2018 [1979]). Note that Givón (2018 [1979]: 154) also explicitly includes intonation into what cues separated vs. integrated elements.

422 Haiman (1978: 564) argues that such subordinate clauses, especially conditionals, have a very similar meaning to topics, because both are "givens which constitute the frame of reference with respect to which the main clause is either true (if a proposition), or felicitous (if not)". He supports this with evidence from languages in which formal marking for both is identical. Note how Haiman's definition and Parker (1976: 160)'s description that such subordinates are "prerequisites" for the proposition denoted by the main clause are also related to the definition for semantic presuppositions by Bas van Fraasen cited in Karttunen (1973: 169), that "sentence A semantically presupposes another sentence B, just in case B is true whenever A is either true or false".

Figure 181 are part of the central proposition, but cued to be non-at-issue. And this also elucidates the difference in interpretation for the two *mankas* in the two utterances in Figure 181: the first is part of the assertion in central position and cued via scaling to be the most prominent part of that, because it is the only new addition in that assertion. The second, on the other hand, is in initial position and a topical subject. In terms of prosodic structure, the proposed model necessarily assumes a higher-level domain which partially determines the tonal make-up and scaling of the phrases it contains, but for which no separate boundary tones have been found. I suggest therefore that it is simply a maximal level of the recursive PhP/AP.

A template generating expectations as conceived in this way aids interpretation, both when individual sequences conform to it and when they diverge from it. Weber (1989: 428–436) suggests a similar relationship between the relative placement of morphologically marked constituents and information structure, showing that the pattern he has identified is preferential and providing an unmarked default, with deviations from it being interpretable as what he calls “rhetorical devices”. Sánchez (2010: 221–228) also observes that intonational form of both left-peripheral and right-peripheral elements can vary in her data, with some of the pitch tracks of right-peripheral elements looking similar to what I have called downstep here. In her data too, intonational form varies irrespectively of whether a left- or right-peripheral constituent is marked with an evidential or *–qa*, or left unmarked, pointing to an independence between cues, as she notes (Sánchez 2010: 224). If morphology, syntax, and prosody thus each interact independently with such a projected preferential default, then they can also either align or misalign, which might result in subtle interpretable effects, perhaps highlighting different aspects of semantics or information structure. My main point here is that the cues to positions should not be considered in isolation. This leads to problems when only morphosyntax is considered without prosody. But we’ve equally also seen that prosodic cues need to be interpreted together with those from other domains. Potential for ambiguous cueing seems to exist in particular in the two initial positions. In a sentence consisting only of a subject and a verb, in this order, the subject might be cued as focal due to its immediate preverbal position, and it is also potentially topical as an initial constituent. This ambiguity is perhaps already reflected in Sánchez (2010: 29, 36, 38–39)’s observation that left-peripheral elements can be either focal or topical, while postverbal elements are topical.[423] We will encounter similarly ambiguous examples below, especially in section 6.4.3.

423 Cf. the brief discussion about fronted constituents in Spanish in section 3.7.3.1 that also suggested that there is a separately phrased initial position in which both topical and focused material can be found.

6.4.2.4 Third sequence: Cueing against context

In this section, before moving on to data from *Conc*, I will discuss one further sequence from the *Maptask* that demonstrates how additional information that is essential for an appropriate interpretation can arise from a conflict between cues from prosody and morphosyntax, on the one hand, and context, on the other.

(187) TP03 & KP04_MT_Q_2329–2398[424] (context for Figures 181 and 182, directly continued after (186))

time (seconds)	KP04 (*with the path on the map*)	TP03 (*without the path on the map*)
232.9	(LH) (HL) ajá o]k	
233.8	*(0.3)*	
234.1	(L H--------) y atuq-qa and fox-TOP *and the fox*	
235.5		(L H L) !(LH-------------L) !(H---------------L) **atuq-qa hana-chaw-mi uh na-chaw ka-yka-n** fox-TOP above-LOC-ASS PSSP-LOC COP-PROG-3 *the fox is above there*
237.4	(L H------) atuq-yaq m- at- fox-TERM *to the fox*	
238.4	(L H----------L)!(LH L)!(H L) **atuq-chaw-mi usha-n naani** fox-LOC-ASS finish-3 path *it is at the fox that the path ends*	

At 232.9, KP04 accepts TP03's description of the relative location of the items via acknowledgement tokens, and proceeds. He then (234.1) introduces a new referent, the fox, topicalizing it via –*qa*, and asking about it.[425] TP03 responds, saying that the

424 https://osf.io/jg2z8/

425 He produces *atuq* with the topic marker –*qa* and in a rising phrase. We could call this a wh-question without wh-word or polar question without polarity (due to the absence of an inflected verb). But it is perhaps better seen as a question whose type is entirely determined by context. Formally it is comparable to similar questions e.g. in Japanese, such as *go-chuumon-wa*? (HON-order-TOP) 'what is your order?' (in a restaurant). Such questions are also possible in European languages, fox is above (235.5, Figure 181), i.e. above the location they had so far been talking about. He realizes the initial topic, *atuq-qa*, with a separate rise-falling phrase. This is followed by a downstepped rise-falling phrase, realizing the focused constituent *hana-chaw-mi*. The following material, *na-chaw ka-yka-n*, is realised in a single downstepped falling phrase aiding to cue it as backgrounded. Even though KP04's question is very broad, TP03 seems to have interpreted it to convey a QUD like WHERE IS THE FOX? and not e.g. WHAT DOES THE FOX WANT FOR DINNER?, since in the *Maptask*, only the first question is relevant for the overall discourse goal (Roberts 2012a) of following the correct path to the end. A possible explanation for why TP03 realizes *atuq-qa* in initial position here, instead of entirely omitting it (permissible since it is part of the question) or realizing it in final position, as done with *manka* earlier in 219.5, involves selection between active referents. Since the directly preceding discourse had been about the relative locations of pot, bat, and lightning, it is cooperative to repeat the somewhat suddenly introduced referent "fox" at the beginning here to avoid confusion.

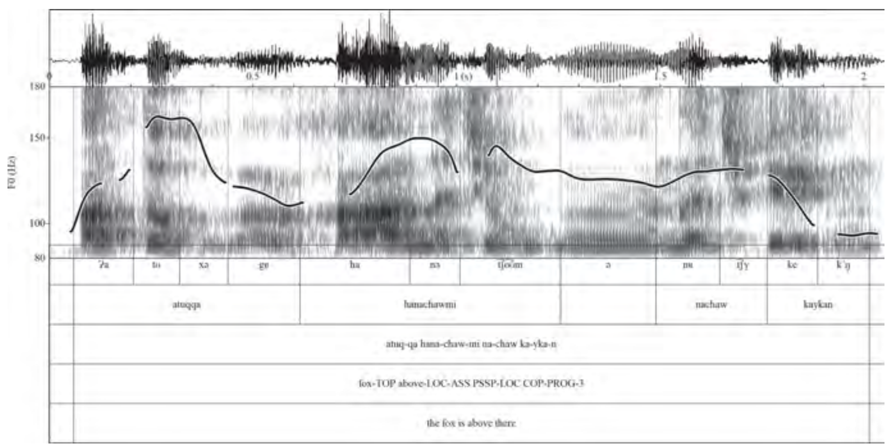


Figure 181: TP03_MT_Q_2355.[426] Cf. (187) for the context.

It is also likely that for similar reasons, TP03 understands KP04's question to imply that the fox should be at the same location as those three, whereas for TP03, it is in a separate location. Realizing *atuq-qa* in initial position and downstepping the

like "and your bike?", "y el zorro?", or "und die Wasserwage?". In our Quechua data, unlike neutral polar and wh-questions, these types of question are probably most often realized with a rise. Cf. also note 48 in section 3.6.1.

426 https://osf.io/7g3vk/

phrase with the focused constituent might be a way of indicating the presence of this perceived expectational contrast in the context.[427]

KP04's response at 233.5 provides further evidence that TP03's utterance in Figure 181 implies a perceived lack of relevance of KP04's question about the fox (234.1) to the previous discourse. It is an assertion that can be understood as justification for bringing up the fox at all: *atuq-chaw-mi usha-n naani* "it is at the fox that the path ends" (238.4, Figure 182). This utterance is remarkable considering the context: from the preceding discourse the fox is active as topic, it was marked so in both KP04's question (234.1) and TP03's response to it (235.5). Applying the criteria developed by Riester and colleagues, the implicit QUD here has to be WHAT HAPPENS AT THE FOX?, so that the fox is backgrounded in the corresponding assertion. An utterance conveying the proposition "the path ends at the fox" at this point would therefore be expected to be structured in such a way as to maintain the fox as topic, and the predicate "the path ends" as the at-issue content. From what we know, such an utterance might for instance be *naani ushan atuqchawqa*, with *atuqchawqa* downstepped; or *atuqchawqa ushan naani,* with it realized as a rise, and *ushan naani* in a falling phrase signaling its status as at-issue content; or just *ushan naani*, with the fox omitted as topical subject. Yet none of this is the case.

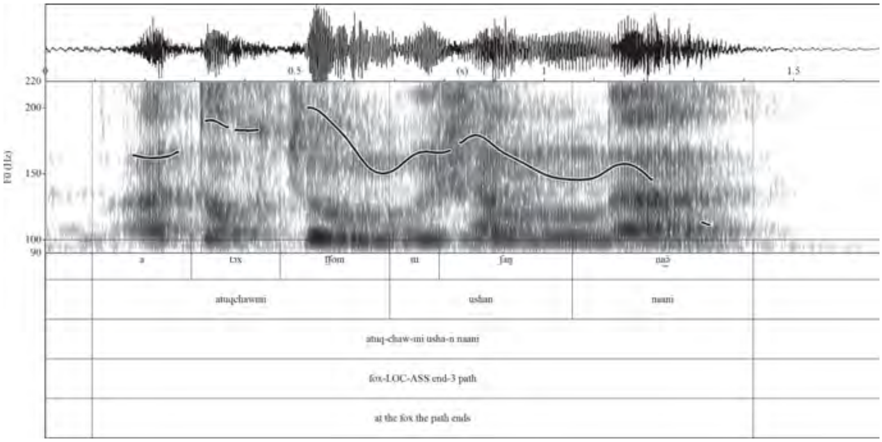


Figure 182: KP04_MT_Q_2384.[428] Cf. (187) for the context.

427 It is tempting to also attribute the substantial rise on the lengthened initial syllable of the focused constituent *hana-chaw-mi* to this contrast, in the manner of iconic prosody like "it's faaar". I can only note this auditory impression by a listener whose hearing conventions are shaped by European languages. Should it turn out to be a real phenomenon in future research, it would be important to investigate what anchors the position of the rise.
428 https://osf.io/xn256/

The constituent denoting the fox, *atuqchawmi*, is realized initially, *–mi*-marked with a rise-falling phrase, cueing it to be focused as answer to the QUD. *Ushan naani* is realized with two downstepped falling phrases, i.e. in a position and with prosody so far seen as cueing backgrounded information. Thus the information structure inferred and expected from the context and that signaled by the utterance itself are in conflict here. This is summarized in (188).

(188) Comparison of information structure of KP04_MT_Q_2384 as inferred from context and as likely cued by morphosyntax and prosody

Inferred from context:	[*not at-issue;* topic]	[*at-issue;* comment]	
	atuq-chaw-mi	ushan	naani
Signaled by scaling relations:	(x		)
	(x)	(x	)
Signaled by contour shape:	[closed]	[closed]	[closed]
Cued by morphology:	[focus]	[	background]
Cued by word order relative to verb:	[focus]	[	background]
	[	comment]	[topic]

From (188), it seems that formal marking suggests a QUD like WHERE DOES THE PATH END?. However, if that were really the implicit QUD, the utterance should be expected to be simply *atuqchawmi*, which would not convey any new information in this context. As in previous examples and according to the proposed template, the fact that constituent(s) that could be omitted are instead realized as downstepped in the postfocal final position seems to indicate that these elements are needed to relate the proposition conveyed by the utterance it to its context, or that it should be understood as answering QUD that is different from the one inferrable from the context. So, if formal marking here aligns to signal a QUD like WHERE DOES THE PATH END?, in spite of this going against the preceding context progression, this gives rise to an implicature that the utterance is relevant with this information structure, even though it does not seem to be so at first sight. I argue that the missing puzzle piece for understanding this lies in the fact that the information where the path ends is necessary for the completion of the game. Without knowing this information, the overall discourse goal cannot be achieved. Thus, WHERE DOES THE PATH END? is a relevant QUD because responses to it constitute a partial answer to the superordinate QUD of the game, WHAT IS THE CORRECT PATH?. Because answering this question is thus by definition relevant for the game, answering it with “at the fox” in turn gives an explanation for why KP04 had brought the fox up at 234.1, and which, as argued, TP03 had implicitly questioned the relevance of at 235.5. It is because the utterance, in utilizing both prosodic and morphosyntactic means against the expectations built up by the context, manages to answer both WHAT

ABOUT THE FOX? and WHERE DOES THE PATH END?, that I have translated it here with an it-cleft, “it is at the fox that the path ends”. In my understanding, such a structure can also be used to express such a complex information structural configuration, in which the fox is both contrastively specified as answer to WHERE DOES THE PATH END? and the topic about which the proposition is predicated.

This section has provided evidence that utterances which from the context can be inferred to have a more complex internal information structure than the broad focus observed in the previous section in *Cuento* do reflect this increased complexity in their prosodic realization, although no simple one-to-one correspondence between prosodic marking and information structural categories can be assumed. I proposed a heuristic tripartite positional template that prosodic cues as well as cues from other domains can refer to, in which material realized to the left of a first relevant prosodic boundary and with a rising contour is linked to a topical interpretation, as is material to the right of a second relevant prosodic boundary. Material in between these two boundaries, realized with a falling contour, is linked to an at-issue interpretation. A reduction in tonal scaling between two adjacent phrases (downstep) has been taken to cue that the second phrase is less prominent than the first one, in opposition to the increase in tonal scaling between units we saw in the previous in *Cuento,* cueing rightmost prominence compatible with broad focus. In these data from *Maptask*, we also continued to observe how prosodic and morphosyntactic cues interact in the signaling of information structure. Mostly, they could be seen to plausibilize an interpretation that is also inferrable from the context. In the final example we also saw how they aligned to contradict the interpretation from the preceding context. It was argued that plausible reasons for this can be found when more than the immediate discourse context is considered. In the final section, utterances with conflicting cues will be examined.

6.4.3 Variability and conflicting cues in Conc

This section will explore further how cues to information structure from prosody, morphology, and word order can interact using data from *Conc*, mainly by ZR29 & HA30, but also in comparison with other speakers. I will begin by recapitulating what we know so far about how word order is related to signaling information structure. In the *Maptask* examples, it was often the case that the constituent directly in preverbal position was (narrowly) focused, which would align with observations on a number of verb-final languages (cf. Kügler & Calhoun 2020: 465). For southern Quechua II varieties, Muysken (1995: 383); Sánchez (2010: 36–37) assert that postverbal elements cannot be focused or marked by one of the evidentials. Sánchez (2010: 12–13, 37, note 10) further claims that the canonical word order is verb-final, and that this is linked to broad focus. According to her, other word orders signal a different information structure. In *Cuento*, utterances with the "non-canonical" order V-O have also been seen to be clearly interpretable as having broad focus from both context and prosody (cf. Figure 174 and (183)). For Huallaga Quechua, which as a variety of Quechua I is more closely related to Huari Quechua, Weber (1989: 15–16) points out that even though it shares many properties typically claimed for SOV-languages (postpositions instead of prepositions, modifiers before their heads, main verbs before auxiliaries, possessors before possessed, cf. Greenberg 1966), main constituent word order is actually not so much one of them: in his count, only slightly less than half of all sentences realizing all three constituents (48 of 99 out of a total of 1309 observed sentences, 714 of which transitive) actually did so in SOV order. He does however also connect the preverbal material to a notion of focus (Weber 1989: 419, 428–431). All three accounts agree that word order is sensitive to information structure, and that there is at least a preference for focal elements to be realized preverbally. In our data, an evidential-marked constituent does not occur postverbally in all seven *Conc*, *Maptask*, and *Cuento* corpora.[429] However, there is evidence that a postverbal constituent can occur in contexts in which it is the answer to the QUD, i.e. focused, and also contrasted, as in some of the following examples.

6.4.3.1 Cues to information structure in five examples from ZR29 & HA30's Conc

I will first introduce examples from *Conc* and then expand the discussion to cues conflicting with each other, and with IS as constructed from the context of the respective examples. To my knowledge, while multiple cues also conflicting with each other have e.g. been discussed in Baumann & Grice (2006); Venditti et al. (2008); Kaiser (2016); Edeleva et al. (2020), among others, this is the first attempt for Quechua to show that there are cases in wich cues conflict both with each other and with expectations from the discourse context. The examples are HA30_Conc_Q_1482 ((189)/Figure 183) and 1576 ((190)/Figure 184), LD20_Conc_Q_1150 ((191)/Figure 185), and HA30_Conc_Q_1836 ((192)/Figure 186).

(189) HA30_Conc_Q_1482

na	washa	chawpi-chaw	ka-yka-n	este	tsuku
PSSP	DEM.DIST	middle-LOC	COP-PROG-3	uhm	hat

"in the middle of that over there there is the hat"

429 Muntendam (2010: 117, 126) reports –*mi*-marked postverbal constituents for an unidentified variety of Quechua not studied by herself, but does not find them in her own data from Ecuadorian and Bolivian varieties.

(190) HA30_Conc_Q_1576

hana	kantu-chaw	ka-yka-n	arash
above	edge-LOC	COP-PROG-3	lizard

"at the border above there is the lizard"

(191) ZR29_Conc_Q_1150

kay	ultimu-chaw-mi	ka-n	pitu
DEM.PROX	last-LOC-ASS	COP-3	whistle

"in the last one here there is the whistle"

(192) HA30_Conc_Q_1836

arash	ka-yka-n	tsay	chawpi-chaw
lizard	COP-PROG-3	DEM.PROX	middle-LOC

"the lizard is in the middle there"

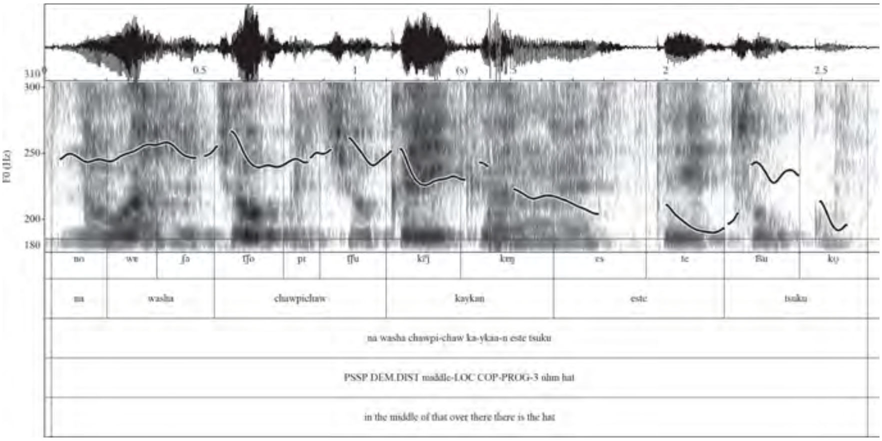


Figure 183: HA30_Conc_Q_1482.[430] Cf. (195) for the context.

All these examples have non-verb-final word order: a form of the copula is surrounded on one side by a noun specifying the location at which an item in the game is to be found, and on the other by another noun specifying the item. As explained in section 6.2.2, speaker pairs mostly follow one of two strategies in *Conc* consistently, either to realize the location-specifying expression before the item-specifying one (pattern a), or the other way around (pattern b). Correspondingly, in three of the four examples they first specify locations and then the items to be found at

430 https://osf.io/4ahqe/

them. HA30_Conc_Q_1836 ((192)/Figure 186) is the exception which reverses the order of the nouns. Section 6.1.1 had an example from this *Conc* (repeated here as (193)) with verb-final word order: location – game item – copula.

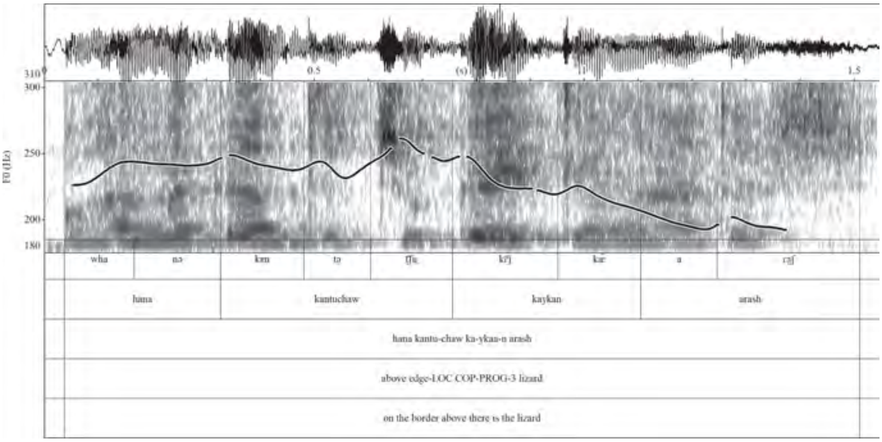


Figure 184: HA30_Conc_Q_1576.[431] Cf. (195) for the context.

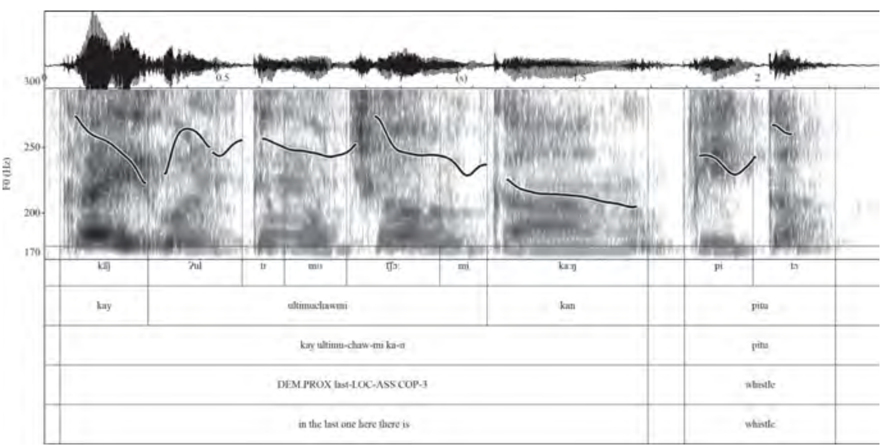


Figure 185: ZR29_Conc_Q_1150.[432]

431 https://osf.io/w5kzy/
432 https://osf.io/bhqxy/

(193) HA30_Conc_Q_0211 (repeated from (108)/Figure 125, see also (125))

washa	kantu-chaw	ukush-mi	ka-ykaa-n
DEM.DIST	edge-LOC	mouse-ASS	COP-PROG-3

"at that border there is a mouse"

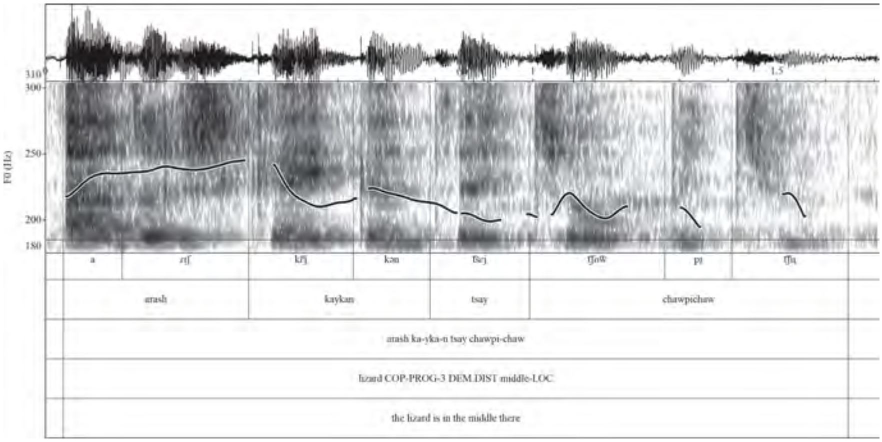


Figure 186: HA30_Conc_Q_1836.[433]

In this example, the location is realized with a rise, and then the -*mi*-marked item-specifying expression plus the locative copula follow in a falling phrase. Word order, morphosyntactic marking and prosody here all act just as the context lets us expect: a location-specifying referent is called up as a topic, and then a predication is made about that location regarding which item is to be found there in a way that also conforms to the predictions made about word order and –*mi*-marking on other Quechua varieties (see above). In the four other examples though, things are different. Word order here is not verbfinal as in HA30_Conc_Q_0211. From the predictions about the relation between word order and information structure by Muysken (1995); Sánchez (2010), and Weber (1989), as well as partially from the analysis of the *Maptask* examples, the preverbal constituent (location-specifying in (189)/Figure 183, (190)/Figure 184, (191)/Figure 185, item-specifying in (192)) would be expected to be focal. It should be -*mi*-marked and realized with a falling phrase, so that it is clearly cued as the at-issue component of the assertion. The postverbal constituent would be expected to be downstepped, cued to be interpreted as post-

433 https://osf.io/28nq9/

focal/non-at-issue. At most it might be a resumptive topic, but not denote a newly introduced referent whose specification is a crucial part of what is asserted. All of these assumptions do not hold.

Instead, the initial constituent is only once realized with *–mi* (in (191)/Figure 185), and also realized with a rise instead of a fall in (190) and (192) (probably also in (189)/Figure 183, but there is interference from consonants). The postverbal constituent is realized as downstepped only in (190) and (192). In (191), it is realized with a rise, and in (189)/Figure 183 with a rise-fall whose high tonal target is just as high as preceding ones, while it is supposedly backgrounded in all of them. The copula however is downstepped or low in all five utterances.

An analysis of their contexts does not fully mirror the divisions made by these formal cues. Instead, context can best be said to group (189)/Figure 183, (191), and (193)/Figure 125 together, with (190) and (192) separate. In particular, (189) and (191) are very similar contextually. In both of them, the item-specifying expression (*tsuku* "hat", and *pitu* "flute", respectively) denotes an item mentioned for the first time in the game. They are first turns after a speaker change, and the location-specifying expression also denotes a new referent. In (193), the location-specifying expression *washa kantuchaw* is repeated from the directly preceding turn at 10.3 consisting only of that phrase (cf. (194)). The item-specifying expression *ukush-mi* "mouse" is lexically given, because HA30 begins by making an enumeration of all the items that are there.

(194) Beginning of ZR29 & HA30_Conc_Q[434]

time (seconds)	HA30
0.8	(L H L-------) tsay-chaw-mi ka-n DEM.DIST-LOC-ASS COP-3 *there is*
2.5	(L H) ukush *mouse*
4.2	(H L) tsuklla *hut*

434 https://osf.io/z6akq/

5.2 (H L)
uusha
sheep

[moderator intervenes]

10.3 (H----- L--------------)
washa kantu-chaw
DEM.DIST edge-LOC
at that edge there

The game moderator intervenes after the first three items, explaining how the game is played, and HA30 continues with the turn at 10.3, leading to the one at 21.1, which is (193). Thus, *ukush* has been mentioned previously in (193), but it is not unambiguously the most salient referent in the discourse. From the context, it makes no sense to interpret the item-specifying constituents as backgrounded in either of (189)/Figure 183, (191), or (193)/Figure 125. The opposite is the case: specifying them means selecting from a limited set of alternatives in all cases (implying a degree of contrast), i.e. all the other available items in the game (twelve in total), or in the case of (193), perhaps from the set of the three mentioned initially. This selection is new information in all three. In (189) and (191), the item-specifying constituent itself is totally new. So it is clearly at-issue in all three examples. Yet it is realized postverbally with N-V-N order in (189) and (191), and preverbally with N-N-V order in (92). If any contextual differences are to be invoked to explain this formal variation, they must be the subtle ones discussed here: the possible differences in the degree of activation of the item referent (*ukush* being more active in (92) than *tsuku* and *pitu* in (189) and (191)), that *ukush* is perhaps selected from a smaller set of referents, and the fact that (92) is directly preceded by a turn that already realizes the location once. But the difference cannot be one in terms of the overall at-issue partition of the utterances, since the item-specifying expression is in focus in all three. Note that the postverbal item-specifying expression is not downstepped in (189) and (191).

The situation is different again in (190). This is the fourth attempt in this *Conc* to correctly guess the location of the lizard (*arash*). In a previous turn, ZR29's at 135.2, its location had last been wrongly guessed (cf. (195)), so that the utterance is perhaps even a kind of reversal/correction of that move.

(195) ZR29 & HA30_Conc_Q_1352–1591[435] (context for (189)/Figure 183 and (190)/ Figure 184)

time (seconds)	HA30	ZR29
135.2		(L H--) (HL-------------------)(L H) kay pitu ladu-n-chaw-na-mi arash DEM.PROX flute side-3-LOC-DISCONT-ASS lizard *beside that flute now there is the lizard*
137.9	*[short discussion between moderator and ZR29]*	
148.2	(L H---------------------------)!(H L)(L H L) **na washa chawpi-chaw ka-yka-n este tsuku** PSSP DEM.DIST middle-LOC COP-PROG-3 uhm hat *there in the middle is the hat*	
151.7	(H----------L---) chawpi-chaw middle-LOC *in the middle*	
156.4	(L H---) y hana *and above*	
157.6	(LH----------------------)!(H L-----------) **hana kantu-chaw ka-yka-n arash** above edge-LOC COP-PROG-3 lizard *at the edge above there is the lizard*	

Thus, *arash* is here both referentially and lexically given, and determining its location is a salient issue in the discourse. This makes it more likely that the utterance has an implicit QUD like WHERE IS THE LIZARD?, as opposed to the previous examples, which answer a QUD like AT [LOCATION], WHAT IS THERE?. This view is supported by the fact that *kaykan arash* is very clearly downstepped and *arash* possibly even part of the phrase-final low,[436] cueing it as backgrounded. However, the preverbal location-specifying constituent is neither *–mi*-marked nor realized with a falling contour, which goes against what we might expect with the QUD WHERE IS

435 https://osf.io/3r5x8/

436 Here it is plausible to see *arash* as not contrastive in the sense of not having any salient alternatives, because it is part of the background of the QUD. This shows again that word order does not suffice as cue, because we just saw that postverbal elements can be focal and contrastive (*pace* Sánchez 2010: 51). Arguably it needs the combination of being postverbal and scaled down to cue backgrounding (and a realization in final position in the proposed template) here.

THE LIZARD?. The case for an implicit QUD in which the item is given instead of the location is even stronger for (192). The move by another speaker directly preceding it in the game, by ZR29 at 169.7, is *mas ladunchaw arash* “further to its side is the lizard”, so that (192) can be seen as a kind of retort,[437] asserting that the item is at a different location. This interpretation is supported by the fact that (91) is the only utterance in the game by these two speakers in which the item-specifying expression *arash* is initial and preverbal, and the location-specifying expression is postverbal. Furthermore, it is also realized with a rise, supporting its status as a topic. Here, word order and prosody thus seem to align in cueing the item as topical. However, both the verb and the postverbal location-specifying expression are also clearly downstepped and thus not prominent despite being at-issue from the context. In both (190) and (91), the cues that signal the referent contextually determined to be non-at-issue as such seem stronger than any that cue the reverse, i.e. signaling the other referent as at-issue. In all of these examples, with the exception of (193)/Figure 125, we see conflicting cues. (196) summarizes these cues to IS from context, word order, morphological marking, and prosody in the five utterances.

(196) Conflicting cues to information structure from context, word order, morphology, and prosody in five utterances from *Conc* by ZR29 and HA30

IS via context	[topic]		[focus]
HA30_Conc_Q_1482 (=(189))	**na washa chawpichaw**	**kaykan**	**este tsuku**
order of elements	location-specifying expr.	verb	item-specifying expr.
IS via word order			
relative to verb	[**preverbal / focus**]		[**postverbal/background**]
overall	[topic]	[	focus]
IS via morphology	--	--	--
IS via prosody			
Contour shape	[rise/open(?)]	[fall/closed]	[rise-fall/closed]
Scaling		[low/downstepped]	
Metrical structure	(x	)	(x)
IS via context	[focus]		[topic]
HA30_Conc_Q_1576 (=(190))	**hana kantuchaw**	**kaykan**	**arash**

437 It is not a direct response to ZR29’s move because that was followed by the game moderator turning the card over and showing that her guess was wrong.

order of elements	location-specifying expr.	verb	item-specifying expr.
IS via word order			
relative to verb	[preverbal / focus]		[postverbal/background]
overall	[topic]	[	focus]
IS via morphology	--	--	--
IS via prosody			
Contour shape	[rise/open]	[fall/closed]	[fall/closed]
Scaling		[low/downstepped]	[low/downstepped]
Metrical structure	(x		)
	(x)	(x	)
IS via context	[topic]		[focus]
LD20_Conc_Q_1150 (= (191))	**kay ultimuchawmi**	**kan**	**pitu**
order of elements	location-specifying expr.	verb	item-specifying expr.
IS via word order			
relative to verb	[preverbal / focus]		[postverbal/background]
overall	[topic]	[	focus]
IS via morphology	[*-mi-marked/focus*]		
IS via prosody			
Contour shape	[	rise-fall/closed]	[rise/open]
Scaling		[low/downstepped]	
Metrical structure	(x	)	(x)
IS via context	[topic]		[focus]
HA30_Conc_Q_1836 (=(192))	**arash**	**kaykan**	**tsay chawpichaw**
order of elements	item-specifying expr.	verb	location-spec. expr.
IS via word order			
relative to verb	[preverbal / focus]		[postverbal/background]
overall	[topic]	[	focus]
IS via morphology	--	--	--
IS via prosody			
Contour shape	[rise/open]	[	fall/closed]
Scaling		[	low/downstepped]
Metrical structure	(x		)

IS via context	[$_{\text{topic}}$]	[$_{\text{focus}}$]	
HA30_Conc_Q_0211 (= (193)/Figure 125)	**washa kantuchaw**	**ukushmi**	**kaykan**
order of elements	location-specifying expr.	item-specifying expr.	verb
IS via word order			
relative to verb	[	$_{\text{preverbal / focus}}$]	
overall	[$_{\text{topic}}$]	[	$_{\text{focus}}$]
IS via morphology	[	$_{\textit{-mi-marked/focus}}$]	
IS via prosody			
Contour shape	[$_{\text{rise/open}}$]	[	$_{\text{fall/closed}}$]
Scaling			[$_{\text{low/downstepped}}$]
Metrical structure	(x)	(x	)

Word order is given as cueing IS in two different ways in (196). As argued in section 6.2.2, the game context likely induces an overall bias to treat the locations as more active/accessible in the discourse. That makes them more likely to be topics about which predications are then made in which the items are focal. This is what plays out in their overall word order, in which the location-specifying expression precedes the item-specifying expression in four out of the five cases here, with (192) being the only exception to that in the entire game. As argued, (192) follows the order topic-comment, but with the item exceptionally being the topic and initial. For (190)/Figure 184, I also argue that the item should be interpreted as more topical or backgrounded than in the other three examples, but here, word order does not reflect that, instead being the same as in (189) and (191), so that the postverbal item is once backgrounded and twice focal. On the other hand, word order relative to the copular verb[438] cues preverbal focus and postverbal background, according to the predictions made for other varieties and to what we have seen in the previous section from the *Maptask* examples. These two different ways in which word order can be seen to cue IS are given as “overall” and “relative to verb”, respectively. This means that in all examples except (193)/Figure 125, overall word order between location-specifying expression and item-specifying expression, and word order relative to the verb, are in conflict with regards to their cues to information structure. Overall word order aligns with context in almost all cases, except (190)/Figure 184. Verb-medial utterances with only one constituent to the left and the right of the verb, respectively,[439] are an inherently

438 Most other utterances from the *Conc* data by all speakers are verbless.
439 This is the preferred realization by HA30 in this corpus, see Table 49 below.

ambiguous case: according to the position template proposed in the previous section, in such cases, the initial constituent will tend towards a topical interpretation, but it will also be likely to receive a focal interpretation because it is directly preverbal.[440] In terms of prosody, three different cues are also distinguished. In (190)/Figure 184, (192), and (193)/Figure 125, the rise on the initial element, cueing openness, aligns with the topical interpretation from context and overall word order. In both (193) and (189), the non-downstepped rise-fall on the item-specifying constituent cues completion, supporting an interpretation as focal comment, but its realization with a rising phrase in (191)/Figure 185 does so less. However, that the postverbal item-specifying constituent is not downstepped in (189) and (90) but in (89) supports the contextual interpretation that it is focal in the two former, but backgrounded in the latter. In both (192) and (89), that the item-specifying constituent is realized as downstepped aligns with the contextual interpretation of it being backgrounded and non-contrastive. None of the utterances unambiguously cue final main prominence. The downstepped verb in all utterances suggests that the preverbal element is at least locally more prominent. The preverbal element is also globally most prominent in the verb-medial utterances (190) and (192), because the postverbal element is also downstepped. However, in (189) and (191), the downstepped verb is surrounded by equally scaled phrases. The phrases realizing the location- and item-specifying expressions are thus relatively balanced in prominence cues, with a rightmost default acting in favour of the final element. In (191), the realization as a falling phrase and the marking with *–mi* on the location-specifying constituent support its interpretation as at-issue, but this clashes with the rising realization of the item-specifying constituent, which makes it difficult to interpret it as backgrounded.[441] Overall however, scaling-based cues seem to align more clearly with the contextual interpretation than contour-based ones, and word order cues seem to also need scaling cues in order to be interpretable for an IS-partition into at-issue- and non-at-issue-material.

