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 oído 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**ié**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.

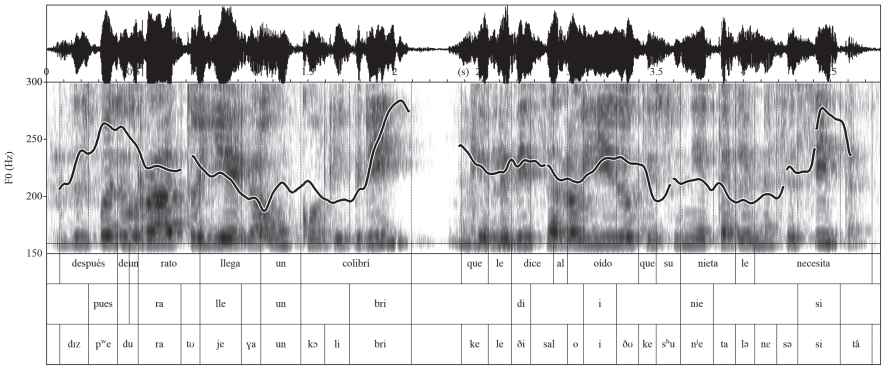


Figure 20[93]: ZR29_Cuento_ES_0083[94,95] (main accentuation declarative) Cf. (37).

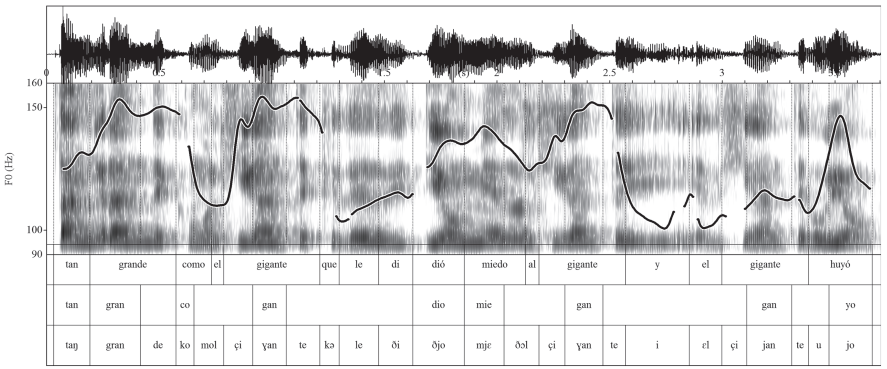


Figure 21: TP03_Cuent_ES_0513[96] (main accentuation declarative). Cf. (38).

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/

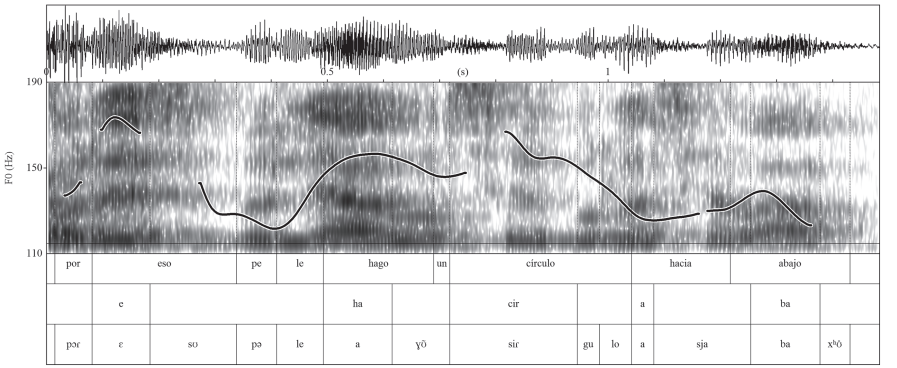


Figure 22: KP04_MT_ES_0799[97] (main accentuation declarative). Cf. (39).

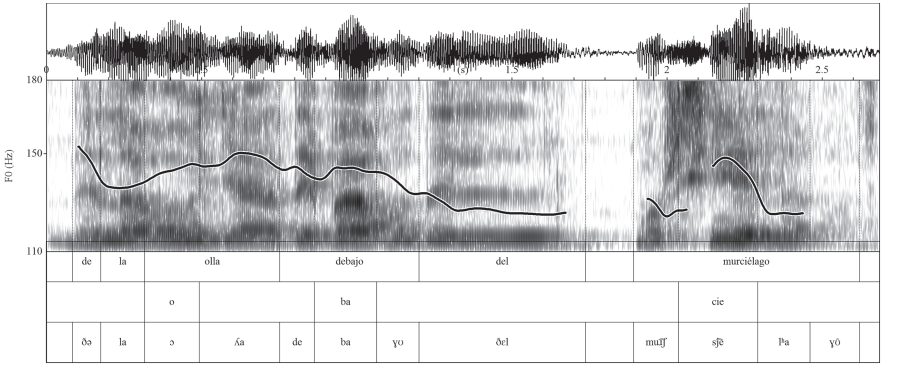


Figure 23: ZZ24_MT_ES_1269[98] (main accentuation declarative). Cf. (40).

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.

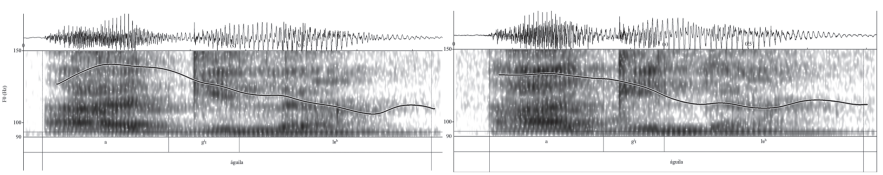


Figure 24: OA32_Conc_ES_1549[102] (left) and 0298[103] (right) (*águila* 'eagle').

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.

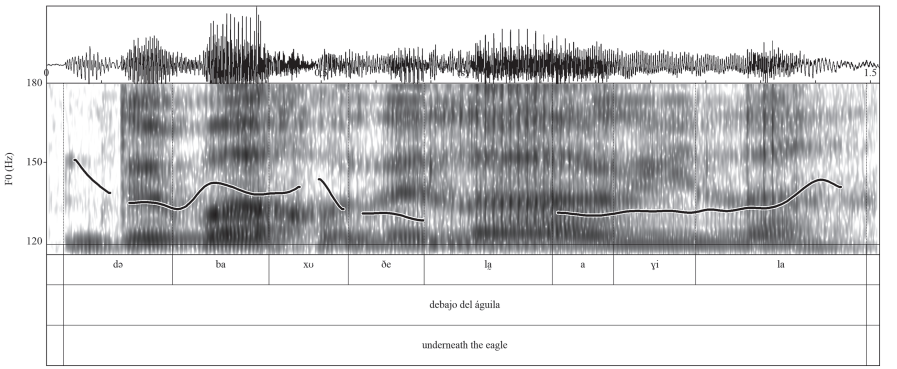


Figure 25: XQ33_Conc_ES_0966[104] (*debajo del águila* 'underneath the eagle').

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/

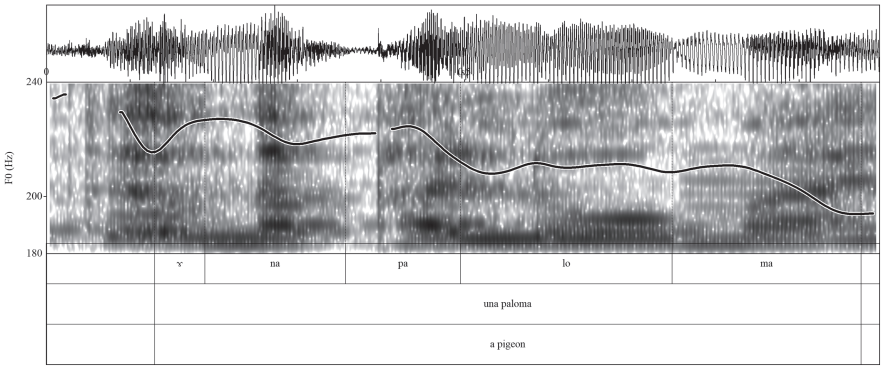


Figure 26: ZR29_Conc_ES_1065[105] (*una paloma* 'a pigeon').

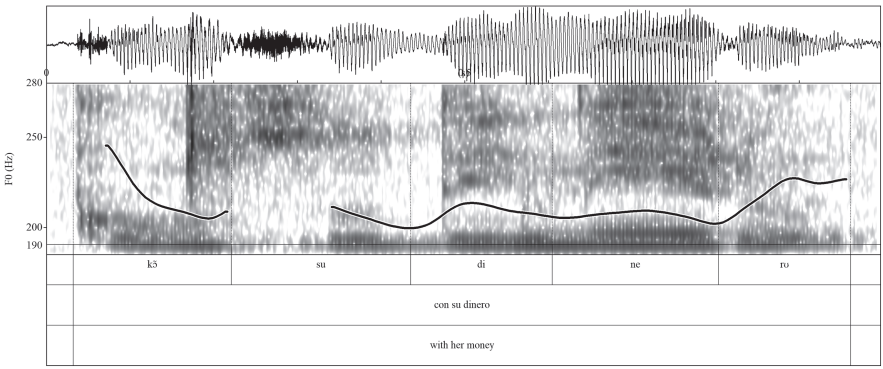


Figure 27: HA30_Conc_ES_0336[106] (*con su dinero* 'with her money').

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/

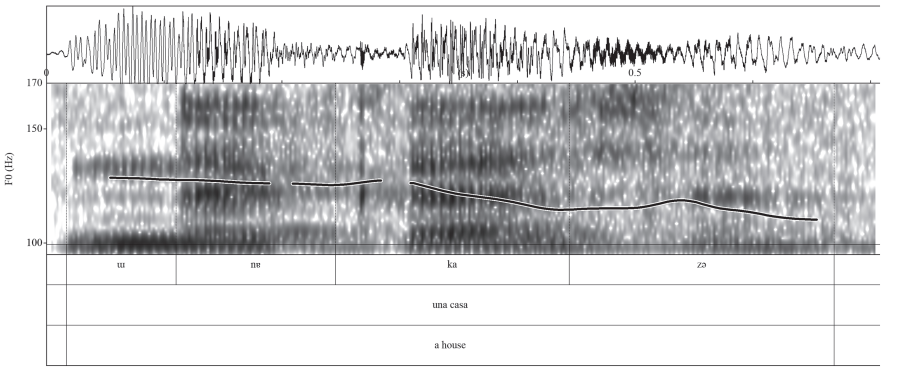


Figure 28: OZ14_Conc_ES_2242[107] (*una casa* 'a house').

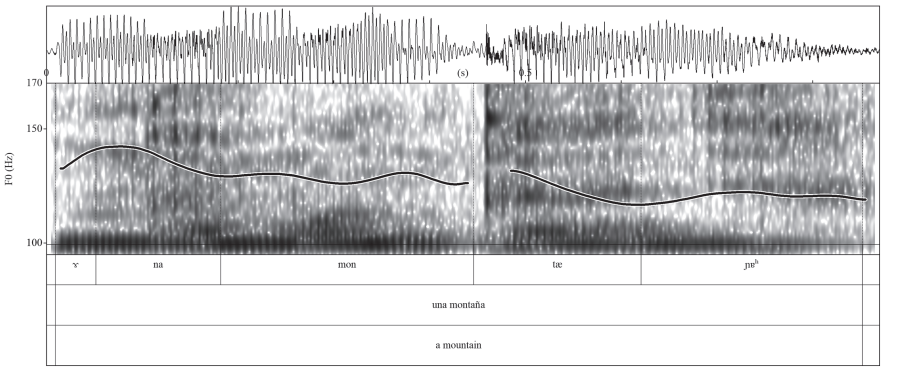


Figure 29: QZ13_Conc_ES_1567[108] (*una montaña* 'a mountain').

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/

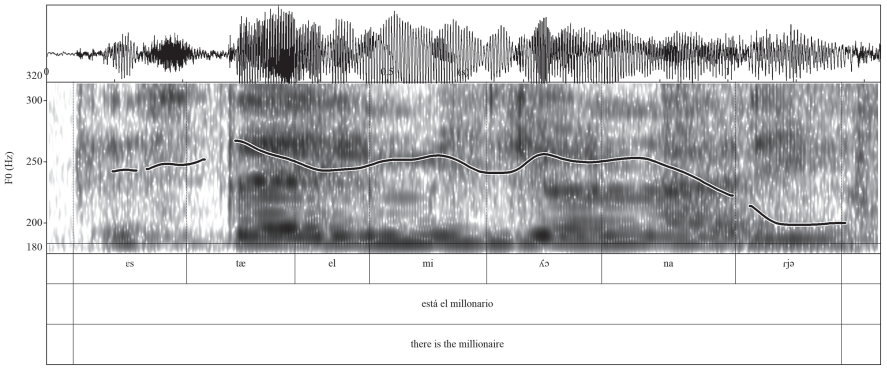


Figure 30: AZ23_Conc_ES_1851[109] (*está el millonario* '(there) is the millionaire').

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/

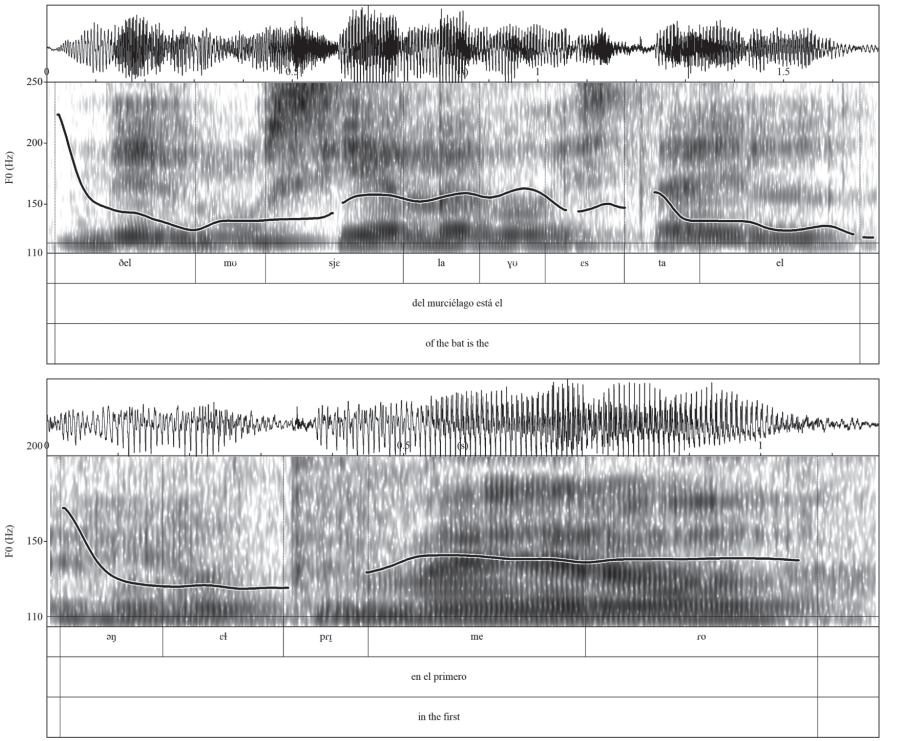


Figure 31: LC34_Conc_ES_1298[110] (*del murciélago está el* 'of the bat is the', top) and ZZ24_Conc_ES_0994[111] (*en el primero* 'in the first', bottom).

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/

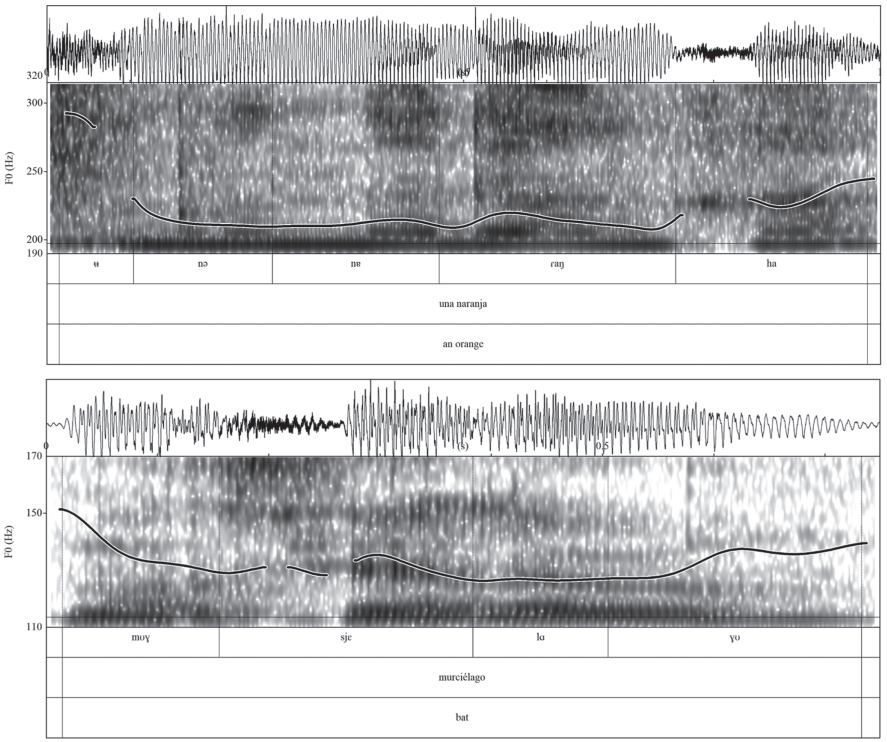


Figure 32: HA30_Conc_ES_0245[112] (*una naranja* 'an orange', top) and QF16_Conc_ES_0315[113] (*murciélago* 'bat', bottom).