440 This analysis is also applied to KP04_MT_Q_2384 ((188)/Figure 182) in the previous section. According to Sánchez (2010), left-peripheral elements can also be either focal or topical in southern varieties of Quechua. The conundrum is rather old: with two proposed principles of pregrammaticalized linearization, the fronting of unexpected information and that of important information, Givón (2018 [1979]: 180, 217, 221) plausibilizes the initial realization of both new topics ("L-dislocation"), and of foci (in clefts).

441 Several other utterances by ZR29 in *Conc* realize the final item-specifying expression as rising. I suggested in section 6.2.2 that this might be part of a stance of uncertainty projected by this speaker.

6.4.3.2 Comparison with further examples

The findings summarised in (196) provide further evidence that prosody, word order and morphosyntactic marking can cue information structure both in conflict with each other and against what is expected from context via QUD analysis. This suggests that they do not cue it directly, and instead do so indirectly, and/or probabilistically. For prosody, this insight is represented in (196), where I take contour shape (rising or falling) and tonal scaling (realization as low or downstepped) to cue openness/closedness (see section 6.2.2) and phrasal prominence, respectively. We have seen in section 6.2.3.3 as well as here that both information status (given-new) and information structure (focus-background, topic-comment) affect prominence, but that prosody does not cue these categories directly. For word order and morphological marking, (196) represents them as cueing information structure directly, but this is only out of convenience until it is better known what exactly they contribute to the meaning of an utterance.[442] We saw that word order alone is not a good cue for an IS-interpretation of constituents. In addition, while we can identify dominant strategies where speakers pick out one of the elements (location or item) to be specified first in order to anchor the predication involving the second of them to it (cf. section 6.2.2), suggesting a kind of topic-comment structure, the nature of what has been called a topic here seems different from that in *Maptask* and also *Cuento*. In those tasks, a topic is often continued across several utterances. In *Conc*, on the other hand, the topic constituent (most often the location) is selected from an available and finite set anew (almost) each time, arguably creating a contrast relative to the other members of this set, and it therefore needs to be just as explicitly specified as the element which is then put into relation with it.[443] In this sense, the two nominal elements here are more similar to each other than was the case for what was discussed as topics and foci in *Maptask* and *Cuento*, which might go some way in explaining the conflicting cues from prosody and morphosyntax.[444] Note that

442 The paradigmatic meaning of the evidentials is of course subject of a number of works. I refer here to their syntagmatic meaning, i.e. what is cued by their placement at some position in an utterance, and not another.

443 The beginning of the game (194) shows that also in *Conc*, a topical location can be omitted in subsequent turns once activated, as HA30 does with *tsaychawmi* 'in there'.

444 I think that introducing some of the terminology specific to copular sentences, like *predicational*, *specificational*, *identificational* (cf. Dikken 2006: 298–300) does not help the issue at this point. On the one hand, the different types of copular sentences are supposed to differ regarding their IS: in specificational sentences, the pre-copular NP cannot be in focus, while this is not the case for predicational sentences (Dikken 2006: 301; Martinović 2013: 137). However, the location- and the item-specifying expressions in *Conc* are arguably both referential, making these copular sentences identificational (Martinović 2013: 139) rather than one of the two other types, which does not help for predicting their information structure. Dikken (2006: 300) also points out that the “topic marker” –*qa* does not occur once in all of the seven *Conc* games analysed here. Whatever topic-related meaning *-qa* signals, there seems to be no need for it in these discourses (it does occur in *Maptask* also in copular sentences, e.g. Figure 181). It seems that the two speakers here can choose varying expressional strategies to express these contextual conditions. Recall again from section 6.2.2 that the majority of speaker pairs, including ZR29 & HA30, follow the order location – item (pattern a) in *Conc*, while one speaker pair out of seven (XU31 & OA32) exclusively prefer the reverse order (pattern b). These are thus two diverging default strategies of ordering the elements, perhaps reflecting two different information structural conceptualizations, with one treating the location as more topical, and the item as more focal, and vice versa. Here we saw that the one time the less frequent order (b) occurs in this *Conc* ((192)), it cooccurs with an information structure that is also other than expected, and probably aids in cueing this. This expectation is shaped by a global distribution, i.e. that the location – item pattern is more frequent across all speakers, but also by the local expectation established in the game between these two speakers. In XU31 & OA32’s *Conc*, the same order of elements cannot be marked in the same way, because it is the exceptionless default in their game (even though it is globally marked across all *Conc* games). This suggests that locally and temporally established conventions between interlocutors are relevant for the analysis of meaning in natural conversation. While ZR29 & HA30 are with the majority in terms of order of elements, suggesting a broadly topical interpretation for the location-specifying expression, prosody in terms of contour shape paints a different picture. As seen from Table 29, five of the seven speaker pairs realize a clear majority of the first elements (the item-specifying expression for XU31 & OA32, the location-specifying expression for all others) with a finally rising contour, and the second element, correspondingly, with a finally falling contour. They thus align the information structural cues from overall word order with those from contour shape. ZR29 & HA30 are the only ones who clearly go against this trend. Table 49 summarizes all of their moves from *Conc*, sorting them according to

copular sentences can be ambiguous between the three readings “on paper”, which seems to mean ‘without context and intonation’, while Martinović (2013: 149) additionally claims that her analysis on data from Wolof puts the claimed information-structural restrictions on specificational sentences, based as usual mostly on European languages, in doubt. I suggest that at this point it is still too early to say anything about what types of sentences the ones here from Huari Quechua are, in which definiteness is not marked and in which word order in copular sentences is variable, as we have seen. However, I would hope that the analysis presented here, aiming to relate formal marking to contexts, can be used to aid such further in-depth studies in the future.

order of elements, *-mi*-marking, and prosody.[445] Their prosodic behaviour against the overall trend shows there as well.

Table 49: Moves in *ZR29&HA30_Conc_Q* sorted according to the order of elements (location-specifying expression, item-specifying expression, verb), marking with *–mi*, final contour on each element, and whether they are downstepped / realized as low.

order of elements	**number of occ.**	**loc *–mi***	**item *–mi***	**loc rise**	**loc fall**	**item rise**	**item fall**	**item flat**	**verb rise**	**verb fall**	**loc downstep/low**	**item downstep/low**	**verb downstep/low**	**number loc. exps.**	**number item exps**	**number verbs**
Location	4	–	–	–	4	–	–	–	–	–	–	–	–			
Location-verb	2	–	–	1	1	–	–	–	1	1	–	–	1			
Location-item	5	2	–	1	4	4	1	–	–	–	–	–	–			
Location-item-verb	2	–	1	1	1	–	2	–	1	1	–	–	1			
Location-verb-item	13	6	–	5	8	4	8	1	3	10	–	2	10			
Item-verb-location	1	–	–	–	1	1	–	–	–	1	1	–	1			
Item	4	–	–	–	–	1	2	1	–	–	–	–	–			
All	**31**	**8**	**1**	**8**	**19**	**10**	**13**	**1**	**5**	**13**	**1**	**2**	**13**	**27**	**25**	**18**

The three rightmost columns are concerned with tonal scaling. They count how often the three elements are realized either with a separate phrase, but downstepped, or as only the low part of a contour. Both speakers here show a clear tendency to realize the verb as scaled low (13 of 18 occurrences). The location-specifying (1 of 27) and the item-specifying expression (2 of 25) are rarely downscaled, and only in final position. In the high number of verbal forms (in 18 out of 31 moves),[446] this *Conc* is also an exception: in the *Conc* corpora by the other six speaker pairs, only five verbal forms occur in total, of which only one is a copula used in the same way as here (the others are part of item-specifying expressions). The frequency of *-mi* (9 times in 31 moves) here is also exceptional compared to the *Conc* games by other speaker pairs: *-mi* only occurs five other times, four times in TP03 & KP04's *Conc*, and once in that by XU31 & OA32. In ZR29 & HA30's game it mostly occurs

445 The table also includes sequences that are not full guessing moves in the game because they do not specify both location and item. They are either specifications of only the item or reformulations after a guess (the "Item" row), or incomplete guessing moves ("Location" and "Location-verb") that are then followed by full. Table 29 did not include any of those, which is why total numbers differ between the two tables.

446 It is mostly HA30 who uses verbal forms at all.

on an element realized with a falling contour, but not always, as Figure 187 shows. There it occurs on a location-specifying expression realized with a rising contour.

While the generalization that *–m(i)*-marked elements do not occur postverbally (Muysken 1995: 387) seems to hold here, utterances with *–m(i)* do not require the presence of a tensed verb in our data, as Table 49 and utterances like (197) demonstrate (*pace* Muysken 1995: 385; Sánchez 2010: 49).

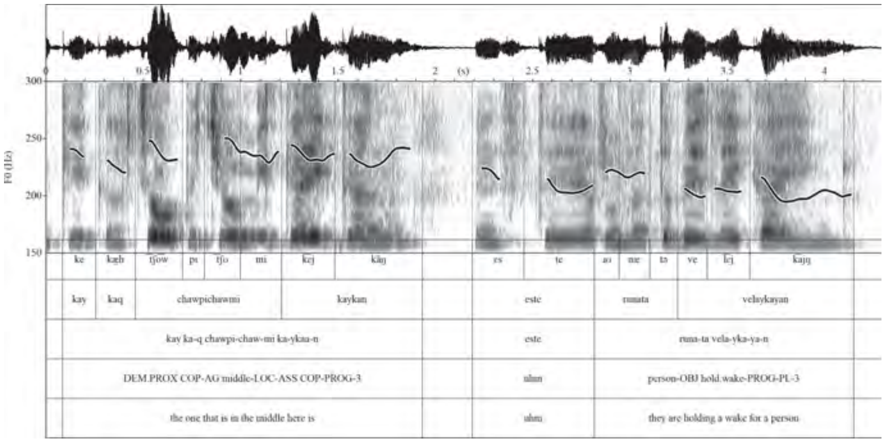


Figure 187: HA30_Conc_Q_0453[447] (voicing threshold 0.9).

(197) KP04_Conc_Q_1653

tsay	keeda-q	anka-m
DEM.DIST	stay-AG	eagle-ASS

“the one that’s left [is the] eagle”

A question for future research is whether the three exceptional behaviours in ZR29 & HA30’s *Conc* (the two speakers are close friends) are related: the use of the copula, the use of *–mi*, and the preference for falling contours in initial position and rising contours in final position. Partly, they seem to constitute individual formal strategies for cueing information structures that are particularly sensitive to the similarity in status of the two elements denoted by the location-specifying and item-specifying expressions in *Conc*, comparable perhaps to what KP04_MT_Q_2834 achieves with the conflicting cues (188) as discussed in the previous section. It also seems plausible that these behaviours are partly the result of their individual and shared multilingual acquisition histories and stylistic stances.

447 https://osf.io/6jfz4/

6.4.4 Conclusions

We have now seen how Quechua prosody and morphosyntax are used to structure utterances in more complex contexts than in those compatible with broad focus in *Cuento*. In the *Maptask*, word order, morphological marking and prosody were seen to mostly align in cueing information structural roles for constituents, relating them in a complex fashion to the context. Yet the *Maptask* and especially the *Conc* data showed that the cues can also not align, leading to ambiguous interpretational possibilities likely pointing to a more differentiated picture than just a separation between focus-background and topic-comment. Possibly, information structure in terms of these two dimensions is best thought of as an interpretational condensation, emerging from heterogeneous cues that contribute other information as well.[448] Regarding prosody, the observations here are difficult to reconcile with a view that aspects of it signal either of these dimensions directly: instead, I have argued that it makes sense to see prosody as signaling 1) whether units are together or apart at different levels of structure (via phrasing and scaling), 2) cueing the relative prominence at each of these levels (via scaling both within and across phrases) and 3) whether a unit should be understood as complete or incomplete (via tonal contour choice). None of this signals focus or topic, but assuming that competent speakers have expectations about how prosodic structures are built, they can exploit them to direct the interpretation of equally competent listeners towards intended information structures. Word order and morphology, which were treated more peripherally here, clearly also contribute to this complex interpretation. They are constrained by their own structures and also build up expectations based on them. Here they have been seen to not straightforwardly signal focus or topic, either. As amply seen, word order is only a good cue for information structure when combined with scaling-based cues. Regarding the evidentials, the prediction that only preverbal elements can be marked with them seems to hold across varieties of Quechua, but that means neither that focus is always preverbal nor always marked by an evidential, nor even that only a constituent that could be marked by an evidential can be interpreted as focal in context (postverbal elements are not -*mi*-marked but were shown to be focal). The constraint against postverbal evidential-marked elements seems

448 Matić & Wedgwood 2013 express the more radical view that focus exists only as an interpretational category for linguistic analysis and has no reality in actual language, especially crosslinguistically. I am thinking more along the lines of what Gunlogson (2001, 2008) has shown for declarative questions in English, in which the cues for polar questionhood from word order (subject-verb inversion) and prosody (final rise) productively misalign for the expression of a speech act category that has more marked contextual specifications than either assertions or neutral polar questions (cf. also section 5.1.2.2).

purely morphosyntactic, not one of information structure. Weber (1989: 427)'s statement that the evidentials do not always mark the focused constituent thus seems to hold, even apart from their likely specialized occurrence in certain types of confirmation-seeking questions reported by Floyd (1996, 1997). Here we've only discussed the syntagmatic function of the evidentials, mostly disregarding their semantics. Studies of them (Faller 2002, 2003, 2011, 2014; Behrens 2012; Bendezú Araujo 2021) show that they likely interact with focus, but that they do not signal it directly as part of their meaning.

It seems that word order and morphology, just like prosody, make contributions cueing constituents towards an interpretation as focal or topical, but this is not a categorical function, and accompanied by other effects they have. Drawing this conclusion is only possible because of the adopted context-based view of information structure proposed by Riester and colleagues (cf. section 3.7.2), allowing for a strict separation of information structure from formal means of expression. The analysis of the data here has shown that occasionally, the formal means suggest a different information structure than that arrived at via implicit QUD analysis from the context. This allows for situations in which it is a speaker's intention to actively influence their interlocutor's interpretation of an utterance *against* what could be expected from the discourse context. The fact that such situations exist consequently means that the model by Riester and colleagues cannot uncover all subtleties of information structure in all contexts. Yet coming to this conclusion is only possible because a context-based approach allows us to first establish default correspondences between contexts and means of expression empirically, so that marked transgressions against them can then be identified. This would be conceptually impossible under an approach that treats information structural categories as formal markers themselves instead of context configurations, because congruence between context and the presence of such marking can then only ever be a coincidence.

All of this leads to a picture in which prosody, word order and morphology can either align or disalign in their cueing of information structure. The latter probably does not mean that no interpretation is possible, but rather that it is more detailed than just imposing a uniform focus-background and topic-comment structure. In order to understand this better, it should be worthwhile distinguishing more accurately at which level of structure each type of expression system contributes cues that are then probabilistically weighted (cf. Calhoun 2010a, 2015) for an interpretation in terms of these two dimensions, but not only in those terms. Venditti et al. (2008: 505) make a similar point regarding different aspects of Japanese prosody:

> [. . .] phonological relationships such as the contrast between an IP boundary and a mere AP boundary are grammatical abstractions that native speakers acquire in the course of extensive exposure to a rich variety of language-specific cues. [. . .] A frequently encountered congruence of cues from multiple sources in the signal induces a stronger abstraction, which allows adult native speakers to produce conflicting cues when necessary to convey the intended information structure [. . .]. The native listener, conversely, can accommodate to such conflicting cues to recover both the intended focus pattern and the syntactic grouping that the prosodic phrasing also cues. [. . .] [T]he next generation of models of intonational phonology needs to do better justice to these complex interactions among discrete contrasts (e.g., between accented and unaccented words) and continuous variation (e.g., more versus less extreme degrees of pitch range expansion or compression at the edges of focus constituents) and the ways that native speakers and listeners take advantage of the statistical dependencies among different patterns.

I would suggest that this view can be extended to the interaction between cues not only from prosody. In a bilingual community such as the one investigated here, the availability of resources from two languages affords speakers with even more possibilities to heterogeneously signal complex structures. I have barely scratched the surface of this, but I am confident that future research will be able to uncover far more here, using more sophisticated kinds of modeling (e.g. Boersma et al. 2020) in order to paint a more realistic picture of how prosody helps structuring information in actual conversation in a multilingual setting.

7 Synthesis

This final chapter first summarizes the results on the prosody of Huari Quechua and Huari Spanish individually, and then presents some conclusions on how they can be brought together. I will begin by summarizing the findings on Huari Quechua and then move on to Huari Spanish.

7.1 Huari Quechua result summary

7.1.1 Word prosodic profile of Huari Quechua

In answer to research question (34)d about the influence of prominent positions at the word level on pitch events in Huari Quechua, the analysis in this study has found Huari Quechua to be an example of a language that can be said to "care" very little about word stress in the sense of Hyman (2014). That is to say, there is evidence for stress, but it is quite subtle. Of the suprasegmental cues for stress, it is known from the literature that in this variety of Quechua duration is employed for lexical distinctions independent of stress, non-culminatively and non-obligatorily. Pitch contours have been found to be assigned at the level of the phonological phrase (PhP) / accentual phrase (AP), as e.g. in Japanese, French, or Korean, with a rate of 1.27 lexical content words per such phrase in a sample of 457 words (cf. section 6.2.3.2). Phrase-final tonal movements show temporal alignment behaviour that is more indicative of boundary tones than of pitch accents associated with word stress (cf. section 6.1.6). Duration does often seem to serve as "phonetic enhancement" of the syllable on which the phrase-final peak is realized, but this is independent of whether that syllable is the word penult or not. Consequently, in accordance with the theoretical framework laid out in section 3.5, the assumption is that the H tone at least sometimes associates with the syllable on which it is realized, but since this is not related to a metrically strong position, alignment rather than association constraints create the relevant differences between the variants in the OT-analysis. Observations of phrase-final devoicing patterns (cf. section 6.1.3) further support the conclusion that tonal events are due to boundary phenomena, since they show the same final tonal movement as fully voiced phrases, but "displaced" to the left, with the putatively stressed word penult position often realized low or entirely devoiced.

However, stress has also been shown not to be entirely disregarded: in a subset of the tonal alignment patterns identified (the word penult-variant), the position at which a tonal transition occurs (from L to H or vice versa) must make reference to the

https://doi.org/10.1515/9783111304595-007

penult of a word, so that in these patterns, the word penult should be seen as prominent or stressed insofar as it serves as the anchoring point for tonal movement. No evidence for prominence on the initial syllable of words could be found. In one of the alignment configurations, initial syllables of a word were found to be high. However, this is better explained (and modeled in the OT analysis as such) via alignment of the H tone to the word boundary. This is a better analysis because falls are also attested at boundaries between words. By assuming alignment to a word boundary, these two phenomena can be explained via a single mechanism in the OT analysis. Furthermore, initial syllables were only ever found to be high as part of a high plateau extending across the entire word, never as the only high syllable, whereas the word penult was found to be the only high syllable in a subset of realizations. In sum, this suggests that the word penult is a possible and variable anchoring site for tonal transitions, which seems to be the only way that regular word stress manifests in this variety of Quechua.

7.1.2 Observed patterns of intonational variation

A lot of intonational variation has been found in the data, much of it without accompanying identifiable differences in meaning. I summarize the main strands of variation that were identified.

7.1.2.1 Three tonal patterns

The Huari Quechua tonal inventory (cf. research question (34)a) consists of three observed PhP/AP-level tonal patterns: a rising (LH), a falling (HL), and a rising-falling (LHL) pattern. In terms of paradigmatic choice for the signaling of pragmatic meanings (cf. research question (35)a), the distribution of these patterns is only marginally influenced by utterance type, with declaratives, wh-questions, and polar questions not mainly differentiated by them. Some evidence (section 6.2.1) indicates that within polar questions, contour choice of either falling or rising interacts with tag particles such as *aw* and suffixes such as –*ku* to differentiate neutral from various types of biased questions.

The rising pattern was found to be linked with what can broadly labeled as "incompleteness" or "continuation" (section 6.2.2): this includes indication that a topic or turn is to be continued, that a coherent unit has not ended, or that no separate discourse commitment is made. The falling patterns (with no differentiation found between the two subpatterns) are related to "completeness" or "finality": indication that a topic, turn, or other coherent unit is finished, or that a discourse commitment has been made. The relation between these meanings and the tonal patterns was only found to be statistical, with context serving for specification and speakers' stylistic choices also playing a role. The rising pattern was found on a little more than half of all phrases in a subset of the data consisting only of nominal elements (section 6.2.3.2). The rising-falling pattern makes up a little less than a third, and the only-falling pattern, the least frequent, the remainder. In all patterns, a tendency exists to create long stretches of pitch effectively at the same level, in particular also longer high plateaus extending across several syllables, with pitch declining not or only very gradually.

7.1.2.2 Alignment variants

The three tonal patterns are all observed to follow several different alignment variants, formalized in an OT-analysis. This analysis (section 6.3) showed that the alignment variants can be differentiated mainly by three factors: 1) whether tones align only with the boundaries of prosodic units or additionally with a word stress position, 2) whether the H tone is allowed to align with more than one edge, 3) whether the word stress position in loan words from Spanish is ignored or not.

In the **word-boundary variant**, tonal transitions (between L and H in the rising pattern, between H and L in the falling contour, and between both in the rising-falling pattern) occur at word or phrase boundaries, i.e. tones are aligned only with prosodic edges. This variant is further differentiated by whether the H tone is allowed to align with more than one edge or not, i.e. whether some of the constraints effecting its alignment to more than one edge dominate the constraints effecting the multiple alignment of the L tone. Contours are here represented schematically, with the arrows with dashed line representing multiple alignment of the tone the arrow originates from. Its head marks the position to which its alignment (and the pitch level it effects) is allowed to extend, based on the constraint ranking underlying each variant.

Rising patterns (without multiple H alignment)
(198)

L------------------------->H
pinkullupa hananpita (*pinkullu-pa hana-n-pita / flute-GEN above-3-ABL /* "from above the flute")

Rising patterns (with multiple H alignment)
(199)

L-------------><----------H
pinkullupa hananpita

(200)

L <-------------------------H
pinkullupa hananpita

Rise-falling patterns (without multiple H alignment)
(201)

L------------------------->HL
pinkullupa hananpita

(202)

L----------->H L------------>
pinkullupa hananpita

Rise-falling patterns (with multiple H alignment)
(203)

L-------------><----------HL
pinkullupa hananpita

(204)

LH----------><------------L
pinkullupa hananpita

(205)

L------------> H L
pinkullupa hanan~~pita~~

Only-falling patterns (with multiple H alignment)
(206)

H------------><------------L
pinkullupa hananpita

(207)

H-------------------------->L
pinkullupa hananpita

If tones cluster at the right edge of a phrase, the falling patterns can create contours in which the phrase-final penult stands out because it is high, while the following (and preceding, in (201)) syllable is low, due to tonal crowding constraints. These cannot normally be differentiated from similar contours arising from the word penult-variant, in which the word penult is specified as the place at which the H tone is aligned. However, cases were observed where phrase-final devoicing takes place, sometimes up to and including the penult of the final word. In those, the pattern was found to be displaced to the left (cf. (205)): this is only explicable with an edge-seeking account. The contour in (201) is is the most frequent rise-falling contour observed (cf. section 6.2.3.2). That is additional evidence that it arises from both the word boundary- and the word penult variant.

In the **word penult-variant**, which constitutes the only instance of detectable word stress on native Quechua words in Huari Quechua, tonal transitions (between L and H in the rising pattern, between H and L in the falling pattern, and between both in the rising-falling pattern) occur at the edges of the penult of a word, which is taken to be the entirely regular word stress position, with the domain of stress assignment including all suffixes attached to a root. In this variant the H tone always aligns with multiple edges if possible.

Rising patterns

(208)

L------------------------>H->
pinkullupa hananpita

(209)

L------>H-------------------->
pinkullupa hananpita

Rise-falling patterns

(210)

L------------------------>H L
pinkullupa hananpita

(211)

L------->H L---------------->
pinkullupa hananpita

Only-falling patterns

(212)

H--------------------------> L
pinkullupa hananpita

Not all contours allow to distinguish which variant they belong to. Contours (210) and (212) are identical in form to (201) and (207), respectively. In contrast, with (198) and (208), and (202) and (211), respectively, the difference is one in alignment of the H tone by only one syllable – and so that either the penult or the final syllable is realized with a H tone. This suggests that the quantitative findings on alignment in section 6.1.6 can be partially explained by the word-boundary and the word-penult variant occurring together in the sample.

No explanation in terms of function or meaning for the two alignment variants on native Quechua words was found (cf. section 6.1). They also do not seem to be preferentially used depending on the type of experiment. Regarding their relative frequency, it has to be noted that the word boundary-variant can in individually phrased words only be identified by the occurrence of the equivalents of (198), (200), and (205), and the word penult-variant only by that of the equivalent of (208). In such phrases, it is thus to a large degree impossible to say which contour is more frequent. In a sample of multi-word phrases (cf. section 6.2.3.2), the word penult-variant was overall found to be more frequent than the word boundary-variant, and considerable differences were found between speaker pairs in separate experiments, ranging from 0% to 74% occurrence of the word boundary-variant in multi-word phrases. Thus it seems that use of the variants is at least partially due to individual preferences.

The **marked variants** on Spanish loanwords are distinguished by tonal alignment oriented according to the position of the syllable bearing word stress in Spanish, instead of alignment to word or phrase boundaries, or the entirely regular word stress position of the penult. Two patterns are distinguished: the "inherited" pattern, and the "grafted" pattern. The "inherited" pattern functions like the word penult-variant, except that the position to which the H tone is aligned is determined by the syllable which bears word stress in Spanish (213). In the "grafted" pattern, two separate tonal movements occur: phrase-finally, a rise takes place, aligned on the penult of the phrase-final word, like a rising pattern in the word penult-variant. Additionally, a rising-falling movement occurs on the syllable bearing word stress in Spanish. This has been labeled as a LH* pitch accent, like in Huari Spanish, with the following low coming from what is ususally the phrase-initial L tone of Quechua (214).

"inherited" pattern

(213)

L H------------>

a**be**jakunapa (*abeja-kuna-pa* / *bee-PL-GEN* / "by the bees")

"grafted" pattern

(214)

LH* L--->H->

a**be**jakunapa

The marked loanword variants are not categorical on Spanish loanwords (cf. section 6.1.8.3). More frequently, loanwords from Spanish are realized with the same intonation as native Quechua words, i.e. either with tonal alignment according to the word boundary-variant or the word penult-variant, instead of aligned according to the position of the stressed syllable in the Spanish word. The incidence of the marked loanword variant ("inherited" and "grafted" together) differs greatly both between speakers and between lexical items. The highest rate of its use was found to be nearly 60% for one speaker in one of the experiments; most other speakers were found to exhibit far lower ratios, some not producing a marked token at all despite using several Spanish loanwords. Overall, Spanish loanwords are realized with a marked variant in somewhat less than a fifth of all cases, with nearly half not aligned according to the Spanish word stress, and the remaining third indeterminable. Amongst the same speakers, different lexical loanword items exhibited considerably different rates of being realized with the marked variants, suggesting that lexical identity also plays a role.

7.1.3 The role of prominence and information structure

Regarding the question of how meaning-related categories, especially IS, are cued by Quechua prosody (question (35)), this study has found that this is mostly achieved syntagmatically. In terms of which word will be realized with a high tone in a multi-word phrase, the findings suggest that this is due to which word is most prominent in such a phrase. Intonational cues for prominence and their relation to information status and structure were investigated on a subset of nominal sequences. In phrases with the rising and rising-falling pattern, final prominence was assumed to be cued only when the entire high stretch was within the final word. In only-falling phrases, final prominence was assumed to be cued only when the tonal transition occurred on the penult in the final word. If a nominal sequence extended across several phrases, prominence was assigned to the phrase with the highest excursion only if all phrases were of the same tonal pattern. This was compared with the relative information status (givenness/newness) of the individual words in the sequence and their information structure (focus/background) as determined from an analysis of the discourse context. The results support the hypothesis that prominence is final by default, as in Spanish. They also suggest that a deviation from this default is a marked strategy for signaling prefinal narrow focus or prefinal newness followed by given material (cf. section 6.2.3.3). The relation between phrasal prominence and information structure does not seem categorical, and instead probabilistic/distributional, as we saw to be the case for Spanish from the literature, and also for Cuzco Quechua (cf. the discussion in section 3.7.3). The relation between prosodic cues and phrasal prominence itself certainly needs further research. Assigning prominence across several phrases of the same tonal pattern according to highest excursion implies the assumption of a larger prosodic unit encompassing those phrases among which prominence is culminative.

The relation between information structure, prosody, and morphosyntax (word order and the use of the so-called evidential suffixes and –*qa*, the topic suffix) was also further investigated in a qualitative analysis of individual utterances in their conversational contexts (section 6.4), to complement the broad results of the preceding quantitative analyses (section 6.2.3.3). The findings from this analysis support the suggestion that tonal scaling at the level of local pitch span serves to demarcate larger prosodic units encompassing several PhPs, in that prefinal PhPs are realized with comparatively smaller pitch span than the final PhP. Such larger prosodic units with the greatest excursion on the final PhP were regularly found on utterances in broad focus contexts in the narrative task *Cuento* (section 6.4.1), again supporting the assumption that default prominence is rightmost, also at that higher prosodic level. No evidence was found that such a larger phrasal unit comes with its own separate boundary tones, so that in answer to questions (34)c and (35)d it is suggested that the PhP is recursive in Quechua and itself the largest unit. The examples analysed there suggest that phrasing and scaling of these larger phrasal units interact with syntax to signal discourse coherence. An analysis of examples from *Maptask*, for which context suggests a more complex utterance-internal information structure, supports the hypothesis that prosody cues information structural roles indirectly. Constituents interpretable as focal from the context were found to be most often aligned with right boundaries of PhPs. The findings suggest that a reduction in scaling of pitch level (downstep) from one phrase to the next cues a deviation from rightmost prominence, so that the last higher scaled phrase has highest prominence. This was found to occur in contexts that suggest an information structural separation such that the boundary between the last higher scaled phrase and the downstepped one aligns with the division between at-issue material and backgrounded material. These findings support the relation between material aligned with a high tone and prominence from section 6.2.3.3 and extend it to units larger than individual PhPs. Contour shape is also shown to cue information structure indirectly in accordance with the proposal made in section 6.2.2 that rising contours signal openness and falling contours completeness. The findings in this section were put into relation with proposals for a pattern or template by Weber (1989) and Sánchez (2010), which connects information structural roles with positions and morphological marking in a sentence, extending those proposals from a prosodic perspective.

In section 6.4.3, these findings were further complemented by an analysis which compared the individual contributions to cueing information structure from prosody, morphology, and word order against what is interpretable from the context using a number of examples from the *Conc* corpora. It was shown that the cues from different domains are quite often in conflict with each other, and suggested that this might reflect that the information structural roles of referents in *Conc* are underspecified by the notions of topic-comment and focus-background. It was found that scaling as cue to prominence is probably quite reliable, even when it goes against contour type or word order. The picture of the relation between information structure and the cues to it from different formal domains that emerges is problematic for a view which takes information structural roles to be formal features that are categorically expressed via some kind of marking, and far more consistent with a view in which information structure is cued only indirectly and probabilistically, as laid out in the theoretical considerations in section 3.7.3. It enhances this view by exploring how each of the domains cueing information structure is subject to formal restrictions particular to that domain, and how the cues from the different domains can be seen to align and misalign to convey complex information structural configurations. Methodologically, it would not have been possible to make these observations if the principled contextual perspective on information structure from Riester et al. (2018); Riester & Shiohara (2018); Riester (2019) had not been adopted to serve as a third of comparison.

7.2 Huari Spanish result summary

7.2.1 Tonal inventory and encoding of pragmatic meaning via paradigmatic contrasts in Huari Spanish

The analysis of Huari Spanish has found a comparatively small inventory of paradigmatic tonal contrasts. Even though declaratives and interrogatives in pragmatically diverse discourse contexts were investigated, it seems likely that only a single bitonal pitch accent LH* is used in all of them. It is realized via a low elbow in the pretonic syllable, followed by a peak early or in the middle of the tonic vowel. This LH* pitch accent is thus regularized to prenuclear and nuclear position in declaratives, neutral and biased polar questions, alternative questions, and wh-questions, in marked difference to what has been described for many other Spanish varieties, including the "Andean Spanish" of Quito in Ecuador (cf. O'Rourke 2010; Hualde & Prieto 2015), but more similar to the findings on Cuzco Spanish by O'Rourke (2005). This study did not systematically investigate a number of pragmatic meanings that have been previously found to exert an influence on prosodic form, like e.g. imperatives and vocatives, so any conclusions about the full paradigmatic inventory of tones must be preliminary. However, we can compare the variation in nuclear configurations in Huari Spanish with those of Table 3 for Peninsular Spanish across the same range of meaning, i.e. statements with finality, neutral polar questions and requests with confirmation bias (corresponding to a, b, f, g, in Table 3). Within this range, Huari Spanish was found to have only two nuclear configurations (LH* L% and LH* LH% under analysis II in section 5.1.2.1), while Peninsular Spanish has four (LH* L%, L* L%, L* H%, LH* HL%).[449] The only non-phonological context tentatively identified as changing pitch accent realization was when an utterance carried an additional non-at-issue meaning of challenging a salient presupposition in the context, in both assertions and clarification questions. In those cases, both the elbow preceding the peak and the peak itself on the constituent denoting the contested referent were considerably delayed. This could possibly be identified as a pitch accent different but derivable from LH*, like L*H or L+<H*, but further research is needed.

Table 50: Interactional meanings/functions and their encoding via formal means that include intonation in Huari Spanish and Quechua.

pragmatic function	Spanish	Quechua
continuation	LH* H-	LH
finality	LH* L%	(L)HL
declarative	LH* L%	LH/(L)HL
neutral polar question	LH* LH%	-*ku* LH/(L)HL
confirmation-seeking question	LH* + tag LH% (tag e.g. *no*)	[(L)HL (-*ku*)]$_{\phi}$ [LH tag]$_{\phi}$ (tag = *aw*)
clarification request for intended content with confirmation bias	LH* L%	LH
question with [alternative 1] [alternative 2] (alternative question)	[(LH*) LH* H-]$_{\text{iP}}$ [(~~LH~~*)~~LH~~* L%]$_{\text{iP}}$	[LH -*ku*]$_{\phi}$![(L)HL -*ku*]$_{\phi}$
wh-question	[LH*$_{\text{wh-word}}$ (LH*)]$_{\text{iP}}$ ([~~LH~~*]$_{\text{iP}}$) L%	. . . [(L)HL]$_{\phi\ \text{wh-word}}$ . . .
presupposition reversal	delayed peak on LH* L%	?

The boundary tone inventory was also found to be comparatively small. Under the analysis which assumes an LH* pitch accent for all types of declaratives and interrogatives considered (cf. section 5.1.2.1), three boundary tones were identified, two monotonal ones (H-/% and L-/%), and a bitonal one (LH-/%). Table 50 summarizes the findings on how different pragmatic meanings apart from information struc-

449 According to O'Rourke (2010: 231–242), "Ecuadorean Andean Spanish" (based mostly on speech from Quito), has five different nuclear contours in this meaning range: L*L%, LH*L%, L*HH%, LH* HH%, L* HL%. That once again puts the homogeneity of what has been called "Andean" Spanish into question (cf. section 2.2).

tural separations are encoded via formal means that at least include intonation in Huari Spanish and Quechua.

Table 50 shows that purely on the basis of paradigmatic choice between what could be called nuclear configurations (cf. research question (35)a), Huari Spanish only has a three-way distinction. The differentiation between neutral and confirmation-seeking polar questions is achieved by the addition of a tag, while that between declaratives (assertions) and clarification requests with confirmation bias is likely purely achieved by context. Quechua makes even less use of paradigmatic tone contrasts, effectively only differentiating between continuation and finality (the results on the types of biased questions being rather preliminary). Instead, further morphosyntactic means like the polar question suffix *–ku* are employed, but context here also plays a crucial role.

7.2.2 Syntagmatic intonational contrasts: Deaccentuation and phrase accentuation

In contrast, syntagmatic intonational contrasts across utterances (cf. research question (35)b) were found to play a very important role in Huari Spanish. On the one hand, deaccentuation after the accent with the strongest metrical position is used to differentiate between the two alternatives in alternative questions in almost all cases and also very often some point after the wh-word in wh-questions (cf. Table 50). Deaccentuation was also found to occur on the noncontested material in partial denials/reversals. It did not occur on material that is only given following new material without a reversal. But it also occurred on other non-at-issue material following at-issue material, like postposed evaluatives or parentheticals. On the other hand, a marked accentuation mode labeled phrase accentuation was observed. On a larger prosodic unit containing several accentable prosodic words, only the final one receives a full pitch event, either in the form of an LH* pitch accent on the final stressed syllable (phrase accentuation I), or as a rising boundary tone (LH) at the right edge of the phrase (phrase accentuation II). This contrasts with previous descriptions of Spanish that assume a pitch accent on virtually every accentable (content) word (Hualde & Prieto 2015: 358) and also with what can be assumed to be the default or main accentuation pattern in the Huari Spanish data, where a pitch accent was found on nearly 90% of all content words in one sample (cf. section 5.1.1.2). It is much more similar to English or German, where a pitch accent usually only occurs on the most prominent prosodic word in a phrase. The phrase accentuation was found to occur in contexts in which it is plausible to assume an information structural partition such that the phrase which only has a final pitch event is either not internally partitioned in a relevant way (broad focus), or partitioned with narrow focus on the final element, both of which correspond to a metrical structure with rightmost prominence. It is thus phrase-optimizing in the sense that it makes the pitch event culminative in a phrase, and again similar to German or English in that this pitch event is located on the most prominent word. A third marked accentuation pattern was found only marginally. It is a plateau-like realization whereby pitch remains at a high level across a number of prosodic words in a phrase after rising on the first accentable syllable. This pattern is a rare variant and was systematically employed only by few speakers. Huari Spanish was thus found to exhibit two complementary syntagmatic intonational strategies for marking prominence contrasts that can be exploited for cueing relevant information structural partitions: deaccentuation, which compresses postnuclear pitch accents and thus can be used to cue at-issue – non-at-issue partitions, and phrase accentuation, which compresses prenuclear pitch accents and can be used to cue non-at-issue – at-issue partitions. They have in common that they impose a partition on an utterance based on postlexical metrical structure, and marking out parts of that utterance as coming before (phrase accentuation) or after (deaccentuation) the most prominent element. These accentuation modes both go at the expense of word-optimizing prosody. The difference between phrase accentuation I and II could in addition be described as moving between head-marking and edge-marking, in Jun (2005d, 2014b)'s typology. The phrase accentuation was not only found as a marked mode for the cueing of information structural contrasts, but for some speakers, it also seemed to constitute a kind of default. This was found to have consequences for how the prosodic units that structure utterances are cued.

7.2.3 Hierarchical scaling contrasts and recursive prosodic structure

Tonal scaling was found to play an important role in Huari Spanish intonational phenomena (cf. research question (35)d). Firstly, it is an essential part of the reality of the two syntagmatic intonational phenomena phrase accentuation and deaccentuation. The nonprominent pitch accent positions that are suppressed in them were found to exhibit variable degrees of tonal compression, ranging from moderate, but with still notably less local excursion than the pitch event at the prominent position, to total reduction, with no identifiable local excursion remaining.

Secondly, tonal scaling was found to be employed systematically in Huari Spanish for the cueing of hierarchical prosodic structure in a set of complex utterances. They were all double-topic constructions, consisting of two sequences of topic-comment each (cf. section 5.2). Tonal scaling is inherently a relative phenomenon, and the scaling of a local pitch event is relative with reference to at least two different values: the scaling of the utterance in which it occurs and other local pitch events in that utterance, and the usual range of a speaker in a given social context. It therefore lends itself quite naturally to the cueing of prominence, which is also inherently relative. Tonal scaling on pitch accents was found to be significantly affected by whether a pitch accent was part of the first topic, the first comment, the second topic, or the second comment, with the first topic scaled higher than the first comment, and the second topic higher than the second comment. Crucially, the scaling difference between first comment and second topic was smaller on average than that between those other groups, but first comment and second topic were clearly separated by a boundary tone in most cases. These findings are incompatible with a simple local downstep model of tonal scaling. Instead, the number of pitch levels distinguished, the domain across which a local level is maintained, and the scaling difference that is sensitive to the type of utterance subpart it separates all suggest that tonal scaling here cues a hierarchical prosodic structure with several levels of embedding. This prosodic structure is subject to constraints that seek to faithfully map it to the information structural units involved on the one hand, and to ones seeking to reduce the number of its levels, on the other. It was found that only with a sufficient degree of length and complexity of the utterances involved did the first of these win out over the second, suggesting that genuinely rhythmic constraints play a crucial role. It was argued that the prosodic structure that best models these phenomena is recursive at the IP-level, to my knowledge the first time that this has been claimed on empirical grounds for Spanish.

In the discussion of a set of examples by three speakers that deviated in their scaling behaviour, it was shown that they employed phrase accentuation in these utterances. While this had the consequence of reducing the number of contrasts that could be achieved via local pitch accent scaling, it was shown that they use other cues, including boundary tones and their scaling as well as systematic differences in the phrasing of the constituents of the first and the second topic, to signal effectively the same prosodic and metrical structure. They were also the only speakers found to make occasional systematic use of the plateau-like realization, in that they used it on the second topic, while the first one was realized with individual pitch events on each constituent. These findings suggest that different cues, in particular pitch level scaling and boundary tones can be equivalent for the signaling of prosodic structure.

7.2.4 Relation between the observed variants

In an OT-analysis, the relationship between the different accentuation variants (main accentuation and the two phrase accentuation variants) was explored. It could be demonstrated that the main accentuation and the first phrase accentuation variant only differed with regards to the number of minimal tone sequences of LH that are assigned per phonological phrase: while the main accentuation has one LH sequence per prosodic word in the PhP, the phrase accentuation has one per PhP. The different phrase accentuation variants, in turn, differed with regards to the ranking of their alignment and association constraints. The H tone that is the starred tone of a pitch accent in the main accentuation and phrase accentuation I is a boundary tone aligned with the right edge of the phonological phrase in phrase accentuation II. The plateau-like realization was shown to result from a reranking of two constraints in the phrase accentuation. The flexible nature of the tones involved was further demonstrated when IP-level boundary tones were considered: in IP-final position does the H tone that is an edge tone in non-final PhPs in the second phrase accentuation variant become the starred tone of a pitch accent again at least in final paroxytones, because it is pushed to that position by the encroaching IP-boundary tone. This result, as well as the findings on the variants observed for Quechua, supports the hypothesis that tones are at some level autonomous and independent of their relations with the segmental chain (cf. sections 5.3, 6.3), and in principle this implies consequences for the convention in much of the literature of giving the intonational inventory of a language as a list of paradigmatical choices of nuclear configurations in which theses relations are fixed. Yet the fact remains that tonal alignment differences can also encode differences in pragmatic meaning, in Huari Spanish as elsewhere (cf. section 5.1.3.3, Fliessbach 2023), but not in Huari Quechua, as far as we can tell. This poses an empirical and theoretical challenge for the future.

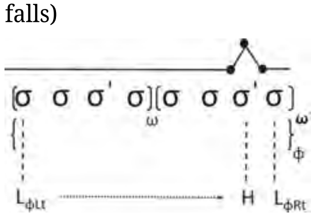
7.3 Conclusions and outlook

After the main results for each language have now been summarized, we are now in a position to synthesize them. The two analytical chapters 5 and 6 aimed at answering the first two main research questions, with in particular the OT-analyses also paving the way for an answer to the third main question of comparing Huari Spanish and Quechua prosody, providing the two halves necessary for it, as it were. In these conclusions, I want to reap the fruits of labour of the foregoing pages and fully focus on what consequences the results have for the third research question. I will integrate the results both from a typological perspective and from one that treats them as evidence of a shared prosodic grammar in a multilingual community.

7.4 Bringing Spanish and Quechua together

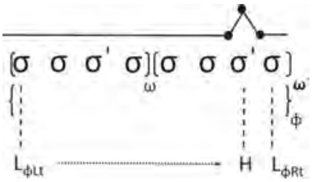
In order to discuss what aspects of Huari Quechua and Spanish prosody are language-specific, and which are shared, I first want to briefly return to the OT-analyses. The descriptions and analyses in the preceding chapters have shown that bilingual Huari speakers employ a range of intonational variants both in their Spanish and their Quechua. Formulated in terms of constraint rankings, it becomes clear that considerable similarities exist between them that cross the boundary between the two languages. (215) and (216) repeat the alignment constraints of what I have called the word-penult and the word-boundary variant of Quechua rise-falls, while (217) gives the ranking for an IP with phrase accentuation of Spanish as exemplified by NQ01, with all association constraints removed. This is done to facilitate comparability, but in fact, as pointed out at the end of section 5.3.2.3, insofar as the phrase accentuation variant of Spanish is concerned that is exemplified in utterances by speaker NQ01, the assumption of association is a somewhat theoretical matter anyway. Linearity, MaxIO(T), NoCrowd are ranked above all other constraints in the rankings here and left out. Between the phrase accentuation of Spanish and the word boundary variant of Quechua, there are some differences that are due to the fact that the tones in Quechua are all taken to belong to the PhP, whereas the Spanish tones are a mixture of PhP- and IP-tones. In the Quechua ranking, there are also constraints referring to the boundary of the prosodic word that do not have a counterpart in the Spanish ranking, and the constraint aligning the H with the stressed syllable, Align (H, σ'), is split up in two constraints separately directing alignment with its right and left edges, Align (H, σ', Lt) and Align (H, σ', Rt). Then there are actual differences in ranking, such as the relative position of Align (L, Lt_{ϕ}) and Align (L, Rt_{ϕ}) in Spanish and their Quechua counterparts Align ($L_{\phi Lt}$, Lt_{ϕ}) and Align ($L_{\phi Lt}$, Rt_{ϕ}).

(215) Alignment constraint ranking for the word-penult variant of Quechua (rise-falls)



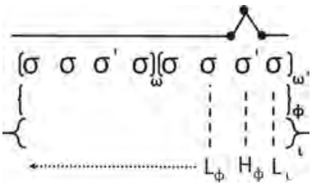
Align ($L_{\phi Rt}$, Rt_{ϕ}) >> Align ($L_{\phi Lt}$, Lt_{ϕ}) >> Align (H, σ', Lt) >> Align (H, σ', Rt) >> Align (H, $Rt_{\omega'}$) >> Align ($L_{\phi Rt}$, Lt_{ϕ}) >> Align (H, Rt_{ϕ}) >> Align ($L_{\phi Lt}$, Rt_{ϕ}) >> Align (H, Lt_{ϕ}) >> Align (H, $Lt_{\omega'}$) >> Align (H, Lt_{ω})

(216) Alignment constraint ranking for the word-boundary variant (without multiple alignment of the H tone) of Quechua (rise-falls)



ALIGN ($L_{\phi Rt}$, Rt_{ϕ}) >> ALIGN ($L_{\phi Lt}$, Lt_{ϕ}) >> ALIGN (H, $Rt_{\omega'}$) >> ALIGN ($L_{\phi Rt}$, Lt_{ϕ}) >> ALIGN (H, Rt_{ϕ}) >> ALIGN ($L_{\phi Lt}$, Rt_{ϕ}) >> ALIGN (H, Lt_{ϕ}) >> ALIGN (H, σ', Lt) >> ALIGN (H, σ', Rt) >> ALIGN (H, $Lt_{\omega'}$) >> ALIGN (H, Lt_{ω})

(217) Alignment constraint ranking for the phrase accentuation variant of Spanish (NQ01's variant) without association constraints, on paroxytones



ALIGN (T_{ι}, Rt_{ι}) >> ALIGN (H, Rt_{ϕ}) >> ALIGN (H, σ') >> ALIGN (L, Rt_{ϕ}) >> ALIGN (L, Lt_{ϕ}) >> ALIGN (H, Lt_{ϕ}) >> ALIGN (T_{ι}, Rt_{ι})

However, there is also considerable overlap: in both rankings, the H tone seeks to right-align and is blocked in that effort by the presence of an L whose right-alignment has precedence; the H tone also seeks to left-align, but the rightwards push of the left L tone is stronger, so that the H forms a peak on the phrase-final penult. Phenotypically, both rankings can be said to converge in that they result in an effectively identical pitch contour of a phrase-final peaked rise-fall in phrases with rightmost prominence. For both rankings, there are also attested variants in which the constraint that right-aligns the L and the one left-aligning the H are reranked relative to each other (cf. sections 5.3.2.1 and 6.3). These result in the attested plateau-like realizations that are more frequent in the Quechua data than in Spanish but are also similar phenotypically. Thus, the plateau-like realizations are also a convergence point for at least two different rankings from the two languages. The phrase-final peaked rise-fall is also generated from the word-penult variant in Quechua, and the main variant in Spanish on paroxytones in IP-final words, so that there are at least four different rankings from the two languages that converge in this contour if only the final word is considered. Equally, the ranking for Quechua rises in the word boundary-variant (cf. sections 6.3.1.1, 6.3.2.1, 6.3.3) and that for the Spanish phrase accentuation in the variant exemplified by XJ45 and NQ01 (cf. section 5.3.2.3) is similar enough that they result in phenotypically very similar contours. Thus both in terms of phenotypical contour shape and in terms of important aspects of the constraint rankings, there is a convergence for some of the described variants for both languages, while other variants are more divergent, independent of which language they belong to.

It is also clear, however, that both the underlying prosodic structures and the detailed phonetic realizations are not all the same. For one, section 6.1.6 showed that in the Quechua sample used there, peak alignment is affected by the relative syllable weight between the final two syllables independent of stress, a subtle interaction of quantity-sensitivity and tonal alignment not observed for Spanish. At a more structural level, for Quechua there has been no need in the analysis to propose more than one phrasal level in the prosodic structure that assigns tones, while for Spanish, both a smaller, PhP-like level and a larger IP-like one are taken to provide tones. Yet for both languages, there is some evidence that the highest level is recursive. Some pitch contour phenotypes are only attested for one of the languages, notably the “main” variant of Spanish and the only-falling contour of Quechua. Thus I do not suggest that Quechua and Spanish are “the same” at a relevant level of representation. Instead, I think that by applying an OT analysis to the observed variants from both languages it has been possible to decompose them into more fundamental variable properties that can assume at least two different values. For some of these properties, all values are shared, i.e. attested in both languages, while for others one or more are specific to one of the two languages. In other words, the variation of some of these properties is orthogonal to the separation between the two languages, extending across both of them, while that of others is parallel to it, keeping it only to one of the languages. In the overall variation space, these properties can assume converging or diverging configurations, which leads to the observed forms that are the result of such configurations being shared or language-specific. In the OT-analysis for each language I tried to model the constraint rankings in such a way that they are maximally specific for each language, making sure that their variation is consistent and plausible mainly within the language itself, by trying to reduce the number of changes from one variant to the next. In this way, what could be called the outer limits of the variation space were traced.

However, that different constraint rankings result in effectively the same contours suggests that a number of intermittent, less peripheral ranking variants also exist, a large number of which is presumably available for both languages. In this way, what I have tried to describe is a part of the shared prosodic resources available to the Huari speaker community as a whole. This can be described as having a core of shared elements, both in terms of constraint rankings and contour phenotypes, which cannot really be said to belong to either language. More peripheral elements that are specific to one of the languages also form part of these resources. Saying that this is a partial description of the linguistic resources of the Huari speaker community does not imply that they are all available to all speakers equally. This is another way in which these resources can be thought of as having a centre and a periphery. Central in this sense are those resources that are available and frequently used by

Table 51: Some of the points of prosodic variation differentiating the described intonational variants in Spanish and Quechua.

Language	Spanish				Quechua				
“Variant”	**main**	**phrase acc. 1**	**phrase acc. 2**	**plateau-like phrase acc.**	**word-boundary w/ mult. H alignment**	**word-boundary w/o mult. H alignment**	**word-penult**	**“inherited” Sp. Accent**	**“grafted” Sp. Accent**
Association of tones	stressed syllables				suboordinate to alignment				stressed syllables
Alignment of H	ALIGN (H, σ') >> ALIGN (H, Rt)		ALIGN (H, Rt) >> ALIGN (H, σ')				ALIGN (H, σ') >> ALIGN (H, Rt)		
Stress effect on pitch	affects pitch alignment				does not affect pitch alignment		affects pitch alignment		
Word stress determined how?	lexical				fully regular			lexical (from Spanish)	
Minimal tone sequence identity (declaratives)	LH				LH, HL, LHL				
Number of possible mininal tone sequences in ϕ	>1	1							> 1

all, while peripheral ones are those that are only available to some, and used infrequently. Table 51 relates some of the variable prosodic properties in both languages to the variants identified in this work. In it, the variable properties are given with the different values they have been shown to assume in the analysis. Gray shading is used to contrast these different values. At the top, the separation between the two languages is marked. For nearly all of the properties given in the table, the separation between the values does not coincide exclusively with the separation between the languages. At least for most of these points of prosodic variation, their variation space is not most usefully described by assigning one of the values to Spanish and the other to Quechua. So at least in this synchronic prosodic description, it is difficult to rigorously differentiate the data using the labels of Quechua and Spanish on a formal level.[450] The different variants as clusters of these values are perhaps more aptly described as prosodic variants available to the Huari speaker community.

7.4.1 Degrees of caring for stress in both languages and typological considerations

The main variational parametres in the OT analysis of both Quechua and Spanish can all be related to different degrees of “caring” for a culminative, obligatory position specified at the word level (word stress), even though they all move along different typological dimensions: alignment of tones with phrasal boundaries clearly serves delimitative, rather than culminative, functions. Alignment to word boundaries also at most serves to distinguish more from less prominent material at the word level, but not at the syllable level. Whether the H tone spreads or not can also be linked to this functional difference, since if a H tone is spreading across several syllables, it automatically loses some of its potential to call attention to a culminative position of syllabic size at the word level. Adjusting the tonal contour around the “inherited” stress position in a Spanish loanword of course aids in cueing the lexical identity of the word via intonation. This function is entirely absent in native Quechua words in any variant, because the fully regular word stress on the penult cannot distinguish between lexical items. For Spanish, it is the default in the main accentuation variant. Finally, minimal tone sequence distribution also plays a crucial role: only if tones are assigned in sufficient numbers

450 Questioning “named languages” as an analytical instrument to describe a synchronic linguistic reality, especially in multilingual contexts, is a central notion in the field of translanguaging (García & Li 2014; Otheguy et al. 2015) as well as others. That does not diminish their usefulness in historical contexts or as heuristic descriptors, or negate that they have a sociolinguistic reality for many speakers.

can they be used to regularly cue the word stress position. The variants in both languages effectively occupy all of these positions. On a cline going from optimizing a culminative, obligatory and lexically specified word stress position to optimizing phrasal units via delimiting tonal movements, the main accentuation variant of Spanish would be at one extreme. The “grafted” and then the “inherited” patterns of Quechua follow after, but at a definitive distance, even though they also cue a lexically specified, instead of a regular, syllabic position. This is on the one hand because in the “grafted” pattern the phrasal tones are expressed in addition to the pitch accent on the word stress, arguably substracting from the culminativity of the pitch accent, and because the “inherited” pattern can still form plateaus. On the other, their optionality and indeed infrequency (cf. section 6.1.8.3) means that the condition of obligatoriness is not even fulfilled for their restricted subset of the lexicon. Further towards the other end we would find the first phrase accentuation variant of Spanish, and the word-penult variant of Quechua, both with only one tone sequence per a PhP containing more than one prosodic word. Fully at the other extreme there are the second phrase accentuation variant of Spanish as exemplified by NQ01, and the word-boundary variant of Quechua without multiple H tone alignment, in both of which constraints making reference to a syllabic prominent position play no role anymore and tonal alignment is purely oriented towards phrase edges. The variants however differ with regards to how they are weighted in terms of frequency: while the main accentuation pattern is the default for Huari Spanish, the “grafted” and “inherited” patterns of Quechua are infrequent and marked because they can only occur on a subset of the lexicon. The word-penult pattern is likely the most frequent pattern of Huari Quechua, and the purely edge-based variants are not the default in either language, but probably more frequent in Quechua. Recall also that while the word-penult variant might be the most frequent one for Quechua, the difference between it and the more edge-based word-boundary variant only really emerges in a subset of configurations (cf. sections 6.1.2, 6.1.3, 6.1.4). In addition, we have seen evidence throughout that Quechua often reduces penults via other prosodic parametres (duration, vowel quality) even if they bear the tonal transition. Thus even though both languages cover a wider range of possibilities with regards to this intonational optimization of word stress or phrasing, all of Huari Quechua as a whole (especially when regarding frequency of variants) can indeed be said to “care” considerably less about stress than Huari Spanish, with the main accentuation of Spanish occupying a position in this respect that could not be observed to be reached by any of the Quechua variants.

Also with regards to other typological parametres that were discussed in chapter 3.4.6, the findings have shown that the variants of both languages occupy positions covering a lot of the ground staked out by them. Huari Spanish in the main accentuation variant is comparable to Egyptian Arabic in that pitch accents of the same regularized form (LH*) are realized on virtually each accentable word. In Jun (2014b)'s typology it would therefore be classed as head-prominent (marking prominence at the AP/PhP-level via pitch events on the heads of phrases), having word stress, and exhibiting strong (tonal) macro-rhythm because it produces the same rising pitch movements at intervals roughly corresponding to the length of a content word (cf. Jun 2014b: 526, 528). Huari Quechua in the word-penult variant marks both the heads of phrases (derived from the regular stress on the strongest word) and their edges, has word stress (even though it "cares" very little about it), and produces both falling and rising pitch movements on phrase heads spaced at intervals usually greater than prosodic words, with stretches of level tones in between. It is thus comparable perhaps to French or Japanese in Jun (2014b: 531), but with a weaker macro-rhythm because the edge tones of the PhP/AP are more variable (both can be either L or H depending on the type of contour). With the less frequent variants, their positions along these parameters shift substantially further in both languages. Arguably, both the second variant of the Spanish phrase accentuation as well as the word boundary-variant of Quechua are only edge-prominence marking, and at least with the word boundary-variant with multiple H tone alignment it is even questionable to class it as showing any evidence of having word stress, so that it is then most comparable to Korean (cf. Jun 2014b: 532). On the other hand, a sequence of words produced in the "grafted" pattern of Quechua would have a stronger macro-rhythm than the word penult-variant. The point here is not so much that applying the parameters is difficult, even though we have seen throughout that classifying Huari Quechua as a language with or without word stress is a complicated issue. I think that these parameters do capture some useful and intersubjectively describable properties, but applying them to the different variants of both Huari Quechua and Spanish puts in doubt that the objects to which these properties belong should really be taken to be "whole" languages. In Jun (2014b), different varieties of e.g. Japanese and Arabic that are spoken by geographically separate populations are also classified differently in the typology, but in our case, a large part of the spectrum is covered by variants of two languages spoken by the same speakers in the same locality. I think nevertheless that this doesn't mean that prosodically all languages or varieties are essentially "the same", just that these parametres perhaps are not fully apt to distinguish between languages in situations of extensive multilingualism.

Consider again pitch accent inventory and stress position optimization. Even though Jun (2014b) introduces them as independent typological parameters, her findings present what is effectively quite a strong correlation between the parameter of prominence-marking (head, edge, or head/edge) and that of macro-rhythm: the edge-marking languages she lists all have strong macro-rhythm, and the head/edge-marking languages have either strong or medium macro-rhythm. Only the head-marking languages cover the full spectrum from strong to weak macro-rhythm, and whether a language has a large paradigmatic inventory of pitch accents is a decisive factor for it to have either medium or weak macro-rhythm. The property of word prosody (stress, lexical pitch accent, tone, or no word prosody) is also correlated: as she points out (Jun 2014b: 530), languages with a lexical pitch accent that occurs only on a subset of the lexicon, including Tokyo Japanese, some varieties of Basque, and arguably also Huari Quechua with Spanish loanwords (cf. section 6.1.8.2), rarely have more than a single pitch accent choice and thus are more likely to have a stronger macro-rhythm. In trying to explain this correlation in search of substantial typological parameters, we could speculate that a linguistic system can only be expected to reliably encode paradigmatic tonal differences at a non-peripheral position if this position is sufficiently present in the mind of speakers – presumably lexicalized – that it is possible to generate expectations associated with it, i.e. that "something" must happen there. Only with such a degree of salience might it then be possible to exploit the expectations connected to this position by varying the kind of pitch event occurring there in a systematic relation to meaning (including by varying tonal alignment relative to the temporal pivot that such a position provides). It seems that in systems where such a position is defined only in a subset of the lexicon, or is overall not very conspicuous (as in Huari Quechua), this threshold of salience for paradigmatic variation is not reached. This is of course a hypothesis as it stands and its claims need to be further corroborated empirically. However, it would point to the size of the paradigmatic tonal inventory and whether it encodes pragmatic meaning differences being one really substantial typological criterion. In this respect, should the findings about Huari Spanish and Quechua presented here hold up in future research, Huari Spanish could then be argued to be just as similar to Huari Quechua as e.g. to Peninsular varieties of Spanish. This is because even though Huari Spanish does have the possibility to express differences in pragmatic meaning beyond continuation and finality, this functional load is almost entirely born by the small boundary tone inventory (also via pitch accent alignment), since likely only a single pitch accent option is available. Thus this possibility is severely reduced compared to what has been described for Peninsular Spanish varieties (Hualde & Prieto 2015; Fliessbach 2023, cf. also section 3.4.3), in practice almost as much as in Huari Quechua, where it was not found (cf. Table 52).

7.4.2 Outlook

Table 52 resumes some further prosodic properties discussed in the analyses that have not been linked to any of the identified variants. The results in both tables should not be taken as definitive. What Table 51 asserts about the property of asso-

Table 52: Further points of prosodic variation between Huari Spanish and Quechua not specifically identified for any of the variants.

Property	Huari Spanish	Huari Quechua
Paradigmatic tone choice signals continuity/finality	✓	✓
Paradigmatic tone choice alone signals additional pragmatic meanings, e.g. utterance type	✓	×
Tone choice plus particles signal additional pragmatic meanings, e.g. utterance type	✓	✓
Syntagmatic tone distribution signals utterance type	✓	×
Syntagmatic tone distribution signals information structure	(more in phrase acc. variant than in main variant)	✓
Alignment differences relative to a stressed position cue additional pragmatic meanings	✓	×
Number of phrasal levels above the prosodic word	2	1
Top level recursive	✓	✓
Devoicing in word- or phrase-final syllables	✓	✓
Devoicing leads to "displacement" of salient pitch movement	×	✓
Interaction of syllable weight and tonal alignment independent of stress	×	✓
Vowel reduction (centralization/devoicing)	✓	✓
Vowel reduction sensitive to word stress	✓	×
Deaccentuation/dephrasing/tonal compression	✓	✓
Pitch scaling employed to express local prominence contrasts	✓	✓
Pitch scaling employed to express nonlocal structure	✓	?

ciation is based on extrapolation from the results of the peak alignment measurements and assumptions in the literature. It is probable that the H tone in Quechua sometimes associates, and sometimes doesn't, but it is unclear whether it does this variably in relation to the identified variants. In Table 52, the absence of a property for one language can only mean that in the data analysed here, no evidence for it was found. It should be clear that further research would be needed to consolidate these results. Nevertheless, here as well there are properties shared by both languages, while others are specific to one of them. It is a separate question which of the shared ones are universal or at least very frequent in the languages of the world, and which are perhaps specific to the Huari speaker community. For example, the signaling of something like continuity and finality via intonation is probably frequent overall. In contrast, in the case of vowel reduction it is well-known that many European varieties of Spanish hardly every reduce vowels, so this is clearly not a universal feature. However, some kind of vowel reduction has also been reported for varieties of both Spanish and Quechua from other regions of Peru (Delforge 2008, 2011; Crignis 2018). That would suggest that some form of it might be a feature crossing the boundaries between the two languages in a wider area. The proposal that a grammar of a multilingual speech community is made up of elements that are specific to one language, elements that are specific to another, and elements that are community-specific, is also made in Höder (2014a, 2014b, 2018) in a construction grammar-based approach. This does not preclude that some of the community-specific features are potentially more broadly universal. Höder's approach rejects "the notion of pre-existing 'languages' in the sense of distinct language systems" (Höder 2014b: 216). He argues that the difference between bi- and multi-lingualism is gradual and that multilingualism in the sense of a speaker having command over several sociolinguistically differentiated varieties is ubiquitous. Grammar as a set of such socially distinct variants made up of specific values that communally relevant variables can assume is then seen as specific to a speaker community. Crucially, observable linguistic differences within a community are taken to be socioindexically meaningful, and the dynamic negotiation of these socioindexical meanings is the driving engine for conventionalization and change (Höder 2014b: 225, 2018: 44). In the present analysis, we have come to identify a number of variants that are specific to either language or the speaker community to different degrees. The results presented here demonstrate that the concept of a repertoire of shared grammatical resources for a multilingual community is meaningfully applicable also in the case of prosody and intonation and even when the two languages in question would normally be considered to be both genetically and typologically distant from each other. What has only been touched upon in passing is the social meaning of using these variants in the Huari speaker community. It has become clear throughout that even within the same elicitation task, there is a considerable degree of speaker-specific variation. This relates not only to the preferential use of the identified intonational variants in either language, but also to other areas of grammar and lexical choice, as well as more global prosodic stylistic choices such as speaking with a comparatively small pitch range. In a sense, this work has only been the first step of inspecting the inventory in terms of exploring the Huari speaker community. I began it by saying that it is a description of the prosody of two varieties of Spanish and Quechua spoken by the same speaker community. Perhaps a more accurate statement would have been that it is a description of the prosody of that community. It is a task for the future to truly relate its variability to community-specific sociopragmatic meanings.