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.

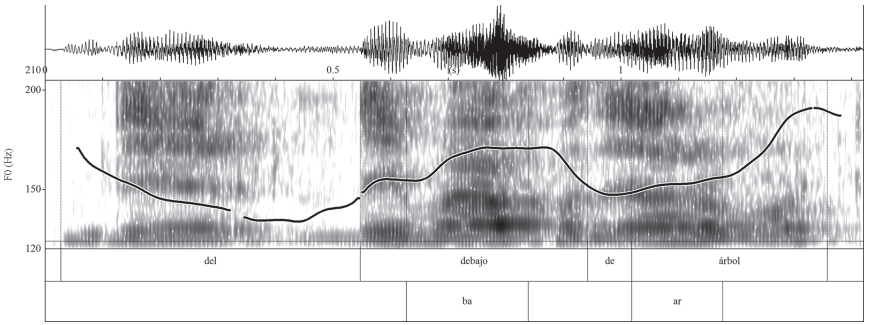


Figure 33: XU31_MT_ES_0871[116] (*debajo de árbol* 'below the tree'; neutral polar question with final paroxytone).

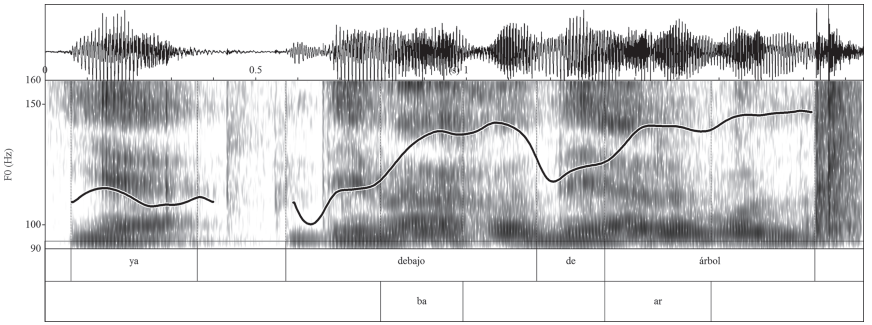


Figure 34: OA32_MT_ES_0890[117] (*ya debajo de árbol* 'yeah below the tree'; declarative with continuation rise, final paroxytone).

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.

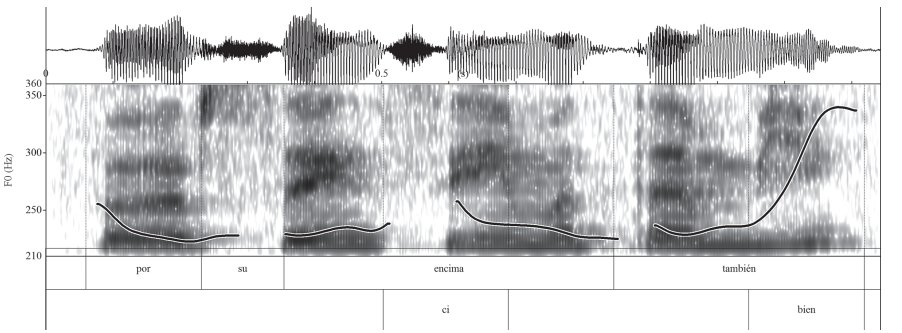


Figure 35: MD40_MT_ES_0324[119] (*por su encima también* 'also above it'; neutral polar question with final oxytone).

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

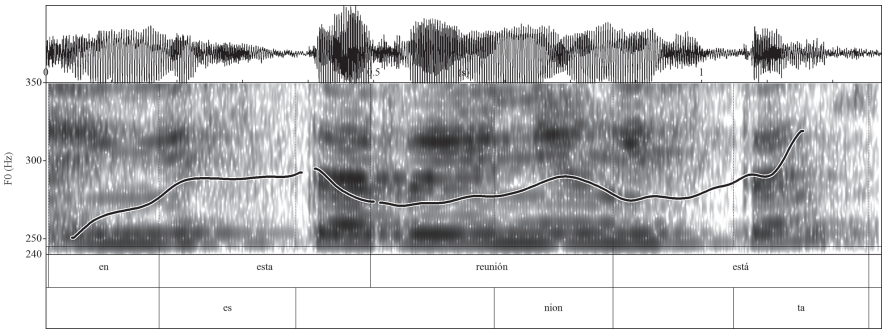


Figure 36: AZ23_Quien_ES_0747[120] (*en esta reunión está* 'is s/he in this meeting'; neutral polar question with final oxytone).

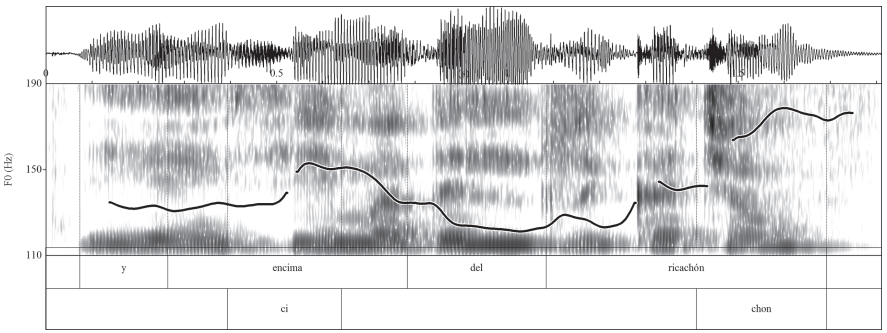


Figure 37: QZ13_MT_ES_0446[121] (*y encima del ricachón* 'and above the rich guy'; declarative with continuation rise, final oxytone).

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/

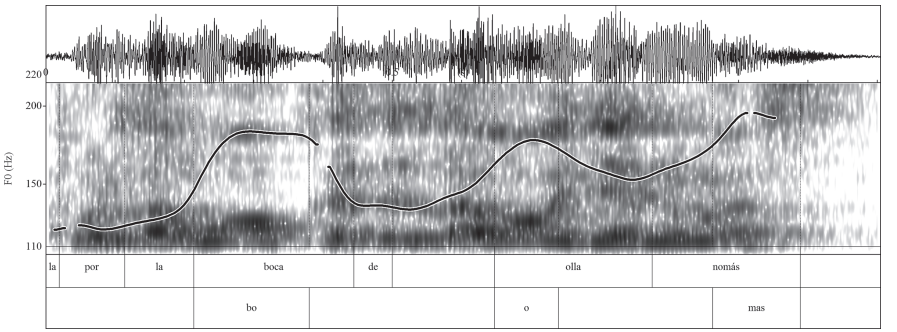


Figure 38: TP03_MT_ES_4425[122] (*por la boca de la olla nomás* 'just by the mouth of the pot'; declarative with continuation rise, final oxytone).

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/

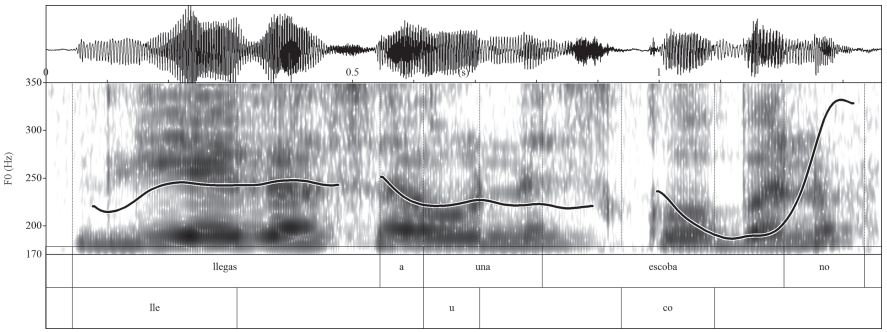


Figure 39: SO39_MT_ES_3237[123] (*llegas a una escoba, no* 'you get to a broom, right'; polar question with question tag after final paroxytone).

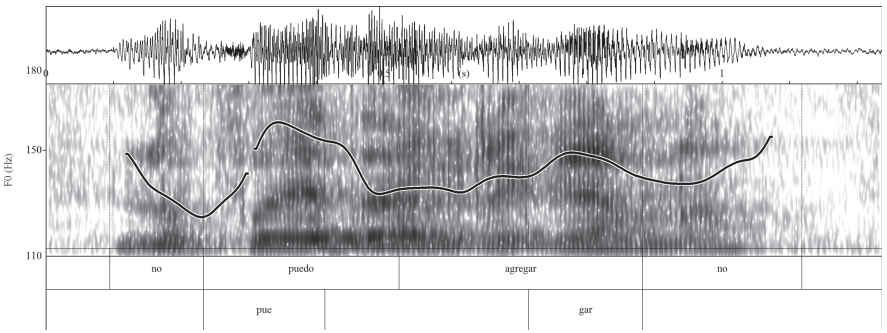


Figure 40: ZZ24_Quien_ES_0788[124] (*no puedo agregar, no* 'I can't add (to that), right'; polar question with question tag after final oxytone).

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.

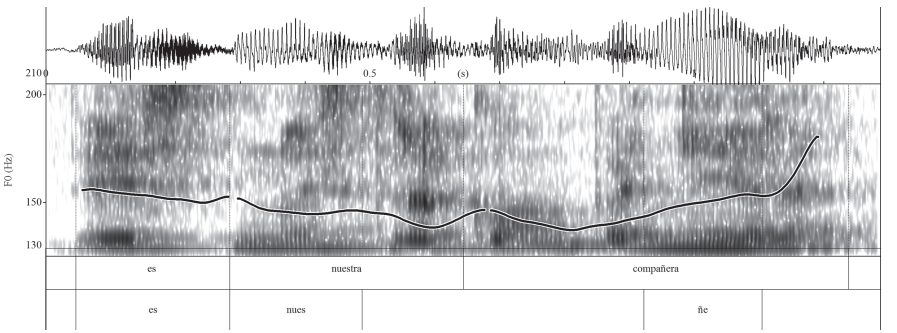


Figure 41: LC34_Quien_ES_1465[125] (*es nuestra compañera* 'is it our classmate'; neutral polar question with final paroxytone).

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 previ-

ously 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
52.1	mhm		sub-QUD How
56.6	su nombre se me ha ido *I've lost the name*		IS X CALLED IN *RELATION Y TO LANDMARK X*?
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**.

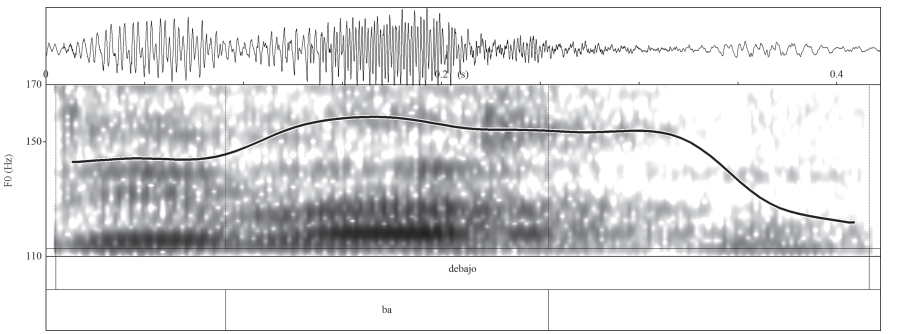


Figure 42: QF16_MT_ES_0615[129] (confirmation-seeking polar question). Cf. context (43).

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.

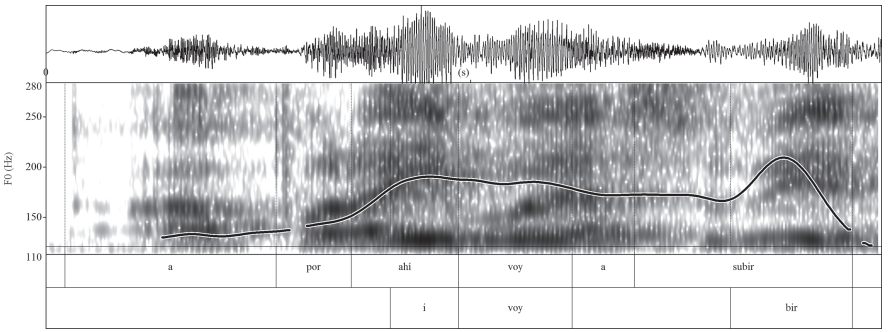


Figure 43: KP04_MT_ES_1975[133] (*a por ahí voy a subir* 'ah I'll go up there'; confirmation-seeking polar question with final oxytone).

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).

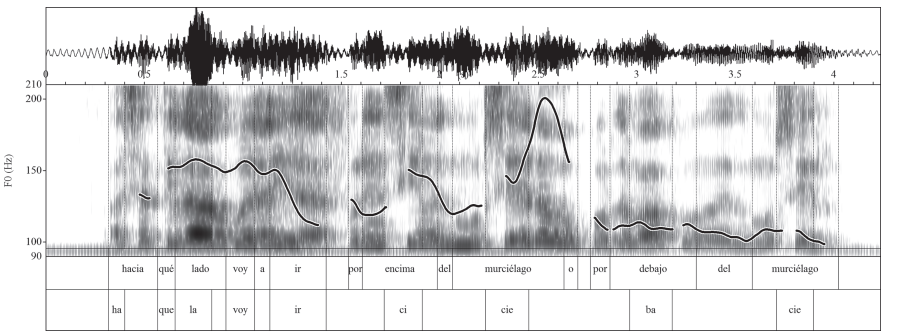


Figure 44: KP04_MT_ES_3219[136] (*hacia qué lado voy a ir por encima del murciélago o por debajo del murciélago* 'which side am I going to go, above the bat or below the bat'; alternative question with two final proparoxytones).

136 https://osf.io/qmtew/

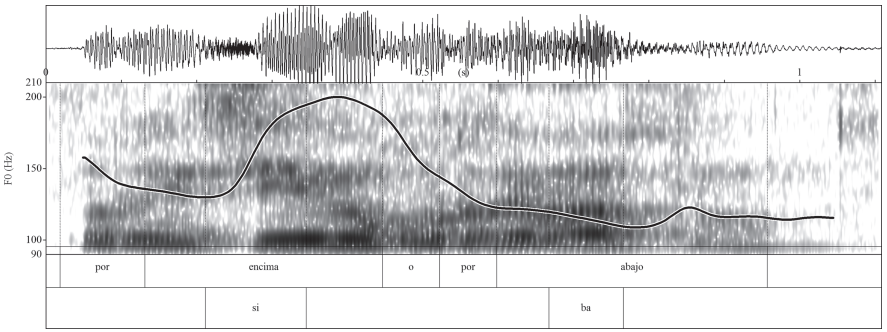


Figure 45: QF16_MT_ES_0136[137] (*por encima o por abajo* 'above or below'; alternative question with two final paroxytones).

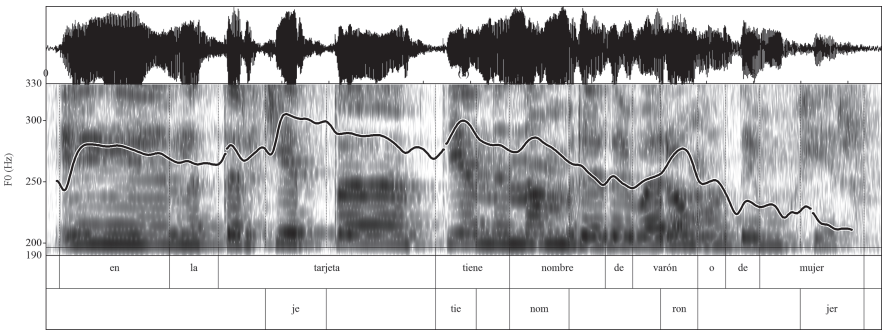


Figure 46: AZ23_Quien_ES_0020[138] (*en la tarjeta tiene nombre de varón o de mujer* 'on the card does it have the name of a man or woman'; alternative question with two final oxytones).

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/

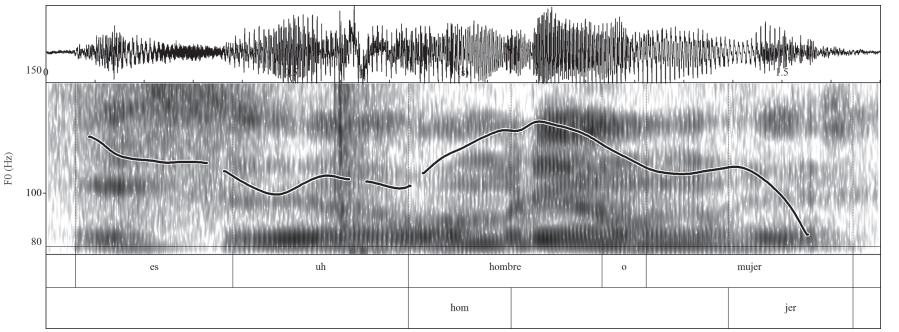


Figure 47: KA36_Quien_ES_0197[139] (*es uh hombre o mujer* 'is it uh a man or woman'; alternative question with a final paroxytone and a final oxytone).