Appendix A – Weblinks to the audio files

Chapter 3

(31)	https://osf.io/d7tzc/

Chapter 5

Figure 20	https://osf.io/vyxtp/
Figure 21	https://osf.io/ypmb4/
Figure 22	https://osf.io/87j56/
Figure 23	https://osf.io/hv4ab/
Figure 24 (left)	https://osf.io/z8vkq/
Figure 24 (right)	https://osf.io/fxujy/
Figure 25	https://osf.io/4685c/
Figure 26	https://osf.io/r73wd/
Figure 27	https://osf.io/g5fs8/
Figure 28	https://osf.io/bhf3e/
Figure 29	https://osf.io/tr5j6/
Figure 30	https://osf.io/3w6sb/
Figure 31 (top)	https://osf.io/cke6p/
Figure 31 (bottom)	https://osf.io/vabz2/
Figure 32 (top)	https://osf.io/rezpf/
Figure 32 (bottom)	https://osf.io/z7eb2/
Figure 33	https://osf.io/tg5r2/
Figure 34	https://osf.io/cp8uk/
(41)	https://osf.io/s8vdr/
Figure 35	https://osf.io/zjfv3/
Figure 36	https://osf.io/8ph9d/
Figure 37	https://osf.io/nyqjr/
Figure 38	https://osf.io/e97u3/
Figure 39	https://osf.io/3v8wy/
Figure 40	https://osf.io/hvw8e/
Figure 41	https://osf.io/q5f3z/
(43)	https://osf.io/jfk4a/
Figure 42	https://osf.io/gby2a/
(44)	https://osf.io/rd3vm/
Figure 43	https://osf.io/bu9tw/
Figure 44	https://osf.io/qmtew/
Figure 45	https://osf.io/xhdu7/
Figure 46	https://osf.io/gt423/
Figure 47	https://osf.io/cdrwf/
Figure 48	https://osf.io/39h2e/
Figure 49	https://osf.io/r4gsd/
Figure 50	https://osf.io/ht96p/

https://doi.org/10.1515/9783111304595-008

Figure 51	https://osf.io/detz6/
Figure 52	https://osf.io/pcn4b/
Figure 53	https://osf.io/8geu6/
(48)	https://osf.io/vn2c4/
Figure 55	https://osf.io/u2bzp/
Figure 56	https://osf.io/xqjkr/
Figure 57	https://osf.io/y75en/
Figure 58	https://osf.io/m47pq/
Figure 59	https://osf.io/zudme/
(50)	https://osf.io/st6d8/
(51)	https://osf.io/xp67y/
(52)	https://osf.io/nvc23/
(53)	https://osf.io/uqptz/
Figure 60	https://osf.io/mx2k7/
Figure 61	https://osf.io/9c6nt/
Figure 62	https://osf.io/eut2x/
Figure 63	https://osf.io/h6wjb/
Figure 64	https://osf.io/ayp93/
Figure 65	https://osf.io/7znmr/
Figure 66	https://osf.io/m7vuf/
(56)	https://osf.io/udz93/
Figure 67	https://osf.io/uqa68/
(57)	https://osf.io/qnjp6/
Figure 68	https://osf.io/vxceg/
Figure 69	https://osf.io/skfvr/
(58)	https://osf.io/ntxpa/
(59)	https://osf.io/4xtd2/
Figure 70	https://osf.io/vun95/
Figure 71	https://osf.io/vht2m/
Figure 72	https://osf.io/j2mpv/
(60)	https://osf.io/9fkwu/
Figure 73	https://osf.io/xdz3q/
Figure 74	https://osf.io/dkawc/
(61)	https://osf.io/pu9jg/
Figure 75	https://osf.io/evu35/
(62)	https://osf.io/94w2h/
Figure 76	https://osf.io/8gk74/
(63)	https://osf.io/cgk7u/
Figure 78	https://osf.io/65yzm/
Figure 80	https://osf.io/hp4fm/
Figure 82	https://osf.io/x9d4r/
Figure 84	https://osf.io/uzwa4/
Figure 87	https://osf.io/ugehf/
Figure 89	https://osf.io/qskvf/
Figure 91	https://osf.io/5ema7/
Figure 94	https://osf.io/rg8zh/
Figure 95	https://osf.io/c3uv9/

Figure 98	https://osf.io/4seky/
Figure 99	https://osf.io/32k64/
Figure 102	https://osf.io/xt2sb/
Figure 103	https://osf.io/4n6ur/
Figure 104	https://osf.io/67ezm/
Figure 106	https://osf.io/h4wgn/
Figure 107	https://osf.io/sb3v9/
Figure 111	https://osf.io/d46st/
Figure 112	https://osf.io/rv8u5/
Figure 113	https://osf.io/uxenr/

Chapter 6

Figure 114	https://osf.io/acp7y/
Figure 115	https://osf.io/mfbse/
Figure 116	https://osf.io/rquxg/
Figure 117	https://osf.io/y97cx/
Figure 118	https://osf.io/xe759/
Figure 119	https://osf.io/g3f56/
Figure 120	https://osf.io/bruj9/
Figure 121	https://osf.io/8tybp/
Figure 122	https://osf.io/ujxny/
Figure 123	https://osf.io/dvc76/
Figure 124	https://osf.io/wrjpn/
Figure 125	https://osf.io/m4ec2/
Figure 126	https://osf.io/bgxzj/
Figure 127	https://osf.io/yxzuw/
Figure 128	https://osf.io/9mbu3/
Figure 129	https://osf.io/8um7v/
Figure 131	https://osf.io/ka8zq/
Figure 132	https://osf.io/pw2ad/
Figure 133	https://osf.io/3c2xa/
Figure 134	https://osf.io/5sfym/
Figure 135	https://osf.io/wj9cv/
Figure 136	https://osf.io/dj8us/
Figure 138	https://osf.io/z7x3s/
Figure 139	https://osf.io/q83es/
Figure 140	https://osf.io/g3x84/
Figure 141 A1	https://osf.io/yuztx/
Figure 141 A2	https://osf.io/jb68t/
Figure 141 A3	https://osf.io/zx4dp/
Figure 141 B1	https://osf.io/uks6e/
Figure 141 B2	https://osf.io/f4ycp/
Figure 141 B3	https://osf.io/2z5wq/
Figure 142 A1	https://osf.io/cqr9m/

Figure 142 A2	https://osf.io/u9ntz/
Figure 142 A3	https://osf.io/f98ph/
Figure 142 B1	https://osf.io/fsd43/
Figure 142 B2	https://osf.io/atdur/
Figure 142 B3	https://osf.io/cjf3s/
Figure 143 1	https://osf.io/cn9aj/
Figure 143 2	https://osf.io/4bkp8/
Figure 143 3	https://osf.io/cnw9x/
Figure 143 4	https://osf.io/c4d2b/
Figure 144 1	https://osf.io/8kjng/
Figure 144 2	https://osf.io/h3rjy/
Figure 144 3	https://osf.io/ugahr/
Figure 144 4	https://osf.io/s6mtg/
Figure 144 5 & 6	https://osf.io/v9epg/
Figure 145 1	https://osf.io/fe78n/
Figure 145 2	https://osf.io/rqj7e/
Figure 145 3	https://osf.io/t98v2/
Figure 145 4	https://osf.io/8d4jt/
Figure 146 1	https://osf.io/qxm84/
Figure 146 2	https://osf.io/asqex/
Figure 146 3	https://osf.io/ga634/
Figure 146 4	https://osf.io/efqnp/
Figure 147	https://osf.io/mxd5q/
Figure 148	https://osf.io/veh3q/
Figure 149	https://osf.io/r56wf/
(126)	https://osf.io/etny4/
Figure 155	https://osf.io/zmbpk/
Figure 156	https://osf.io/z4re9/
Figure 157	https://osf.io/9u4s6/
Figure 158	https://osf.io/e2b4y/
Figure 159	https://osf.io/3ejw9/
Figure 160	https://osf.io/judpy/
Figure 161	https://osf.io/qbuk3/
Figure 162	https://osf.io/rcqxa/
Figure 163	https://osf.io/fr2py/
Figure 164	https://osf.io/g6qux/
Figure 165	https://osf.io/wq9s5/
Figure 166	https://osf.io/kh29b/
Figure 167	https://osf.io/unpge/
(144)	https://osf.io/tay7p/
Figure 168	https://osf.io/9fw5u/
Figure 169	https://osf.io/hrxyq/
(145)	https://osf.io/rnq5y/
(146)	https://osf.io/5j46n/
Figure 170	https://osf.io/wsb8y/
Figure 171	https://osf.io/4aqbd/
Figure 172	https://osf.io/8nvrq/

Figure 173	https://osf.io/uxqtc/
(183)	https://osf.io/m397j/
Figure 174	https://osf.io/dmf6u/
Figure 175	https://osf.io/9h8u3/
(184)	https://osf.io/s9kqr/
Figure 176	https://osf.io/4as7m/
Figure 177	https://osf.io/wur83/
(185)	https://osf.io/pcu9b/
Figure 178	https://osf.io/ptjnu/
Figure 179	https://osf.io/8v7rn/
(186)	https://osf.io/bg78v/
Figure 180	https://osf.io/ajm2n/
(187)	https://osf.io/jg2z8/
Figure 181	https://osf.io/7g3vk/
Figure 182	https://osf.io/xn256/
Figure 183	https://osf.io/4ahqe/
Figure 184	https://osf.io/w5kzy/
Figure 185	https://osf.io/bhqxy/
Figure 186	https://osf.io/28nq9/
(194)	https://osf.io/z6akq/
(195)	https://osf.io/3r5x8/
Figure 187	https://osf.io/6jfz4/

Appendix B – Maptask maps

Maptask maps as used by TP03 & KP04_MT_Q (section 6.4.2)

Figure 188: Maptask map without the path as used by TP03 in TP03 & KP04_MT_Q.

https://doi.org/10.1515/9783111304595-009

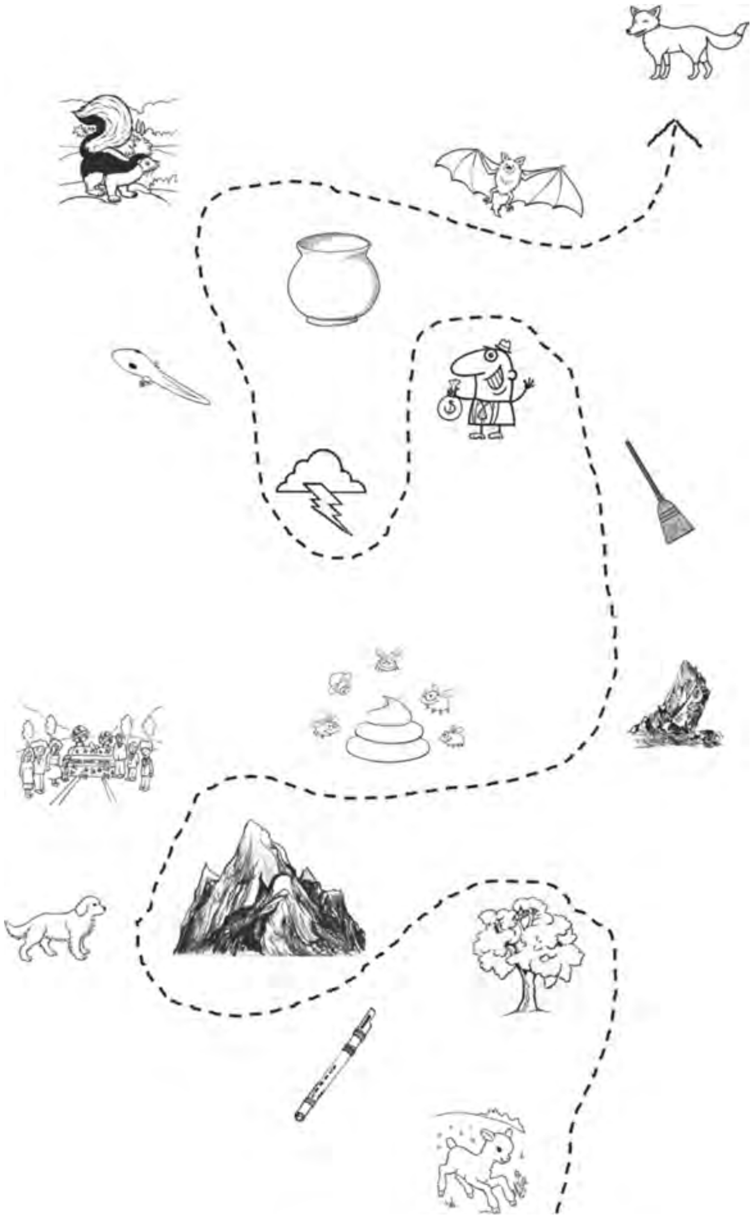


Figure 189: Maptask map with the path as used by KP04 in TP03 & KP04_MT_Q.

Maptask maps as used by ZR29 & HA30_MT_ES, SO39 & MD40_MT_ES, and TP03 & KP04_MT_ES (section 5.1.3.3), and SG15 & QF16_MT_ES (section 5.1.3.2.1 and 5.1.3.2.2)

Figure 190: Maptask map without the path as used by HA30, MD40, KP04, and QF16 in ZR29 & HA30_MT_ES, SO39 & MD40_MT_ES, TP03 & KP04_MT_ES, and SG15 & QF16_MT_ES.

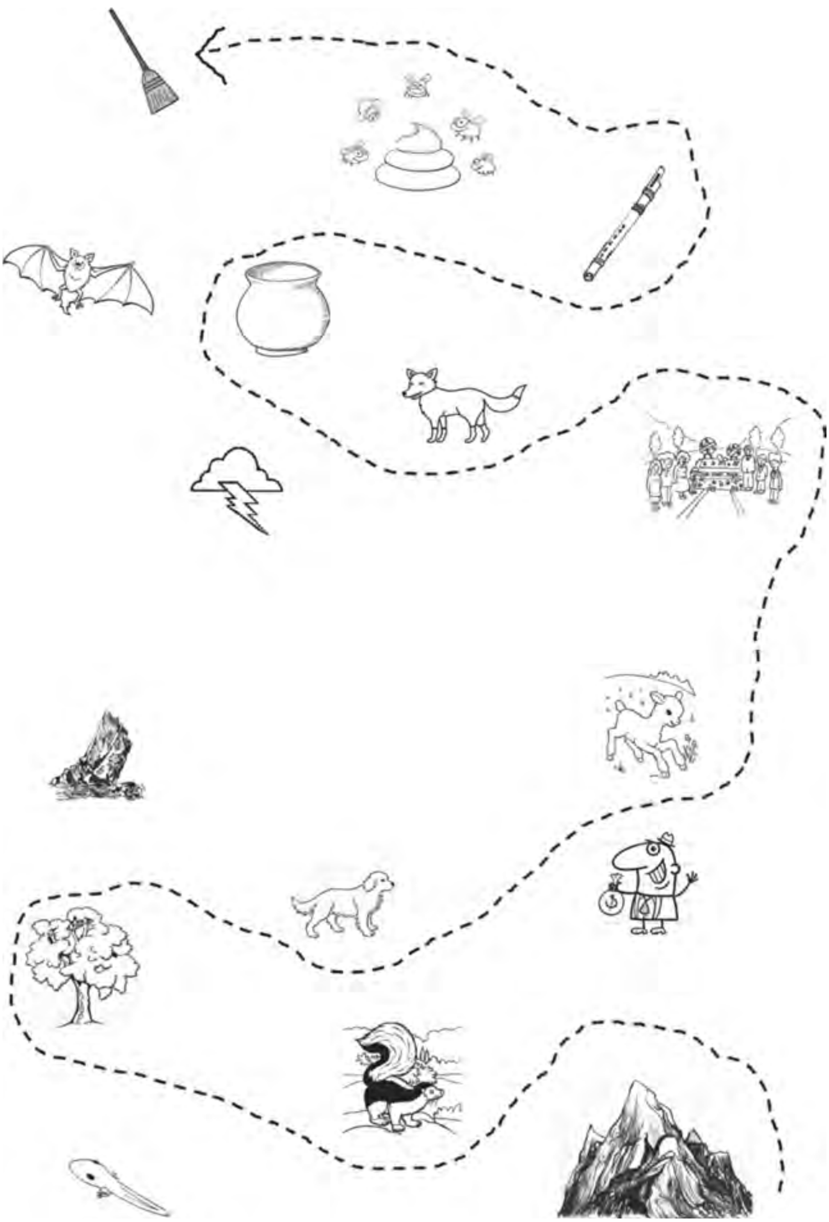


Figure 191: Maptask map with the path as used by as used by ZR29, SO39, TP03, and SG15 in ZR29 & HA30_MT_ES, SO39 & MD40_MT_ES, TP03 & KP04_MT_ES, and SG15 & QF16_MT_ES.

Appendix C – Praat scripts

Script used in section 5.2.3

```
################################################################
### relative_peaks_elqud.praat
###
### Timo Buchholz
### timobuch@zedat.fu-berlin.de
### April 2021
###
###
### Praat 6.1.16
################################################################
form Select directories (without final slash)
        comment Where are the WAV files kept?
        sentence wav_dir C:\User\Folder
        comment Where are the TextGrid files kept?
        sentence txt_dir C:\User\Folder
        comment Where should the results file be kept?
        sentence directory C:\User\Folder
        comment Where are the pitch files kept?
        sentence pitch_dir C:\User\Folder

endform

# Create TXT file with header
filedelete 'directory$'/results.txt
header_row$ = "File" + tab$ + "Word.label" + tab$ +
... "number.all.acc.syls.utterance" + tab$ +
... "start.time.utterance" + tab$ + "end.time.utterance" + tab$ +
... "length.utterance" + tab$ + "acc.syl.label" + tab$ + "contrast.on.acc.syl.
word" + tab$ + "number.this.acc.syl.utterance" + tab$ + "utterance.part" + tab$ +
"wordclass" + tab$ + "start.acc.syl" + tab$ +
... "end.acc.syl" + tab$ + "length.acc.syl" + tab$ + "meanf0.accsyl" + tab$ +
... "position.max.acc.syl" + tab$ + "position.min.acc.syl" + tab$ +
... "maximum.acc.syl" + tab$ + "minimum.acc.syl" + tab$ +
... "mean.first.pretonic.fifth" + tab$ + "mean.second.pretonic.fifth" + tab$ +
... "mean.third.pretonic.fifth" + tab$ + "mean.fourth.pretonic.fifth" + tab$ +
... "mean.fifth.pretonic.fifth" + tab$ + "trans.meanf0.accsyl" + tab$ +
... "trans.maxf0.accsyl" + tab$ + "trans.minf0.accsyl" + tab$ +
... "trans.first.pretonic.fifth" + tab$ + "trans.second.pretonic.fifth" + tab$ +
... "trans.third.pretonic.fifth" + tab$ + "trans.fourth.pretonic.fifth" + tab$ +
... "trans.fifth.pretonic.fifth" + tab$ +
... "start.time.word" + tab$ + "end.time.word" + tab$ +
... "length.word" + tab$ + "label.first.utt.extreme" + tab$ +
... "label.second.utt.extreme" + tab$ + "utterance.maximum" + tab$ +
... "utterance.minimum" + tab$ + "pos.utterance.maximum" + tab$ +
... "pos.utterance.minimum" + tab$ + "word.end.topic1" + tab$ +
... "start.endsyl.topic1" + tab$ + "end.endsyl.topic1" + tab$ +
```


https://doi.org/10.1515/9783111304595-010

```
... “meanf0.endsyl.topic1” + tab$ + “maxf0.endsyl.topic1” + tab$ +
... “pos.max.endsyl.topic1” + tab$ + “minf0.endsyl.topic1” + tab$ + “pos.min.
endsyl.topic1” + tab$ + “trans.meanf0.endsyl.topic1” + tab$ +
... “trans.maxf0.endsyl.topic1” + tab$ + “trans.minf0.endsyl.topic1” + tab$ +
“word.end.comment1” + tab$ +
... “start.endsyl.comment1” + tab$ + “end.endsyl.comment1” + tab$ +
... “meanf0.endsyl.comment1” + tab$ + “maxf0.endsyl.comment1” + tab$ +
... “pos.max.endsyl.comment1” + tab$ + “minf0.endsyl.comment1” + tab$ + “pos.min.
endsyl.comment1” + tab$ + “trans.meanf0.endsyl.comment1” + tab$ +
... “trans.maxf0.endsyl.comment1” + tab$ + “trans.minf0.endsyl.comment1” + tab$ +
“word.end.topic2” + tab$ +
... “start.endsyl.topic2” + tab$ + “end.endsyl.topic2” + tab$ +
... “meanf0.endsyl.topic2” + tab$ + “maxf0.endsyl.topic2” + tab$ +
... “pos.max.endsyl.topic2” + tab$ + “minf0.endsyl.topic2” + tab$ + “pos.min.
endsyl.topic2” + tab$ + “trans.meanf0.endsyl.topic2” + tab$ +
... “trans.maxf0.endsyl.topic2” + tab$ + “trans.minf0.endsyl.topic2” + tab$ +
“word.end.comment2” + tab$ +
... “start.endsyl.comment2” + tab$ + “end.endsyl.comment2” + tab$ +
... “meanf0.endsyl.comment2” + tab$ + “maxf0.endsyl.comment2” + tab$ +
... “pos.max.endsyl.comment2” + tab$ + “minf0.endsyl.comment2” + tab$ + “pos.min.
endsyl.comment2” + tab$ + “trans.meanf0.endsyl.comment2” + tab$ +
... “trans.maxf0.endsyl.comment2” + tab$ + “trans.minf0.endsyl.comment2” + tab$ +
... newline$
header_row$ > ‘directory$’/results.txt

#Check whether there are pitch files already that have been used and corrected
manually before WARNING: You must have one pitch file per each wav file for this
to work. If you don’t, the script will overwrite your existing pitch files
# List all pitch files

Create Strings as file list... list ‘pitch_dir$’/*.Pitch
numberofPitchfiles = Get number of strings

# List all WAV files
Create Strings as file list... list ‘wav_dir$’/*.wav
numberOfFiles = Get number of strings

# Loop that goes through all files

for ifile to numberOfFiles
        select Strings list
        fileName$ = Get string... ifile
        baseName$ = fileName$ - “.wav”
        # Read in the Sound files with that base name
        Read from file... ‘wav_dir$’/’baseName$’.wav
        Read from file... ‘txt_dir$’/’baseName$’.TextGrid
```

```
## hash the if-clause optionally if you know all the files have pitch files
anyway, and only keep the “read from file...” line

#        if ‘numberofPitchfiles’ = ‘numberOfFiles’

                  Read from file... ‘txt_dir$’/’baseName$’.Pitch
#        else

# # Create Pitch object
#                 select Sound ‘baseName$’
#                 To Pitch (ac)... 0.005 75 15 no 0.03 0.6 0.01 0.35 0.30 600
#        endif

# Select the individual words

         selectObject: “TextGrid ‘baseName$’”
         plusObject: “Sound ‘baseName$’”
         Scale times

         select TextGrid ‘baseName$’


# Get start and end times for utterance (including pauses)


         nintervalbottomtier = Get number of intervals... 4
         startutterance = Get starting point... 4 2
         endutterance = Get end point... 4 ‘nintervalbottomtier’-1
         lengthutterance = (endutterance - startutterance)*1000

#Get number of accented syllables in utterance


         naccsyltier = Get number of intervals... 2
         nemptyintervalaccsyltier = Count intervals where: 2, “is equal to”, “”
         naccsyls = ‘naccsyltier’ - ‘nemptyintervalaccsyltier’


# look at pitch object and see if it has any obvious errors

         select Pitch ‘baseName$’

         Shift times to: “start time”, 0

# hash out the following view & edit and pause lines if you already checked all
the pitch files and just want to generate the results

#        View & Edit

#        pause Edit pitch object, then press “continue” to save and proceed to the
next file
```

```
# determine positions for the final syllables in each of the four parts

         select TextGrid 'baseName$'

         top1endlabel$ = Get label of interval... 3 2
         com1endlabel$ = Get label of interval... 3 4
         top2endlabel$ = Get label of interval... 3 6
         com2endlabel$ = Get label of interval... 3 8

         starttop1end = Get starting point... 3 2
         startcom1end = Get starting point... 3 4
         starttop2end = Get starting point... 3 6
         startcom2end = Get starting point... 3 8

         endtop1end = Get end point... 3 2
         endcom1end = Get end point... 3 4
         endtop2end = Get end point... 3 6
         endcom2end = Get end point... 3 8

# Get word label of final word at each part

         wordattop1end = Get interval at time... 1 ('starttop1end'+0.0001)
         top1endwordlabel$ = Get label of interval... 1 'wordattop1end'

         wordatcom1end = Get interval at time... 1 ('startcom1end'+0.0001)
         com1endwordlabel$ = Get label of interval... 1 'wordatcom1end'

         wordattop2end = Get interval at time... 1 ('starttop2end'+0.0001)
         top2endwordlabel$ = Get label of interval... 1 'wordattop2end'

         wordatcom2end = Get interval at time... 1 'startcom2end'+0.0001
         com2endwordlabel$ = Get label of interval... 1 'wordatcom2end'

# determine positions and labels for the highest and lowest annotated points


         bottomtierlabel$ = Get label of interval... 4 2

                  if bottomtierlabel$ <> ""
                           firstextremelabel$ = Get label of interval... 4 2
                           startfirstextreme = Get starting point... 4 2
                           endfirstextreme = Get end point... 4 2

                           secondextremelabel$ = Get label of interval... 4 4
                           startsecondextreme = Get starting point... 4 4
                           endsecondextreme = Get end point... 4 4

                  endif

                  if bottomtierlabel$ == ""
                           firstextremelabel$ = Get label of interval... 4 3
```

```
                  startfirstextreme = Get starting point... 4 3
                  endfirstextreme = Get end point... 4 3

                  secondextremelabel$ = Get label of interval... 4 5
                  startsecondextreme = Get starting point... 4 5
                  endsecondextreme = Get end point... 4 5

         endif

# get f0 values for the endpoints of the parts and the extremes

     select Pitch ‘baseName$’

     meanfirstextreme = Get mean... ‘startfirstextreme’ ‘endfirstextreme’
Hertz

     tMaxF0firstextreme = Get time of maximum... ‘startfirstextreme’
‘endfirstextreme’ Hertz Parabolic
     tMinF0firstextreme = Get time of minimum... ‘startfirstextreme’
‘endfirstextreme’ Hertz Parabolic

     f0maxfirstextreme = Get value at time... ‘tMaxF0firstextreme’ Hertz
Linear
     f0minfirstextreme = Get value at time... ‘tMinF0firstextreme’ Hertz
Linear

     meansecondextreme = Get mean... ‘startsecondextreme’ ‘endsecondextreme’
Hertz

     tMaxF0secondextreme = Get time of maximum... ‘startsecondextreme’
‘endsecondextreme’ Hertz Parabolic
     tMinF0secondextreme = Get time of minimum... ‘startsecondextreme’
‘endsecondextreme’ Hertz Parabolic

     f0maxsecondextreme = Get value at time... ‘tMaxF0secondextreme’ Hertz
Linear
     f0minsecondextreme = Get value at time... ‘tMinF0secondextreme’ Hertz
Linear


     if firstextremelabel$ == “H” and secondextremelabel$ == “L”
         utterancemax = ‘f0maxfirstextreme’
         utterancemin = ‘f0minsecondextreme’
         posutterancemax = ‘tMaxF0firstextreme’
         posutterancemin = ‘tMinF0secondextreme’
     endif

     if firstextremelabel$ == “L” and secondextremelabel$ == “H”
         utterancemax = ‘f0maxsecondextreme’
         utterancemin = ‘f0minfirstextreme’
         posutterancemax = ‘tMaxF0secondextreme’
```

```
                posutterancemin = 'tMinF0firstextreme'
        endif

        meantop1end = Get mean... 'starttop1end' 'endtop1end' Hertz

        tMaxF0top1end = Get time of maximum... 'starttop1end' 'endtop1end' Hertz
Parabolic
        tMinF0top1end = Get time of minimum... 'starttop1end' 'endtop1end' Hertz
Parabolic

        f0maxtop1end = Get value at time... 'tMaxF0top1end' Hertz Linear
        f0mintop1end = Get value at time... 'tMinF0top1end' Hertz Linear

        meancom1end = Get mean... 'startcom1end' 'endcom1end' Hertz

        tMaxF0com1end = Get time of maximum... 'startcom1end' 'endcom1end' Hertz
Parabolic
        tMinF0com1end = Get time of minimum... 'startcom1end' 'endcom1end' Hertz
Parabolic

        f0maxcom1end = Get value at time... 'tMaxF0com1end' Hertz Linear
        f0mincom1end = Get value at time... 'tMinF0com1end' Hertz Linear

        meantop2end = Get mean... 'starttop2end' 'endtop2end' Hertz

        tMaxF0top2end = Get time of maximum... 'starttop2end' 'endtop2end' Hertz
Parabolic
        tMinF0top2end = Get time of minimum... 'starttop2end' 'endtop2end' Hertz
Parabolic

        f0maxtop2end = Get value at time... 'tMaxF0top2end' Hertz Linear
        f0mintop2end = Get value at time... 'tMinF0top2end' Hertz Linear

        meancom2end = Get mean... 'startcom2end' 'endcom2end' Hertz

        tMaxF0com2end = Get time of maximum... 'startcom2end' 'endcom2end' Hertz
Parabolic
        tMinF0com2end = Get time of minimum... 'startcom2end' 'endcom2end' Hertz
Parabolic

        f0maxcom2end = Get value at time... 'tMaxF0com2end' Hertz Linear
        f0mincom2end = Get value at time... 'tMinF0com2end' Hertz Linear

        transformedmeantop1end = ('meantop1end'-'utterancemin')/('utterancemax'-
'utterancemin')
        transformedmeancom1end = ('meancom1end'-'utterancemin')/('utterancemax'-
'utterancemin')
        transformedmeantop2end = ('meantop2end'-'utterancemin')/('utterancemax'-
'utterancemin')
        transformedmeancom2end = ('meancom2end'-'utterancemin')/('utterancemax'-
'utterancemin')
```

```
	transformedmaxtop1end = ('f0maxtop1end'-'utterancemin')/('utterancemax'-
'utterancemin')
	transformedmaxcom1end = ('f0maxcom1end'-'utterancemin')/('utterancemax'-
'utterancemin')
	transformedmaxtop2end = ('f0maxtop2end'-'utterancemin')/('utterancemax'-
'utterancemin')
	transformedmaxcom2end = ('f0maxcom2end'-'utterancemin')/('utterancemax'-
'utterancemin')
	transformedmintop1end = ('f0mintop1end'-'utterancemin')/('utterancemax'-
'utterancemin')
	transformedmincom1end = ('f0mincom1end'-'utterancemin')/('utterancemax'-
'utterancemin')
	transformedmintop2end = ('f0mintop2end'-'utterancemin')/('utterancemax'-
'utterancemin')
	transformedmincom2end = ('f0mincom2end'-'utterancemin')/('utterancemax'-
'utterancemin')


# Go through each individual accented syllable

	for a from 1 to naccsyltier
		select TextGrid 'baseName$'
		syllabel$ = Get label of interval... 2 'a'

		if 'a' = 1
			counter = 0
		endif

		if syllabel$ <> ""
			startsyl = Get starting point... 2 'a'
			endsyl = Get end point... 2 'a'
			lengthsyl = (endsyl - startsyl)*1000

			counter = 'counter'+1

# Get label of word from which the accented syllable comes

			wordatsyl = Get interval at time... 1 'startsyl'+0.0001
			wordlabel$ = Get label of interval... 1 'wordatsyl'

# Get label of which part and word class accented word belongs to

			if syllabel$ == "T1N1" or syllabel$ == "T1A1" or
syllabel$ == "T1P1"
				sylpartlabel$ = "topic1"
			endif

			if syllabel$ == "C1Aux1" or syllabel$ == "C1V1" or
syllabel$ == "C1N1" or syllabel$ == "C1A1" or syllabel$ == "C1P1" or syllabel$ ==
"C1Art1" or syllabel$ == "C1Aux2" or syllabel$ == "C1V2" or syllabel$ == "C1N2" or
syllabel$ == "C1A2" or syllabel$ == "C1P2" or syllabel$ == "C1Art2"
```

```
				sylpartlabel$ = “comment1”
			endif

			if syllabel$ == “T2N1” or syllabel$ == “T2A1”
				sylpartlabel$ = “topic2”
			endif

			if syllabel$ == “C2Aux1” or syllabel$ == “C2V1” or
syllabel$ == “C2N1” or syllabel$ == “C2A1” or syllabel$ == “C2P1” or syllabel$ ==
“C2Art1” or syllabel$ == “C2Aux2” or syllabel$ == “C2V2” or syllabel$ == “C2N2” or
syllabel$ == “C2A2” or syllabel$ == “C2P2” or syllabel$ == “C2Art2”

				sylpartlabel$ = “comment2”
			endif


			if syllabel$ == “T1N1” or syllabel$ == “T2N1” or
syllabel$ == “C1N1” or syllabel$ == “C1N2” or syllabel$ == “C2N1” or syllabel$ ==
“C2N2”
				sylwordclasslabel$ = “N”
			endif

			if syllabel$ == “T1A1” or syllabel$ == “T2A1” or
syllabel$ == “C1A1” or syllabel$ == “C1A2” or syllabel$ == “C2A1” or syllabel$ ==
“C2A2”
				sylwordclasslabel$ = “A”
			endif

			if syllabel$ == “T1V1” or syllabel$ == “T2V1” or
syllabel$ == “C1V1” or syllabel$ == “C1V2” or syllabel$ == “C2V1” or syllabel$ ==
“C2V2”
				sylwordclasslabel$ = “V”
			endif

			if syllabel$ == “T1Aux1” or syllabel$ == “T2Aux1”
or syllabel$ == “C1Aux1” or syllabel$ == “C1Aux2” or syllabel$ == “C2Aux1” or
syllabel$ == “C2Aux2”
				sylwordclasslabel$ = “Aux”
			endif

			if syllabel$ == “T1P1” or syllabel$ == “T2P1” or
syllabel$ == “C1P1” or syllabel$ == “C1P2” or syllabel$ == “C2P1” or syllabel$ ==
“C2P2”
				sylwordclasslabel$ = “P”
			endif

			if syllabel$ == “C1Art1” or syllabel$ == “C1Art2” or
syllabel$ == “C2Art1” or syllabel$ == “C2Art2”
				sylwordclasslabel$ = “Art”
			endif
```

```
# Get duration of word

                           startword = Get starting point... 1 ‘wordatsyl’
                           endword = Get end point... 1 ‘wordatsyl’
                           lengthword = (endword - startword)*1000

# find out if the word is contrasted in the utterance or not

                           contratsyl = Get interval at time... 5
‘startsyl’+0.0001
                           contrlabel$ = Get label of interval... 5 ‘contratsyl’
                           if contrlabel$ <> “”
                                    contrast$ = “yes”
                           else
                                    contrast$ = “no”
                           endif

# Find out what’s before the accented syl to determine minimum points

                           if ‘startword’ = ‘startsyl’
                                    pretonic = Get interval at time... 1
‘startword’-0.0001
                                    pretoniclabel$ = Get label of interval... 1
‘pretonic’

                                    if pretoniclabel$ <> “”
                                             firstpretonicfifth = ‘startsyl’-0.15
                                             secondpretonicfifth =
‘startsyl’-(0.15-0.15/5)
                                             thirdpretonicfifth =
‘startsyl’-(0.15-2*0.15/5)
                                             fourthpretonicfifth =
‘startsyl’-(0.15-3*0.15/5)
                                             fifthpretonicfifth =
‘startsyl’-(0.15-4*0.15/5)
                                    endif

                           endif

                           if ‘startword’ < ‘startsyl’
                                             firstpretonicfifth = ‘startword’
                                             secondpretonicfifth =
‘startword’+(‘startsyl’-’startword’)/5
                                             thirdpretonicfifth =
‘startword’+2*((‘startsyl’-’startword’)/5)
                                             fourthpretonicfifth =
‘startword’+3*((‘startsyl’-’startword’)/5)
                                             fifthpretonicfifth =
‘startword’+4*((‘startsyl’-’startword’)/5)
                           endif
```

```
# move on to pitch object

                              select Pitch ‘baseName$’

                              meansyl = Get mean... ‘startsyl’ ‘endsyl’ Hertz

                              tMaxF0syl = Get time of maximum... ‘startsyl’ ‘endsyl’
Hertz Parabolic
                              tMinF0syl = Get time of minimum... ‘startsyl’ ‘endsyl’
Hertz Parabolic

                              f0maxsyl = Get value at time... ‘tMaxF0syl’ Hertz
Linear
                              f0minsyl = Get value at time... ‘tMinF0syl’ Hertz
Linear

                              meanfirstpretonicfifth = Get mean...
‘firstpretonicfifth’ ‘secondpretonicfifth’ Hertz
                              meansecondpretonicfifth = Get mean...
‘secondpretonicfifth’ ‘thirdpretonicfifth’ Hertz
                              meanthirdpretonicfifth = Get mean...
‘thirdpretonicfifth’ ‘fourthpretonicfifth’ Hertz
                              meanfourthpretonicfifth = Get mean...
‘fourthpretonicfifth’ ‘fifthpretonicfifth’ Hertz
                              meanfifthpretonicfifth = Get mean...
‘fifthpretonicfifth’ ‘startsyl’ Hertz

                              transformedmeansyl = (‘meansyl’-’utterancemin’)/
(‘utterancemax’-’utterancemin’)
                              transformedmaxsyl = (‘f0maxsyl’-’utterancemin’)/
(‘utterancemax’-’utterancemin’)
                              transformedminsyl = (‘f0minsyl’-’utterancemin’)/
(‘utterancemax’-’utterancemin’)
                              transformedmeanfirstpretonicfifth =
(‘meanfirstpretonicfifth’-’utterancemin’)/(‘utterancemax’-’utterancemin’)
                              transformedmeansecondpretonicfifth =
(‘meansecondpretonicfifth’-’utterancemin’)/(‘utterancemax’-’utterancemin’)
                              transformedmeanthirdpretonicfifth =
(‘meanthirdpretonicfifth’-’utterancemin’)/(‘utterancemax’-’utterancemin’)
                              transformedmeanfourthpretonicfifth =
(‘meanfourthpretonicfifth’-’utterancemin’)/(‘utterancemax’-’utterancemin’)
                              transformedmeanfifthpretonicfifth =
(‘meanfifthpretonicfifth’-’utterancemin’)/(‘utterancemax’-’utterancemin’)

# save edited pitch object

                              Save as text file: “’directory$’\’baseName$’.Pitch”
```

```
# Append values in the TXT file

                            select TextGrid 'baseName$'

                            fileappend "'directory$'/results.txt"
'baseName$''tab$''wordlabel$''tab$'
                            ...'naccsyls''tab$'
                            ...'startutterance''tab$''endutterance''tab$'
                            ...'lengthutterance''tab$''syllabel$''tab$''contrast$'
'tab$''counter''tab$''sylpartlabel$''tab$'
'sylwordclasslabel$''tab$'
                            ...'startsyl''tab$''endsyl''tab$''lengthsyl''tab$'
                            ...'meansyl''tab$''tMaxF0syl''tab$'
                            ...'tMinF0syl''tab$''f0maxsyl''tab$''f0minsyl''tab$'
                            ...'meanfirstpretonicfifth''tab$'
'meansecondpretonicfifth''tab$'
                            ...'meanthirdpretonicfifth''tab$'
'meanfourthpretonicfifth''tab$'
                            ...'meanfifthpretonicfifth''tab$'
'transformedmeansyl''tab$'
                            ...'transformedmaxsyl''tab$''transformedminsyl''tab$'
                            ...'transformedmeanfirstpretonicfifth''tab$'
'transformedmeansecondpretonicfifth''tab$'
                            ...'transformedmeanthirdpretonicfifth''tab$'
'transformedmeanfourthpretonicfifth''tab$'
                            ...'transformedmeanfifthpretonicfifth''tab$'
'startword''tab$'
                            ...'endword''tab$''lengthword''tab$'
                            ...'firstextremelabel$''tab$''secondextremelabel$'
'tab$'
                            ...'utterancemax''tab$''utterancemin''tab$'
                            ...'posutterancemax''tab$''posutterancemin''tab$'
                            ...'top1endwordlabel$''tab$''starttop1end'
'tab$''endtop1end''tab$'
                            ...'meantop1end''tab$''f0maxtop1end''tab$'
'tMaxF0top1end''tab$'
                            ...'f0mintop1end''tab$''tMinF0top1end''tab$'
                            ...'transformedmeantop1end''tab$'
'transformedmaxtop1end''tab$''transformedmintop1end''tab$'
                            ...'com1endwordlabel$''tab$''startcom1end''tab$'
'endcom1end''tab$'
                            ...'meancom1end''tab$''f0maxcom1end''tab$'
'tMaxF0com1end''tab$'
                            ...'f0mincom1end''tab$''tMinF0com1end''tab$'
                            ...'transformedmeancom1end''tab$'
'transformedmaxcom1end''tab$''transformedmincom1end''tab$'
                            ...'top2endwordlabel$''tab$''starttop2end''tab$'
'endtop2end''tab$'
                            ...'meantop2end''tab$''f0maxtop2end''tab$'
'tMaxF0top2end''tab$'
                            ...'f0mintop2end''tab$''tMinF0top2end''tab$'
```

```
                            ...'transformedmeantop2end''tab$'
'transformedmaxtop2end''tab$''transformedmintop2end'
'tab$'
                            ...'com2endwordlabel$''tab$''startcom2end''tab$'
'endcom2end''tab$'
                            ...'meancom2end''tab$''f0maxcom2end''tab$'
'tMaxF0com2end''tab$'
                            ...'f0mincom2end''tab$''tMinF0com2end''tab$'
                            ...'transformedmeancom2end''tab$'
'transformedmaxcom2end''tab$''transformedmincom2end'
'tab$'
                            ...'newline$'


                  endif
         endfor

# Clean up before going on with next file
         select Sound 'baseName$'
         plus TextGrid 'baseName$'
         plus Pitch 'baseName$'
         Remove

endfor

# Final clean up
select Strings list
Remove
clearinfo
print All files done!
```