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
      $[(LH^*)\ LH^*\ H\text{-}/\%]_{\text{alternative 1}}\ [\sout{(LH^*)\ LH}^*\ L\%]_{\text{alternative 2}}$
   b. Without deaccentuation
      $[(LH^*)\ LH^*\ H\text{-}/\%]_{\text{alternative 1}}\ [(LH^*)\ LH^*\ L\%]_{\text{alternative 2}}$

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/

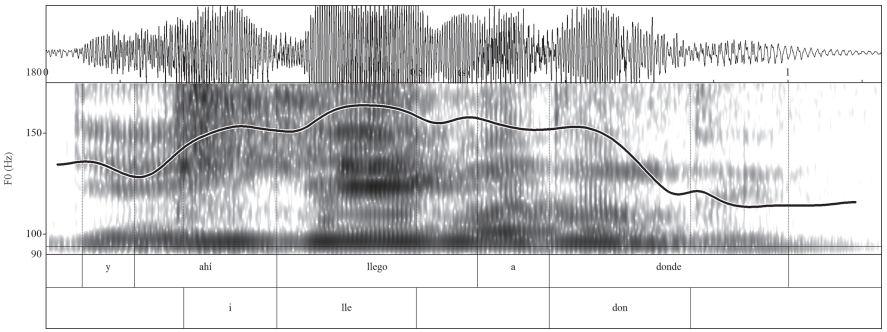


Figure 48: QF16_MT_ES_2110[140] (*y ahí llego a dónde* 'and there I get to where'; wh-question with the wh-word finally and preceding accented words).

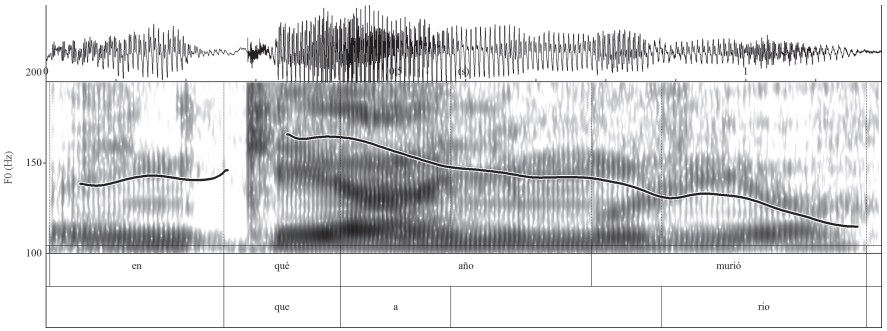


Figure 49: OV37_Quien_0155[141] (*en qué año murió* 'what year did s/he die'; wh-question with the wh-word followed by accentable words without identifiable pitch accents).

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/

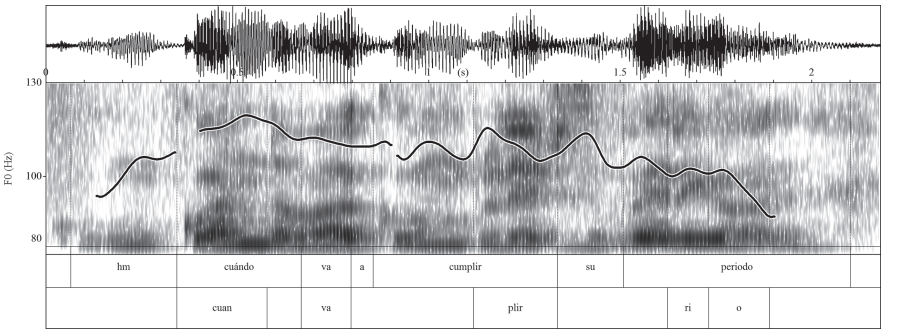


Figure 50: KA36_Quien_1075[142] (*cuándo va a cumplir su periodo* 'when is their term going to end'; wh-question with the wh-word followed by accentable words with pitch accents).

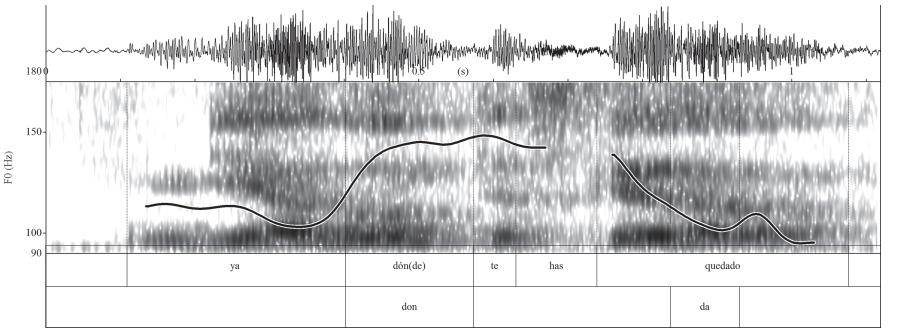


Figure 51: TP03_MT_ES_4108[143] (*ya dónde te has quedado* 'so where did you get left'; wh-question with a post-wh-word fall between an auxiliary and a verb).

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/

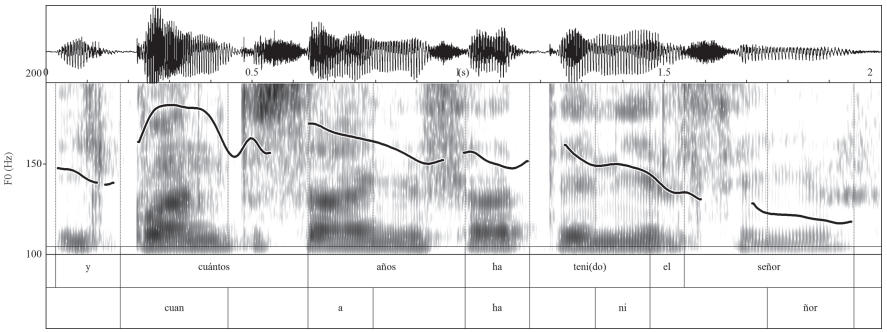


Figure 52: OV37_Quien_0446[144] (*y cuántos años ha tenido el señor* 'and how old was the gentleman'; wh-question with the wh-word followed by accentable words with and without accents).

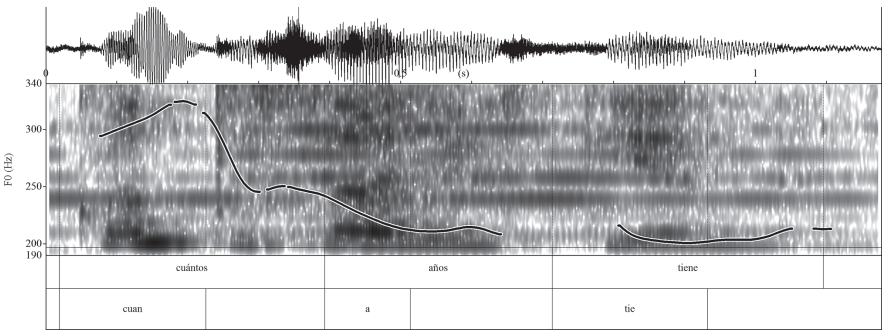


Figure 53: HA30_Quien_ES_0571[145] (*cuántos años tiene* 'how old is s/he'; wh-question with two deaccented words after the wh-word).

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/

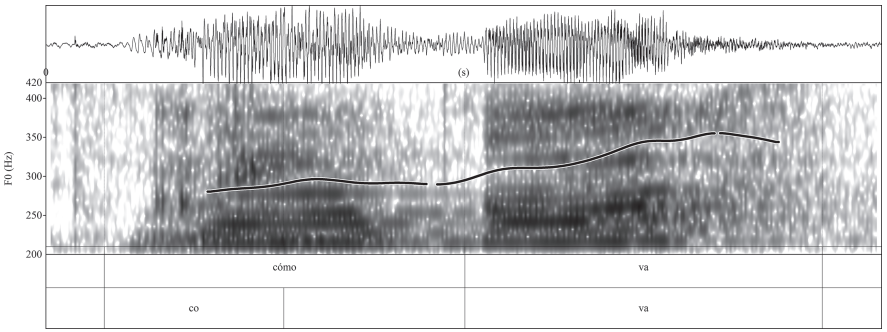


Figure 54: AZ23_MT_ES_0153 (wh-question with a final rise, a request for repetition).

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.

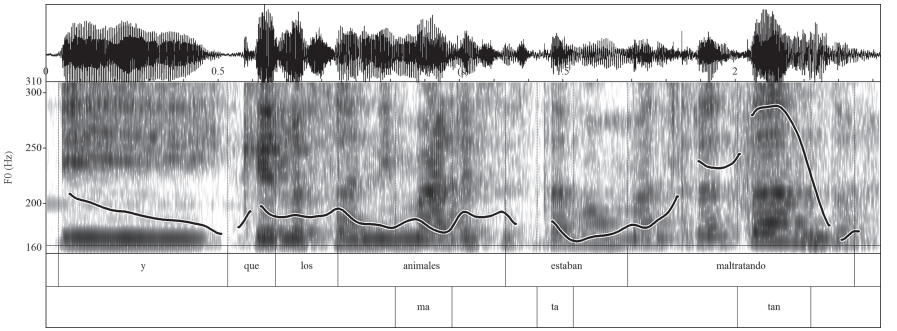


Figure 55: MS27_Cuent_ES_1520[147] (*y que los animales estaban maltratando* 'and which the animals were mistreating'; declarative with strong phrase accentuation) Cf. context (53).

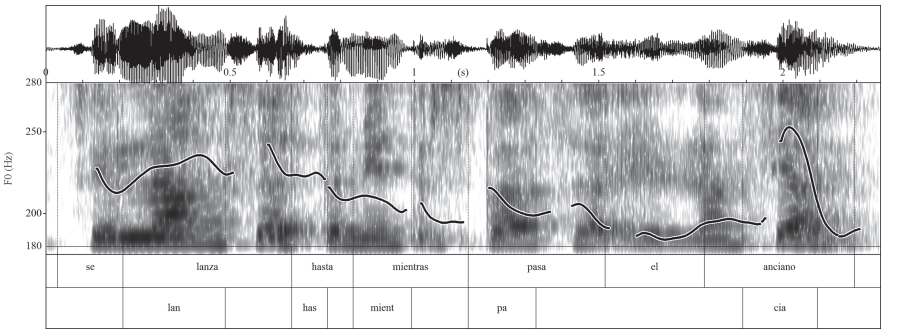


Figure 56: ZR29_Cuent_ES_1572[148] (*se lanza hasta mientras pasa el anciano* 'it launches itself while the old man passes by'; declarative with strong phrase accentuation). Cf. (52) for context.

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/

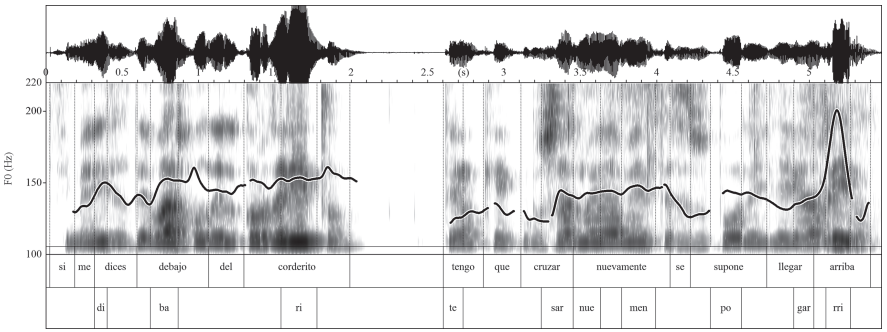


Figure 57: QF16_MT_ES_1011[149] (*si me dices debajo del corderito tengo que cruzar nuevamente se supone llegar arriba* 'if you tell me below the little lamb I have to cross again supposedly get to above'*;* declarative with weak phrase accentuation).

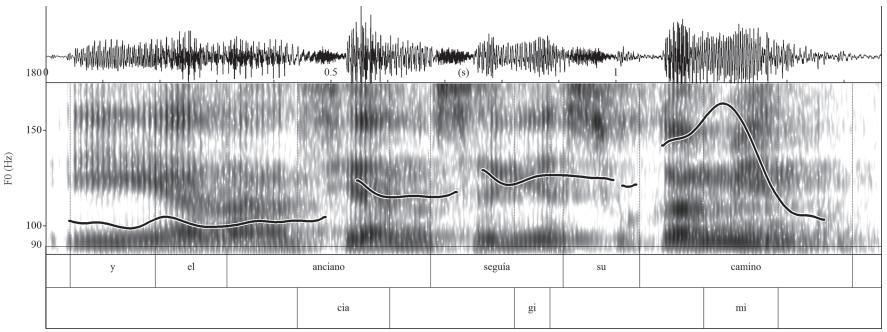


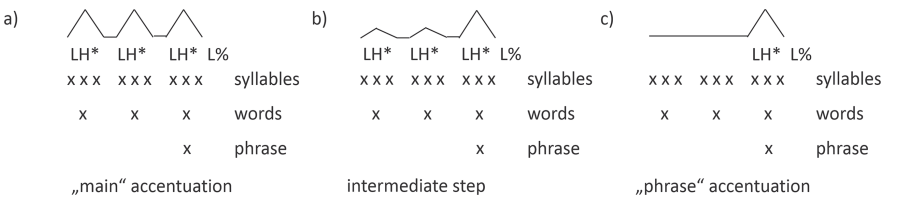
Figure 58: TP03_Cuent_ES_0561[150] (*y el anciano seguía su camino* 'and the old man continued on his way'; declarative with weak "phrase accentuation"). Cf. context (51).

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

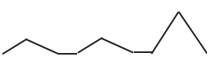


The "main" accentuation (49)a) and full "phrase" accentuation (49)c) versions of the same hypothetical utterance here do not differ in their metrical representation, in which the final word in the phrase is assigned the highest prominence in the phrase. This predicts that both versions should cue ambiguously 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

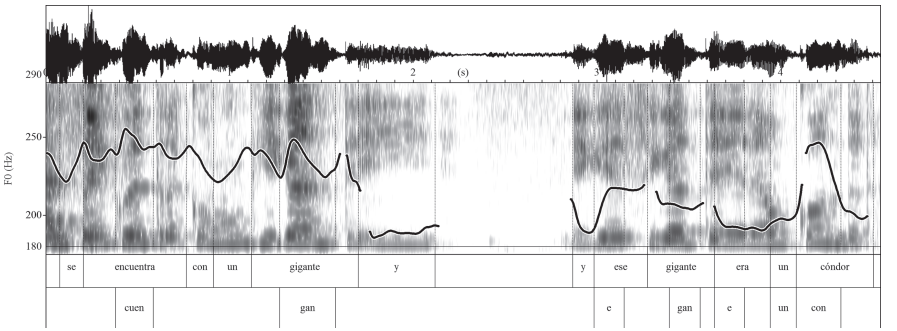


Figure 59: HA30_Cuent_ES_2285[152] (two declaratives, the first with with main, the second with phrase accentuation).

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.

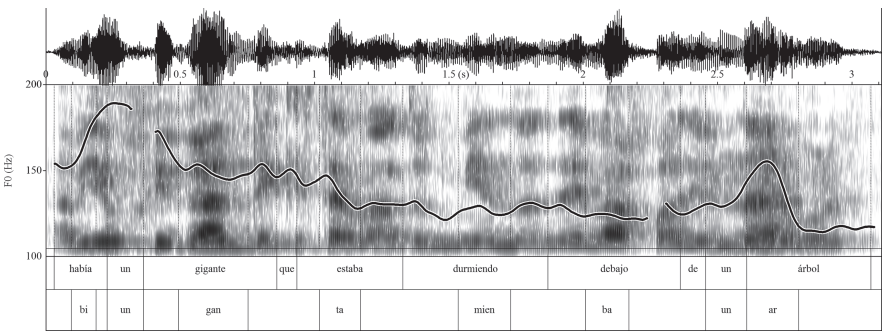


Figure 60: CF28_Cuent_ES_0012[162] (*había un gigante que estaba durmiendo debajo de un árbol* 'there was a giant who was sleeping under a tree'; declarative with phrase accentuation).

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/

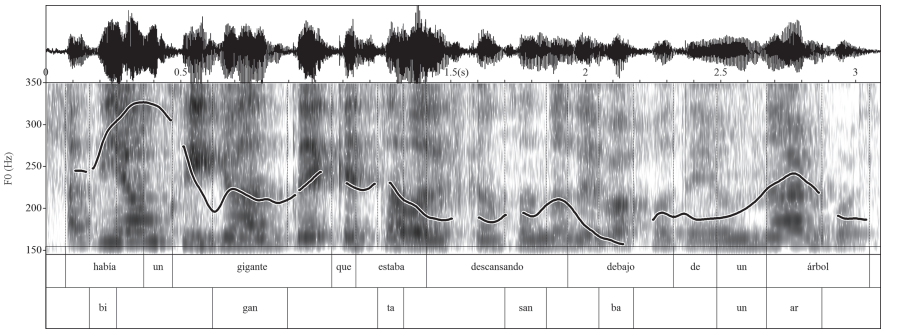


Figure 61: MS27_Cuent_1207[163] (*había un gigante que estaba descansando debajo de un árbol* 'there was a giant who was resting under a tree'; declarative with phrase accentuation).

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.

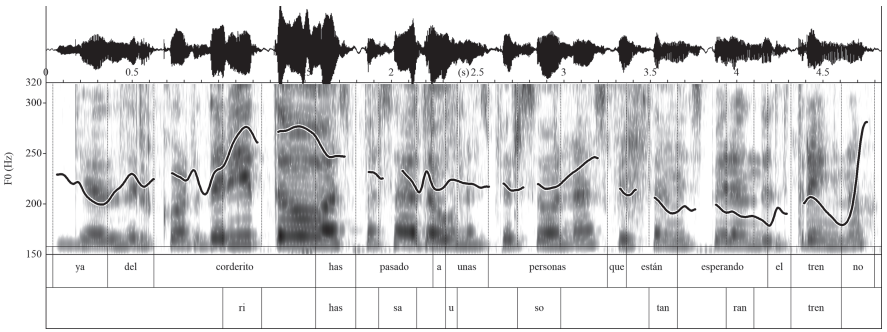


Figure 62: SO39_MT_ES_3351[164] (*ya del corderito has pasado a unas personas que están esperando el tren, no* 'ok from the little lamb you passed some people that are waiting for the train, right'; confirmation-seeking question with "phrase accentuation").

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 pers*o*nas] que es*tán* esper*an*do el **tren** no]
LH*H- H- LH* LH%

(55) ha**bí**a] *un* gi*gan*te] que es*ta*ba descans*an*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).

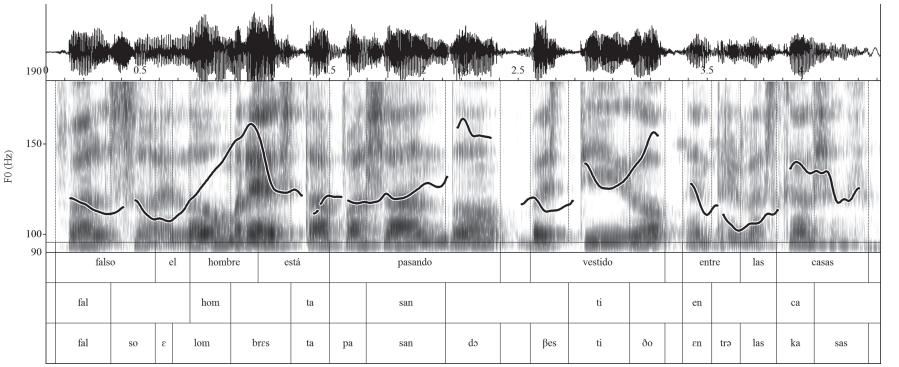


Figure 63: XJ45_ELQUD_ES_16[167] (*falso el hombre está pasando vestido entre las casas* 'wrong the man is passing between the houses with clothes on'; declarative with phrase accentuation).

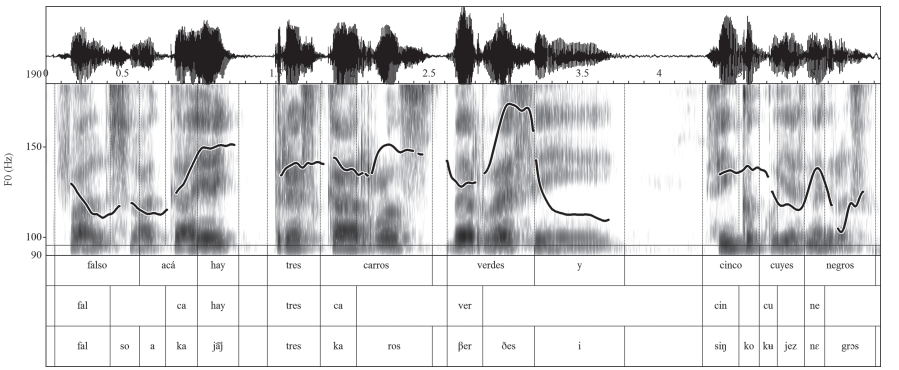


Figure 64: XJ45_ELQUD_ES_19B[168] (*falso acá hay tres carros verdes y cinco cuyes negros* 'wrong here there are three green cars and five black guinea pigs'; declarative with "phrase accentuation").

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.

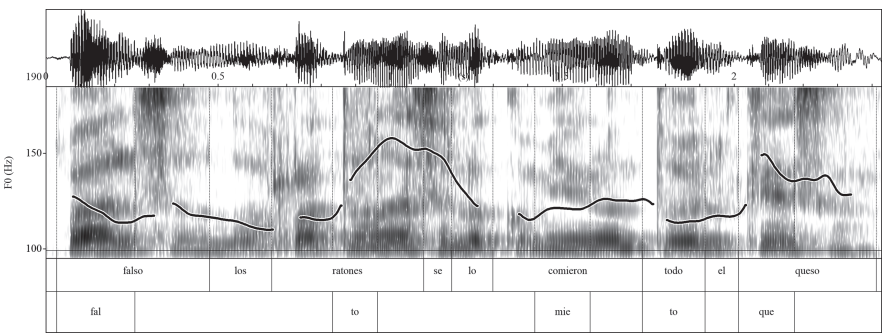


Figure 65: XJ45_ELQUD_ES_17[171] (*falso los ratones se lo comieron todo el queso* 'wrong the mice ate all the cheese'; declarative with "phrase accentuation").

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.

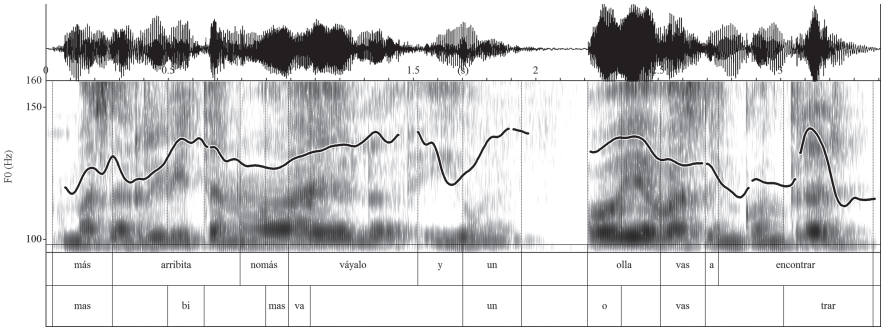


Figure 66: OA32_MT_ES_1443[173] (declarative with main accentuation). Cf. context (56).

(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*.

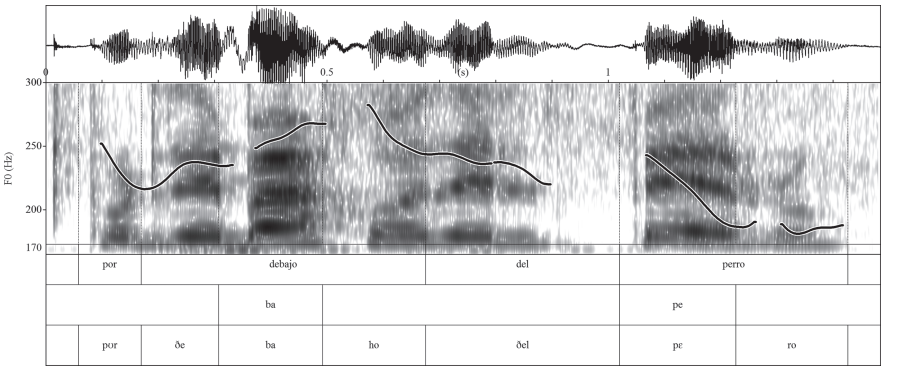


Figure 67: SO39_MT_ES_0743[176] (declarative with main accentuation). Cf. (57).

(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

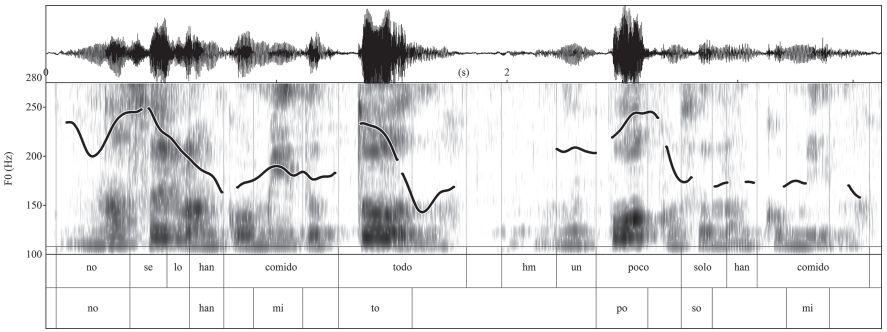


Figure 68: ZE55_ELQUD_ES_55[180] (*no se lo han comido todo hm un poco solo han comido* 'they haven't eaten it all hm only a bit they've eaten'; declarative with deaccentuation after *poco*).

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/

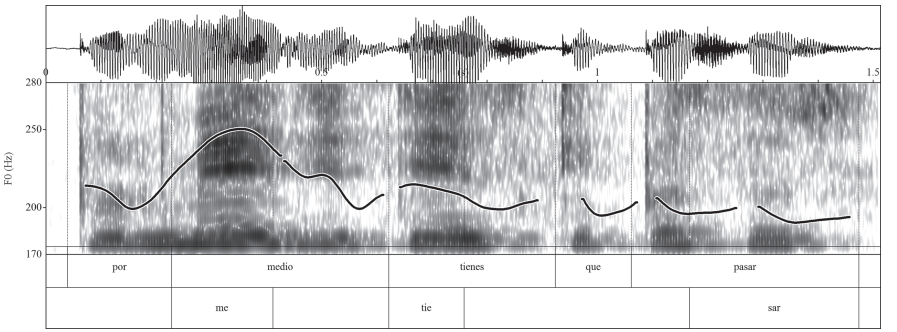


Figure 69: SG15_MT_ES_1815[181] (*por medio tienes que pasar* 'through the middle you have to go'; declarative with deaccentuation after *medio*). Cf. (58) for context.

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

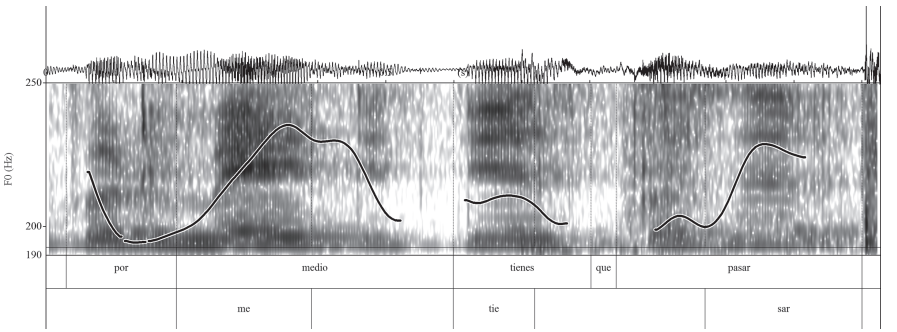


Figure 70: SG15_MT_ES_2272[185] (declarative with no deaccentuation). Cf. context (59).

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.

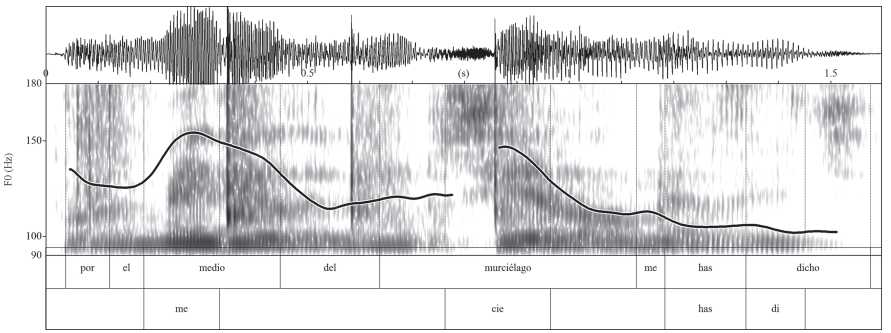


Figure 71: QF16_MT_ES_2251[186] (declarative with deaccentuation after *murciélago*). Cf. context (59).

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).

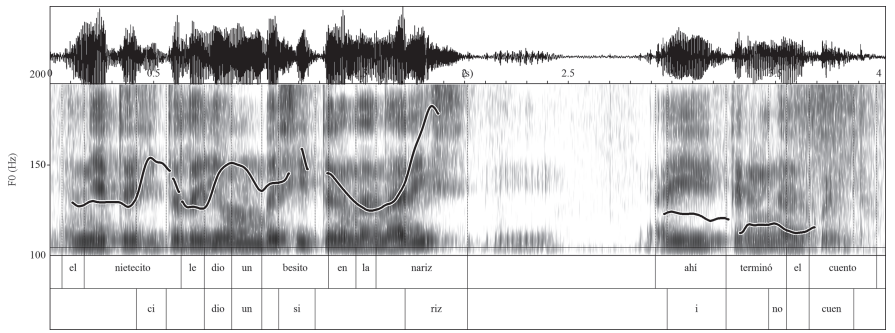


Figure 72: KP04_Cuent_ES_2317[191] (*el nietecito le dio un besito en la naríz / ahí terminó el cuento* 'the little grandson gave him a peck on the nose / there the story ended'; declarative with deaccentuation after *naríz*).

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**	

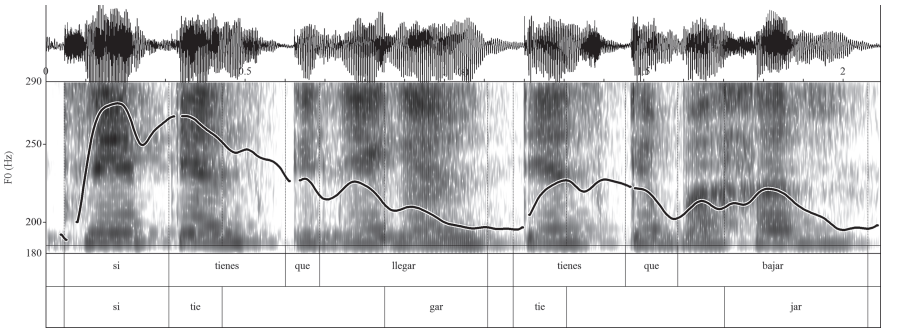


Figure 73: SG15_MT_ES_2396[193] (declarative with accentuation reset after *llegar*).

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.

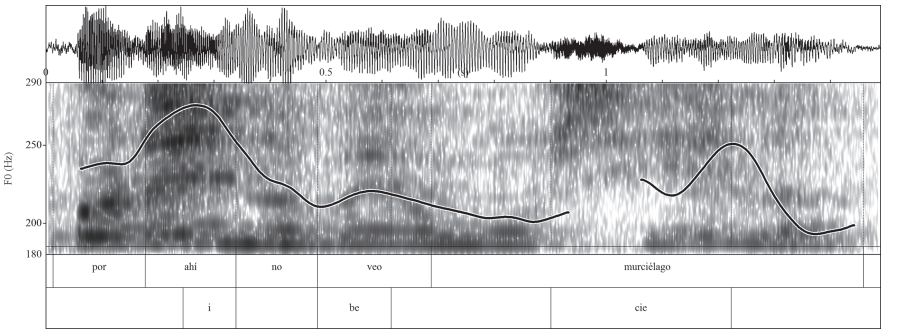


Figure 74: HA30_MT_ES_1372[194] (declarative with delayed peak on *murciélago*).

(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.

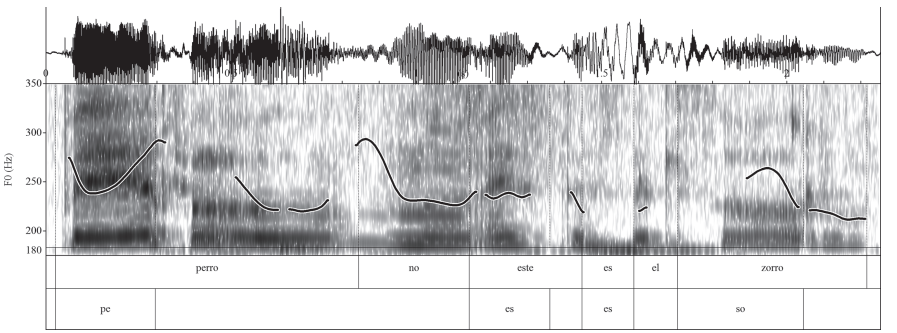


Figure 75: MD40_MT_ES_1542[197] (clarification question with delayed peak on *perro*, and declarative with normally aligned peak).

(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.

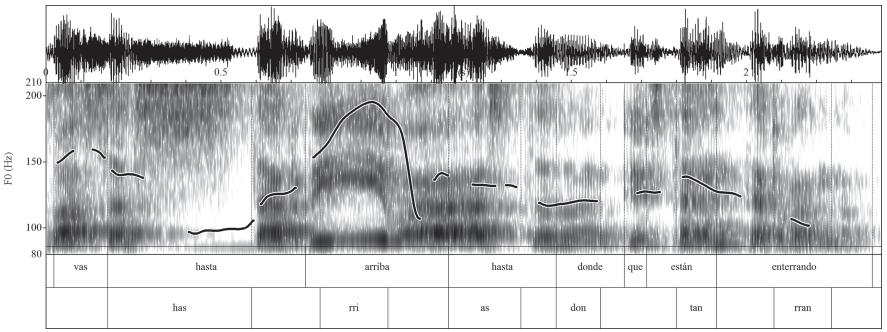


Figure 76: TP03_MT_ES_1312[199] (declarative with deaccentuation after and delayed peak on *arriba*).

(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.