Script used in section 6.1.6

```
#################################################################
### intonation_slopes.praat
###
### Timo Buchholz
### timobuch@zedat.fu-berlin.de
### March 2018
###
###
### Praat 5.3.56
#################################################################
form Select directories (without final slash)
	comment Where are the WAV files kept?
	sentence wav_dir C:\User\Folder
	comment Where are the TextGrid files kept?
	sentence txt_dir C:\User\Folder
	comment Where should the results file be kept?
	sentence directory C:\User\Folder
	comment Where are the pitch files kept?
	sentence pitch_dir C:\User\Folder
	comment What is the name of the first (main) speaker?
	sentence speakerOne XU33
	comment Which is the tier with the syllables when the main speaker is
speaking?
	integer speakersyltier 4
	comment Which is the tier with the word labels?
	integer wordtier 5
	comment Which language
	sentence language Quechua
	comment Which type of utterance
	sentence typeutterance pos_resp
endform

# Create TXT file with header
filedelete ‘directory$’/results.txt
header_row$ = “Word.label” + tab$ + “File” + tab$ + “Speaker.name” + tab$ +
... “language” + tab$ + “type.utterance” + tab$ + “number.words.utterance” + tab$
+
... “start.time.utterance” + tab$ + “end.time.utterance” + tab$ +
... “length.utterance” + tab$ + “number.syls.word” + tab$ + “position” + tab$ +
... “initial.slope” + tab$ + “final.slope” + tab$ +
... “start.time.word” + tab$ + “end.time.word” + tab$ +
... “length.word” + tab$ +
... “initial.pitchrange” + tab$ + “final.pitchrange” + tab$ +
... “initial.tone.distance” + tab$ + “final.tone.distance” + tab$ + “initial.Max”
+ tab$ + “final.Max” + tab$ +
... “initial.Min” + tab$ + “final.Min” + tab$ + “initial.pos.Max” + tab$ + “final.
pos.Max” + tab$ + “initial.pos.Min” + tab$ + “final.pos.Min” + tab$ +
... “firstsyluttmean” + tab$ + “lastsyluttmean” + tab$ +
```

```
... “meanF0.firstsyl” + tab$ + “meanF0.secondsyl” + tab$ + “meanF0.penult” + tab$
+ “meanF0.finalsyl” + tab$ +
... “distance.init.max.from.end.of.firstsyl” + tab$ + “distance.init.min.from.end.
of.firstsyl” + tab$ + “distance.final.max.from.end.of.penult” + tab$ + “distance.
final.min.from.end.of.penult” + tab$ +
... “start.firstsyl” + tab$ + “end.firstsyl” + tab$ + “end.secondsyl” + tab$ +
“start.antepenult” + tab$ + “start.penult” + tab$ + “end.penult” + tab$ + “end.
finalsyl” + tab$ +
... “length.firsttwosyls” + tab$ + “length.lasttwosyls” + tab$ +
... “initial.peak.pos” + tab$ + “initial.low.pos” + tab$ + “final.peak.pos” + tab$
+ “final.low.pos” + tab$ + “last.syl.in.word.voiceless” + tab$ +
... newline$
header_row$ > ‘directory$’/results.txt

#Check whether there are pitch files already that have been used and corrected
manually before WARNING: You must have one pitch file per each wav file for this
to work. If you don’t, the script will overwrite your existing pitch files
# List all pitch files

Create Strings as file list... list ‘pitch_dir$’/*.Pitch
numberofPitchfiles = Get number of strings

# List all WAV files
Create Strings as file list... list ‘wav_dir$’/*.wav
numberOfFiles = Get number of strings

# Loop that goes through all files

for ifile to numberOfFiles
         select Strings list
         fileName$ = Get string... ifile
         baseName$ = fileName$ - “.wav”
         # Read in the Sound files with that base name
         Read from file... ‘wav_dir$’/’baseName$’.wav
         Read from file... ‘txt_dir$’/’baseName$’.TextGrid

         if ‘numberofPitchfiles’ = ‘numberOfFiles’

                  Read from file... ‘txt_dir$’/’baseName$’.Pitch
         else

# # Create Pitch object
                  select Sound ‘baseName$’
                  To Pitch (ac)... 0.005 75 15 no 0.03 0.45 0.01 0.35 0.30 600
         endif

# Select the individual words

         select TextGrid ‘baseName$’
```

```
# Determine the number of non-empty intervals on the word tier, i.e. the number of
words in the utterance

         nwordintervaltier = Get number of intervals... ‘wordtier’
         nemptyintervalswordtier = Count intervals where: ‘wordtier’, “is equal
to”, “”
         nwords = ‘nwordintervaltier’ - ‘nemptyintervalswordtier’

# Get start and end times for utterance (including pauses)

         startutterance = Get starting point... ‘wordtier’ 2
         endutterance = Get end point... ‘wordtier’ ‘nwordintervaltier’-1
         lengthutterance = (endutterance - startutterance)*1000

         startfile = Get starting point... ‘wordtier’ 1

         firstsyl = Get interval at time... ‘speakersyltier’
‘startutterance’+0.0001
         lastsyl = Get interval at time... ‘speakersyltier’ ‘endutterance’-0.0001
         firstsyluttend = Get end point... ‘speakersyltier’ ‘firstsyl’
         lastsylbegin = Get start point... ‘speakersyltier’ ‘lastsyl’

         select Pitch ‘baseName$’

         View & Edit

         pause Edit pitch object, then press “continue” to save and proceed to the
next file

         select TextGrid ‘baseName$’

# Go through each individual word

         for a from 1 to nwordintervaltier
                  select TextGrid ‘baseName$’
                  wordlabel$ = Get label of interval... ‘wordtier’ ‘a’

                  if wordlabel$ <> “”
                           startword = Get starting point... ‘wordtier’ ‘a’
                           endword = Get end point... ‘wordtier’ ‘a’
                           lengthword = (endword - startword)*1000

# Get number of syllables on the word

                           sylintervalstartword = Get interval at time...
‘speakersyltier’ ‘startword’+0.0001
                           sylintervalendword = Get interval at time...
‘speakersyltier’ ‘endword’-0.0001
                           numbersyls = ‘sylintervalendword’ -
‘sylintervalstartword’+1
```

```
                    firstsylwordend = Get end point... ‘speakersyltier’
‘sylintervalstartword’
                    secondsylend = Get end point... ‘speakersyltier’
‘sylintervalstartword’+1
                    penultbegin = Get start point... ‘speakersyltier’
‘sylintervalendword’-1
                    penultend = Get end point... ‘speakersyltier’
‘sylintervalendword’-1

                    if ‘sylintervalendword’-2 >= 1
                            antepenultbegin = Get start point...
‘speakersyltier’ ‘sylintervalendword’-2
                    endif

                    lengthfirstsecondsyl = ‘secondsylend’-’startword’
                    lengthpenultfinalsyl = ‘endword’-’penultbegin’

# Find out whether the word is initial medial or final

                    leftwordinterval = Get interval at time... ‘wordtier’
‘startword’-0.0001
                    rightwordinterval = Get interval at time... ‘wordtier’
‘endword’+0.0001
                    leftwordlabel$ = Get label of interval... ‘wordtier’
‘leftwordinterval’
                    rightwordlabel$ = Get label of interval... ‘wordtier’
‘rightwordinterval’

                    if leftwordlabel$ <> “” and rightwordlabel$ <> “”
                            wordposition$ = “medial”
                    endif

                    if leftwordlabel$ <> “” and rightwordlabel$ = “”
                            wordposition$ = “final”
                    endif

                    if leftwordlabel$ = “” and rightwordlabel$ <> “”
                            wordposition$ = “initial”
                    endif

                    if leftwordlabel$ = “” and rightwordlabel$ = “”
                            wordposition$ = “standalone”
                    endif

# Move on to pitch object

                    select Pitch ‘baseName$’

                    meanF0_firstsyl = Get mean... (‘startutterance’-
’startfile’) (‘firstsyluttend’-’startfile’) Hertz
```

```
				meanF0_lastsyl = Get mean... ('lastsylbegin'-
'startfile') ('endutterance'-'startfile') Hertz

# test whether the whole word is voiceless and then proceed accordingly

				testmeanF0_word = Get mean... ('startword'-'startfile')
('endword'-'startfile') Hertz
				testmeanF0_wordstring$ = "'testmeanF0_word'"

#THIS IS THE !!!OPTIONAL!!! IF-CONDITION CHECKING WHETHER THE WORD HAS ANY VOICING
AT ALL - IT'S NOT TABSPACED THE WAY IT SHOULD BE BECAUSE THAT WOULD HAVE MEANT
TABSPACING EVERYTHING ELSE AGAIN BUT IT IS SURROUNDED BY THIS WARNING TO MAKE YOU
AWARE
#				if testmeanF0_wordstring$ <> "--undefined--"
#THIS IS THE !!!OPTONAL!!! IF-CONDITION CHECKING WHETHER THE WORD HAS ANY VOICING
AT ALL - IT'S NOT TABSPACED THE WAY IT SHOULD BE BECAUSE THAT WOULD HAVE MEANT
TABSPACING EVERYTHING ELSE AGAIN BUT IT IS SURROUNDED BY THIS WARNING TO MAKE YOU
AWARE

# test whether last syllable is voiceless and then proceed accordingly

				testlastsylF0 = Get mean... 'penultend'-'startfile'
'endword'-'startfile' Hertz
				testlastsylF0string$ = "'testlastsylF0'"

				if testlastsylF0string$ = "--undefined--"

					reducednumbersyls = 'numbersyls'-1
				else
					reducednumbersyls = 'numbersyls'

				endif


				if 'reducednumbersyls' = 'numbersyls'


					if 'numbersyls' < 2

						tMaxF0init = Get time of maximum...
'startword'-'startfile' 'endword'-'startfile' Hertz Parabolic

						tMinF0init = Get time of minimum...
'startword'-'startfile' 'endword'-'startfile' Hertz Parabolic

						f0maxsylinit = Get value at time...
'tMaxF0init' Hertz Linear
						f0minsylinit = Get value at time...
'tMinF0init' Hertz Linear
```

```
                                             maxdistfromendoffirstsyl =
'tMaxF0init' + 'startfile' - 'firstsylwordend'
                                             mindistfromendoffirstsyl =
'tMinF0init' + 'startfile' -
'firstsylwordend'

                                             maxdistfromendofpenult =
'maxdistfromendoffirstsyl'
                                             mindistfromendofpenult =
'mindistfromendoffirstsyl'

                                             initialpitchrange =
'f0maxsylinit'-'f0minsylinit'
                                             initialtonedistance = ('tMaxF0init' -
'tMinF0init')*1000

                                             initialslope = 'initialpitchrange' /
'initialtonedistance'
                                             finalslope = 'initialpitchrange' /
'initialtonedistance'

                                             finalpitchrange = 'initialpitchrange'
                                             finaltonedistance =
'initialtonedistance'

                                             finalpeaksyl$ = "'NA'"
                                             finallowsyl$ = "'NA'"
                                             initialpeaksyl$ = "'NA'"
                                             initiallowsyl$ = "'NA'"

                                             tMaxF0initreal = 'tMaxF0init'+'
startfile'
                                             tMinF0initreal = 'tMinF0init'+'
startfile'

                                             tMaxF0finreal = 'tMaxF0initreal'
                                             tMinF0finreal = 'tMinF0initreal'

                                             f0maxsylfin = 'f0maxsylinit'
                                             f0minsylfin = 'f0minsylinit'

                                             meanF0firstsyl = Get mean...
'startword'-'startfile' 'firstsylwordend'-'startfile' Hertz

                                             meanF0firstsylstring$ =
"'meanF0firstsyl'"
                                             meanF0secondsylstring$ = "'NA'"
                                             meanF0penultstring$ = "'NA'"
                                             meanF0finalsylstring$ = "'NA'"
```

```
                                    initialslopestring$ =
"'initialslope'"
                                    finalslopestring$ = "'finalslope'"
                                    initialpitchrangestring$ =
"'initialpitchrange'"
                                    finalpitchrangestring$ =
"'finalpitchrange'"
                                    initialtonedistancestring$ =
"'initialtonedistance'"
                                    finaltonedistancestring$ =
"'finaltonedistance'"
                                    f0maxsylinitstring$ =
"'f0maxsylinit'"
                                    f0minsylinitstring$ =
"'f0minsylinit'"
                                    f0maxsylfinstring$ = "'f0maxsylfin'"
                                    f0maxsylfinstring$ = "'f0maxsylfin'"
                                    f0minsylinitstring$ =
"'f0minsylinit'"
                                    f0minsylfinstring$ = "'f0minsylfin'"
                                    tMaxF0initrealstring$ =
"'tMaxF0initreal'"
                                    tMinF0initrealstring$ =
"'tMinF0initreal'"
                                    tMaxF0finrealstring$ =
"'tMaxF0finreal'"
                                    tMinF0finrealstring$ =
"'tMinF0finreal'"


                                    secondsylendstring$ = "'NA'"
                                    antepenultbeginstring$ = "'NA'"
                                    penultbeginstring$ = "'NA'"
                                    penultendstring$ = "'NA'"

                                    maxdistfromendoffirstsylstring$ =
"'maxdistfromendoffirstsyl'"
                                    mindistfromendoffirstsylstring$ =
"'mindistfromendoffirstsyl'"
                                    maxdistfromendofpenultstring$ =
"'maxdistfromendofpenult'"
                                    mindistfromendofpenultstring$ =
"'mindistfromendofpenult'"
                            endif

                            if 'numbersyls' = 2

                                    tMaxF0init = Get time of maximum...
'startword'-'startfile' 'secondsylend'-'startfile' Hertz Parabolic
```

```
                                             tMinF0init = Get time of minimum...
'startword'-'startfile' 'secondsylend'-'startfile' Hertz Parabolic

                                             f0maxsylinit = Get value at time...
'tMaxF0init' Hertz Linear
                                             f0minsylinit = Get value at time...
'tMinF0init' Hertz Linear

                                                      if 'tMaxF0init' <=
('secondsylend'-'startfile') and 'tMaxF0init' > ('firstsylwordend'-'startfile')
                                                               initialpeaksyl$ =
"second"
                                                      endif

                                                      if 'tMaxF0init' <= ('first-
sylwordend'-'startfile') and 'tMaxF0init' >= ('startword'-'startfile')
                                                               initialpeaksyl$ =
"first"
                                                      endif

                                                      if 'tMinF0init' <= ('sec-
ondsylend'-'startfile') and 'tMinF0init' > ('firstsylwordend'-'startfile')
                                                               initiallowsyl$ =
"second"
                                                      endif

                                                      if 'tMinF0init' <= ('first-
sylwordend'-'startfile') and 'tMinF0init' >= ('startword' - 'startfile')
                                                               initiallowsyl$ =
"first"
                                                      endif

                                             maxdistfromendoffirstsyl =
'tMaxF0init' + 'startfile' - 'firstsylwordend'
                                             mindistfromendoffirstsyl =
'tMinF0init' + 'startfile' - 'firstsylwordend'

                                             initialpitchrange =
'f0maxsylinit'-'f0minsylinit'
                                             initialtonedistance = ('tMaxF0init' -
'tMinF0init')*1000

                                             finalpeaksyl$ = "'initialpeaksyl$'"
                                             finallowsyl$ = "'initiallowsyl$'"

                                             initialslope = 'initialpitchrange' /
'initialtonedistance'

                                             tMaxF0fin = Get time of maximum...
'penultbegin'-'startfile' 'endword'-'startfile' Hertz Parabolic
```

```
                                             tMinF0fin = Get time of minimum...
'penultbegin'-'startfile' 'endword'-'startfile' Hertz Parabolic

                                             f0maxsylfin = Get value at time...
'tMaxF0fin' Hertz Linear
                                             f0minsylfin = Get value at time...
'tMinF0fin' Hertz Linear

                                             finalpitchrange =
'f0maxsylfin'-'f0minsylfin'
                                             finaltonedistance = ('tMaxF0fin' -
'tMinF0fin')*1000

                                             finalslope = 'finalpitchrange' /
'finaltonedistance'

                                             tMaxF0initreal = 'tMaxF0init'+
'startfile'
                                             tMinF0initreal = 'tMinF0init'+
'startfile'
                                             tMaxF0finreal = 'tMaxF0fin'+
'startfile'
                                             tMinF0finreal = 'tMinF0fin'+
'startfile'

                                             meanF0firstsyl = Get mean...
'startword'-'startfile' 'firstsylwordend'-'startfile' Hertz
                                             meanF0finalsyl = Get mean...
'penultend'-'startfile' 'endword'-'startfile' Hertz

                                             maxdistfromendofpenult =
'maxdistfromendoffirstsyl'
                                             mindistfromendofpenult =
'mindistfromendoffirstsyl'

                                             meanF0firstsylstring$ =
"'meanF0firstsyl'"
                                             meanF0secondsylstring$ = "'NA'"
                                             meanF0penultstring$ = "'NA'"
                                             meanF0finalsylstring$ =
"'meanF0finalsyl'"

                                             initialslopestring$ =
"'initialslope'"
                                             finalslopestring$ = "'finalslope'"
                                             initialpitchrangestring$ =
"'initialpitchrange'"
                                             finalpitchrangestring$ =
"'finalpitchrange'"
                                             initialtonedistancestring$ =
"'initialtonedistance'"
```

```
                                             finaltonedistancestring$ =
"'finaltonedistance'"
                                             f0maxsylinitstring$ =
"'f0maxsylinit'"
                                             f0minsylinitstring$ =
"'f0minsylinit'"
                                             f0maxsylfinstring$ = "'f0maxsylfin'"
                                             f0maxsylfinstring$ = "'f0maxsylfin'"
                                             f0minsylinitstring$ =
"'f0minsylinit'"
                                             f0minsylfinstring$ = "'f0minsylfin'"
                                             tMaxF0initrealstring$ =
"'tMaxF0initreal'"
                                             tMinF0initrealstring$ =
"'tMinF0initreal'"
                                             tMaxF0finrealstring$ =
"'tMaxF0finreal'"
                                             tMinF0finrealstring$ =
"'tMinF0finreal'"


                                             secondsylendstring$ =
"'secondsylend'"
                                             antepenultbeginstring$ = "'NA'"
                                             penultbeginstring$ = "'NA'"
                                             penultendstring$ = "'NA'"

                                             maxdistfromendoffirstsylstring$ =
"'maxdistfromendoffirstsyl'"
                                             mindistfromendoffirstsylstring$ =
"'mindistfromendoffirstsyl'"
                                             maxdistfromendofpenultstring$ =
"'maxdistfromendofpenult'"
                                             mindistfromendofpenultstring$ =
"'mindistfromendofpenult'"

                                    endif

                                    if 'numbersyls' > 2

                                             tMaxF0init = Get time of maximum...
'startword'-'startfile' 'secondsylend'-'startfile' Hertz Parabolic

                                             tMinF0init = Get time of minimum...
'startword'-'startfile' 'secondsylend'-'startfile' Hertz Parabolic

                                             f0maxsylinit = Get value at time...
'tMaxF0init' Hertz Linear
                                             f0minsylinit = Get value at time...
'tMinF0init' Hertz Linear
```

```
                                        if ‘tMaxF0init’ <=
(‘secondsylend’-’startfile’) and ‘tMaxF0init’ > (‘firstsylwordend’-’startfile’)
                                                  initialpeaksyl$ =
“second”
                                        endif

                                        if ‘tMaxF0init’ <=
(‘firstsylwordend’-’startfile’) and ‘tMaxF0init’ >= (‘startword’-’startfile’)
                                                  initialpeaksyl$ =
“first”
                                        endif

                                        if ‘tMinF0init’ <=
(‘secondsylend’-’startfile’) and ‘tMinF0init’ > (‘firstsylwordend’-’startfile’)
                                                  initiallowsyl$ =
“second”
                                        endif

                                        if ‘tMinF0init’ <=
(‘firstsylwordend’-’startfile’) and ‘tMinF0init’ >= (‘startword’-’startfile’)
                                                  initiallowsyl$ =
“first”
                                        endif

                              maxdistfromendoffirstsyl =
‘tMaxF0init’ + ‘startfile’ - ‘firstsylwordend’
                              mindistfromendoffirstsyl =
‘tMinF0init’ + ‘startfile’ - ‘firstsylwordend’

                              initialpitchrange = ‘f0maxsylinit’-
’f0minsylinit’
                              initialtonedistance = (‘tMaxF0init’ -
‘tMinF0init’)*1000

                              initialslope = ‘initialpitchrange’ /
‘initialtonedistance’

                              tMaxF0fin = Get time of maximum...
‘penultbegin’-’startfile’ ‘endword’-’startfile’ Hertz Parabolic

                              tMinF0fin = Get time of minimum...
‘penultbegin’-’startfile’ ‘endword’-’startfile’ Hertz Parabolic

                              f0maxsylfin = Get value at time...
‘tMaxF0fin’ Hertz Linear
                              f0minsylfin = Get value at time...
‘tMinF0fin’ Hertz Linear

                                        if ‘tMaxF0fin’ <=
(‘endword’-’startfile’) and ‘tMaxF0fin’ > (‘penultend’-’startfile’)
```

```
                                                  finalpeaksyl$ =
"final"
                                        endif

                                        if 'tMaxF0fin' <=
('penultend'-'startfile') and 'tMaxF0fin' >= ('penultbegin'-'startfile')
                                                  finalpeaksyl$ =
"penult"
                                        endif

                                        if 'tMinF0fin' <=
('endword'-'startfile') and 'tMinF0fin' > ('penultend'-'startfile')
                                                  finallowsyl$ =
"final"
                                        endif

                                        if 'tMinF0fin' <=
('penultend'-'startfile') and 'tMinF0fin' >= ('penultbegin' - 'startfile')
                                                  finallowsyl$ =
"penult"
                                        endif

                              maxdistfromendofpenult = 'tMaxF0fin'
+ 'startfile' - 'penultend'
                              mindistfromendofpenult = 'tMinF0fin'
+ 'startfile' - 'penultend'

                              finalpitchrange = 'f0maxsylfin'-
'f0minsylfin'
                              finaltonedistance = ('tMaxF0fin' -
'tMinF0fin')*1000

                              finalslope = 'finalpitchrange' /
'finaltonedistance'

                              tMaxF0initreal =
'tMaxF0init'+'startfile'
                              tMinF0initreal =
'tMinF0init'+'startfile'
                              tMaxF0finreal =
'tMaxF0fin'+'startfile'
                              tMinF0finreal =
'tMinF0fin'+'startfile'

                              meanF0firstsyl = Get mean...
'startword'-'startfile' 'firstsylwordend'-'startfile' Hertz
                              meanF0secondsyl = Get mean...
'firstsylwordend'-'startfile' 'secondsylend'-'startfile' Hertz
                              meanF0penult = Get mean...
'penultbegin'-'startfile' 'penultend'-'startfile' Hertz
```

```
meanF0finalsyl = Get mean...
'penultend'-'startfile' 'endword'-'startfile' Hertz

meanF0firstsylstring$ =
"'meanF0firstsyl'"
meanF0secondsylstring$ =
"'meanF0secondsyl'"
meanF0penultstring$ =
"'meanF0penult'"
meanF0finalsylstring$ =
"'meanF0finalsyl'"

initialslopestring$ =
"'initialslope'"
finalslopestring$ = "'finalslope'"
initialpitchrangestring$ =
"'initialpitchrange'"
finalpitchrangestring$ =
"'finalpitchrange'"
initialtonedistancestring$ =
"'initialtonedistance'"
finaltonedistancestring$ =
"'finaltonedistance'"
f0maxsylinitstring$ =
"'f0maxsylinit'"
f0minsylinitstring$ =
"'f0minsylinit'"
f0maxsylfinstring$ = "'f0maxsylfin'"
f0maxsylfinstring$ = "'f0maxsylfin'"
f0minsylinitstring$ =
"'f0minsylinit'"
f0minsylfinstring$ = "'f0minsylfin'"
tMaxF0initrealstring$ =
"'tMaxF0initreal'"
tMinF0initrealstring$ =
"'tMinF0initreal'"
tMaxF0finrealstring$ =
"'tMaxF0finreal'"
tMinF0finrealstring$ =
"'tMinF0finreal'"

secondsylendstring$ =
"'secondsylend'"
antepenultbeginstring$ =
"'antepenultbegin'"
penultbeginstring$ = "'penultbegin'"
penultendstring$ = "'penultend'"

maxdistfromendoffirstsylstring$ =
"'maxdistfromendoffirstsyl'"
```

```
                                        mindistfromendoffirstsylstring$ =
"'mindistfromendoffirstsyl'"
                                        maxdistfromendofpenultstring$ =
"'maxdistfromendofpenult'"
                                        mindistfromendofpenultstring$ =
"'mindistfromendofpenult'"

                              endif

                              lastsylvoiceless$ = "NO"

                    endif

                    if 'reducednumbersyls' = 'numbersyls'-1

                              if 'reducednumbersyls' < 2

                                        tMaxF0init = Get time of maximum...
'startword'-'startfile' 'firstsylwordend'-'startfile' Hertz Parabolic

                                        tMinF0init = Get time of minimum...
'startword'-'startfile' 'firstsylwordend'-'startfile' Hertz Parabolic

                                        tMaxF0initstring$ = "'tMaxF0init'"
                                        tMinF0initstring$ = "'tMinF0init'"

                                        if tMaxF0initstring$ <>
"--undefined--"

                                                  f0maxsylinit = Get value at
time... 'tMaxF0init' Hertz Linear
                                                  f0minsylinit = Get value at
time... 'tMinF0init' Hertz Linear

                                                  maxdistfromendoffirstsyl =
'tMaxF0init' + 'startfile' - 'firstsylwordend'
                                                  mindistfromendoffirstsyl =
'tMinF0init' + 'startfile' - 'firstsylwordend'

                                                  maxdistfromendofpenult =
'maxdistfromendoffirstsyl'
                                                  mindistfromendofpenult =
'mindistfromendoffirstsyl'

                                                  initialpitchrange =
'f0maxsylinit'-'f0minsylinit'
                                                  initialtonedistance =
('tMaxF0init' - 'tMinF0init')*1000

                                                  initialslope =
'initialpitchrange' / 'initialtonedistance'
```

```
                                                    finalslope =
'initialpitchrange' / 'initialtonedistance'

                                                    finalpitchrange =
'initialpitchrange'
                                                    finaltonedistance =
'initialtonedistance'

                                                    tMaxF0initreal =
'tMaxF0init'+'startfile'
                                                    tMinF0initreal =
'tMinF0init'+'startfile'

                                                    tMaxF0finreal =
'tMaxF0initreal'
                                                    tMinF0finreal =
'tMinF0initreal'

                                                    f0maxsylfin = 'f0maxsylinit'
                                                    f0minsylfin = 'f0minsylinit'

                                                    meanF0firstsyl = Get mean...
'startword'-'startfile' 'firstsylwordend'-'startfile' Hertz

                                                    meanF0firstsylstring$ =
"'meanF0firstsyl'"
                                                    meanF0secondsylstring$ =
"'NA'"
                                                    meanF0penultstring$ = "'NA'"
                                                    meanF0finalsylstring$ =
"'NA'"

                                                    initialslopestring$ =
"'initialslope'"
                                                    finalslopestring$ =
"'finalslope'"
                                                    initialpitchrangestring$ =
"'initialpitchrange'"
                                                    finalpitchrangestring$ =
"'finalpitchrange'"
                                                    initialtonedistancestring$ =
"'initialtonedistance'"
                                                    finaltonedistancestring$ =
"'finaltonedistance'"
                                                    f0maxsylinitstring$ =
"'f0maxsylinit'"
                                                    f0minsylinitstring$ =
"'f0minsylinit'"
                                                    f0maxsylfinstring$ =
"'f0maxsylfin'"
```

```
                                                                      f0maxsylfinstring$ =
“’f0maxsylfin’”
                                                                      f0minsylinitstring$ =
“’f0minsylinit’”
                                                                      f0minsylfinstring$ =
“’f0minsylfin’”
                                                                      tMaxF0initrealstring$ =
“’tMaxF0initreal’”
                                                                      tMinF0initrealstring$ =
“’tMinF0initreal’”
                                                                      tMaxF0finrealstring$ =
“’tMaxF0finreal’”
                                                                      tMinF0finrealstring$ =
“’tMinF0finreal’”

maxdistfromendoffirstsylstring$ = “’maxdistfromendoffirstsyl’”

mindistfromendoffirstsylstring$ = “’mindistfromendoffirstsyl’”

maxdistfromendofpenultstring$ = “’maxdistfromendofpenult’”

mindistfromendofpenultstring$ = “’mindistfromendofpenult’”


                                                             else
                                                                      initialslopestring$ = “’NA’”
                                                                      finalslopestring$ = “’NA’”
                                                                      initialpitchrangestring$ =
“’NA’”
                                                                      finalpitchrangestring$ =
“’NA’”
                                                                      initialtonedistancestring$ =
“’NA’”
                                                                      finaltonedistancestring$ =
“’NA’”
                                                                      f0maxsylinitstring$ = “’NA’”
                                                                      f0minsylinitstring$ = “’NA’”
                                                                      f0maxsylfinstring$ = “’NA’”
                                                                      f0maxsylfinstring$ = “’NA’”
                                                                      f0minsylinitstring$ = “’NA’”
                                                                      f0minsylfinstring$ = “’NA’”
                                                                      tMaxF0initrealstring$ =
“’NA’”
                                                                      tMinF0initrealstring$ =
“’NA’”
                                                                      tMaxF0finrealstring$ =
“’NA’”
                                                                      tMinF0finrealstring$ =
“’NA’”
                                                                      meanF0firstsylstring$ =
“’NA’”
```

```
                                                meanF0secondsylstring$ =
“’NA’”
                                                meanF0penultstring$ = “’NA’”
                                                meanF0finalsylstring$ =
“’NA’”

maxdistfromendoffirstsylstring$ = “’NA’”

mindistfromendoffirstsylstring$ = “’NA’”

maxdistfromendofpenultstring$ = “’NA’”

mindistfromendofpenultstring$ = “’NA’”
                                        endif

                                                finalpeaksyl$ = “’NA’”
                                                finallowsyl$ = “’NA’”
                                                initialpeaksyl$ = “’NA’”
                                                initiallowsyl$ = “’NA’”


                                        secondsylendstring$ = “’NA’”
                                        antepenultbeginstring$ = “’NA’”
                                        penultbeginstring$ = “’NA’”
                                        penultendstring$ = “’NA’”

                                endif

                                if ‘reducednumbersyls’ = 2

                                        tMaxF0init = Get time of maximum...
‘startword’-’startfile’ ‘secondsylend’-’startfile’ Hertz Parabolic

                                        tMinF0init = Get time of minimum...
‘startword’-’startfile’ ‘secondsylend’-’startfile’ Hertz Parabolic

                                        f0maxsylinit = Get value at time...
‘tMaxF0init’ Hertz Linear
                                        f0minsylinit = Get value at time...
‘tMinF0init’ Hertz Linear

                                                if ‘tMaxF0init’ <=
(‘secondsylend’-’startfile’) and ‘tMaxF0init’ > (‘firstsylwordend’-’startfile’)
                                                        initialpeaksyl$ =
“second”
                                                endif

                                                if ‘tMaxF0init’ <=
(‘firstsylwordend’-’startfile’) and ‘tMaxF0init’ >= (‘startword’-’startfile’)
                                                        initialpeaksyl$ =
“first”
```

```
                                                            endif

                                                            if ‘tMinF0init’ <=
(‘secondsylend’-’startfile’) and ‘tMinF0init’ > (‘firstsylwordend’-’startfile’)
                                                                    initiallowsyl$ =
“second”
                                                            endif

                                                            if ‘tMinF0init’ <=
(‘firstsylwordend’-’startfile’) and ‘tMinF0init’ >= (‘startword’ - ‘startfile’)
                                                                    initiallowsyl$ =
“first”
                                                            endif

                                                    finalpeaksyl$ = “’initialpeaksyl$’”
                                                    finallowsyl$ = “’initiallowsyl$’”

                                                    maxdistfromendoffirstsyl =
‘tMaxF0init’ + ‘startfile’ - ‘firstsylwordend’
                                                    mindistfromendoffirstsyl =
‘tMinF0init’ + ‘startfile’ - ‘firstsylwordend’

                                                    initialpitchrange = ‘f0maxsylinit’-
’f0minsylinit’
                                                    initialtonedistance = (‘tMaxF0init’ -
‘tMinF0init’)*1000

                                                    initialslope = ‘initialpitchrange’ /
‘initialtonedistance’

                                                    tMaxF0fin = ‘tMaxF0init’
                                                    tMinF0fin = ‘tMinF0init’

                                                    f0maxsylfin = Get value at time...
‘tMaxF0fin’ Hertz Linear
                                                    f0minsylfin = Get value at time...
‘tMinF0fin’ Hertz Linear

                                                    finalpitchrange = ‘f0maxsylfin’-
’f0minsylfin’
                                                    finaltonedistance = (‘tMaxF0fin’ -
‘tMinF0fin’)*1000

                                                    finalslope = ‘finalpitchrange’ /
‘finaltonedistance’

                                                    tMaxF0initreal =
‘tMaxF0init’+’startfile’
                                                    tMinF0initreal =
‘tMinF0init’+’startfile’
```

```
                                        tMaxF0finreal =
'tMaxF0fin'+'startfile'
                                        tMinF0finreal =
'tMinF0fin'+'startfile'

                                        maxdistfromendofpenult =
'maxdistfromendoffirstsyl'
                                        mindistfromendofpenult =
'mindistfromendoffirstsyl'

                                        meanF0firstsyl = Get mean...
'startword'-'startfile' 'firstsylwordend'-'startfile' Hertz
                                        meanF0finalsyl = Get mean...
'firstsylwordend'-'startfile' 'secondsylend'-'startfile' Hertz

                                        meanF0firstsylstring$ =
"'meanF0firstsyl'"
                                        meanF0secondsylstring$ = "'NA'"
                                        meanF0penultstring$ = "'NA'"
                                        meanF0finalsylstring$ =
"'meanF0finalsyl'"

                                        initialslopestring$ =
"'initialslope'"
                                        finalslopestring$ = "'finalslope'"
                                        initialpitchrangestring$ =
"'initialpitchrange'"
                                        finalpitchrangestring$ =
"'finalpitchrange'"
                                        initialtonedistancestring$ =
"'initialtonedistance'"
                                        finaltonedistancestring$ =
"'finaltonedistance'"
                                        f0maxsylinitstring$ =
"'f0maxsylinit'"
                                        f0minsylinitstring$ =
"'f0minsylinit'"
                                        f0maxsylfinstring$ = "'f0maxsylfin'"
                                        f0maxsylfinstring$ = "'f0maxsylfin'"
                                        f0minsylinitstring$ =
"'f0minsylinit'"
                                        f0minsylfinstring$ = "'f0minsylfin'"
                                        tMaxF0initrealstring$ =
"'tMaxF0initreal'"
                                        tMinF0initrealstring$ =
"'tMinF0initreal'"
                                        tMaxF0finrealstring$ =
"'tMaxF0finreal'"
                                        tMinF0finrealstring$ =
"'tMinF0finreal'"
```

```
                                             secondsylendstring$ =
"'secondsylend'"
                                             antepenultbeginstring$ = "'NA'"
                                             penultbeginstring$ = "'NA'"
                                             penultendstring$ = "'NA'"

                                             maxdistfromendoffirstsylstring$ =
"'maxdistfromendoffirstsyl'"
                                             mindistfromendoffirstsylstring$ =
"'mindistfromendoffirstsyl'"
                                             maxdistfromendofpenultstring$ =
"'maxdistfromendofpenult'"
                                             mindistfromendofpenultstring$ =
"'mindistfromendofpenult'"

                                    endif

                                    if 'reducednumbersyls' > 2

                                             tMaxF0init = Get time of maximum...
'startword'-'startfile' 'secondsylend'-'startfile' Hertz Parabolic

                                             tMinF0init = Get time of minimum...
'startword'-'startfile' 'secondsylend'-'startfile' Hertz Parabolic

                                             f0maxsylinit = Get value at time...
'tMaxF0init' Hertz Linear
                                             f0minsylinit = Get value at time...
'tMinF0init' Hertz Linear

                                                      if 'tMaxF0init' <=
('secondsylend'-'startfile') and 'tMaxF0init' > ('firstsylwordend'-'startfile')
                                                               initialpeaksyl$ =
"second"
                                                      endif

                                                      if 'tMaxF0init' <=
('firstsylwordend'-'startfile') and 'tMaxF0init' >= ('startword'-'startfile')
                                                               initialpeaksyl$ =
"first"
                                                      endif

                                                      if 'tMinF0init' <=
('secondsylend'-'startfile') and 'tMinF0init' > ('firstsylwordend'-'startfile')
                                                               initiallowsyl$ =
"second"
                                                      endif

                                                      if 'tMinF0init' <=
('firstsylwordend'-'startfile') and 'tMinF0init' >= ('startword' - 'startfile')
```

```
                                                            initiallowsyl$ =
"first"
                                                    endif

                                            maxdistfromendoffirstsyl =
'tMaxF0init' + 'startfile' - 'firstsylwordend'
                                            mindistfromendoffirstsyl =
'tMinF0init' + 'startfile' - 'firstsylwordend'

                                            initialpitchrange = 'f0maxsylinit'-
'f0minsylinit'
                                            initialtonedistance = ('tMaxF0init' -
'tMinF0init')*1000

                                            initialslope = 'initialpitchrange' /
'initialtonedistance'

                                            tMaxF0fin = Get time of maximum...
'antepenultbegin'-'startfile' 'penultend'-'startfile' Hertz Parabolic

                                            tMinF0fin = Get time of minimum...
'antepenultbegin'-'startfile' 'penultend'-'startfile' Hertz Parabolic

                                            f0maxsylfin = Get value at time...
'tMaxF0fin' Hertz Linear
                                            f0minsylfin = Get value at time...
'tMinF0fin' Hertz Linear

                                                    if 'tMaxF0fin' <=
('penultend'-'startfile') and 'tMaxF0fin' > ('penultbegin'-'startfile')
                                                            finalpeaksyl$ =
"penult"
                                                    endif

                                                    if 'tMaxF0fin' <=
('penultbegin'-'startfile') and 'tMaxF0fin' >= ('antepenultbegin'-'startfile')
                                                            finalpeaksyl$ =
"antepenult"
                                                    endif

                                                    if 'tMinF0fin' <=
('penultend'-'startfile') and 'tMinF0fin' > ('penultbegin' - 'startfile')
                                                            finallowsyl$ =
"penult"
                                                    endif

                                                    if 'tMinF0fin' <=
('penultbegin'-'startfile') and 'tMinF0fin' >= ('antepenultbegin'-'startfile')
                                                            finallowsyl$ =
"antepenult"
                                                    endif
```

```
                                             finalpitchrange = ‘f0maxsylfin’-
’f0minsylfin’
                                             finaltonedistance = (‘tMaxF0fin’ -
‘tMinF0fin’)*1000

                                             maxdistfromendofpenult = ‘tMaxF0fin’
+ ‘startfile’ - ‘penultbegin’
                                             mindistfromendofpenult = ‘tMinF0fin’
+ ‘startfile’ - ‘penultbegin’

                                             finalslope = ‘finalpitchrange’ /
‘finaltonedistance’

                                             tMaxF0initreal =
‘tMaxF0init’+’startfile’
                                             tMinF0initreal =
‘tMinF0init’+’startfile’
                                             tMaxF0finreal =
‘tMaxF0fin’+’startfile’
                                             tMinF0finreal =
‘tMinF0fin’+’startfile’

                                             meanF0firstsyl = Get mean...
‘startword’-’startfile’ ‘firstsylwordend’-’startfile’ Hertz
                                             meanF0secondsyl = Get mean...
‘firstsylwordend’-’startfile’ ‘secondsylend’-’startfile’ Hertz
                                             meanF0penult = Get mean...
‘antepenultbegin’-’startfile’ ‘penultbegin’-’startfile’ Hertz
                                             meanF0finalsyl = Get mean...
‘penultbegin’-’startfile’ ‘penultend’-’startfile’ Hertz

                                             meanF0firstsylstring$ =
“’meanF0firstsyl’”
                                             meanF0secondsylstring$ =
“’meanF0secondsyl’”
                                             meanF0penultstring$ =
“’meanF0penult’”
                                             meanF0finalsylstring$ =
“’meanF0finalsyl’”

                                             initialslopestring$ =
“’initialslope’”
                                             finalslopestring$ = “’finalslope’”
                                             initialpitchrangestring$ =
“’initialpitchrange’”
                                             finalpitchrangestring$ =
“’finalpitchrange’”
                                             initialtonedistancestring$ =
“’initialtonedistance’”
                                             finaltonedistancestring$ =
“’finaltonedistance’”
```

```
                                            f0maxsylinitstring$ =
"'f0maxsylinit'"
                                            f0minsylinitstring$ =
"'f0minsylinit'"
                                            f0maxsylfinstring$ = "'f0maxsylfin'"
                                            f0maxsylfinstring$ = "'f0maxsylfin'"
                                            f0minsylinitstring$ =
"'f0minsylinit'"
                                            f0minsylfinstring$ = "'f0minsylfin'"
                                            tMaxF0initrealstring$ =
"'tMaxF0initreal'"
                                            tMinF0initrealstring$ =
"'tMinF0initreal'"
                                            tMaxF0finrealstring$ =
"'tMaxF0finreal'"
                                            tMinF0finrealstring$ =
"'tMinF0finreal'"

                                            secondsylendstring$ =
"'secondsylend'"
                                            antepenultbeginstring$ =
"'antepenultbegin'"
                                            penultbeginstring$ = "'penultbegin'"
                                            penultendstring$ = "'penultend'"

                                            maxdistfromendoffirstsylstring$ =
"'maxdistfromendoffirstsyl'"
                                            mindistfromendoffirstsylstring$ =
"'mindistfromendoffirstsyl'"
                                            maxdistfromendofpenultstring$ =
"'maxdistfromendofpenult'"
                                            mindistfromendofpenultstring$ =
"'mindistfromendofpenult'"

                                   endif

                                   lastsylvoiceless$ = "YES"

                          endif

#THIS IS THE !!!OPTIONAL!!! IF-CONDITION CHECKING WHETHER THE WORD HAS ANY VOICING
AT ALL - IT'S NOT TABSPACED THE WAY IT SHOULD BE BECAUSE THAT WOULD HAVE MEANT
TABSPACING EVERYTHING ELSE AGAIN BUT IT IS SURROUNDED BY THIS WARNING TO MAKE YOU
AWARE
#                         endif
#THIS IS THE !!!OPTIONAL!!! IF-CONDITION CHECKING WHETHER THE WORD HAS ANY VOICING
AT ALL - IT'S NOT TABSPACED THE WAY IT SHOULD BE BECAUSE THAT WOULD HAVE MEANT
TABSPACING EVERYTHING ELSE AGAIN BUT IT IS SURROUNDED BY THIS WARNING TO MAKE YOU
AWARE
```

```
#                                    print "initialslope" 'initialslope',
"finalslope" 'finalslope', "initialpitchrange" 'initialpitchrange', "initialtone-
distance" 'initialtonedistance', "finalpitchrange" 'finalpitchrange', "finaltone-
distance" 'finaltonedistance'

# save edited pitch object

                           Save as text file: "'directory$'\'baseName$'.Pitch"

# Append values in the TXT file

                           select TextGrid 'baseName$'

                           fileappend "'directory$'/results.txt" 'wordlabel$''tab$
''baseName$''tab$''speakerOne$''tab$'
                           ...'language$''tab$''typeutter-
ance$''tab$''nwords''tab$'
                           ...'startutterance''tab$''endutterance''tab$'
                           ...'lengthutterance''tab$''numbersyls''tab$'
                           ...'wordposition$''tab$''initialslopestring$''tab$'
                           ...'finalslopestring$''tab$''startword''tab$'
                           ...'endword''tab$''lengthword''tab$'
                           ...'initialpitchrangestring$''tab$''finalpitchrange-
string$''tab$'
                           ...'initialtonedistancestring$''tab$'
'finaltonedistancestring$''tab$''f0maxsylinitstring$''tab$'
                           ...'f0maxsylfinstring$''tab$''f0minsylinit-
string$''tab$'
                           ...'f0minsylfinstring$''tab$''tMaxF0initreal-
string$''tab$'
                           ...'tMaxF0finrealstring$''tab$''tMinF0initreal-
string$''tab$'
                           ...'tMinF0finrealstring$''tab$'
                           ...'meanF0_firstsyl''tab$''meanF0_lastsyl''tab$'
                           ...'meanF0firstsylstring$''tab$''meanF0secondsyl-
string$''tab$'
                           ...'meanF0penultstring$''tab$''meanF0finalsyl-
string$''tab$'
                           ...'maxdistfromendoffirstsylstring$''tab$'
'mindistfromendoffirstsylstring$''tab$''maxdistfromendofpenultstring$''tab$'
'mindistfromendofpenultstring$''tab$'
                           ...'startword''tab$''firstsylwordend''tab$'
'secondsylendstring$''tab$''antepenultbeginstring$''tab$''penultbeginstring$'
'tab$''penultendstring$''tab$''endword''tab$'
                           ...'lengthfirstsecondsyl''tab$''lengthpenultfinal-
syl''tab$''initialpeaksyl$''tab$'
                           ...'initiallowsyl$''tab$''finalpeaksyl$''tab$'
'finallowsyl$''tab$''lastsylvoiceless$''tab$'
                           ...'newline$'
```

```
                endif
        endfor

# Clean up before going on with next file
        select Sound 'baseName$'
        plus TextGrid 'baseName$'
        plus Pitch 'baseName$'
        Remove

endfor

# Final clean up
select Strings list
Remove
clearinfo
print All files done!
```

References

Adelaar, Willem F. H. 1977. *Tarma Quechua*: *Grammar, Texts, Dictionary*. Lisse: Peter de Ridder.

Adelaar, Willem F. H. & Pieter Muysken. 2004. *The languages of the Andes* (Cambridge language surveys). Cambridge, New York: Cambridge University Press.

Aguilar, Lourdes, Carme De-la-Mota & Pilar Prieto. 2009. Sp_ToBI Training Materials. http://prosodia.upf.edu/sp_tobi/ (16 January, 2019).

Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson & Regina Weinert. 1991. The Hcrc Map Task Corpus. *Language and Speech* 34(4). 351–366.

Andrade Ciudad, Luis. 2016. *The Spanish of the northern Peruvian Andes*: *A sociohistorical and dialectological account* (Historical sociolinguistics 3). Bern: Peter Lang.

Andrade Ciudad, Luis. 2021. La construcción del castellano andino en el Perú. In Luis Andrade Ciudad & Sandro Sessarego (eds.), *Los castellanos del Perú*: *Historia, variación y contacto lingüístico* (Routledge studies in Hispanic and Lusophone linguistics), 118–155. London: Routledge.

Armstrong, Meghan E. 2010. Puerto Rican Spanish Intonation. In Pilar Prieto & Paolo Roseano (eds.), *Transcription of intonation of the Spanish language*, 155–189. München: LINCOM.

Arnhold, Anja. 2014. Prosodic structure and focus realization in West Greenlandic. In Sun-Ah Jun (ed.), *Prosodic typology II*: *The phonology of intonation and phrasing*, 216–251. Oxford: Oxford University Press.

Arvaniti, Amalia & Mary Baltazani. 2005. Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora. In Sun-Ah Jun (ed.), *Prosodic typology*: *The phonology of intonation and phrasing* (Oxford linguistics), 84–117. Oxford, New York: Oxford University Press.

Arvaniti, Amalia, D. R. Ladd & Ineke Mennen. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26. 3–25.

Arvaniti, Amalia, D. R. Ladd & Ineke Mennen. 2000. What is a starred tone?: Evidence from Greek. In Michael Broe & Janet B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 119–131. Cambridge: Cambridge University Press.

Arvaniti, Amalia, D. R. Ladd & Ineke Mennen. 2006. Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. *Language and Speech* 49(4). 421–450.

Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2010. Turning points, tonal targets, and the English L- phrase accent. *Language and Cognitive Processes* 25(7–9). 982–1023.

Bat-El, Outi. 2005. The emergence of the trochaic foot in Hebrew hypocoristics. *Phonology* 22(2). 115–143.

Baumann, Stefan & Martine Grice. 2006. The intonation of accessibility. *Journal of Pragmatics* 38(10). 1636–1657.

Baumann, Stefan & Arndt Riester. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In Gorka Elordieta & Pilar Prieto (eds.), *Prosody and Meaning*: *[based on the Workshop on Prosody and Meaning in Barcelona on September 17–18, 2009]* (Interface Explorations / IE] 25), 119–161. Berlin: de Gruyter.

Beaver, David I., Brady Z. Clark, Edward S. Flemming, T. F. Jaeger & Maria. Wolters. 2007. When Semantics Meets Phonetics: Acoustical Studies of Second-Occurrence Focus. *Language* 83(2). 245–276.

Beaver, David I., Craige Roberts, Mandy Simons & Judith Tonhauser. 2017. Questions Under Discussion: Where Information Structure Meets Projective Content. *Annual Review of Linguistics* 3(1). 265–284.

https://doi.org/10.1515/9783111304595-011

Becker, Michael. 2003. Hebrew Stress: Can't You Hear Those Trochees? *University of Pennsylvania Working Papers in Linguistics* 9(1). 45–57.
Beckman, Mary E. 1986. *Stress and non-stress accent* (Netherlands Phonetic Archives 7). Dordrecht: Foris.
Beckman, Mary E., Manuel Díaz-Campos, Julia T. McGory & Terrell A. Morgan. 2002. Intonation across Spanish, in the Tones and Break Indices framework. *Probus* 14. 9–36.
Beckman, Mary E. & Jan Edwards. 1990. Lengthenings and shortenings and the nature of prosodic constituency. In John Kingston & Mary E. Beckman (eds.), *Between the grammar and physics of speech* (Papers in laboratory phonology 1), 152–178. Cambridge: Cambridge University Press.
Beckman, Mary E. & Julia Hirschberg. 1994. The ToBI Annotation Conventions. https://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html (10 January, 2019).
Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel. 2005. The Original ToBi System and the Evolution of the ToBi Framework. In Sun-Ah Jun (ed.), *Prosodic typology*: *The phonology of intonation and phrasing* (Oxford linguistics), 9–54. Oxford, New York: Oxford University Press.
Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational Structure in Japanese and English. *Phonology Yearbook* 3. 255–309.
Beckman, Mary E. & Jennifer J. Venditti. 2010. Tone and Intonation. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds.), *The Handbook of phonetic sciences*, 2nd edn. (Blackwell handbooks in linguistics), 603–652. Oxford: Blackwell Publishers.
Behnstedt, Peter & Manfred Woidich. 1985a. *Die ägyptisch-arabischen Dialekte*: *Band 1: Einleitung und Anmerkungen zu den Karten* (Tübinger Atlas des Vorderen Orients 50,1). Wiesbaden: L. Reichert.
Behnstedt, Peter & Manfred Woidich. 1985b. *Die ägyptisch-arabischen Dialekte*: *Band 2: Dialektatlas von Ägypten* (Tübinger Atlas des Vorderen Orients 50,2). Wiesbaden: L. Reichert.
Behrens, Leila. 2012. Evidentiality, modality, focus and other puzzles. In Andrea C. Schalley (ed.), *Practical Theories and Empirical Practice*: *A linguistic perspective* (Human cognitive processing), 185–243. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Bendezú Araujo, Raúl. 2021. *Identificación y aserción en la marcación de foco del quechua de Conchucos (Áncash, Perú).* Berlin: Freie Universität Dissertation.
Bendezú Araujo, Raúl, Timo Buchholz & Uli Reich. 2019. Corpora amerikanischer Sprachen: Interaktive Sprachspiele aus dem mehrsprachigen Lateinamerika (Quechua 1). https://refubium.fu-berlin.de/handle/fub188/25747.
Bendezú Araujo, Raúl, Timo Buchholz & Uli Reich. 2021. Corpora amerikanischer Sprachen: Interaktive Sprachspiele aus dem mehrsprachigen Lateinamerika (Huari Spanisch 1). https://refubium.fu-berlin.de/handle/fub188/29497.
Bianchi, Valentina, Giuliano Bocci & Silvio Cruschina. 2016. Focus fronting, unexpectedness, and evaluative implicatures. *Semantics and Pragmatics* 9(3). 1–54.
Bickel, Balthasar, Kristine A. Hildebrandt & René Schiering. 2009. The distribution of phonological word domains: A probabilistic typology. In Janet Grijzenhout & Barış Kabak (eds.), *Phonological domains*: *Universals and deviations* (Interface Explorations 16), 47–75. Berlin, New York: de Gruyter.
Bishop, Jason. 2012. Information structural expectations in the perception of prosodic prominence. In Gorka Elordieta & Pilar Prieto (eds.), *Prosody and Meaning*: *[based on the Workshop on Prosody and Meaning in Barcelona on September 17–18, 2009]* (Interface Explorations /IE] 25), 239–269. Berlin: de Gruyter.
Boersma, Paul. 1998. *Functional Phonology*: *Formalizing the interaction between articulatory and perceptual drives.* Amsterdam: University of Amsterdam Dissertation.
Boersma, Paul, Titia Benders & Klaas Seinhorst. 2020. Neural network models for phonology and phonetics. *Journal of Language Modelling* 8(1). 103–177.

Boersma, Paul & Bruce Hayes. 2001. Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32(1). 45–86.

Boersma, Paul & David Weenink. 2020. *Praat 6.1.16*: *Doing Phonetics by Computer*.

Bögels, Sara, Kobin H. Kendrick & Stephen C. Levinson. 2015. Never Say No… How the Brain Interprets the Pregnant Pause in Conversation. *PLoS ONE* 10(12). 1–15.

Bögels, Sara & Francisco Torreira. 2015. Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics* 52. 46–57.

Bolinger, Dwight L. 1978. Intonation across languages. In Joseph Greenberg (ed.), *Universals of Human Language.*: *Vol II: Phonology*, 471–524. Palo Alto: Stanford University Press.

Braun, Bettina, Audra Dainora & Mirjam Ernestus. 2011. An unfamiliar intonation contour slows down online speech comprehension. *Language and Cognitive Processes* 26(3). 350–375.

Brown, Meredith, Anne P. Salverda, Christine Gunlogson & Michael K. Tanenhaus. 2015. Interpreting prosodic cues in discourse context. *Language, Cognition and Neuroscience* 30(1–2). 149–166.

Bruggeman, Anna, Timo B. Roettger & Martine Grice. 2017. Question Word Intonation in Tashlhiyt Berber: Is 'high' good enough? *Laboratory Phonology* 8(1). 1–27.

Buchholz, Timo. 2021a. Dissertation_Intonation_between_phrasing_and_accent. https://doi.org/10.17605/OSF.IO/MAH8Z (14 August, 2021).

Buchholz, Timo. 2021b. *No stress in Conchucos Quechua?*: *Evidence from peak alignment data* (Phonetics and Phonology in Europe (PAPE)). Barcelona, online.

Buchholz, Timo & Uli Reich. 2017. *Conflicting functions and phonetic cues in bilingual Spanish and Quechua* (Phonetics and Phonology in Europe (PAPE)). Köln.

Buchholz, Timo & Uli Reich. 2018. The realizational coefficient: Devising a method for empirically determining prominent positions in Conchucos Quechua. In Ingo Feldhausen, Jan Fliessbach & Maria d. M. Vanrell (eds.), *Methods in prosody*: *A Romance language perspective* (Studies in Laboratory Phonology 6), 123–164. Berlin: Language Science Press.

Büring, Daniel. 2003. On D-Trees, Beans, And B-Accents. *Linguistics and Philosophy* 26(5). 511–545.

Büring, Daniel. 2007. Semantics, Intonation, and Information Structure. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces* (Oxford handbooks in linguistics), 445–474. Oxford: Oxford University Press.

Calhoun, Sasha. 2010a. How does informativeness affect prosodic prominence? *Language and Cognitive Processes* 25(7–9). 1099–1140.

Calhoun, Sasha. 2010b. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language* 86(1). 1–42.

Calhoun, Sasha. 2015. The interaction of prosody and syntax in Samoan focus marking. *Lingua* 165. 205–229.

Calhoun, Sasha, Erwin La Cruz & Ana Olssen. 2018. The interplay of information structure, semantics, prosody, and word ordering in Spanish intransitives. *Laboratory Phonology* 9(1). 1–30.

Calhoun, Sasha & Antje Schweitzer. 2012. Can intonation contours be lexicalised?: Implications for discourse meanings. In Gorka Elordieta & Pilar Prieto (eds.), *Prosody and Meaning*: *[based on the Workshop on Prosody and Meaning in Barcelona on September 17–18, 2009]* (Interface Explorations / IE] 25), 271–327. Berlin: de Gruyter.

Cangemi, Francesco & Martine Grice. 2016. The Importance of a Distributional Approach to Categoriality in Autosegmental-Metrical Accounts of Intonation. *Laboratory Phonology* 7(1.9). 1–20.

Cardinaletti, Anna & Lori Repetti. 2009. Phrase-level and word-level syllables: Resyllabification and prosodization of clitics. In Janet Grijzenhout & Barış Kabak (eds.), *Phonological domains*: *Universals and deviations* (Interface Explorations 16), 77–102. Berlin, New York: de Gruyter.

Carranza Romero, Francisco J. 2003. *Diccionario quechua ancashino – castellano*. Edición y prólogo de Wolf Lustig. Madrid, Frankfurt am Main: Iberoamericana; Vervuert.

Cerrón-Palomino, Rodolfo. 1987. *Lingüística quechua* (Biblioteca de la tradición oral andina 8). Lima: Centro de Estudios Rurales Andinos "Bartolomé de las Casas".

Cerrón-Palomino, Rodolfo. 2008. *Quechumara*: *Estructuras paralelas del quechua y del aimara*. La Paz: Plural Editores.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles N. Li (ed.), *Subject and topic*, 27–55. New York: Academic Press.

Chafe, Wallace L. 1994. *Discourse, consciousness, and time*: *The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Chahal, Dana & Sam Hellmuth. 2014. The intonation of Lebanese and Egyptian Arabic. In Sun-Ah Jun (ed.), *Prosodic typology II*: *The phonology of intonation and phrasing*, 365–404. Oxford: Oxford University Press.

Chen, Aoju. 2007. Language-specificity in the perception of continuation intonation. In Carlos Gussenhoven & Tomas Riad (eds.), *Tones and tunes*: *Volume 2: Experimental studies in word and sentence prosody* (Phonology and phonetics 12-2), 107–141. Berlin, New York: Mouton de Gruyter.

Cho, Hyesun & Edward Flemming. 2015. Compression and truncation: The case of Seoul Korean accentual phrase. *Studies in Phonetics, Phonology, and Morphology* 21(2). 359–382.

Cho, Taehong & Patricia Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37(4). 466–485.

Chomsky, Noam & Morris Halle. 1991 [1968]. *The sound pattern of English*, 4th edn. Cambridge, Mass.: MIT Press.

Clopper, Cynthia G., Rory Turnbull & Rachel S. Burdin. 2018. Assessing predictability effects in connected read speech. *Linguistics Vanguard* 4(2). 1–13.

Cole, Jennifer, José I. Hualde, Caroline L. Smith, Christopher Eager, Timothy Mahrt & Ricardo Napoleão de Souza. 2019. Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics* 75. 113–147.

Cole, Peter. 1982. *Imbabura quechua*. Amsterdam: North-Holland.

Crignis, Patricia de. 2018. *Vokalschwächung im peruanischen Spanisch*. Dissertartion (Elektronische Hochschulschriften). München: Ludwig-Maximilians-Universität München.

Crosswhite, Katherine M. 2004. Vowel reduction. In Bruce Hayes, Robert M. Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 191–231. Cambridge, New York: Cambridge University Press.

Cruschina, Silvio. 2019. Focus Fronting in Spanish: Mirative implicature and information structure. *Probus* 31(1). 119–146.

Cruttenden, Alan. 1997. *Intonation* (Cambridge textbooks in linguistics), 2nd edn. Cambridge: Cambridge University Press.

Cruttenden, Alan. 2006. The de-accenting of given information: A cognitive universal. In Giuliano Bernini & Marcia L. Schwartz (eds.), *Pragmatic organization of discourse in the languages of Europe* (Empirical approaches to language typology EUROTYP 20-8), 311–355. Berlin, New York: Mouton de Gruyter.

Cusihuamán, Antonio. 2001. *Gramática Quechua: Cuzco-Collao*, 2nd edn. Lima: Centro Bartolomé de las Casas.

Dainora, Audra. 2001. *An empirically based probabilistic model of intonation in American English.* Chicago: University of Chicago Dissertation.

Dainora, Audra. 2002. Does Intonational Meaning Come From Tones or Tunes?: Evidence Against a Compositional Approach. In Speech Prosody (ed.), *Proceedings of Speech Prosody 2002*, 1–4.

Dainora, Audra. 2006. Modeling intonation in English: a probabilistic approach to phonological competence. In Louis Goldstein, D. H. Whalen & Catherine T. Best (eds.), *Laboratory Phonology 8*: *[papers from Eight Conference on Laboratory Phonology, held in New Haven, CT]* (Phonology and phonetics 4,2), 107–132. Berlin: de Gruyter.
Davis, Stuart, Natsuko Tsujimura & Jung-yueh Tu. 2012. Toward a Taxonomy of Loanword Prosody. *Catalan Journal of Linguistics* 11. 13–39.
Dawson, Caitlin, Daniel Aalto, Juraj Simko & Martti Vainio. 2017. The influence of fundamental frequency on perceived duration in spectrally comparable sounds. *PeerJ* 5. 1–18.
Dayal, Veneeta. 2016. *Questions* (Oxford linguistics 4). Oxford, United Kingdom: Oxford University Press.
De-la-Mota, Carme, Pedro M. Butragueño & Pilar Prieto. 2010. Mexican Spanish Intonation. In Pilar Prieto & Paolo Roseano (eds.), *Transcription of intonation of the Spanish language*, 319–350. München: LINCOM.
Delforge, Ann M. 2008. Unstressed Vowel Reduction in Andean Spanish. In Laura Colantoni & Jeffrey Steele (eds.), *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, 107–124. Somerville, MA: Cascadilla Proceedings Project.
Delforge, Ann M. 2011. Vowel Devoicing in Cusco Collao Quechua. In Wai-Sum Lee & Eric Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 556–559. Hong Kong: City University of Hong Kong.
Dikken, Marcel d. 2006. Specificational Copular Sentences and Pseudoclefts. In Martin Everaert & Henk C. van Riemsdijk (eds.), *The Wiley Blackwell companion to syntax*, 292–409. Hoboken: John Wiley & Sons.
D'Imperio, Mariapaola, Robert Espesser, Hélène Loevenbruck, Caroline Menezes, Noël Nguyen & Pauline Welby. 2007. Are tones aligned to articulatory events?: Evidence from Italian and French. In Jennifer Cole & José I. Hualde (eds.), *Laboratory phonology 9* (Phonology and phonetics 4.3), 577–608. Berlin: Mouton de Gruyter.
D'Imperio, Mariapaola & David House. 1997. Perception of Questions and Statements in Neapolitan Italian. *5th European Conference on Speech Communication and Technology*. 1–4.
Dingemanse, Mark, Seán G. Roberts, Julija Baranova, Joe Blythe, Paul Drew, Simeon Floyd, Rosa S. Gisladottir, Kobin H. Kendrick, Stephen C. Levinson, Elizabeth Manrique, Giovanni Rossi & N. J. Enfield. 2015. Universal Principles in the Repair of Communication Problems. *PLoS ONE* 10(9). 1–15.
Downing, Laura J. & Annie Rialland (eds.). 2017. *Intonation in African Tone Languages* (Phonology and phonetics Volume 24). Berlin, Boston: de Gruyter.
Dowty, David. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67(3). 547–619.
Dufter, Andreas & Christoph Gabriel. 2016. Information structure, prosody, and word order. In Susann Fischer & Christoph Gabriel (eds.), *Manual of Grammatical Interfaces in Romance* (Manuals of Romance Linguistics 10), 419–455. Berlin, Boston: de Gruyter Mouton.
Eddington, David. 2000. Spanish stress assignment within the analogical modeling of language. *Language* 76(1). 92–109.
Edeleva, Julia, Anna Chrabaszcz & Valeriia Demareva. 2020. Resolving conflicting cues in processing of ambiguous words: The role of case, word order, and animacy. *Quarterly journal of experimental psychology* 73(8). 1173–1188.
Elfner, Emily. 2015. Recursion in prosodic phrasing: evidence from Connemara Irish. *Natural Language & Linguistic Theory* 33(4). 1169–1208.
Elordieta, Gorka. 1998. Intonation in a pitch accent variety of Basque. *Anuario del Seminario de Filología Vasca Julio de Urquijo. International journal of basque linguistics and philology* 32(2). 511–569.

Elvira-García, Wendy. 2017. *Create pictures with tiers*: *Praat script*.
Enfield, Nicholas J. 2014. *Natural causes of language*: *Frames, biases, and cultural transmission* (Conceptual foundations of language science 1). Berlin: Language Science Press.
Escandell-Vidal, Victoria. 2017. Intonation and Evidentiality in Spanish Polar Interrogatives. *Language and Speech* 60(2). 224–241.
Escobar, Anna-Maria. 2000. *Contacto social y lingüístico: El español en contacto con el quechua en el Perú*. Lima: Pontificia Universidad Católica del Perú.
Escobar, Anna-Maria. 2011. Spanish in contact with Quechua. In Manuel Díaz-Campos (ed.), *The handbook of Hispanic sociolinguistics*, 323–352. Malden, MA: Wiley-Blackwell.
Estebas Vilaplana, Eva & Pilar Prieto. 2008. La notación prosódica del español: una revisión del Sp_ToBI. *Estudios de fonética experimental*(17). 263–283.
Estebas Vilaplana, Eva & Pilar Prieto. 2010. Castilian Spanish Intonation. In Pilar Prieto & Paolo Roseano (eds.), *Transcription of intonation of the Spanish language*, 17–48. München: LINCOM.
Face, Timothy L. 2001. Focus and early peak alignment in Spanish intonation. *Probus* 13(2). 223–246.
Face, Timothy L. 2003. Intonation in Spanish declaratives: differences between lab speech and spontaneous speech. *Catalan Journal of Linguistics* 2. 115–131.
Face, Timothy L. & Pilar Prieto. 2007. Rising Accents in Castilian Spanish: A Revision of Sp_ToBI. *Journal of Portuguese Linguistics* 5/6(1). 117–146.
Fahey, Richard P. & Randy Diehl. 1996. The missing fundamental in vowel height perception. *Perception & Psychophysics* 58(5). 725–733.
Faller, Martina. 2002. *Semantics and Pragmatics of Evidentials in Cuzco Quechua.* Stanford: Stanford University Dissertation.
Faller, Martina. 2003. The evidential and validational licensing conditions for the Cusco Quechua enclitic -mi. *Belgian Journal of Linguistics* 16. 7–21.
Faller, Martina. 2011. A possible worlds semantics for Cuzco Quechua evidentials. *Proceedings of SALT* 20. 660–683.
Faller, Martina. 2014. Reportativity, (not-)at-issueness, and assertion. *Annual Meeting of the Berkeley Linguistics Society* 40. 62–84.
Farkas, Donka F. & Kim B. Bruce. 2010. On Reacting to Assertions and Polar Questions. *Journal of Semantics* 27(1). 81–118.
Farkas, Donka F. & Floris Roelofsen. 2017. Division of Labor in the Interpretation of Declaratives and Interrogatives. *Journal of Semantics* 34. 237–289.
Fastl, Hugo & Eberhard Zwicker. 2007. *Psychoacoustics*: *Facts and models* (Springer series in information sciences 22). Berlin: Springer.
Féry, Caroline. 2013. Focus as prosodic alignment. *Natural Language & Linguistic Theory* 31(3). 683–734.
Féry, Caroline. 2015. Extraposition and Prosodic Monsters in German. In Lyn Frazier & Edward Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing*: *Studies in Honor of Janet Dean Fodor* (Studies in theoretical psycholinguistics 46), 11–37. Cham: Springer-Verlag.
Féry, Caroline. 2017. *Intonation and prosodic structure* (Key topics in phonology). Cambridge: Cambridge University Press.
Féry, Caroline & Shinichiro Ishihara. 2010. How Focus and Givenness Shape Prosody. In Caroline Féry & Malte Zimmermann (eds.), *Information structure*: *Theoretical, typological, and experimental perspectives*, 36–63. Oxford, New York: Oxford University Press.
Féry, Caroline & Frank Kügler. 2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36(4). 680–703.
Féry, Caroline & Fabian Schubö. 2010. Hierarchical prosodic structures in the intonation of center-embedded relative clauses. *The Linguistic Review* 27(3). 293–317.

Féry, Caroline & Hubert Truckenbrodt. 2005. Sisterhood and tonal scaling. *Studia Linguistica* 59(2–3). 223–243.
Field, Andy P., Jeremy Miles & Zoë Field. 2012. *Discovering statistics using R*. London, Thousand Oaks: Sage.
Flemming, Edward. 2002. *Auditory Representations in Phonology* (Outstanding dissertations in linguistics). New York: Routledge.
Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, Robert M. Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 232–276. Cambridge, New York: Cambridge University Press.
Flemming, Edward. 2008. Speech Perception and Phonological Contrast. In David B. Pisoni & Robert E. Remez (eds.), *The handbook of speech perception*, 1st edn. (Blackwell handbooks in linguistics), 156–181. Malden, Mass.: Blackwell Publ.
Fletcher, Janet. 2010. The Prosody of Speech: Timing and Rhythm. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds.), *The Handbook of phonetic sciences*, 2nd edn. (Blackwell handbooks in linguistics), 523–602. Oxford: Blackwell Publishers.
Fliessbach, Jan. 2023. *The intonation of expectations: On marked declaratives, exclamatives, and discourse particles in Castilian Spanish*. Berlin: Language Science Press.
Floyd, Rick. 1996. Experience, certainty and control, and the direct evidential in Wanka Quechua questions. *Functions of Language* 3(1). 69–93.
Floyd, Rick. 1997. *La estructura categorial de los evidenciales en el quechua wanka* (Serie Lingüistica Peruana 44). Lima: Ministerio de Educación.
Fodor, Janet D. 2002. Prosodic Disambiguation in Silent Reading. In M. Hirotani (ed.), *Proceedings of the North-East Linguistic Society 32*, 113–132. Amherst: University of Massachusetts.
Foster, Erin D. & Ariel Deardorff. 2017. Open Science Framework (OSF). *Journal of the Medical Library Association* 105(2). 203–206.
Fougeron, Cécile & Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101(6). 3728.
Frazier, Lyn & Edward Gibson (eds.). 2015. *Explicit and Implicit Prosody in Sentence Processing*: *Studies in Honor of Janet Dean Fodor* (Studies in theoretical psycholinguistics 46). Cham: Springer-Verlag.
Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe. 2004. Similarity Avoidance and the OCP. *Natural Language & Linguistic Theory* 22(1). 179–228.
Frota, Sónia. 2000. *Prosody and focus in European Portuguese*: *Phonological phrasing and intonation* (Outstanding dissertations in linguistics). London: Routledge.
Frota, Sónia. 2012. Prosodic Structure, Constituents and their Implementation. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.), *The Oxford handbook of laboratory phonology* (Oxford handbooks in linguistics), 255–265. Oxford, New York: Oxford University Press.
Frota, Sónia. 2014. The intonational phonology of European Portuguese. In Sun-Ah Jun (ed.), *Prosodic typology II*: *The phonology of intonation and phrasing*, 6–42. Oxford: Oxford University Press.
Frota, Sónia, Marisa Cruz, Flaviane Svartman, Gisela Collischonn, Aline Fonseca, Carolina Serra, Pedro Oliveira & Marina Vigário. 2015. Intonational variation in Portuguese: European and Brazilian varieties. In Sónia Frota & Pilar Prieto (eds.), *Intonation in Romance*, 235–283. Oxford: Oxford University Press.
Frota, Sónia, Mariapaola D'Imperio, Gorka Elordieta, Pilar Prieto & Marina Vigário. 2007. The Phonetics and Phonology of Intonational Phrasing in Romance. In Pilar Prieto, Joan Mascaró & Maria J. Solé (eds.), *Segmental and prosodic issues in romance phonology* (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory v. 282), 131–153. Amsterdam, Philadelphia: John Benjamins Pub.