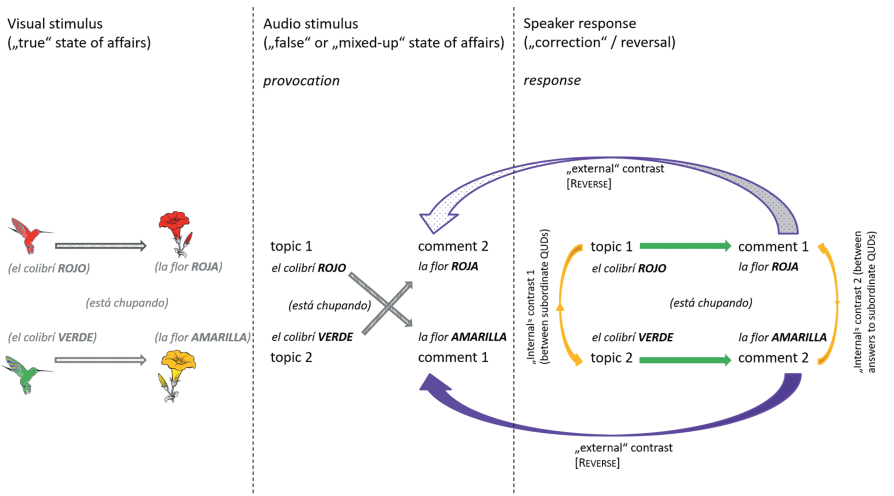
Item number	Text of audio stimulus	Description of visual stimulus / mismatch
14	*El HOMBRE está COMIENDO y la MUJER DURMIENDO* ‘The man is eating and the woman is sleeping’	The man is sleeping while the woman is eating
26	*El PERRO está DEBAJO de la roca y el GATO está ENCIMA de la roca* ‘The dog is 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.



superordinate QUD: Who$_i$ is drinking what$_j$?
context-given sets of referents: {*colibrí rojo, colibrí verde*}$_i$ {*flor roja, flor amarilla*}$_j$

subordinate QUD 1: What about the red hummingbird, what$_j$ is it drinking from?
subordinate QUD 2: What about the green hummingbird, what$_j$ is it drinking from?

at-issue: {contrastive elements}$_i$ in comment 1 & comment 2

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).

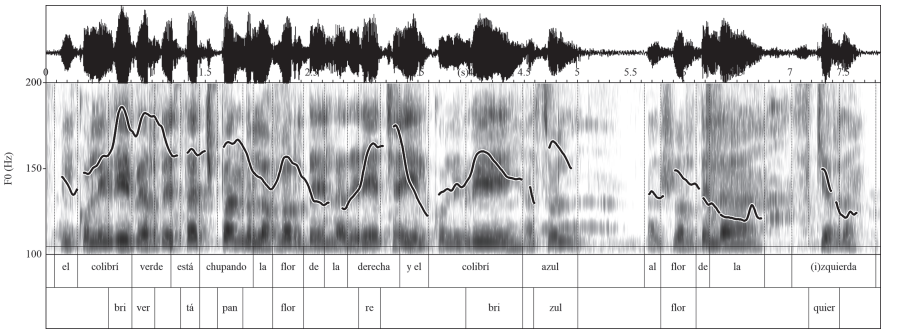


Figure 78: QZ13_ELQUD_ES_52[206] (*el colibrí verde está chupando la flor de la derecha y el colibrí azul al*[207] *flor de la izquierda* ‘the green hummingbird is drinking from the flower to the right and the blue hummingbird from the flower to the left’).

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).

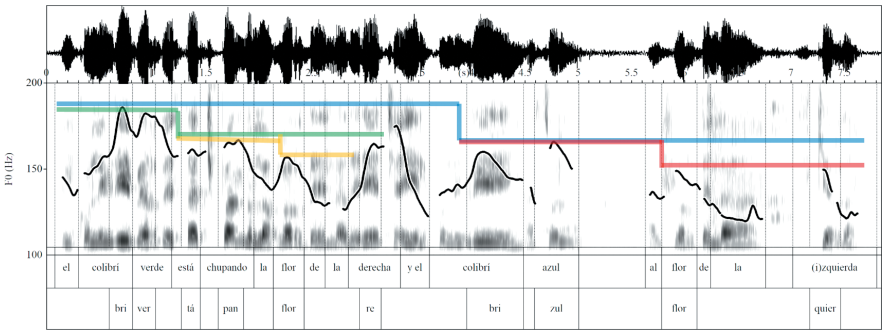


Figure 79: QZ13_ELQUD_ES_52 with reference lines added.

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

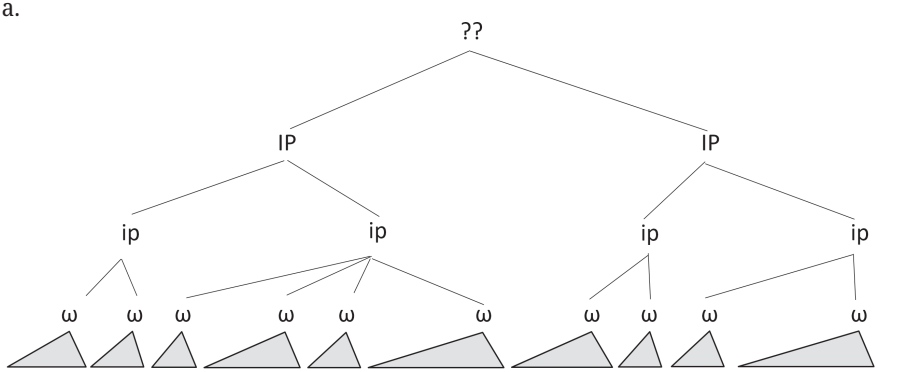
it has been shown to do in English (Liberman & Pierrehumbert 1984).[208] Coming back to the example at hand, this consideration opens an alternative analysis for the downstep described by the yellow reference line in the figure: we might also suppose that the downstep between *chupando* and *flor* is 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.

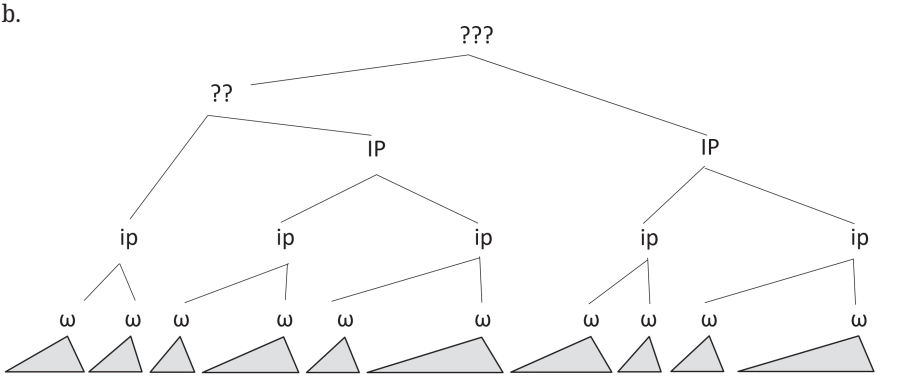
208 The reading stimuli for the Prieto et al. (1996) study expanded the noun phrase *rayo de luna* ‘ray of moonlight’ (two accents) to the maximal (and ambiguous in terms of parsing) *rayo de la luna de mi mayo de la gala de la Lola* ‘Lola’s gala’s moonlight ray of my May’ (five accents), while in the 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

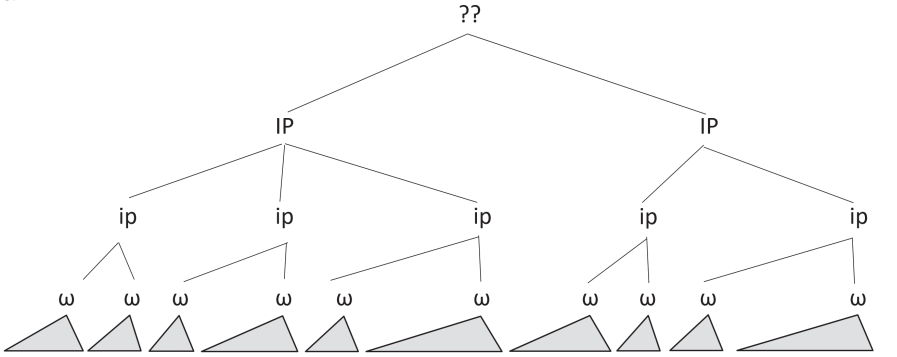
a.



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.

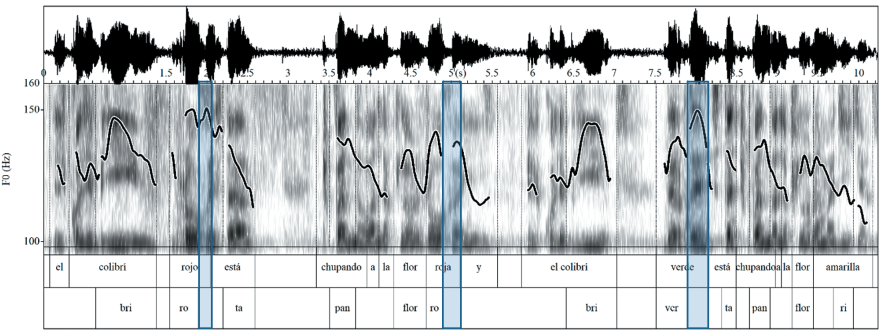


Figure 80: OZ14_ELQUD_ES_43[209] (*el colibrí rojo está chupando a la flor roja y el colibrí verde está chupando a la flor amarilla* 'the red hummingbird is drinking from the red flower and the green hummingbird is drinking from the yellow flower'). Blue marking highlights high posttonic pitch indicating 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.

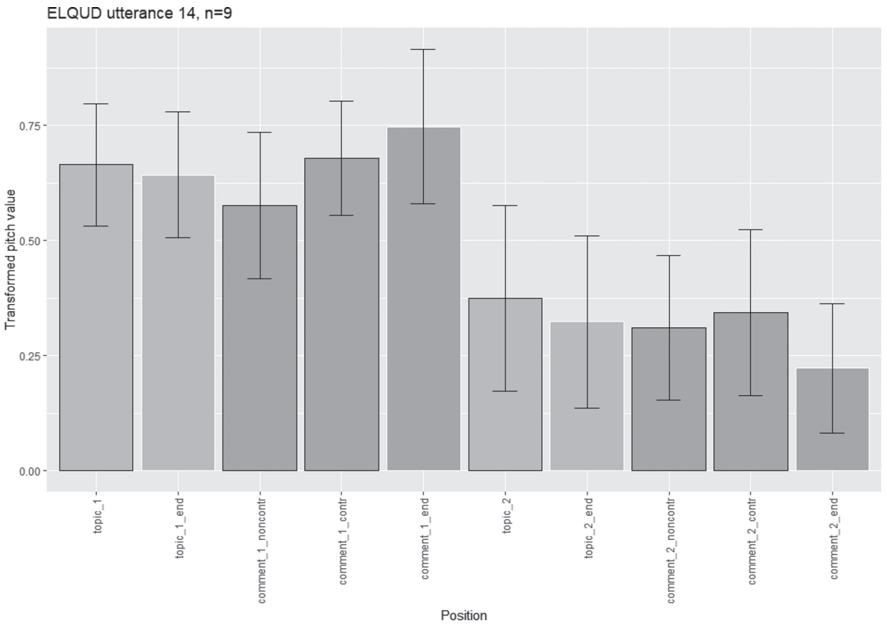


Figure 81: Barplot of pooled transformed mean pitch values from all utterances of experimental item 14 from Spanish *Elqud*.

Example: [*el* ***hom****bre*]$_{topic1}$ [*es****tá***]$_{comment1\_noncontr}$ [*dur****mien****do*]$_{comment1\_contr}$ [*la* ***ni****ña*]$_{topic2}$ [*es****tá***]$_{comment2\_noncontr}$ [*co****mien****do*]$_{comment2\_contr}$ (cf. Figure 82).

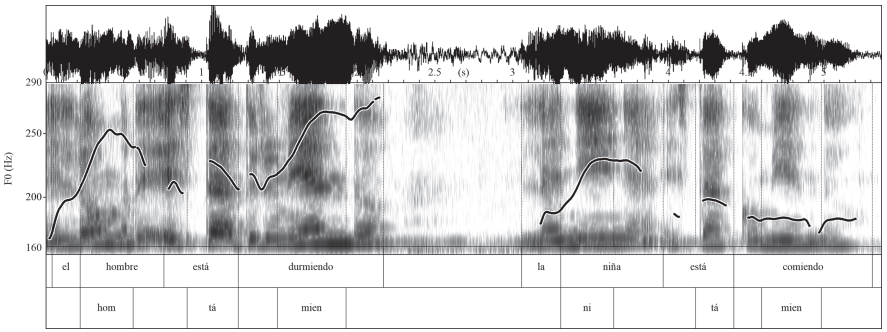


Figure 82: SG15_ELQUD_ES_14[211] (*el hombre está durmiendo la niña está comiendo* 'the man is sleeping the girl is eating').

211 https://osf.io/x9d4r/

5.2.3.2 ELQUD-item 26

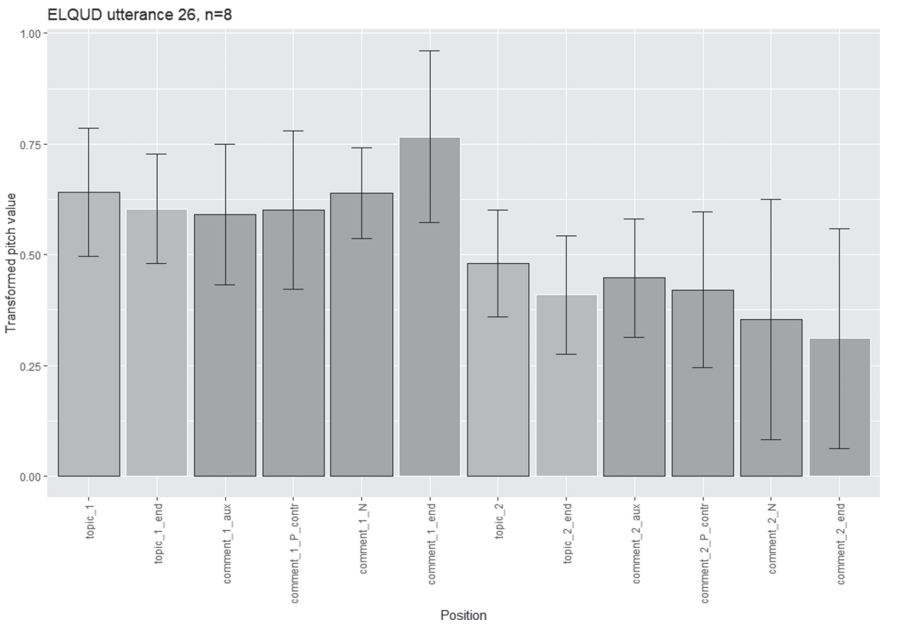


Figure 83: Barplot of pooled transformed mean pitch values from all utterances of experimental item 26 from Spanish *Elqud*.

Example: [*el* ***per****ro*]$_{\text{topic1}}$[*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.

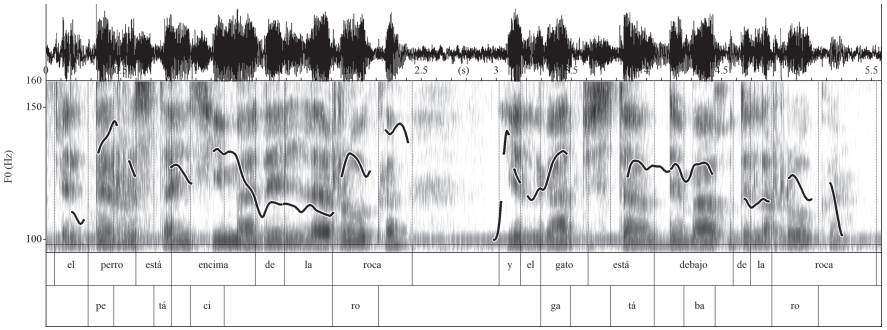


Figure 84: OZ14_ELQUD_ES_26[213] (*el perro está encima de la roca y el gato está debajo de la roca* 'the dog is above the rock and the cat is below the rock').

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.

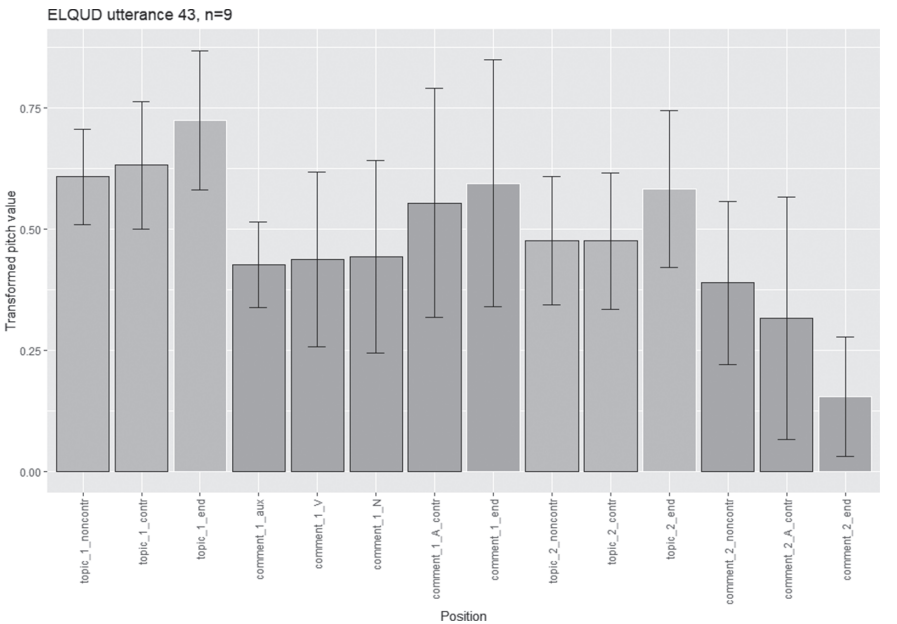


Figure 85: Barplot of pooled transformed mean pitch values from all utterances of experimental item 43 from Spanish *Elqud*.

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***]$_{\text{comment2\_noncontr}}$[*amari**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.

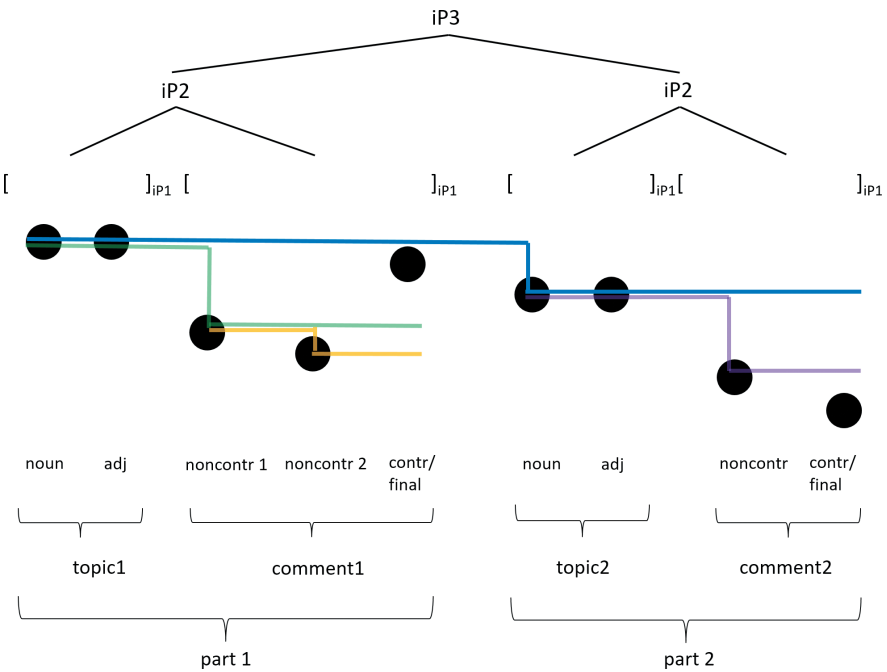


Figure 86: Schematized scaling model with register lines and phrasal prosodic structure for complex Spanish double topic utterances from *Elqud.*

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.

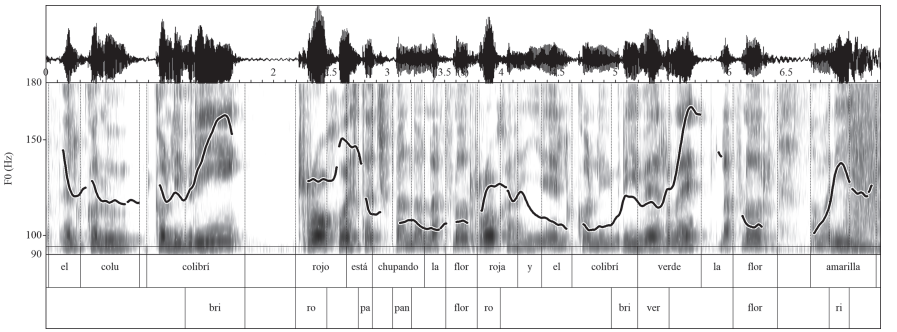


Figure 87: XJ45_ELQUD_ES_43[214] (*el colu- colibrí rojo está chupando la flor roja y el colibrí verde la flor amarilla* ‘the red hummingbird is drinking the red flower and the green hummingbird the yellow flower’).

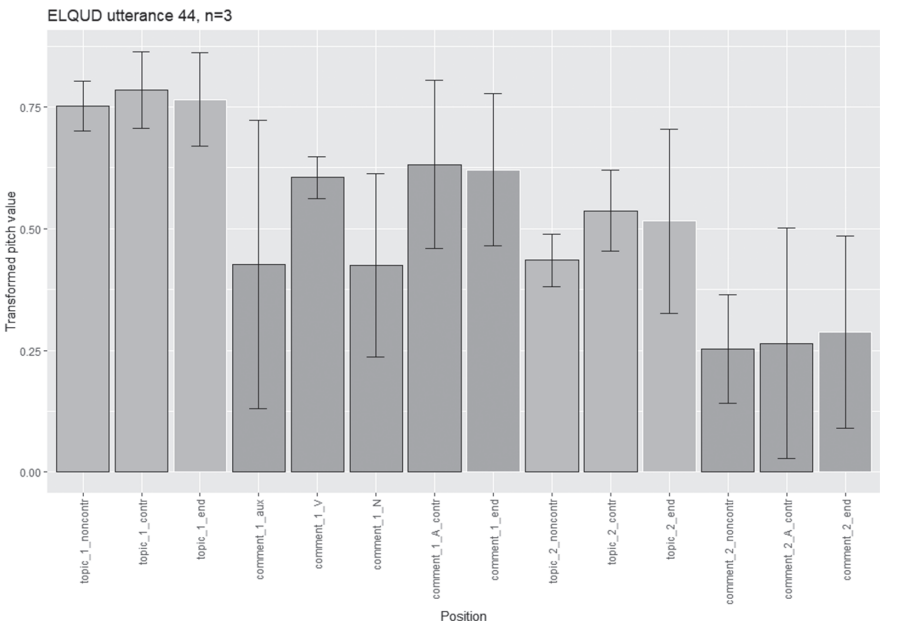


Figure 88: Barplot of pooled transformed mean pitch values from all utterances of experimental item 44 from Spanish *Elqud*.

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*]$_{\text{comment1\_N}}$[***neg****ra*]$_{\text{comment1\_A\_contr}}$[***pe****rro*]$_{\text{topic2\_noncontr}}$[***blan****co*]$_{\text{topic2\_contr}}$ [*con la pe****lo****ta*]$_{\text{comment2\_noncontr}}$[***azul***]$_{\text{comment2\_A\_contr}}$ (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.

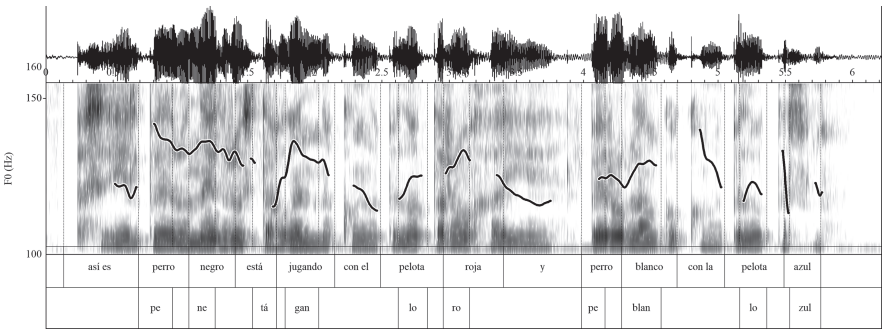


Figure 89: ZZ24_ELQUD_ES_44[215] (*así es perro negro está jugando con el pelota roja y perro blanco con la pelota azul* 'that's it black dog is playing with the red ball and white dog with the blue ball').

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/

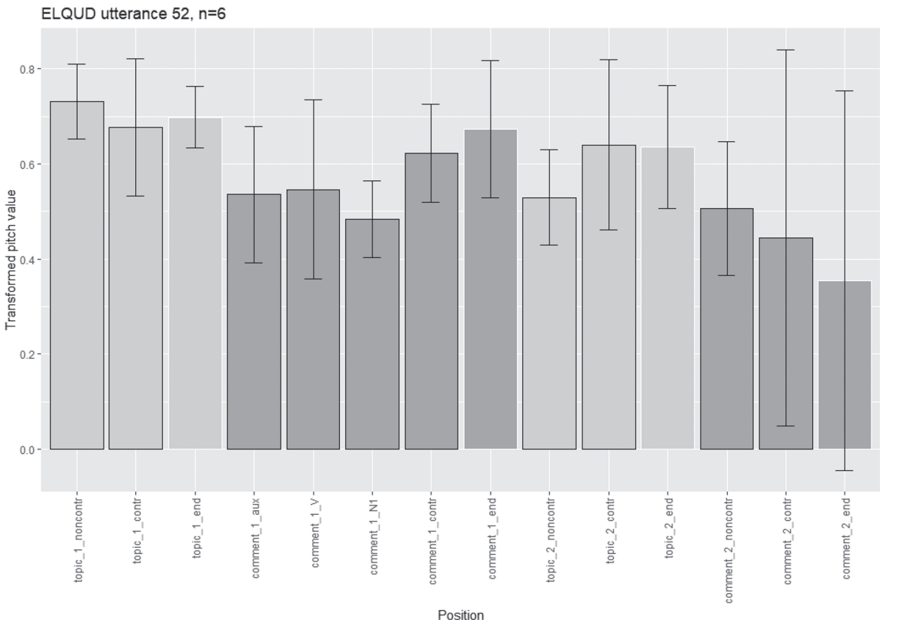


Figure 90: Barplot of pooled transformed mean pitch values from all utterances of experimental item 52 from Spanish *Elqud*.

Example: [*coli**brí***]$_{topic1\_noncontr}$[***ver**de*]$_{topic1\_contr}$[*es**tá***]$_{comment1\_aux}$[*chu**pan**do*]$_{comment1\_V}$[*la **flor***]$_{comment1\_N1}$[*que está a a la de**re**cha*]$_{comment1\_contr}$[*coli**brí***]$_{topic2\_noncontr}$[***azul***]$_{topic2\_contr}$ [*está chupando la **flor***]$_{comment2\_noncontr}$[*que está a la iz**quier**da*]$_{comment2\_A\_contr}$ (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

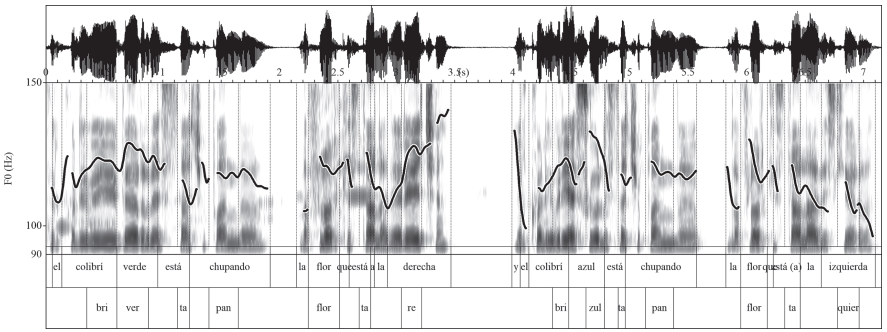


Figure 91: QP44_ELQUD_ES_52[216] (*el colibrí verde está chupando la flor que está a la derecha y el colibrí azul está chupando la flor que está a la izquierda* 'the green hummingbird is drinking from the flower that's on the right and the blue hummingbird is drinking from the flower that's on the left').

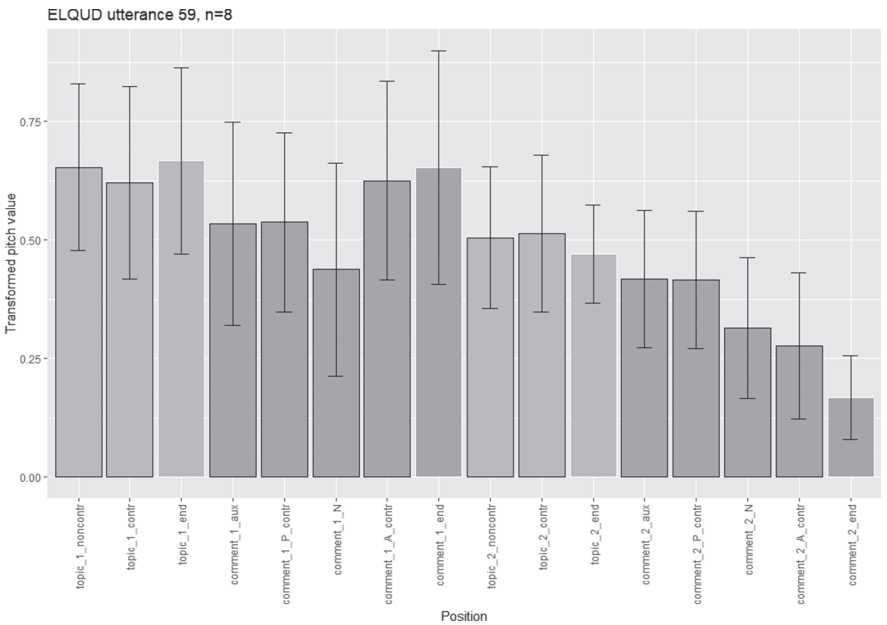


Figure 92: Barplot of pooled transformed mean pitch values from all utterances of experimental item 59 from Spanish *Elqud*.

Example: [*el zo**rri**llo*]$_{topic1\_noncontr}$[*pe**que**ño*]$_{topic1\_contr}$[*es**tá***]$_{comment1\_aux}$[*al **fren**te*]$_{comment1\_P\_contr}$ [*de la **ca**sa*]$_{comment1\_N}$[***gran**de*]$_{comment1\_A\_contr}$ [*el zo**rri**llo*]$_{topic2\_noncontr}$[***gran**de*]$_{topic2\_contr}$[*es**tá***]$_{comment2\_aux}$ [*de**trás***]$_{comment2\_P\_contr}$[*de la **ca**sa*]$_{comment2\_N}$[*pe**que**ña*]$_{comment2\_A\_contr}$ (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.

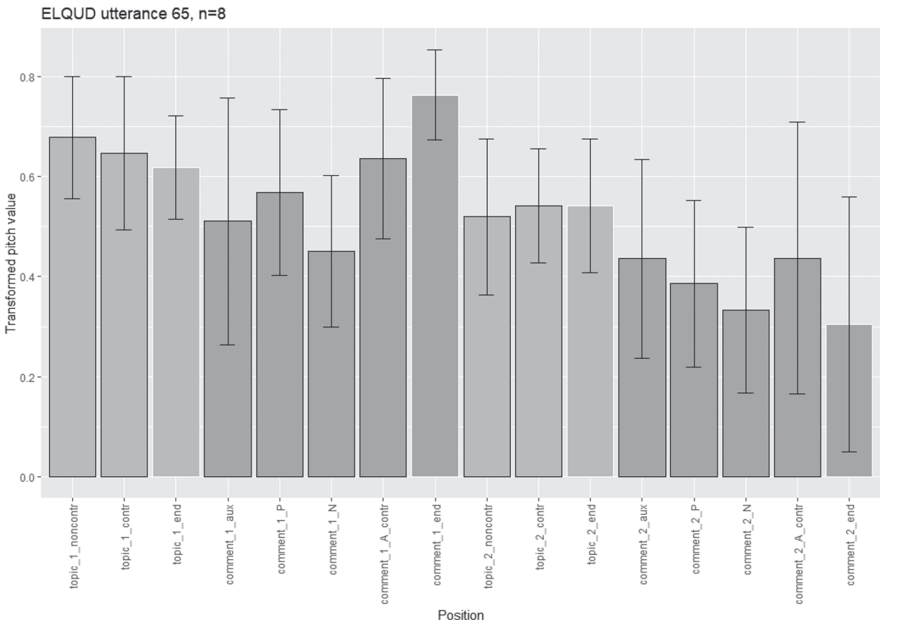


Figure 93: Barplot of pooled transformed mean pitch values from all utterances of experimental item 65 from Spanish *Elqud*.

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.

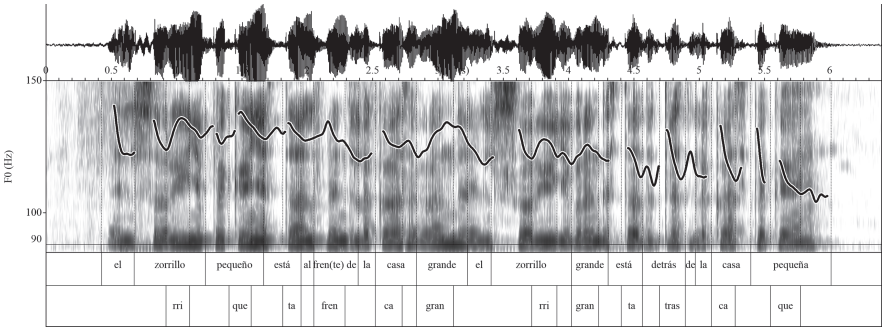


Figure 94: LJ22_ELQUD_ES_59[217] (*el zorrillo pequeño está al frente de la casa grande el zorrillo grande está detrás de la casa pequeña* 'the small skunk is in front of the large house the large skunk is behind the small house').