Frota, Sónia & Pilar Prieto (eds.). 2015. *Intonation in Romance*. Oxford: Oxford University Press.
Fry, D. B. 1955. Duration and Intensity as Physical Correlates of Linguistic Stress. *The Journal of the Acoustical Society of America* 27(4). 765–768.
Gabriel, Christoph. 2007. *Fokus im Spannungsfeld von Phonologie und Syntax*: *Eine Studie zum Spanischen*. Zugl.: Osnabrück, Univ., Habil.-Schr., 2005 (Editionen der Iberoamericana Ser. B, Sprachwissenschaft 7). Frankfurt am Main: Vervuert.
Gabriel, Christoph, Ingo Feldhausen & Andrea Pešková. 2011. Prosodic Phrasing in Porteño Spanish. In Christoph Gabriel & Conxita Lleó (eds.), *Intonational phrasing in Romance and Germanic*: *Cross-linguistic and bilingual studies* (Hamburg studies on multilingualism 10), 153–182. Amsterdam, Philadelphia: John Benjamins.
Gabriel, Christoph, Ingo Feldhausen, Andrea Pešková, Laura Colantoni, Su-Ar Lee, Valeria Arana & Leopoldo Labastía. 2010. Argentinian Spanish Intonation. In Pilar Prieto & Paolo Roseano (eds.), *Transcription of intonation of the Spanish language*, 285–317. München: LINCOM.
García, Miguel. 2016. *The Intonation of Peruvian Amazonian Spanish: Rising Accents and Segmental Factors.* Columbus: Ohio State University Dissertation.
García, Ofelia & Wei Li. 2014. *Translanguaging*: *Language, Bilingualism and Education*. Basingstoke, New York: Palgrave Macmillan.
Garrido, Juan-María, Joaquim Llisterri, Rafael Marín, Carme De-la-Mota & Antonio Ríos. 1995. Prosodic markers at syntactic boundaries in Spanish. In Kjell Elenius & Peter Branderud (eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences*: *Vol. 2*, 370–373. Stockholm: Stockholm University.
Georgeton, L. & C. Fougeron. 2014. Domain-initial strengthening on French vowels and phonological contrasts: Evidence from lip articulation and spectral variation. *Journal of Phonetics* 44. 83–95.
Gerrits, E. & M. E. H. Schouten. 2004. Categorical perception depends on the discrimination task. *Perception & Psychophysics* 66(3). 363–376.
Geurts, Bart. 2019. Communication as commitment sharing: speech acts, implicatures, common ground. *Theoretical Linguistics* 45(1–2). 1–30.
Gili-Fivela, Barbara, Cinzia Avesani, Marco Barone, Giuliano Bocci, Claudia Crocco, Mariapaola D'Imperio, Rosa Giordano, Giovanna Marotta, Michelina Savino & Patrizia Sorianello. 2015. Intonational phonology of the regional varieties of Italian. In Sónia Frota & Pilar Prieto (eds.), *Intonation in Romance*, 140–197. Oxford: Oxford University Press.
Ginzburg, Jonathan. 2012. *The interactive stance*. Oxford: Oxford University Press.
Ginzburg, Jonathan, Raquel Fernández & David Schlangen. 2014. Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics* 7(9). 1–64.
Givón, Talmy. 2018 [1979]. *On Understanding Grammar*: *Revised Edition*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Goedemans, Rob, Harry van der Hulst & Ellen van Zanten (eds.). 2010. *A survey of word accentual patterns in the languages of the world*. Berlin: de Gruyter.
Göksel, Aslı & Celia Kerslake. 2005. *Turkish*: *A comprehensive grammar* (Routledge comprehensive grammars). London: Routledge.
Goldsmith, John A. 1976. *Autosegmental Phonology.* Boston: MIT Dissertation.
Gordon, Matthew & Timo Roettger. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard* 3(1). 41–51.
Gordon, Matthew J. 2014. Disentangling stress and pitch accent: A typology of prominence at different prosodic levels. In Harry van der Hulst (ed.), *Word stress*: *Theoretical and typological issues*, 83–118. Cambridge: Cambridge University Press.

Greenberg, Joseph H. 1966. Some Universals of Grammar With Particular Reference to the Order of Meaningful Elements. In Joseph H. Greenberg (ed.), *Universals of language*: *Report of a conference held at Dobbs Ferry, New York April 13–15, 1961*, 2nd edn., 73–113. Cambridge, Mass: MIT Press.

Gretsch, Petra. 2003. Omission impossible? Topic and focus in focal ellipsis. In Kerstin Schwabe & Susanne Winkler (eds.), *The interfaces*: *Deriving and interpreting omitted structures* (Linguistik aktuell v. 61), 341–365. Amsterdam, Philadelphia: J. Benjamins Pub. Co.

Grice, Herbert P. 1989. Logic and Conversation. In Herbert P. Grice (ed.), *Studies in the way of words*, 22–40. Cambridge, Mass: Harvard University Press.

Grice, Martine. 1995. Leading tones and downstep in English. *Phonology* 12(2). 183–233.

Grice, Martine, Stefan Baumann & Ralf Benzmüller. 2005. German Intonation in Autosegmental-Metrical Phonology. In Sun-Ah Jun (ed.), *Prosodic typology*: *The phonology of intonation and phrasing* (Oxford linguistics), 55–83. Oxford, New York: Oxford University Press.

Grice, Martine, D. R. Ladd & Amalia Arvaniti. 2000. On the place of phrase accents in intonational phonology. *Phonology* 17(2). 143–185.

Grice, Martine, Rachid Ridouane & Timo B. Roettger. 2015. Tonal association in Tashlhiyt Berber: evidence from polar questions and contrastive statements. *Phonology* 32(2). 241–266.

Grice, Martine & M. Savino. 2004. Information Structure and Questions – Evidence from Task-Oriented Dialogues in a Variety of Italian. In Peter Gilles & Jörg Peters (eds.), *Regional variation in intonation* (Linguistische Arbeiten 492), 161–190. Tübingen: Niemeyer.

Grice, Martine & Michelina Savino. 1997. Can pitch accent type convey information status in yes-no questions? In Kai Alter, Hannes Pirker & Wolfgang Finkler (eds.), *Concept to Speech Generation Systems*: *Proceedings of a Workshop in conjunction with 35th Annual Meeting of the Association for Computational Linguistics*, 29–38. Madrid: Universidad Nacional de Educación a Distancia.

Grice, Martine & Michelina Savino. 2003. Map Tasks in Italian: Asking Questions about Given, Accessible and New Information. *Catalan Journal of Linguistics* 2. 153–180.

Grijzenhout, Janet & Barış Kabak. 2009. Prosodic Phonology: An appraisal. In Janet Grijzenhout & Barış Kabak (eds.), *Phonological domains*: *Universals and deviations* (Interface Explorations 16), 1–14. Berlin, New York: de Gruyter.

Guion, Susan G. 2003. The Vowel Systems of Quichua-Spanish Bilinguals. *Phonetica* 60(2). 98–128.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2). 274–307.

Gunlogson, Christine. 2001. *True to Form: Rising and Falling Declaratives as Questions in English.* Santa Cruz: University of California Santa Cruz Dissertation.

Gunlogson, Christine. 2008. A question of commitment. *Belgian Journal of Linguistics* 22. 101–136.

Gussenhoven, Carlos. 2000a. The boundary tones are coming: on the nonperipheral realization of boundary tones. In Michael Broe & Janet B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 132–151. Cambridge: Cambridge University Press.

Gussenhoven, Carlos. 2000b. The Lexical Tone Contrast of Roermond Dutch in Optimality Theory. In Merle Horne (ed.), *Prosody: Theory and Experiment*: *Studies Presented to Gösta Bruce* (Text, speech and language technology 14), 129–168. Dordrecht: Springer.

Gussenhoven, Carlos. 2002a. Intonation and interpretation: phonetics and phonology. In Bernard Bel & Isabelle Marlien (eds.), *Speech Prosody 2002*: *Proceedings of the First International Conference on Speech Prosody*, 47–57. Aix-en-Provence: Laboratoire Parole et Langage.

Gussenhoven, Carlos. 2002b. Phonology of Intonation. *Glot International* 6(9/10). 271–284.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation* (Research surveys in linguistics). Cambridge, New York: Cambridge University Press.

Gussenhoven, Carlos. 2005. Transcription of Dutch Intonation. In Sun-Ah Jun (ed.), *Prosodic typology*: *The phonology of intonation and phrasing* (Oxford linguistics), 118–145. Oxford, New York: Oxford University Press.
Gussenhoven, Carlos. 2018. Prosodic typology meets phonological representations. In Larry M. Hyman & Frans Plank (eds.), *Phonological typology* (Phonology and phonetics Volume 23), 389–418. Berlin, New York: de Gruyter Mouton.
Gussenhoven, Carlos & Peter van der Vliet. 1999. The phonology of tone and intonation in the Dutch dialect of Venlo. *Journal of Linguistics* 35(1). 99–135.
Haiman, John. 1978. Conditionals are Topics. *Language* 54(3). 564–589.
Halliday, Michael. 1967. Notes on Transitivity and Theme in English: Part 1 and 2. *Journal of Linguistics* 3(1–2). 37–81, 199–244.
Hamblin, C. L. 1958. Questions. *Australasian Journal of Philosophy* 36(3). 159–168.
Hamblin, C. L. 1973. Questions in Montague English. *Foundations of Language* 10(1). 41–53.
Hanson, Helen M. 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America* 125(1). 425–441.
Harris, James W. 1983. *Syllabic Structure and stress in Spanish*: *A Nonlinear Analysis* (Linguistic Inquiry Monographs 8). Cambridge, Mass: MIT Press.
Harris, James W. 1987. The accentual patterns of verb paradigms in Spanish. *Natural Language & Linguistic Theory* 5(1). 61–90.
Hayes, Bruce. 1995. *Metrical stress theory*: *Principles and case studies*. Chicago: University of Chicago Press.
Hayes, Bruce & Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9(1). 47–96.
Hellmuth, Sam. 2005. No de-accenting in (or of) phrases: Evidence from Arabic for cross-linguistic and cross-dialectal prosodic variation. In Sónia Frota, Marina C. Vigário & Maria J. Freitas (eds.), *Prosodies*: *With special reference to Iberian languages* (Phonology and phonetics 9), 99–121. Berlin, New York: Mouton de Gruyter.
Hellmuth, Sam. 2006. *Intonational pitch accent distribution in Egyptian Arabic.* London: School of Oriental and African Studies Dissertation.
Hellmuth, Sam. 2007. The relationship between prosodic structure and pitch accent distribution: Evidence from Egyptian Arabic. *The Linguistic Review* 24(2–3). 291–316.
Hellmuth, Sam. 2013. Phonology. In Jonathan Owens (ed.), *The Oxford handbook of Arabic linguistics* (The Oxford handbooks in linguistics), 45–70. Oxford: Oxford University Press.
Henriksen, Nicholas. 2013. Style, prosodic variation, and the social meaning of intonation. *Journal of the International Phonetic Association* 43(02). 153–193.
Heusinger, Klaus v. 2007. Discourse Structure and Intonational Phrasing. In Chungmin Lee, Matthew K. Gordon & Daniel Büring (eds.), *Topic and focus*: *Cross-linguistic perspectives on meaning and intonation*, 265–290. Dordrecht: Springer.
Himmelmann, Nikolaus P., Meytal Sandler, Jan Strunk & Volker Unterladstetter. 2018. On the universality of intonational phrases: a cross-linguistic interrater study. *Phonology* 35(2). 207–245.
Hintz, Daniel. 2000. *Caracteristicas distintivas del quechua de Corongo* (Seria Lingüistica Peruana 50). Lima: Instituto Lingüistico de Verano.
Hintz, Daniel J. 2014. South Conchucos Quechua. In Carol Genetti, Allison Adelman & Alexandra Y. Aikhenvald (eds.), *How languages work*: *An introduction to language and linguistics*, 463–481. Cambridge: Cambridge Univ. Press.
Hintz, Daniel J. & Diane M. Hintz. 2017. The evidential category of mutual knowledge in Quechua. *Lingua* 186–187. 88–109.

Hintz, Diane M. 2006. Stress in South Conchucos Quechua: A Phonetic and Phonological Study. *International Journal of American Linguistics* 72(4). 477–521.
Hintz, Diane M. 2007. *Past tense forms and their functions in South Conchucos Quechua: time, evidentiality, discourse structure, and affect.* Santa Barbara: University of California Santa Barbara Dissertation.
Hirschberg, Julia & Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics* 20(2). 241–251.
Höder, Steffen. 2014a. Constructing diasystems: Grammatical organisation in bilingual groups. In Tor A. Åfarli & Brit Maehlum (eds.), *The Sociolinguistics of Grammar* (Studies in language companion series v.154), 137–152. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Höder, Steffen. 2014b. Phonological elements and Diasystematic Construction Grammar. *Constructions and Frames* 6(2). 202–231.
Höder, Steffen. 2018. Grammar is community-specific: Background and basic concepts of Diasystematic Construction Grammar. In Hans C. Boas & Steffen Höder (eds.), *Constructions in Contact*: *Constructional Perspectives on Contact Phenomena in Germanic Languages* (Constructional Approaches to Language Ser v.24), 37–70. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Hombert, Jean-Marie, John J. Ohala & William G. Ewan. 1979. Phonetic Explanations for the Development of Tones. *Language* 55(1). 37–58.
Hoot, Bradley. 2012. *Presentational focus in heritage and monolingual Spanish.* Chicago: University of Illinois at Chicago Dissertation.
Hoot, Bradley. 2016. Narrow presentational focus in Mexican Spanish: Experimental evidence. *Probus* 28(2).
Hopper, Paul J. 1979. Aspect and Foregrounding in Discourse. In Talmy Givón (ed.), *Discourse and syntax* (Syntax and semantics 12), 213–241. New York: Academic Press.
Horne, Merle, Petra Hansson, Gösta Bruce & Johan Frid. 2001. Accent Patterning on Domain-Related Information in Swedish Travel Dialogues. *International Journal of Speech Technology* 4. 93–102.
Hualde, José I. 1991. Aspiration and Resyllabification in Chinato Spanish. *Probus* 3(1). 55–76.
Hualde, José I. 1993. On The Historical Origin of Basque Accentuation. *Diachronica* 10(1). 13–50.
Hualde, José I. 1999. Basque Accentuation. In Harry van der Hulst (ed.), *Word prosodic systems in the languages of Europe* (Empirical approaches to language typology 20-4), 947–993. Berlin, New York: Mouton de Gruyter.
Hualde, José I. 2002. Intonation in Spanish and the other Ibero-Romance Languages: Overview and status quaestionis. In Caroline R. Wiltshire & Joaquim Camps (eds.), *Romance phonology and variation*: *Selected papers from the 30th Linguistic Symposium on Romance Languages, Gainesville, Florida, February 2000* (Amsterdam studies in the theory and history of linguistic science Series 4, Current issues in linguistic theory 217), 101–115. Amsterdam: Benjamins.
Hualde, José I. 2007. Stress removal and stress addition in Spanish. *Journal of Portuguese Linguistics* 5/6. 59–89.
Hualde, José I. 2009. Unstressed words in Spanish. *Language Sciences* 31(2–3). 199–212.
Hualde, José I. & Alder Beristain. 2018. Acoustic correlates of word-accent in Basque. In Hansjörg Mixdorff, Angelika Hönemann, Oliver Jokisch, Petra Wagner & Frank Kügler (eds.), *TAL2018, Sixth International Symposium on Tonal Aspects of Languages*. Boston: ISCA.
Hualde, José I. & Sonia Colina. 2014. *Los sonidos del español*. Cambridge: Cambridge University Press.
Hualde, José I., Gorka Elordieta, Iñaki Gaminde & Rajka Smiljanic. 2002. From pitch accent to stress accent in Basque. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7* (Phonology and phonetics 4-1), 547–584. Berlin, New York: Mouton de Gruyter.

Hualde, José I. & Pilar Prieto. 2015. Intonational variation in Spanish: European and American varieties. In Sónia Frota & Pilar Prieto (eds.), *Intonation in Romance*, 350–391. Oxford: Oxford University Press.
Huttenlauch, Clara, Sophie Egger, Daniela Wochner & Ingo Feldhausen. 2016. The intonation of echo wh-questions in Ecuadorian Spanish. In Jon Barnes, Alejna Brugos, Stefanie Shattuck-Hufnagel & Nanette Veilleux (eds.), *Proceedings of the 8th International Conference on Speech Prosody 2016* (SpeechProsody), 385–389. Boston: ISCA.
Huttenlauch, Clara, Ingo Feldhausen & Bettina Braun. 2018. The purpose shapes the vocative: Prosodic realisation of Colombian Spanish vocatives. *Journal of the International Phonetic Association* 48(1). 33–56.
Hyman, Larry M. 2006. Word-prosodic typology. *Phonology* 23(2). 225–257.
Hyman, Larry M. 2009. How (not) to do phonological typology: the case of pitch-accent. *Language Sciences* 31(2–3). 213–238.
Hyman, Larry M. 2011. Does Gokana really have no syllables? Or: what‘s so great about being universal? *Phonology* 28(1). 55–85.
Hyman, Larry M. 2014. Do All Languages Have Word Accent? In Harry van der Hulst (ed.), *Word stress*: *Theoretical and typological issues*, 56–82. Cambridge: Cambridge University Press.
Hyman, Larry M. 2015. Does Gokana really have syllables?: A postscript. *Phonology* 32(2). 303–306.
Hyman, Larry M. 2018. What is phonological typology? In Larry M. Hyman & Frans Plank (eds.), *Phonological typology* (Phonology and phonetics Volume 23), 1–20. Berlin, New York: de Gruyter Mouton.
Igarashi, Yosuke. 2014. Typology of intonational phrasing in Japanese dialects. In Sun-Ah Jun (ed.), *Prosodic typology II*: *The phonology of intonation and phrasing*, 464–492. Oxford: Oxford University Press.
Igarashi, Yosuke. 2015. Intonation. In Haruo Kubozono (ed.), *Handbook of Japanese phonetics and phonology* (Handbooks of Japanese language and linguistics 2), 525–568. Boston: de Gruyter Mouton.
Igarashi, Yosuke, Ken‘ya Nishikawa, Kuniyoshi Tanaka & Reiko Mazuka. 2013. Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech. *Journal of the Acoustical Society of America* 134(2). 1283–1294.
Instituto Nacional de Estadística e Informática. 2018. Censos nacionales 2017: XII de población, VII de vivienda, y III de comunidades indígenas. http://censos2017.inei.gob.pe/redatam/ (27 June, 2021).
International Phonetic Asssociation. 2005. The International Phonetic Alphabet (revised to 2005). http://upload.wikimedia.org/wikipedia/en/8/8f/IPA_chart_%28C%292005.pdf (24 September, 2013).
Ishihara, Shinichiro. 2011. Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua* 121(13). 1870–1889.
Ishihara, Shinichiro. 2016. Japanese downstep revisited. *Natural Language and Linguistic Theory* 34(4). 1389–1443.
Ishihara, Shinichiro. 2017. The Intonation of Wh- and Yes/No-Questions in Tokyo Japanese. In Chungmin Lee, Ferenc Kiefer & Manfred Krifka (eds.), *Contrastiveness in Information Structure, Alternatives and Scalar Implicatures* (Studies in natural language and linguistic theory 91), 399–415. Cham: Springer.
Ishihara, Takeshi. 2006. *Tonal alignment in Tokyo Japanese.* Edinburgh: University of Edinburgh Dissertation.

Ito, Junko & Armin Mester. 2007. Prosodic Adjunction in Japanese Compounds. In Yoichi Miyamoto & Masao Ochi (eds.), *Formal Approaches to Japanese Linguistics (FAJL 4)* (MIT Working Papers in Linguistics 55), 97–111. Cambridge, Mass.: MIT Press.

Ito, Junko & Armin Mester. 2012. Recursive prosodic phrasing in Japanese. In Toni Borowsky, Shigeto Kawahara, Mariko Sugahara & Shinya Takahito (eds.), *Prosody matters*: *Essays in honor of Elisabeth Selkirk* (Advances in optimality theory), 280–303. London: Equinox.

Ito, Kiwako. 2002. *The Interaction of Focus and Lexical Pitch Accent in Speech Production and Dialogue Comprehension: Evidence from Japanese and Basque.* Urbana-Champaign: University of Illinois at Urbana-Champaign Dissertation.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar* (Current studies in linguistics series 2). Cambridge, Mass.: MIT Press.

Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý & Vít Suchomel. 2013. The TenTen Corpus Family. *7th International Corpus Linguistics Conference CL*. 125–127.

Jun, Sun-Ah. 1993. *The Phonetics and Phonology of Korean Prosody.* Ohio: Ohio State University Dissertation.

Jun, Sun-Ah. 2005a. Introduction. In Sun-Ah Jun (ed.), *Prosodic typology*: *The phonology of intonation and phrasing* (Oxford linguistics), 1–8. Oxford, New York: Oxford University Press.

Jun, Sun-Ah. 2005b. Korean Intonational Phonology and Prosodic Transcription. In Sun-Ah Jun (ed.), *Prosodic typology*: *The phonology of intonation and phrasing* (Oxford linguistics), 201–229. Oxford, New York: Oxford University Press.

Jun, Sun-Ah (ed.). 2005c. *Prosodic typology*: *The phonology of intonation and phrasing* (Oxford linguistics). Oxford, New York: Oxford University Press.

Jun, Sun-Ah. 2005d. Prosodic Typology. In Sun-Ah Jun (ed.), *Prosodic typology*: *The phonology of intonation and phrasing* (Oxford linguistics), 430–458. Oxford, New York: Oxford University Press.

Jun, Sun-Ah. 2007. The intermediate phrase in Korean: Evidence from sentence processing. In Carlos Gussenhoven & Tomas Riad (eds.), *Tones and tunes*: *Volume 2: Experimental studies in word and sentence prosody* (Phonology and phonetics 12-2), 143–167. Berlin, New York: Mouton de Gruyter.

Jun, Sun-Ah (ed.). 2014a. *Prosodic typology II*: *The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Jun, Sun-Ah. 2014b. Prosodic typology: by prominence type, word prosody, and macrorhythm. In Sun-Ah Jun (ed.), *Prosodic typology II*: *The phonology of intonation and phrasing*, 520–540. Oxford: Oxford University Press.

Jun, Sun-Ah & Janet Fletcher. 2014. Methodology of studying intonation: from data collection to data analysis. In Sun-Ah Jun (ed.), *Prosodic typology II*: *The phonology of intonation and phrasing*, 493–519. Oxford: Oxford University Press.

Jun, Sun-Ah & Cécile Fougeron. 2000. A Phonological Model of French Intonation. In Antonis Botinis (ed.), *Intonation*: *Analysis, Modelling and Technology* (Text, speech and language technology 15), 209–242. Dordrecht: Springer.

Jun, Sun-Ah & Cécile Fougeron. 2002. Realizations of accentual phrase in French intonation. *Probus* 14(1). 147–172.

Kager, René. 1999. *Optimality theory* (Cambridge textbooks in linguistics). Cambridge: Cambridge University Press.

Kaiser, Elsi. 2016. Information Structure and Language Comprehension: Insights from Psycholinguistics. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure* (Oxford handbooks in linguistics), 1–34. Oxford, New York: Oxford University Press.

Kaisse, Ellen M. 1999. Resyllabification Precedes All Segmental Rules: Evidence from Argentinian Spanish. In Jean-Marc Authier, Barbara E. Bullock & Lisa E. Reed (eds.), *Formal perspectives on*

Romance linguistics: *Selected papers from the 28th Linguistic Symposium on Romance Languages (LSRL XXVIII); University Park, 16–19 April 1998* (Amsterdam studies in the theory and history of linguistic science Serie 4, Current issues in linguistic theory 185), 197–210. Amsterdam: John Benjamins.

Kamali, Beste. 2011. *Topics at the PF Interface of Turkish.* Cambridge, Mass.: Harvard University Dissertation.

Kang, Yoonjung. 2010. Tutorial overview: Suprasegmental adaptation in loanwords. *Lingua* 120(9). 2295–2310.

Karttunen, Lauri. 1973. Presuppositions of Compound Sentences. *Linguistic Inquiry* 4(2). 169–193.

Katsika, Argyro. 2012. *Coordination of Prosodic Gestures at Boundaries in Greek.* Yale: Yale University Dissertation.

Katsika, Argyro, Jelena Krivokapic, Christine Mooshammer, Mark Tiede & Louis Goldstein. 2014. The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics* 44. 62–82.

Kawahara, Shigeto. 2015. The phonology of Japanese accent. In Haruo Kubozono (ed.), *Handbook of Japanese phonetics and phonology* (Handbooks of Japanese language and linguistics 2), 445–492. Boston: de Gruyter Mouton.

Keating, Patricia, Taehong Cho, Cécile Fougeron & Chai-Shune Hsu. 2004. Domain-initial articulatory strengthening in four languages. In John Local, Richard Ogden & Rosalind A. M. Temple (eds.), *Phonetic interpretation*: *The Sixth Conference in Laboratory Phonology* (Papers in laboratory phonology 6), 145–163. Cambridge: Cambridge University Press.

Keating, Patricia A. 2006. Phonetic Encoding of Prosodic Structure. In Jonathan Harrington & Marija Tabain (eds.), *Speech production*: *Models, phonetic processes, and techniques*, 167–187: Psychology Press UK.

Kendrick, Kobin H. & Francisco Torreira. 2015. The Timing and Construction of Preference: A Quantitative Study. *Discourse Processes* 52(4). 255–289.

Kilgarriff, Adam, Pavel Richlý, Pavel Smrz & David Tugwell. 2004. The Sketch Engine. In Geoffrey Williams & Sandra Vessier (eds.), *Proceedings of the eleventh EURALEX international congress, EURALEX 2004*, 105–115. Lorient: Université de Bretagne-Sud, Faculté des lettres et des sciences humaines; Universite de Bretagne Sud.

Kim, Kyunghee. 2013. *Tone, pitch accent and intonation of Korean*: *A synchronic and diachronic view.* Köln: Universität Köln Dissertation.

Kiparsky, Paul. 1982. Lexical Morphology and Phonology. In The Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*: *Selected Papers from SICOL-1981*, 3–91. Seoul: Hanshin Publishing Company.

Kiparsky, Paul. 2015. Stratal OT: A synopsis and FAQs. In Lian-Hee Wee & Yuchau E. Hsiao (eds.), *Capturing phonological shades within and across languages*, 2–44. Newcastle upon Tyne: Cambridge Scholars Publishing.

Kirby, James P. & D. R. Ladd. 2016. Effects of obstruent voicing on vowel F0: Evidence from “true voicing” languages. *Journal of the Acoustical Society of America* 140(4). 2400–2411.

Klatt, Dennis H. 1973. Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. *Journal of the Acoustical Society of America* 53(1). 8–16.

Kohler, Klaus J. 1987. Categorical Pitch Perception. *Proceedings of ICPhS* 11. 331–333.

Kohler, Klaus J. 2005. Timing and communicative functions of pitch contours. *Phonetica* 62(2–4). 88–105.

König, Ekkehard & Peter Siemund. 2007. Speech act distinctions in grammar. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*: *Volume 1: Clause Structure*, 2nd edn., 276–322. Cambridge: Cambridge University Press.

Krifka, Manfred. 2007. Basic Notions of Information Structure. In Caroline Féry, Gisbert Fanselow & Manfred Krifka (eds.), *Interdisciplinary Studies on Information Structure*: *The notions of information structure* (Working Papers of the SFB 632 Interdisciplinary Studies on Information Structure 6), 13–56. Potsdam: Universitäts-Verlag.

Krifka, Manfred. 2011. Questions. In Claudia Maienborn, Klaus v. Heusinger & Paul Portner (eds.), *Semantics*: *An international handbook of natural language meaning*. Volume 2 (Handbücher zur Sprach- und Kommunikationswissenschaft 33.2), 1742–1785. Berlin: de Gruyter.

Krifka, Manfred & Renate Musan. 2012. Information structure: Overview and linguistic issues. In Manfred Krifka & Renate Musan (eds.), *The expression of information structure* (The expression of cognitive categories 5), 1–43. Berlin: de Gruyter Mouton.

Kuang, Jianjing & Mark Liberman. 2016. Pitch-Range Perception: The Dynamic Interaction Between Voice Quality and Fundamental Frequency. *Interspeech*. 1350–1354.

Kubozono, Haruo. 1989. Syntactic and Rhythmic Effects on Downstep in Japanese. *Phonology* 6(1). 39–67.

Kubozono, Haruo. 2006. Where does loanword prosody come from? *Lingua* 116(7). 1140–1170.

Kubozono, Haruo. 2008. Japanese Accent. In Shigeru Miyagawa & Mamoru Saito (eds.), *The Oxford Handbook of Japanese Linguistics*, 165–191. Oxford & New York: Oxford University Press.

Kubozono, Haruo. 2015. Introduction to Japanese phonetics and phonology. In Haruo Kubozono (ed.), *Handbook of Japanese phonetics and phonology* (Handbooks of Japanese language and linguistics 2), 1–40. Boston: de Gruyter Mouton.

Kügler, Frank & Sasha Calhoun. 2020. Prosodic encoding of information structure: A typological perspective. In Carlos Gussenhoven & Aoju Chen (eds.), *The Oxford Handbook of Language Prosody*, 454–467. Oxford: Oxford University Press.

Kügler, Frank & Caroline Féry. 2017. Postfocal Downstep in German. *Language and Speech* 60(2). 260–288.

Kügler, Frank & Susanne Genzel. 2012. On the Prosodic Expression of Pragmatic Prominence: The Case of Pitch Register Lowering in Akan. *Language and Speech* 55(3). 331–359.

Kuno, Susumu. 1982. The focus of the question and the focus of the answer. In Robinson Schneider, Kevin Tuite & Robert Chametzky (eds.), *Papers from the Parasession on Nondeclaratives, April 17th, 1982*, 134–157. Chicago: Chicago Linguistic Society.

Kurumada, Chigusa, Meredith Brown, Sarah Bibyk, Daniel F. Pontillo & Michael K. Tanenhaus. 2014. Is it or isn‘t it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition* 133(2). 335–342.

Ladd, D. R. 1986. Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook* 3. 311–340.

Ladd, D. R. 1988. Declination “reset” and the hierarchical organization of utterances. *Journal of the Acoustical Society of America* 84(2). 530–544.

Ladd, D. R. 1990. Metrical representation of pitch register. In John Kingston & Mary E. Beckman (eds.), *Between the grammar and physics of speech* (Papers in laboratory phonology 1), 35–57. Cambridge: Cambridge University Press.

Ladd, D. R. 1993. In Defense of a Metrical Theory of Intonational Downstep. In Keith Snider & Harry van der Hulst (eds.), *The Phonology of Tone*: *The Representation of Tonal Register* (Linguistic Models 17), 109–132. Berlin: de Gruyter.

Ladd, D. R. 1994. Constraints on the gradient variability of pitch range, or, Pitch level 4 lives! In Patricia Keating (ed.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*, 43–63. Cambridge: Cambridge University Press.

Ladd, D. R. 2006. Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? *Rivista di Linguistica* 18(1). 19–38.

Ladd, D. R. 2008. *Intonational phonology*, 2nd edn. Cambridge, New York: Cambridge University Press.

Ladd, D. R. 2014. *Simultaneous structure in phonology*. Oxford: Oxford University Press.

Ladd, D. R. & Stephan Schmid. 2018. Obstruent voicing effects on F0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops. *Journal of Phonetics* 71. 229–248.

Ladd, D. R., Rory Turnbull, Charlotte Browne, Catherine Caldwell-Harris, Lesya Ganushchak, Kate Swoboda, Verity Woodfield & Dan Dediu. 2013. Patterns of Individual Differences in the Perception of Missing-Fundamental Tones. *Journal of Experimental Psychology: Human Perception and Performance* 39(5). 1386–1397.

Ladefoged, Peter. 2005. *Vowels and consonants*: *An introduction to the sounds of languages*, 2nd edn. Malden: Blackwell.

Ladefoged, Peter & Keith Johnson. 2011. *A course in phonetics*, 6th edn. Boston: Wadsworth/Cengage Learning.

Lambrecht, Knud. 1994. *Information structure and sentence form*: *Topic, focus, and the mental representations of discourse referents* (Cambridge studies in linguistics). Cambridge, New York: Cambridge University Press.

Leben, William R. 1973. *Suprasegmental phonology.* Boston, MA: MIT Dissertation.

Lefebvre, Claire & Pieter Muysken. 1988. *Mixed categories*: *Nominalizations in Quechua* (Studies in natural language and linguistic theory). Dordrecht: Kluwer.

Leonetti, Manuel & Victoria Escandell-Vidal. 2009. Fronting and Verum Focus in Spanish. In Andreas Dufter & Daniel Jacob (eds.), *Focus and background in Romance languages*: *[papers presented at a Workshop, Focus and Background in Romance Languages, convened during the 30th Romanistentag, in Vienna, Austria, 23–27 September 2007]* (Studies in language companion series 112), 155–204. Amsterdam: John Benjamins.

Levi, Susannah V. 2005. Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association* 35(1). 73–97.

Levinsohn, Stephen H. 1976. *The Inga language* (Janua linguarum: ser. practica 188). The Hague, Paris: Mouton.

Levinson, Stephen C. 2016. Turn-taking in Human Communication – Origins and Implications for Language Processing. *Trends in Cognitive Sciences* 20(1). 6–14.

Li, Charles N. & Sandra A. Thompson. 1976. Subject and Topic: A new typology of language. In Charles N. Li (ed.), *Subject and topic*, 457–489. New York: Academic Press.

Liberman, Mark. 1975. *The Intonational System of English.* Boston, MA: MIT Dissertation.

Liberman, Mark & Janet B. Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In Mark Aronoff, Richard T. Oehrle, Frances Kelley & Bonnie W. Stephens (eds.), *Language sound structure*: *Studies in phonology*, 157–233. Cambridge, Mass.: MIT Press.

Liberman, Mark & Alan S. Prince. 1977. On Stress and Linguistic Rhythm. *Linguistic Inquiry* 8(2). 249–336.

Liberman, Mark & Ivan Sag. 1974. Prosodic Form and Discourse Function. In Michael W. La Galy, Robert A. Fox & Anthony Bruck (eds.), *Papers from the Tenth Regional Meeting Chicago Linguistic Society*: *April 19–21, 1974*, 416–427. Chicago: Chicago Linguistic Society.

Lipski, John M. 1997. Spanish word stress: the interaction of moras and minimality. In Fernando Martínez-Gil & Alfonso Morales-Front (eds.), *Issues in the phonology and morphology of the major*

Iberian languages (Georgetown studies in Romance linguistics), 559–593. Washington, DC: Georgetown University Press.

López-Ortega, Nuria R. 2000. Tense, Aspect, and Narrative Structure in Spanish as a Second Language. *Hispania* 83(3). 488–502.

Łupkowski, Pawel & Jonathan Ginzburg. 2016. Query responses. *Journal of Language Modelling* 4(2). 245–292.

Marneffe, Marie-Catherine de & Judith Tonhauser. 2019. Inferring meaning from indirect answers to polar questions: The contribution of the rise-fall-rise contour. In Malte Zimmermann, Klaus v. Heusinger & Edgar Onea Gáspár (eds.), *Questions in discourse*: *Volume 2: Pragmatics* (Current research in the semantics/pragmatics interface Volume 36), 132–163. Leiden, Boston: Brill.

Martín Butragueño, Pedro & Érika Mendoza. 2018. Prosodic nuclear patterns in narrow and broad focus utterances: Pragmatic and social factors in Central Mexican Spanish. In Marco García García & Melanie Uth (eds.), *Focus realization in Romance and beyond* (Studies in language companion series Volume 201), 132–172. Amsterdam, Philadelphia: John Benjamins.