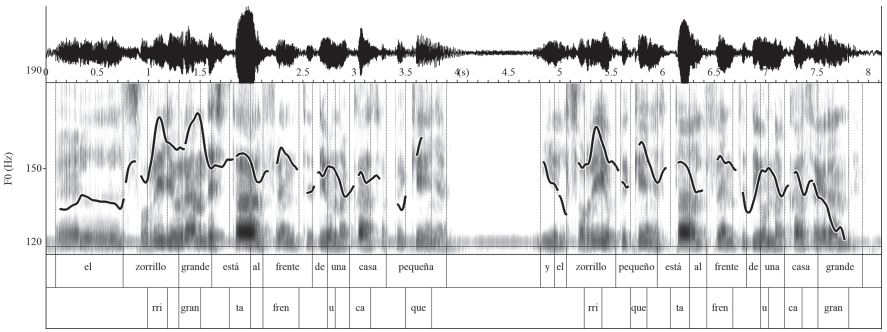


Figure 95: QZ13_ELQUD_ES_65[218] (*el zorrillo grande está al frente de una casa pequeña y el zorrillo pequeño está al frente de una casa grande* 'the large skunk is in front of a small house and the small skunk is in front of a large house').

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

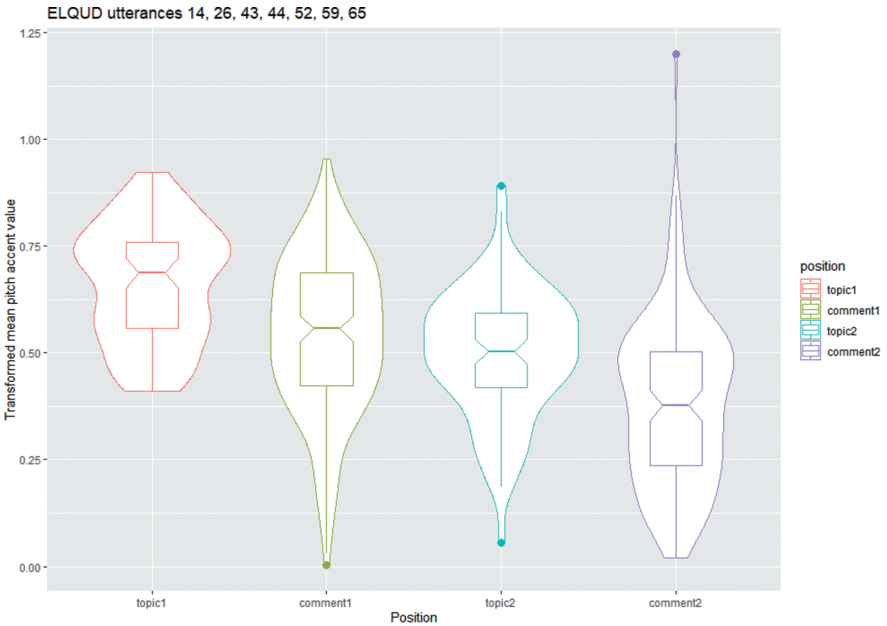


Figure 96: Combined violin- and boxplots for the transformed mean pitch value measured on accentable syllables in the double-topic utterances in Spanish *Elqud*, sorted according to utterance subpart. Violin shape indicates density. Horizontal bar in boxplot indicates median, notches indicate confidence 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.

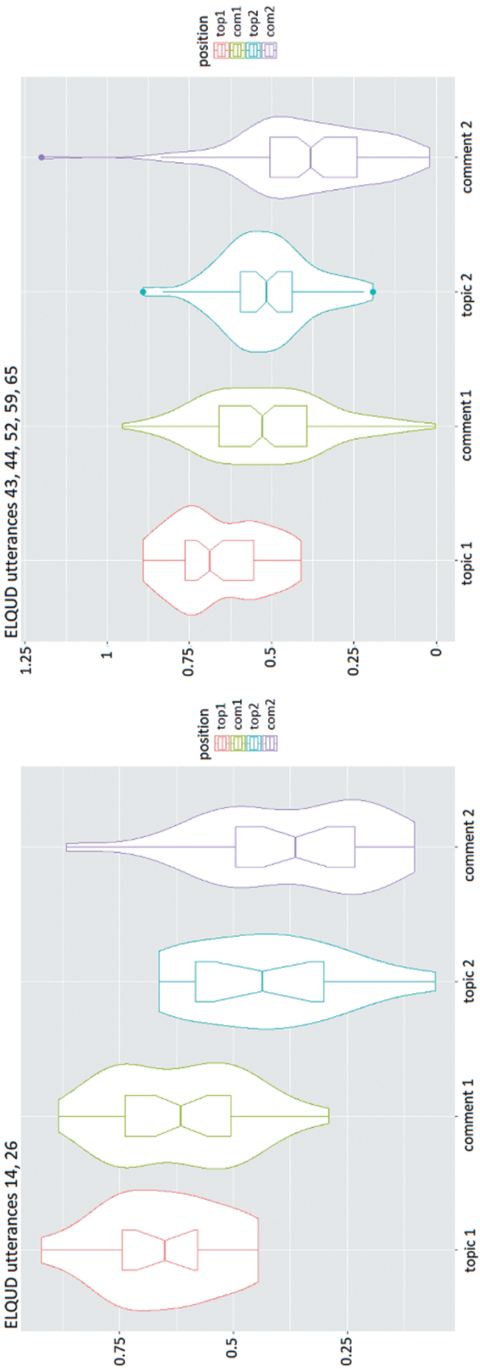


Figure 97: Same data as in Figure 96, but separated into the data from the shorter experimental items 14, 26 on the left, and those from the longer items 43, 44, 52, 59, 65 on the right. Number of observations per utterance subpart: 17, 41, 17, 35 (left); 67, 140, 67, 111 (right).

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.

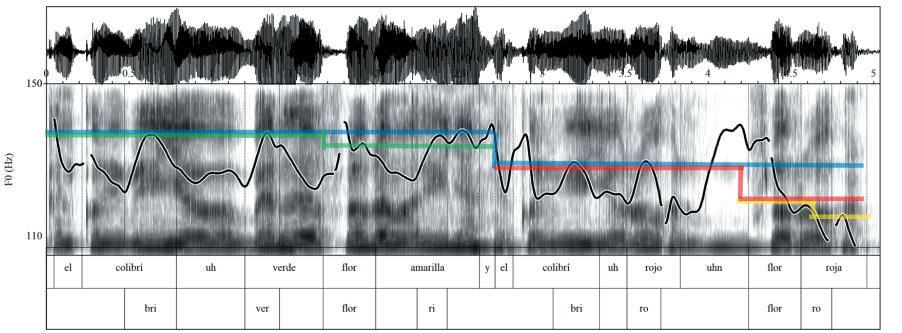


Figure 98: LJ22_ELQUD_ES_43[223] (*el colibrí uh verde flor amarilla y el colibrí uh rojo uhm flor roja* ‘the green hummingbird yellow flower and the red hummingbird red flower’). Cf. (65).

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

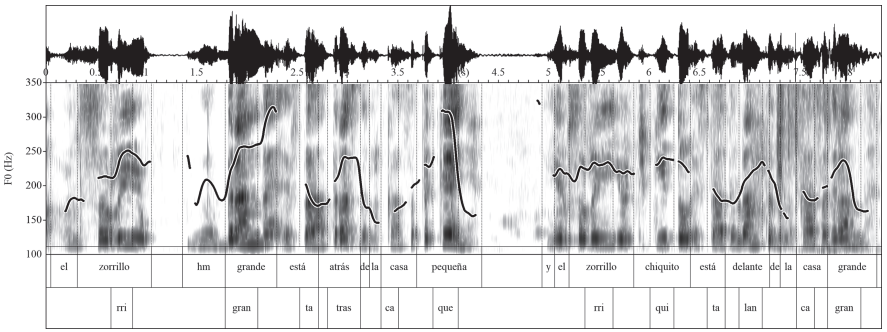


Figure 99: ZE55_ELQUD_ES_59[225] (*el zorrillo hm grande está atrás de la casa pequeña y el zorrillo chiquito está delante de la casa grande* 'the big skunk is behind the small house and the little skunk is in front of the big house').

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/

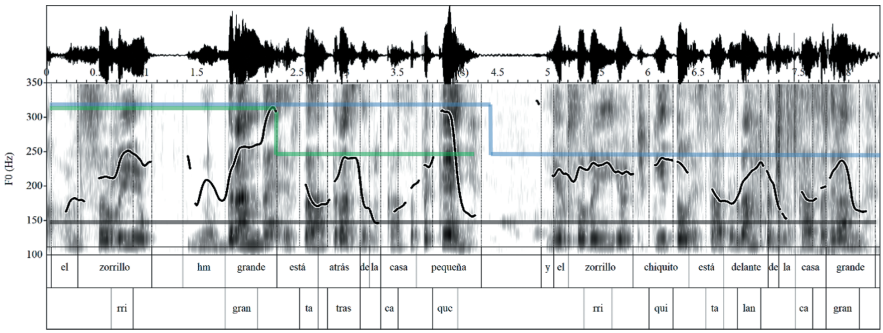


Figure 100: ZE55_ELQUD_ES_59 with added reference lines for levels of scaling.

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.

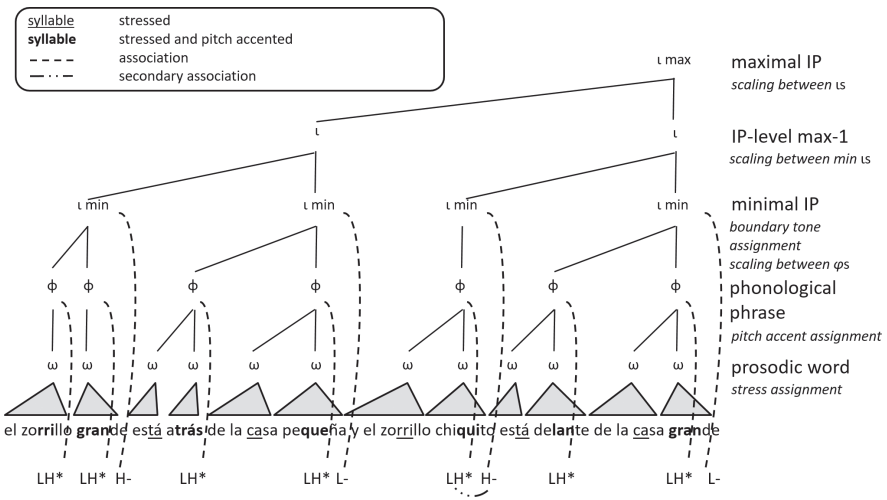


Figure 101: Proposed recursive prosodic structure for utterance ZE55_ELQUD_ES_59.

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.

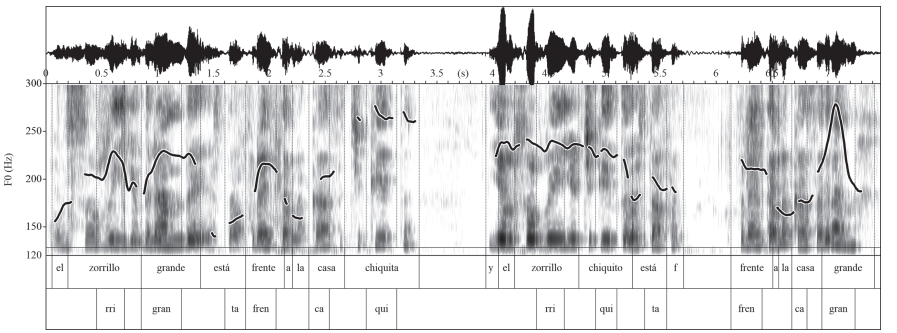


Figure 102: ZE55_ELQUD_ES_65[226] (*el zorrillo grande está frente a la casa chiquita y el zorrillo chiquito está f- frente a la casa grande* ‘the large skunk is in front of the small house and the small skunk is in front of the large house’).

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

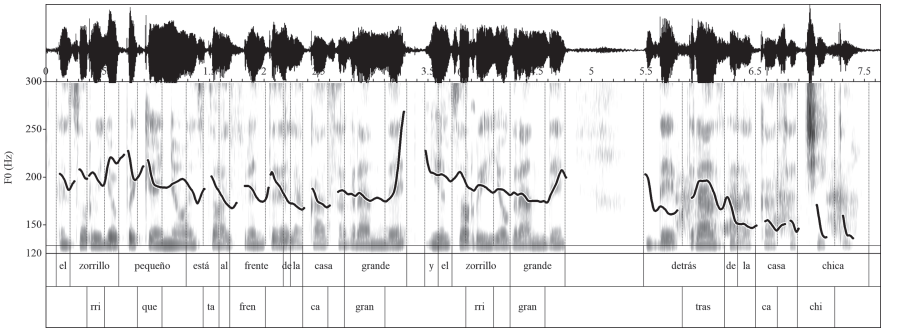


Figure 103: NQ01_ELQUD_ES_59[228] (*el zorrillo pequeño está al frente de la casa grande y el zorrillo grande detrás de la casa chica* ‘the small skunk is in front of the large house and the large skunk behind the small house’).

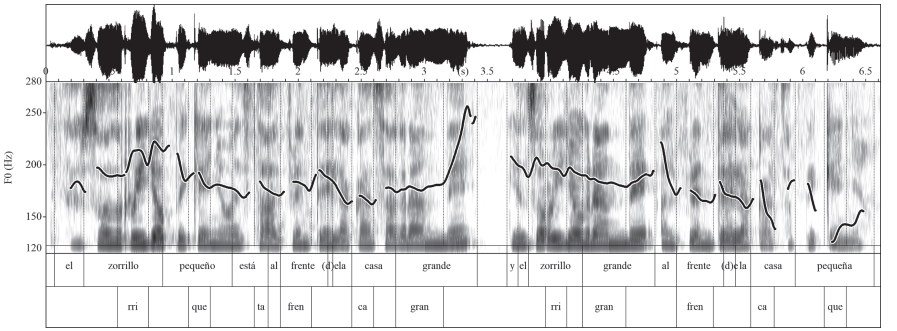


Figure 104: NQ01_ELQUD_ES_65[229] (*el zorrillo pequeño está al frente (d)e la casa grande y el zorrillo grande al frente (d)e la casa pequeña* ‘the small skunk is in from of the large house and the large skunk in front of the small house’).

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.

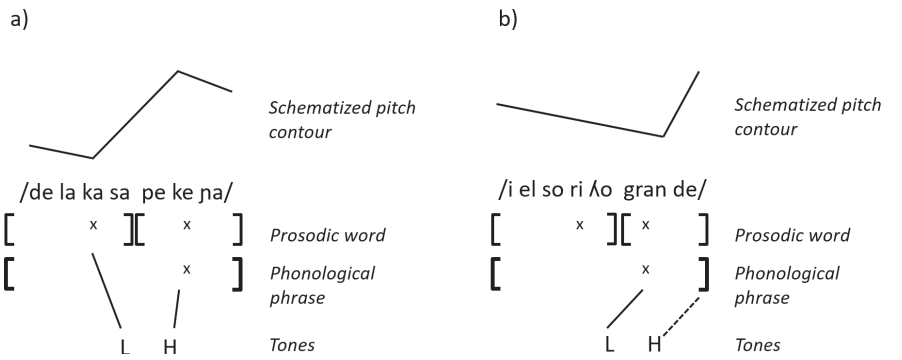


Figure 105: Proposed association model for „phrase accentuated“ realizations by ZE55 (***a***, e.g. Figure 99) and NQ01 (***b***, e.g. Figure 104), given exemplarily using the phrases *de la casa pequeña* for ZE55 and *y el zorrillo grande* for NQ01. Solid lines between tones and positions in the structure 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 phra-

sing 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.

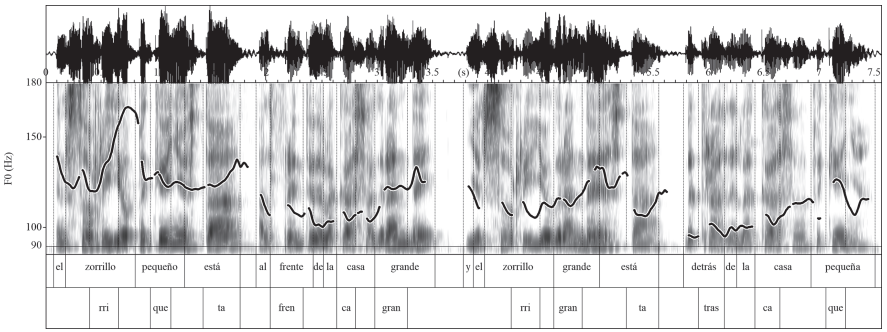


Figure 106: XJ45_ELQUD_ES_59[231] (*el zorrillo pequeño está al frente de la casa grande y el zorrillo grande está detrás de la casa pequeña* 'the small skunk is in front of the big house and the large skunk is behind the small house').

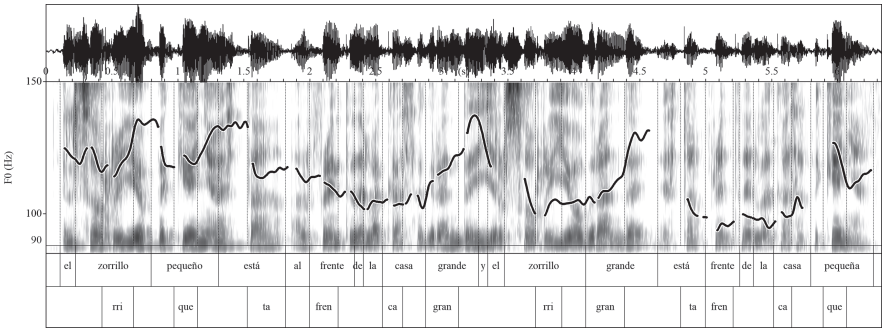


Figure 107: XJ45_ELQUD_ES_65[232] (*el zorrillo pequeño está al frente de la casa grande y el zorrillo grande está frente de la casa pequeña* 'the small skunk is in front of the big house and the large skunk is in front of the small house').

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

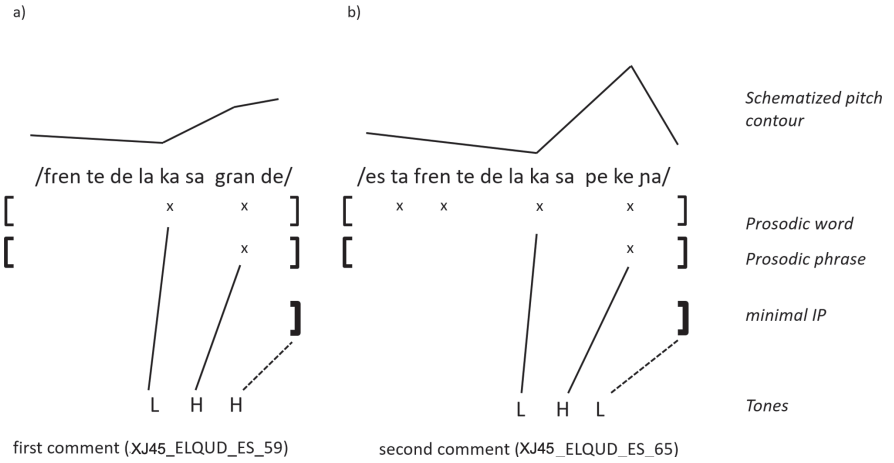


Figure 108: Proposed alignment and association model for the last phrase in the first comment of 59 (a) and the second of 65 (b) by XJ45. Solid lines between tones and positions in the structure indicate association, dashed lines alignment.

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-

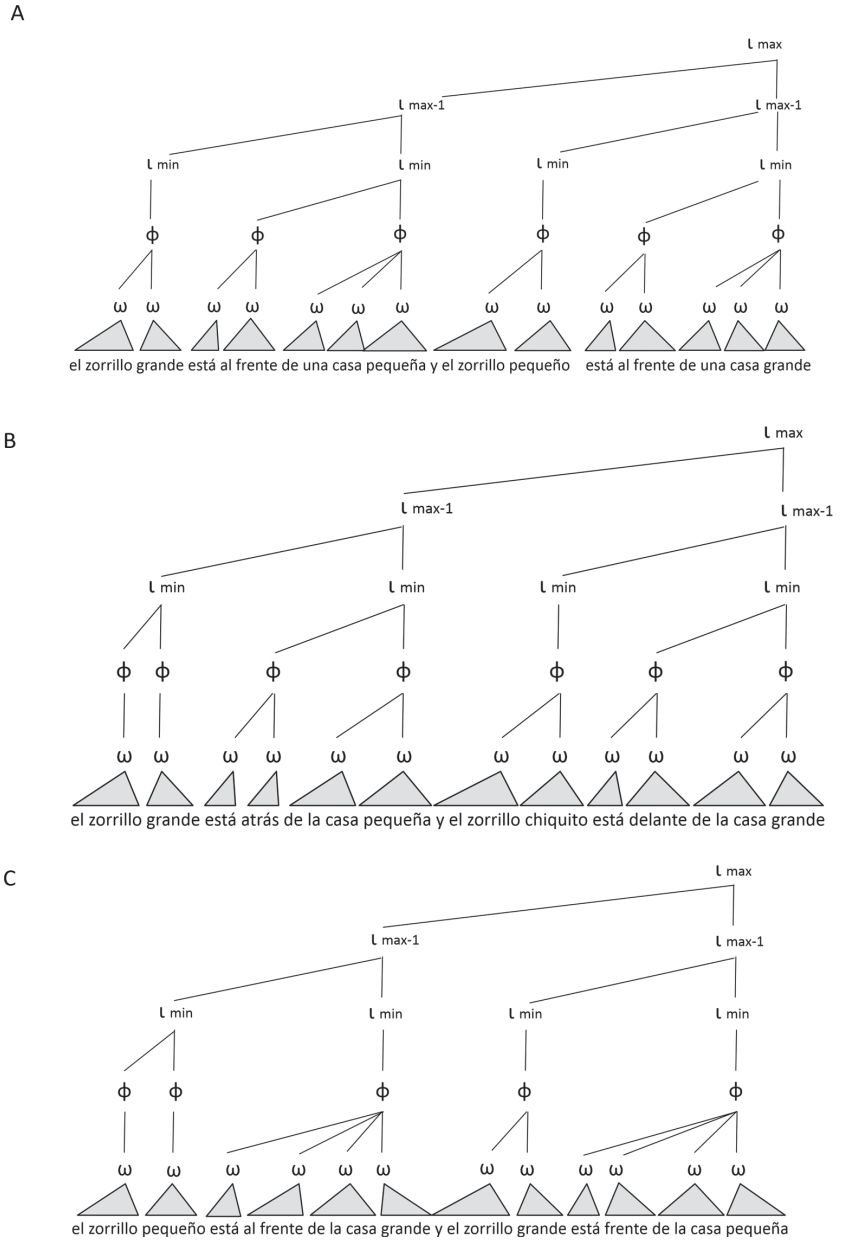


Figure 109: Prosodic structures for QZ13_ELQUD_ES_65 (A), ZE55_ELQUD_59 (B), and XJ45_ELQUD_65 (C), intended to exemplify the range of variation in these structures for the Spanish “double topic” constructions, including the “main” variant (A) and the different variants within the “phrase accentuation” 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.

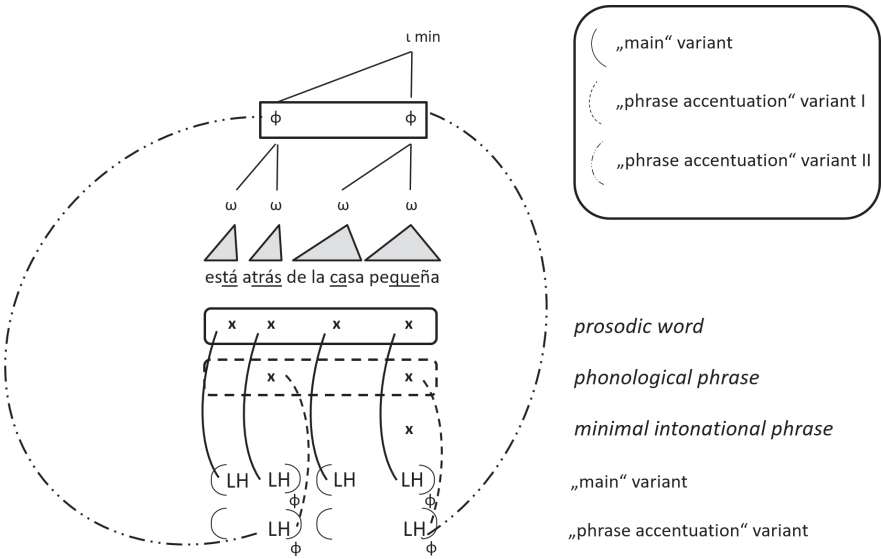


Figure 110: Difference in tone assignment between the „main“ and the „phrase accentuation“ variants. Phrase accentuation variant I is the one exemplified by ZE55, variant II that by NQ01 & XJ45.

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 cate-

gorical 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
[ ]$_{IP\_n}$
[ ]$_{IP\_n-1}$ [ ]$_{IP\_n-1}$
b. unbalanced recursive structure
*[ ]$_{IP\_n}$
[ ]$_{\phi}$ [ ]$_{IP\_n-1}$

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)

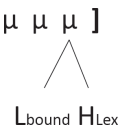
```
x  x  x  x  xx  x
      x      x
             x
          LH*
```

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 → TBU: the H tone is associated with a TBU.

(83) L → TBU: the L tone is associated with a TBU.

In our case, a relative ranking of $(\sigma')_{\omega'} \leftarrow$ T >> $\sigma' \leftarrow$ T >> H → TBU >> NoAssoc >> $\sigma \leftarrow$ T >> L → 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 align-

ment 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, σ́) >> $(\acute{\sigma})_{\omega}$← T >> H → TBU >> σ́← T >> ALIGN (L, Rt_{ϕ}) >> ALIGN (L, Lt_{ϕ}) >> ALIGN (H, Lt_{ϕ}) >> NOASSOC >> NOTARGET >> ALIGN (H, Rt_{ϕ})

In this constraint ranking and with the tone sequences we are dealing with here, ALIGN (L, Rt_{ϕ}) and ALIGN (L, Lt_{ϕ}) 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 σ́← 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, σ́) together with NOCROWD ensures that an H associates with a stressed syllable, rather than an L. ALIGN (H, Rt_{ϕ}), 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).

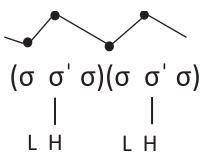
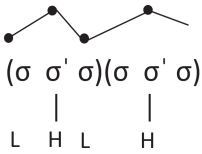
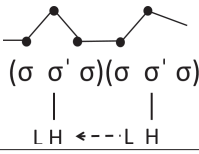
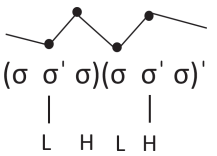
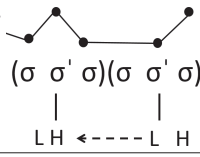
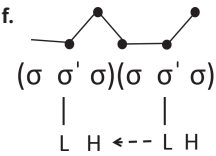
were indeed far less frequently found (cf. section 5.1.1.2). They are therefore not the main focus here, but of some interest because the plateau-forming behaviour is very frequent in the Quechua data. Whether they are allowed or disallowed is decided by the relative ranking of ALIGN (L, Rt_ϕ) and ALIGN (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.

$[\,(\sigma\sigma'\sigma)_{\omega}\,(\sigma\sigma'\sigma)_{\omega}{}'\,]_{\phi}$ LHLH	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. 					**7***	***!		**	****	*****
b. 					**8*!**	**		**	****	*****
☞ **c.** 					**7***	**		**	*****	*****
d. 	*!		*		**6***	****		**	****	****
e. 	*!		*		**6***	**		**	**6***	****
f. 	**!		**		*****	****		**	*****	***

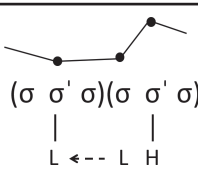
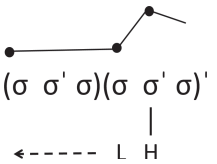
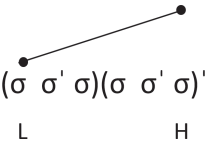
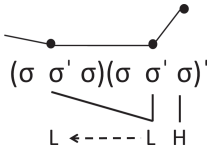
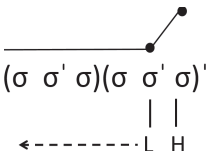
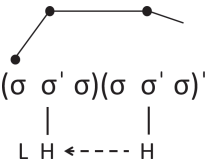
low until right before the next one. The last three candidates (d-f) all fail to be selected because they violate ALIGN (H, σ') and H → TBU, thus preventing a realization with the middle L and H as floating central tones (d), H as right-aligned boundary tone (e) or one that could be annotated as L*H (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.

[(σσˈσ)$_{\omega}$ (σσˈσ)$_{\omega}$ˈ]$_{\phi}$ LH	Align (H, σˈ)	(σˈ)$_{\omega}$←T	H → TBU	σˈ←T	Align (L, Rt$_{\phi}$)	Align (L, Lt$_{\phi}$)	Align (H, Lt$_{\phi}$)	NoAssoc	NoTarget	Align (H, Rt$_{\phi}$)
☞ **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 $(\sigma')_\omega$← 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).

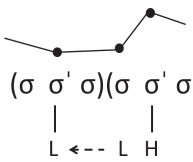
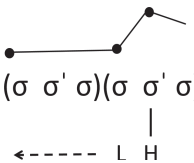
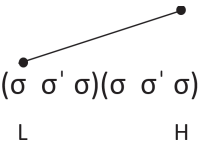
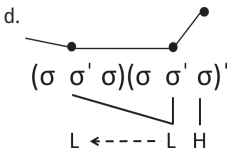
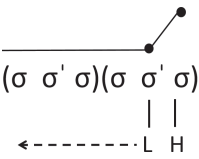
(93) Association and alignment constraint ranking for XJ45's "phrase accentuation" variant
ALIGN (H, Rt_ϕ) >> $(\sigma')_\omega$← T >> H → TBU >> σ'← T >> ALIGN (H, σ') >> ALIGN (L, Rt_ϕ) >> ALIGN (L, Lt_ϕ) >> ALIGN (H, Lt_ϕ) >> NOASSOC >> NOTARGET

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 $(\sigma')_\omega$← T and σ'← 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.

[(σσ'σ)$_{\omega}$ (σσ'σ)$_{\omega}$']$_{\phi}$ LH	**Align (H, Rt$_{\phi}$)**	**(σ')$_{\omega}$ ← T**	**σ' ← T**	**H → TBU**	**Align (H, σ')**	**Align (L, Rt$_{\phi}$)**	**Align (L, Lt$_{\phi}$)**	**Align (H, Lt$_{\phi}$)**	**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, σ'← T is moved further down the ranking, below NoAssoc. This positions this ranking even closer to Quechua, where σ'← 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_{ϕ}) >> $(\sigma')_{\omega}$← T >> H → TBU >> ALIGN (H, σ') >> ALIGN (L, Rt_{ϕ}) >> ALIGN (L, Lt_{ϕ}) >> ALIGN (H, Lt_{ϕ}) >> NOASSOC >> NOTARGET >> σ'← 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.

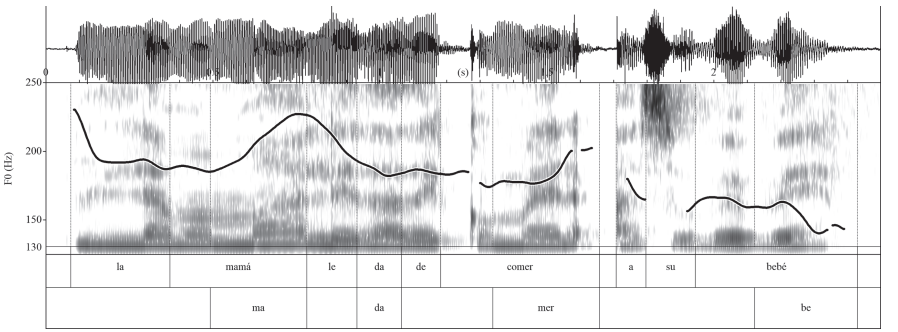


Figure 111: NQ01_ELQUD_ES_60[246] (*la mamá le da de comer a su bebé* ‘the mother gives her baby to eat’).

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 $(\sigma')_\omega$← T and H → 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/

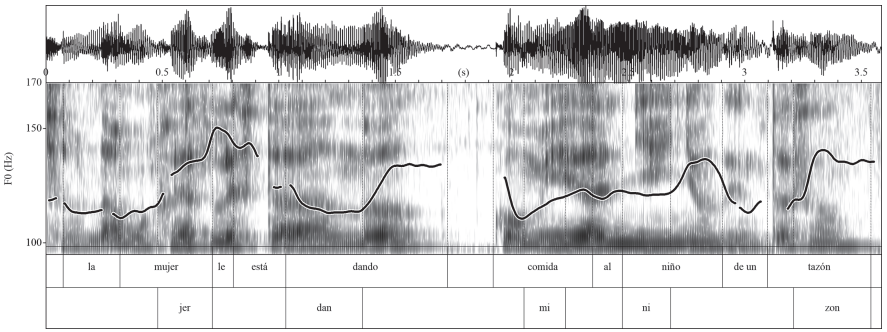


Figure 112: XJ45_ELQUD_ES_60[247] (*la mujer le está dando comida al niño de un tazón* 'the woman is giving food to the child from a mug').

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 cons-

traint 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.

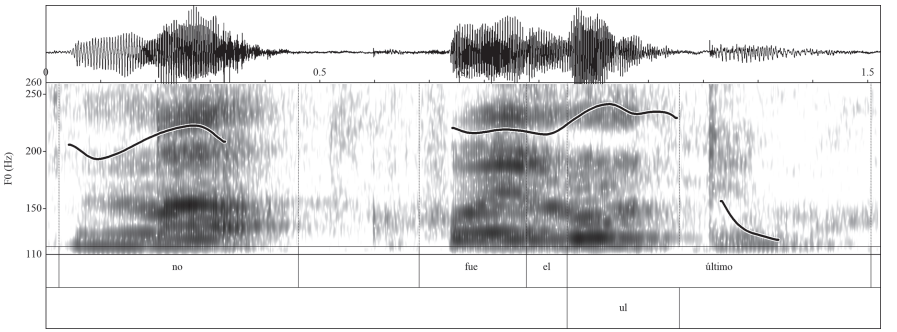


Figure 113: ZE55_ELQUD_ES_51[250] (*no fue el último* ‘no it was the last one’).

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).