Martinović, Martina. 2013. The Topic-Comment Structure in Copular Sentences: Evidence from Wolof. *Annual Meeting of the Berkeley Linguistics Society* 39(1). 137.

Maskikit-Essed, Raechel & Carlos Gussenhoven. 2016. No stress, no pitch accent, no prosodic focus: the case of Ambonese Malay. *Phonology* 33(2). 353–389.

Matić, Dejan & Daniel Wedgwood. 2013. The meanings of focus: The significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49(01). 127–163.

Max Planck Institute for Evolutionary Anthropology. 2015. The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses. https://www.eva.mpg.de/lingua/resources/glossing-rules.php (11 October, 2018).

McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17(2). 207–263.

McCarthy, John J. 2016. The theory and practice of Harmonic Serialism. http://works.bepress.com/john_j_mccarthy/112/.

McCarthy, John J. & Alan S. Prince. 1993. Generalized alignment. *Yearbook of Morphology* 12. 1–69.

McCarthy, John J. & Alan S. Prince. 1995. Faithfulness and Reduplicative Identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *Papers in Optimality Theory*, 249–384. Amherst: Graduate Linguistic Student Association University of Massachusetts.

McCarthy, John J. & Alan S. Prince. 2001[1993]. *Prosodic Morphology*: *Constraint Interaction and Satisfaction* (Rutgers University Center for Cognitive Science Technical Report).

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague u.a.: Mouton.

Meinschaefer, Judith. in press. Effects of Stress. In Adam Ledgeway & Martin Maiden (eds.), *The Cambridge Handbook of Romance Linguistics*, 32 pages. Cambridge: Cambridge University Press.

Mücke, Doris, Martine Grice, Johannes Becker & Anne Hermes. 2009. Sources of variation in tonal alignment: Evidence from acoustic and kinematic data. *Journal of Phonetics* 37(3). 321–338.

Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge, UK, New York: Cambridge University Press.

Muntendam, Antje. 2010. *Linguistic Transfer in Andean Spanish: Syntax or Pragmatics?* Urbana-Champaign: University of Illinois at Urbana-Champaign Dissertation.

Muntendam, Antje. 2015. Discourse Deixis in Southern Quechua: A Case Study on Topic and Focus. In Marilyn S. Manley & Antje Muntendam (eds.), *Quechua expressions of stance and deixis* (Brill's Studies in the Indigenous Languages of the Americas 11), 208–258. Leiden, Boston: Brill.

Muntendam, Antje. 2017. *Cross-linguistic transfer in the prosodic domain: Experimental studies on bilingual Spanish and Quechua* (XXXV Romanistentag Zürich). Zürich.

Muntendam, Antje G. & Francisco Torreira. 2016. Focus and prosody in Spanish and Quechua: Insights from an interactive task. In Meghan E. Armstrong, Nicholas Henriksen & Maria d. M. Vanrell (eds.), *Intonational Grammar in Ibero-Romance*: *Approaches across linguistic subfields* (Issues in Hispanic and Lusophone linguistics), 69–89. Amsterdam: John Benjamins Publishing Company.

Muysken, Pieter. 1995. Focus in Quechua. In Katalin É. Kiss (ed.), *Discourse configurational languages* (Oxford studies in comparative syntax), 375–393. New York: Oxford University Press.

Myrberg, Sara. 2013. Sisterhood in prosodic branching. *Phonology* 30(1). 73–124.

Nance, Claire, Sam Kirkham & Eve Groarke. 2018. Studying Intonation in Varieties of English: Gender and Individual Variation in Liverpool. In Natalie Braber & Sandra Jansen (eds.), *Sociolinguistics in England*, 275–295. London: Palgrave Macmillan Limited.

Navarro Tomás, Tomás. 1968 [1944]. *Manual de Entonación Española*. La Habana: Edición Revolucionaria.

Nespor, Marina & Irene Vogel. 2007. *Prosodic Phonology*: *With a New Foreword* (Studies in generative grammar 28). Berlin: de Gruyter.

Nibert, Holly J. 1999. A Perception Study of Intermediate Phrasing in Spanish Intonation. In Javier Gutiérrez-Rexach & Fernando Martínez-Gil (eds.), *Advances in Hispanic linguistics*: *Papers from the 2nd Hispanic Linguistics Symposium; [held on October 9–11, 1998 at the Ohio State University in Columbus, Ohio]*, 231–247. Somerville, Mass.: Cascadilla Press.

Nibert, Holly J. 2000. *Phonetic and Phonological Evidence for Intermediate Phrasing in Spanish Intonation.* Urbana-Champaign: University of Illinois at Urbana-Champaign Dissertation.

Niebuhr, Oliver, Mariapaola D'Imperio, Barbara Gili-Fivela & Francesco Cangemi. 2011. Are there "shapers" and "aligners"? Individual differences in signalling pitch accent category. In Wai-Sum Lee & Eric Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 120–123. Hong Kong: City University of Hong Kong.

Ocampo, Francisco. 2003. The expression of topic in spoken Spanish: An empirical study. In Rafael A. Núñez Cedeño, Luis López & Richard Cameron (eds.), *A Romance perspective on language knowledge and use*: *Selected papers from the 31st Linguistics Symposium on Romance Languages (LSRL), Chicago, 19–22 April 2001*, 195–208. Amsterdam: John Benjamins.

Ocampo, Francisco. 2010. The Place of Conversational Data in Spanish Syntax: Topic, Focus, and Word Order. *Studies in Hispanic and Lusophone Linguistics* 3(2). 533–543.

Odden, David. 1986. On the Role of the Obligatory Contour Principle in Phonological Theory. *Language* 62(2). 353–383.

Ohala, John J. 1978. Production of Tone. In V. A. Fromkin (ed.), *Tone: A Linguistic Survey*, 5–39. New York: Academic Press.

Onaka, Akiko. 2003. Domain-Initial Strengthening in Japanese: An Acoustic and Articulatory Study. In M. J. Solé & Daniel Recasens i Vives (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences [also] 15th IcPhS, Barcelona 3–9 August 2003*, 2091–2094. Barcelona: Universitat Autónoma de Barcelona.

O'Rourke, Erin. 2005. *Intonation and Language Contact: A Case Study of two Varieties of Peruvian Spanish.* Urbana-Champaign: University of Illinois Dissertation.

O'Rourke, Erin. 2009. Phonetics and phonology of Cuzco Quechua declarative intonation: An instrumental analysis. *Journal of the International Phonetic Association* 39(3). 291–312.

O'Rourke, Erin. 2010. Ecuadorian Andean Spanish Intonation. In Pilar Prieto & Paolo Roseano (eds.), *Transcription of intonation of the Spanish language*, 227–253. München: LINCOM.

O'Rourke, Erin. 2012. The realization of contrastive focus in Peruvian Spanish intonation. *Lingua* 122(5). 494–510.

O'Rourke, Erin & Tod D. Swanson. 2013. Tena Quichua. *Journal of the International Phonetic Association* 43(1). 107–120.

Ortega-Llebaria, Marta & Pilar Prieto. 2007. Disentangling Stress from Accent in Spanish: Production Patterns of the Stress Contrast in Deaccented Syllables. In Pilar Prieto, Joan Mascaró & Maria J. Solé (eds.), *Segmental and prosodic issues in romance phonology* (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory v. 282), 155–176. Amsterdam, Philadelphia: John Benjamins Pub.

Ortega-Llebaria, Marta & Pilar Prieto. 2011. Acoustic Correlates of Stress in Central Catalan and Castilian Spanish. *Language and Speech* 54(1). 73–97.

Otheguy, Ricardo L., Ofelia García & Wallis Reid. 2015. Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review* 6(3). 281–307.

Parker, Gary. 1976. *Gramática quechua: Ancash-Huailas*. Lima: Ministerio de Educación.

Parker, Gary & Amancio Chavez Reyes. 1976. *Diccionario quechua: Ancash-Huailas*. Lima: Ministerio de Educación.

Parker, Steve. 1997. An OT Account of Laryngealization in Cuzco Quechua. *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 41(1). 1–11.

Pérez Silva, Jorge I., Jorge Acurio Palma & Raúl Bendezú Araujo. 2008. *Contra el prejuicio lingüístico de la motosidad*: *Un estudio de las vocales del castellano andino desde la fonética acústica* (Publicación del Instituto Riva-Agüero 232). Lima: Pontificia Universidad Católica del Perú.

Peters, Jörg. 2008. Tone and intonation in the dialect of Hasselt. *Linguistics* 46(5). 983–1018.

Peters, Jörg, Judith Hanssen & Carlos Gussenhoven. 2015. The timing of nuclear falls: Evidence from Dutch, West Frisian, Dutch Low Saxon, German Low Saxon, and High German. *Laboratory Phonology* 6(1). 1–52.

Petrone, Caterina, Hubert Truckenbrodt, Caroline Wellmann, Julia Holzgrefe-Lang, Isabell Wartenburger & Barbara Höhle. 2017. Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics* 61. 71–92.

Pierrehumbert, Janet B. 1979. The perception of fundamental frequency declination. *The Journal of the Acoustical Society of America* 66(2). 363–369.

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation.* Boston: MIT PhD / Dissertation.

Pierrehumbert, Janet B. & Mary E. Beckman. 1988. *Japanese tone structure* (Linguistic Inquiry Monographs 15). Cambridge, Mass: MIT Press.

Pierrehumbert, Janet B. & Shirley A. Steele. 1989. Categories of Tonal Alignment in English. *Phonetica* 46. 181–196.

Pompino-Marschall, Bernd. 2009. *Einführung in die Phonetik* (De Gruyter Studienbuch), 3rd edn. Berlin: de Gruyter.

Post, Brechtje. 2000. Pitch accents, Liaison and the Phonological Phrase in French. *Probus* 12(1). 127–164.

Post, Brechtje. 2002. French Tonal Structures. In Bernard Bel & Isabelle Marlien (eds.), *Speech Prosody 2002*: *Proceedings of the First International Conference on Speech Prosody*, 583–586. Aix-en-Provence: Laboratoire Parole et Langage.

Potts, Christopher. 2005. *The logic of conventional implicatures* (Oxford studies in theoretical linguistics 7). Oxford: Oxford University Press.

Potts, Christopher. 2015. Presupposition and Implicature. In Chris Fox & Shalom Lappin (eds.), *The handbook of contemporary semantic theory* (Blackwell handbooks in linguistics), 168–202. Malden: Wiley-Blackwell.

Prieto, Pilar. 2006. Phonological phrasing in Spanish. In Fernando Martínez-Gil & Sonia Colina (eds.), *Optimality-theoretic studies in Spanish phonology* (Linguistik aktuell/Linguistics today 99), 39–61. Amsterdam, Philadelphia: John Benjamins.

Prieto, Pilar. 2015. Intonational meaning. *Wiley interdisciplinary reviews. Cognitive science* 6(4). 371–381.

Prieto, Pilar, Joan Borràs-Comes, Teresa Cabré, Crespo-Sendra Verònica, Ignasi Mascaró, Paolo Roseano, Rafèu Sichel-Bazin & Maria d. M. Vanrell. 2015. Intonational phonology of Catalan and its dialectal varieties. In Sónia Frota & Pilar Prieto (eds.), *Intonation in Romance*, 9–62. Oxford: Oxford University Press.

Prieto, Pilar & Marta Ortega-Llebaria. 2009. Do complex pitch gestures induce syllable lengthening in Catalan and Spanish? In Marina C. Vigário, Sónia Frota & Maria J. Freitas (eds.), *Phonetics and phonology*: *Interactions and interrelations* (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory v. 306), 51–69. Amsterdam, Philadelphia: John Benjamins Pub. Co.

Prieto, Pilar & Paolo Roseano. 2009–2013. Atlas interactivo de la entonación del español. http://prosodia.upf.edu/atlasentonacion/ (7 July, 2019).

Prieto, Pilar & Paolo Roseano (eds.). 2010. *Transcription of intonation of the Spanish language*. München: LINCOM.

Prieto, Pilar, Chilin Shih & Holly Nibert. 1996. Pitch downtrend in Spanish. *Journal of Phonetics* 24(4). 445–473.

Prieto, Pilar & Francisco Torreira. 2007. The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35(4). 473–500.

Prieto, Pilar, Jan van Santen & Julia Hirschberg. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23. 429–451.

Prince, Alan S. & Paul Smolensky. 2004 [1993]. *Optimality Theory*: *Constraint Interaction in Generative Grammar*. Oxford: Blackwell.

Prince, Ellen F. 1981. Toward a Taxonomy of Given-New Information. In Peter Cole (ed.), *Radical pragmatics*: *Edited by Peter Cole*, 223–255. New York: Academic Press.

Prince, Ellen F. 1992. The ZPG Letter: Subjects, Definiteness, and Information-status. In William C. Mann & Sandra A. Thompson (eds.), *Discourse description*: *Diverse linguistic analyses of a fund-raising text* (Pragmatics & beyond New series 16), 295–326. Amsterdam, Philadelphia: John Benjamins.

Quené, Hugo. 2014. hqmisc: Miscellaneous convenience functions and dataset: R package version 0.1-1. https://CRAN.R-project.org/package=hqmisc.

Quilis, Antonio. 1993. *Tratado de fonología y fonética españolas*. Madrid: Editorial Gredos.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Wien: R Foundation for Statistical Computing.

Rao, Rajiv. 2007. On the Phonological Phrasing Patterns in the Spanish of Lima, Péru. *Southwest Journal of Linguistics* 26(1). 81–111.

Rao, Rajiv. 2008. Observations on the Roles of Prosody and Syntax in the Phonological Phrasing of Barcelona Spanish. *The Linguistics Journal* 3(3). 84–131.

Rao, Rajiv. 2009. Deaccenting in Spontaneous Speech in Barcelona Spanish. *Studies in Hispanic and Lusophone Linguistics* 2(1). 1–45.

Rao, Rajiv. 2010. Final Lengthening and Pause Duration in Three Dialects of Spanish. In Marta Ortega-Llebaria (ed.), *Selected proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, 69–82. Somerville, MA: Cascadilla Press.

Rathcke, Tamara. 2013. On the neutralizing status of truncation in intonation: A perception study of boundary tones in German and Russian. *Journal of Phonetics* 41(3–4). 172–185.

Rathcke, Tamara. 2016. How Truncating Are 'Truncating Languages'?: Evidence from Russian and German. *Phonetica* 73(3–4). 194–228.
Reich, Uli. 2018. Presupposed Modality. In Marco García García & Melanie Uth (eds.), *Focus realization in Romance and beyond* (Studies in language companion series Volume 201), 203–227. Amsterdam, Philadelphia: John Benjamins.
Reich, Uli & Martin Rohrmeier. 2014. Batidas latinas: On rhythm and meter in Spanish and Portuguese and other forms of music. In Javier Caro Reina & Renata Szczepaniak (eds.), *Syllable and word languages* (Linguae & Litterae 40), 391–420. Berlin: de Gruyter.
Riad, Tomas. 1998. Towards a Scandinavian accent typology. In Wolfgang Kehrein & Richard Wiese (eds.), *Phonology and Morphology of the Germanic Languages* (Linguistische Arbeiten 386), 77–109. Berlin: de Gruyter.
Riesberg, Sonja, Janina Kalbertodt, Stefan Baumann & Nikolaus P. Himmelmann. 2018. On the perception of prosodic prominences and boundaries in Papuan Malay. In Sonja Riesberg, Asako Shiohara & Atsuko Utsumi (eds.), *Perspectives on information structure in Austronesian languages* (Studies in Diversity Linguistics 21), 389–414. Berlin: Language Science Press.
Riester, Arndt. 2019. Constructing QUD trees. In Malte Zimmermann, Klaus v. Heusinger & Edgar Onea Gáspár (eds.), *Questions in discourse*: *Volume 2: Pragmatics* (Current research in the semantics/pragmatics interface Volume 36), 164–192. Leiden, Boston: Brill.
Riester, Arndt, Lisa Brunetti & Kordula de Kuthy. 2018. Annotation guidelines for Questions under Discussion and information structure. In Evangelia Adamou, Katharina Haude & Martine Vanhove (eds.), *Information structure in lesser-described languages*: *Studies in prosody and syntax* (Studies in language companion series 199), 403–443. Amsterdam, Philadelphia: John Benjamins.
Riester, Arndt & Asako Shiohara. 2018. Information structure in Sumbawa: A QUD analysis. In Sonja Riesberg, Asako Shiohara & Atsuko Utsumi (eds.), *Perspectives on information structure in Austronesian languages* (Studies in Diversity Linguistics 21), 285–311. Berlin: Language Science Press.
Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. In Liliane Haegeman (ed.), *Elements of grammar*: *Handbook in generative syntax* (Kluwer international handbooks of linguistics 1). Dordrecht: Kluwer.
Roberts, Craige. 2011. Topics. In Claudia Maienborn, Klaus v. Heusinger & Paul Portner (eds.), *Semantics*: *An international handbook of natural language meaning*. Volume 2 (Handbücher zur Sprach- und Kommunikationswissenschaft 33.2), 1908–1934. Berlin: de Gruyter.
Roberts, Craige. 2012a. Information Structure: Afterword. *Semantics and Pragmatics* 5(Article 7). 1–19.
Roberts, Craige. 2012b. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(Article 6). 1–69.
Roberts, Craige. 2015. The Character of Epistemic Modality: Evidentiality, Indexicality, and What's at Issue. http://www.ling.ohio-state.edu/~roberts.21/Roberts.EpistemicModality.pdf (7 February, 2017).
Roberts, Craige. 2017. Agreeing and Assessing: Epistemic Modals and the Question Under Discussion. https://www.asc.ohio-state.edu/roberts.21/Roberts.A&A.pdf.
Roca, Iggy. 1997. On the role of accent in stress systems: Spanish evidence. In Fernando Martínez-Gil & Alfonso Morales-Front (eds.), *Issues in the phonology and morphology of the major Iberian languages* (Georgetown studies in Romance linguistics), 619–664. Washington, DC: Georgetown University Press.
Roca, Iggy. 1999. Stress in the Romance languages. In Harry van der Hulst (ed.), *Word prosodic systems in the languages of Europe* (Empirical approaches to language typology 20-4), 659–811. Berlin, New York: Mouton de Gruyter.

Roca, Iggy. 2006. The Spanish Stress Window. In Fernando Martínez-Gil & Sonia Colina (eds.), *Optimality-theoretic studies in Spanish phonology* (Linguistik aktuell/Linguistics today 99), 239–277. Amsterdam, Philadelphia: John Benjamins.

Roelofsen, Floris & Donka F. Farkas. 2015. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language* 91(2). 359–414.

Roessig, Simon. 2021. *Categoriality and continuity in prosodic prominence* (Studies in Laboratory Phonology 10). Berlin: Language Science Press.

Roettger, Timo & Matthew Gordon. 2017. Methodological issues in the study of word stress correlates. *Linguistics Vanguard* 3(1). 86–96.

Roettger, Timo B. 2017. *Tonal placement in Tashlhiyt* (Studies in Laboratory Phonology 3). Berlin: Language Science Press.

Roll, Mikael, Merle Horne & Magnus Lindgren. 2009. Left-edge boundary tone and main clause verb effects on syntactic processing in embedded clauses – An ERP study. *Journal of Neurolinguistics* 22(1). 55–73.

Romera, Magdalena & Gorka Elordieta. 2013. Prosodic accommodation in language contact: Spanish intonation in Majorca. *International Journal of the Sociology of Language*(221). 127–151.

Rooth, Mats. 1985. *Association with Focus.* Amherst: University of Massachusetts Dissertation.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116.

RStudio, Inc. 2009–2020. *RStudio*: RStudio, Inc.

Sadock, Jerrold M. & Arnold M. Zwicky. 1985. Speech act distinctions in syntax. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*: *Volume 1: Clause Structure*, 155–196. Cambridge: Cambridge University Press.

Sánchez, Liliana. 2010. *The morphology and syntax of topic and focus*: *Minimalist inquiries in the Quechua periphery* (Linguistik aktuell / Linguistics Today 169). Amsterdam: Benjamins.

Savino, Michelina. 2012. Degrees of (un)certainty in Bari Italian yes-no question intonation: perceptual evidence. In Andrzej Zuczkowski, Ramona Bongelli, Ilaria Riccioni & Carla Canestrari (eds.), *The Communication of Certainty and Uncertainty: Linguistic, Psychological, Philosophical Aspects*. Proceedings of the International Conference. University of Macerata (Italy), 3rd-5th October 2012. Under the auspices of the University of Verona (Italy) (iada.online.series 1), 52–67: International Association for Dialogue Analysis.

Schiering, René, Balthasar Bickel & Kristine A. Hildebrandt. 2010. The prosodic word is not universal, but emergent. *Journal of Linguistics* 46(3). 657–709.

Schmeißer, Anika, Malin Hager, Laia Arnaus Gil, Veronika Jansen, Jasmin Geveler, Nadine Eichler, Marisa Patuto & Natascha Müller. 2015. Related but different: The two concepts of language dominance and language proficiency. In Carmen Silva-Corvalán & Jeanine Treffers-Daller (eds.), *Language dominance in bilinguals*: *Issues of measurement and operationalization*, 36–65. Cambridge: Cambridge University Press.

Schweitzer, Katrin. 2012. *Frequency Effects on Pitch Accents: Towards an Exemplar-Theoretic Approach to Intonation.* Stuttgart: Universität Stuttgart Dissertation.

Schweitzer, Katrin, Michael Walsh, Sasha Calhoun, Hinrich Schütze, Bernd Möbius, Antje Schweitzer & Grzegorz Dogil. 2015. Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation. *Speech Communication* 66. 65–81.

Seifart, Frank, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nivja H. de Jong & Balthasar Bickel. 2018. Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences of the United States of America* 115(22). 5720–5725.

Selkirk, Elisabeth. 1996. The Prosodic Structure of Function Words. In James L. Morgan & Katherine Demuth (eds.), *Signal to Syntax*: *Bootstrapping From Speech To Grammar in Early Acquisition*, 187–214. Mahwah, NJ: Erlbaum.

Selkirk, Elisabeth. 2011. The Syntax-Phonology Interface. In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The Handbook of Phonological Theory* (Blackwell handbooks in linguistics), 435–484. Oxford: Wiley-Blackwell.

Selkirk, Elisabeth, Takahito Shinya & Mariko Sugahara. 2003. Degree of Initial Lowering in Japanese as a Reflex of Prosodic Structure Organization. In M. J. Solé & Daniel Recasens i Vives (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences [also] 15th IcPhS, Barcelona 3–9 August 2003*, 491–494. Barcelona: Universitat Autónoma de Barcelona.

Selkirk, Elisabeth O. 1984. *Phonology and syntax*: *The relation between sound and structure* (Current studies in linguistics series 10). Cambridge: MIT Press.

Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzluff, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock & Susanne Uhmann. 2009. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10. 353–402.

Serrano, Celia T. 2010. *Prosodia y sintaxis en el procesamiento oracional.* Madrid: Universidad Autónoma de Madrid Dissertation.

Shannon, C. E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27(4). 623–656.

Simons, Mandy, Judith Tonhauser, David Beaver & Craige Roberts. 2010. What projects and why. *Semantics and Linguistic Theory* 20. 309–327.

Siptár, Péter & Miklós Törkenczy. 2007. *The Phonology of Hungarian* (The phonology of the world's languages). Oxford: Oxford University Press UK.

Solé, Maria-Josep. 2012. Natural and unnatural patterns of sound change? In Maria-Josep Solé & Daniel Recasens (eds.), *The Initiation of Sound Change*: *Perception, production, and social factors* (Current issues in linguistic theory), 123–145. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Solé, Maria-Josep. 2014. The perception of voice-initiating gestures. *Laboratory Phonology* 5(1). 37–68.

Sosa, Juan M. 1999. *La entonación del español*: *Su estructura fónica, variabilidad y dialectología* (Lingüística). Madrid: Edicíon Cátedra.

Sosa, Juan M. 2003. La notación tonal del español en el modelo SP-ToBI. In Pilar Prieto (ed.), *Teorías de la entonación* (Ariel lingüística), 185–208. Barcelona: Ed. Ariel.

Stalnaker, Robert C. 1974. Pragmatic Presuppositions. In Milton K. Munitz & Peter Unger (eds.), *Semantics and Philosophy*, 197–213. New York: New York University Press.

Stalnaker, Robert C. 2002. Common Ground. *Linguistics and Philosophy* 25(701–721).

Steedman, Mark. 1991. Structure and Intonation. *Language* 67(2). 260–296.

Steedman, Mark. 2000. Information Structure and the Syntax-Phonology Interface. *Linguistic Inquiry* 31(4). 649–689.

Stewart, Anne M. 1984. New approaches to coping with stress: A case study in Conchucos Quechua. *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 28. 193–212.

Stewart, Anne M. 1987. *Clause-combining in Conchucos Quechua discourse.* Los Angeles: UC Los Angeles Dissertation.

Stivers, Tanya, N. J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan P. d. Ruiter, Kyung-Eun Yoon & Stephen C. Levinson.

2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106(26). 10587–10592.

Strycharczuk, Patrycja & Martin Kohlberger. 2016. Resyllabification Reconsidered: On the Durational Properties of Word-Final /s/ in Spanish. *Laboratory Phonology* 7(1). 1–24.

Taff, Alice. 1999. *Phonetics and Phonology of Unangan (Eastern Aleut) Intonation:* University of Washington Dissertation.

Terken, Jacques & Dik Hermes. 2000. The Perception of Prosodic Prominence. In Merle Horne (ed.), *Prosody: Theory and Experiment*: *Studies Presented to Gösta Bruce* (Text, speech and language technology 14), 89–128. Dordrecht: Springer.

Tilsen, Sam. 2013. A Dynamical Model of Hierarchical Selection and Coordination in Speech Planning. *PLoS ONE* 8(4). 1–18.

Tilsen, Sam. 2016. Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics* 55. 53–77.

Tilsen, Sam. 2019. Motoric Mechanisms for the Emergence of Non-local Phonological Patterns. *Frontiers in psychology* 10. 1–25.

Tokar, Alexander. 2017. *Stress variation in English* (Language in Performance (LIP) 50). Tübingen: Narr Francke Attempto.

Tonhauser, Judith, David Beaver & Judith Degen. 2018. How Projective is Projective Content?: Gradience in Projectivity and At-issueness. *Journal of Semantics* 35(3). 495–542.

Tonhauser, Judith, David Beaver, Craige Roberts & Mandy Simons. 2013. Toward a Taxonomy of Projective Content. *Language* 89(1). 66–109.

Torreira, Francisco & Martine Grice. 2018. Melodic constructions in Spanish: Metrical structure determines the association properties of intonational tones. *Journal of the International Phonetic Association* 48(1). 9–32.

Torreira, Francisco, Miquel Simonet & José I. Hualde. 2014. Quasi-neutralization of stress contrasts in Spanish. In Nick Campbell, Dafydd Gibbon & Daniel Hirst (eds.), *Proceedings of the 7th international conference on Speech Prosody*, 197–201. Dublin.

Traunmüller, Hartmut. 1990. Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88(1). 97–100.

Traunmüller, Hartmut. 2005. Conversion between acoustic and auditory units of pitch and calculation of auditory intervals and distances. http://www2.ling.su.se/staff/hartmut/umrechnung.htm (24 April, 2018).

Traunmüller, Hartmut. 2017. Auditory scales of frequency representation. http://www2.ling.su.se/staff/hartmut/bark.htm (24 April, 2018).

Treffers-Daller, Jeanine. 2019. What Defines Language Dominance in Bilinguals? *Annual Review of Linguistics* 5(1). 375–393.

Truckenbrodt, Hubert. 2012. Semantics of intonation. In Claudia Maienborn, Klaus v. Heusinger & Paul Portner (eds.), *Semantics*: *An international handbook of natural language meaning*. Volume 3 (Handbücher zur Sprach- und Kommunikationswissenschaft 33.3), 2039–2070. Berlin [u.a.]: de Gruyter.

Truckenbrodt, Hubert & Caroline Féry. 2015. Hierarchical organisation and tonal scaling. *Phonology* 32(1). 19–47.

Turnbull, Rory, Rachel S. Burdin, Cynthia G. Clopper & Judith Tonhauser. 2015. Contextual predictability and the prosodic realisation of focus: A cross-linguistic comparison. *Language, Cognition and Neuroscience* 30(9). 1061–1076.

Uth, Melanie. 2014. Spanish preverbal subjects in contexts of narrow information focus: Non-contrastive focalization or epistemic-evidential marking? *Grazer Linguistische Studien* 81. 87–104.

Vaissière, Jacqueline. 1983. Language-Independent Prosodic Features. In Anne Cutler & D. R. Ladd (eds.), *Prosody: Models and Measurements* (Springer Series in Language and Communication 14), 53–66. Berlin, Heidelberg: Springer Berlin Heidelberg.

van de Ven, Marco & Carlos Gussenhoven. 2011. On the timing of the final rise in Dutch falling–rising intonation contours. *Journal of Phonetics* 39(2). 225–236.

van Deemter, Kees. 1994. What's New?: A Semantic Perspective on Sentence Accent. *Journal of Semantics* 11. 1–31.

van den Berg, Rob, Carlos Gussenhoven & Toni Rietveld. 1992. Downstep in Dutch: implications for a model. In Gerard Docherty & D. R. Ladd (eds.), *Gesture, segment, prosody: Papers in Laboratory Phonology II* (Papers in laboratory phonology 2), 335–359. Cambridge: Cambridge University Press.

van der Hulst, Harry. 1999a. Word accent. In Harry van der Hulst (ed.), *Word prosodic systems in the languages of Europe* (Empirical approaches to language typology 20-4), 3–115. Berlin, New York: Mouton de Gruyter.

van der Hulst, Harry (ed.). 1999b. *Word prosodic systems in the languages of Europe* (Empirical approaches to language typology 20-4). Berlin, New York: Mouton de Gruyter.

van der Hulst, Harry. 2010. A note on recursion in phonology. In Harry van der Hulst (ed.), *Recursion and Human Language* (Studies in generative grammar 104), 301–341. Berlin New York: de Gruyter.

van Santen, Jan & Julia Hirschberg. 1994. Segmental Effects on Timing and Height of Pitch Contours. *International Conference on Spoken Language Processing (ICSLP)*. 719–722.

van Zanten, Ellen & Rob Goedemans. 2009. Prominence in Indonesian: Stress, phrases, and boundaries. *Wacana* 11(2). 197–225.

van Zanten, Ellen, Rob Goedemans & Jos Pacilly. 2003. The status of word stress in Indonesian. In Jeroen M. van de Weijer, Vincent J. van Heuven & Harry van der Hulst (eds.), *The Phonological Spectrum Volume II: Suprasegmental Structure* (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory 234), 151–175. Amsterdam, Philadelphia: Benjamins.

Vanrell, Maria d. M., Meghan E. Armstrong & Pilar Prieto. 2017. Experimental Evidence for the Role of Intonation in Evidential Marking. *Language and Speech* 60(2). 242–259.

Vanrell, Maria d. M. & Olga Fernández Soriano. 2018. Language variation at the prosody-syntax interface: Focus in European Spanish. In Marco García García & Melanie Uth (eds.), *Focus realization in Romance and beyond* (Studies in language companion series Volume 201), 33–70. Amsterdam, Philadelphia: John Benjamins.

Vanrell, Maria d. M., Ignasi Mascaró, Francesc Torres-Tamarit & Pilar Prieto. 2013. Intonation as an Encoder of Speaker Certainty: Information and Confirmation Yes-No Questions in Catalan. *Language and Speech* 56(2). 163–190.

Venables, William N. & Brian D. Ripley. 2007. *Modern applied statistics with S* (Statistics and computing), 4th edn. New York, NY: Springer.

Venditti, Jennifer J. 2005. The J_ToBi Model of Japanese Intonation. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing* (Oxford linguistics), 172–200. Oxford, New York: Oxford University Press.

Venditti, Jennifer J., Sun-Ah Jun & Mary E. Beckman. 1996. Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English. In James L. Morgan & Katherine Demuth (eds.), *Signal to Syntax: Bootstrapping From Speech To Grammar in Early Acquisition*, 287–311. Mahwah, NJ: Erlbaum.

Venditti, Jennifer J., Kazuaki Maeda & Jan van Santen. 1998. Modeling Japanese Boundary Pitch Movements for Speech Synthesis. In Mike Edgington (ed.), *Proceedings of the 3rd ESCA Workshop on Speech Synthesis*, 317–322. New York: Institute of Electrical and Electronics Engineers.

Venditti, Jennifer J., Kikuo Maekawa & Mary E. Beckman. 2008. Prominence Marking in the Japanese Intonation System. In Shigeru Miyagawa & Mamoru Saito (eds.), *The Oxford Handbook of Japanese Linguistics*, 456–512. Oxford & New York: Oxford University Press.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change*: *With special reference to German, Germanic, Italian, and Latin*. Berlin, New York: Mouton de Gruyter.

Ward, Gregory & Julia Hirschberg. 1985. Implicating Uncertainty: The Pragmatics of Fall-Rise Intonation. *Language* 61(4). 747–776.

Weber, David J. 1989. *A grammar of Huallaga (Huánuco) Quechua* (Linguistics 112). Berkeley: University of California Press.

Welby, Pauline. 2006. French intonational structure: Evidence from tonal alignment. *Journal of Phonetics* 34(3). 343–371.

Welby, Pauline & Hélène Loevenbruck. 2006. Anchored down in Anchorage: Syllable structure, rate, and segmental anchoring in French. *Rivista di Linguistica* 18(1). 73–124.

Wetzels, Leo & Sérgio Meira. 2010. A Survey of South American stress systems. In Rob Goedemans, Harry van der Hulst & Ellen van Zanten (eds.), *A survey of word accentual patterns in the languages of the world*, 313–380. Berlin: de Gruyter.

Wickham, Hadley. 2016. *ggplot2*: *Elegant graphics for data analysis* (Use R!). Cham: Springer.

Wikipedia. 2021a. Languages of Peru. https://en.wikipedia.org/wiki/Languages_of_Peru (27 June, 2021).

Wikipedia. 2021b. Ruta departamental AN-110. https://es.wikipedia.org/wiki/Ruta_departamental_AN-110 (29 July, 2021).

Wiltschko, Martina. 2021. *The grammar of interactional language*. Cambridge: Cambridge University Press.

Wölck, Wolfgang. 1972. *Especificación y foco en quechua* (Documento de Trabajo). San Marcos.

Wollermann, Charlotte, Ulrich Schade, Bernhard Fisseni & Bernhard Schröder. 2010. Accentuation, Uncertainty and Exhaustivity – Towards a Model of Pragmatic Focus Interpretation. *Speech Prosody 2010*(100063). 1–4.

Wollermann, Charlotte, Bernhard Schröder & Ulrich Schade. 2014. Audiovisual Prosody of Uncertainty: an Overview. *Ricerche di Pedagogia e Didactica – Journal of Theories and Research in Education* 9(1). 137–157.

Zeileis, Achim. 2004. Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software* 11(10).

Zeileis, Achim & T. Hothorn. 2002. Diagnostic Checking in Regression Relationships. *R News* 2(3). 7–10.

Zeileis, Achim, Susanne Köll & Nathaniel Graham. 2020. Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R. *Journal of Statistical Software* 95(1).

Zheng, Xiaoju & Janet B. Pierrehumbert. 2010. The effects of prosodic prominence and serial position on duration perception. *The Journal of the Acoustical Society of America* 128(2). 851–859.

Zimmermann, Malte. 2008. Contrastive focus and emphasis. *Acta Linguistica Hungarica* 55(3–4). 347–360.

Zimmermann, Malte. 2011. Discourse Particles. In Claudia Maienborn, Klaus v. Heusinger & Paul Portner (eds.), *Semantics*: *An international handbook of natural language meaning*. Volume 2 (Handbücher zur Sprach- und Kommunikationswissenschaft 33.2), 2012–2038. Berlin: de Gruyter.

Zubizarreta, Maria L. 1998. *Prosody, focus, and word order* (Linguistic Inquiry Monographs 33). Cambridge, Mass.: MIT Press.

Zuckermann, Ghilʿad. 2003. *Language contact and lexical enrichment in Israeli Hebrew* (Palgrave studies in language history and language change). Basingstoke: Palgrave Macmillan.