

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:

2018-06-20

Deposited version:

Post-print

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Mata, A. I., Moniz, H. & Batista, F. (2016). Stylistic variation in the intonation of European Portuguese teenagers and adults. In Meghan E. Armstrong and Nicholas C Henriksen and Maria del Mar Vanrell (Ed.), *Intonational Grammar in Ibero-Romance: Approaches across linguistic subfields*. (pp. 45-68). Amsterdam: John Benjamins Publishing Company.

Further information on publisher's website:

<https://benjamins.com/#catalog/books/ihll.6.03mat/details>

Publisher's copyright statement:

This is the peer reviewed version of the following article: Mata, A. I., Moniz, H. & Batista, F. (2016). Stylistic variation in the intonation of European Portuguese teenagers and adults. In Meghan E. Armstrong and Nicholas C Henriksen and Maria del Mar Vanrell (Ed.), *Intonational Grammar in Ibero-Romance: Approaches across linguistic subfields*. (pp. 45-68). Amsterdam: John Benjamins Publishing Company.. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Running head: STYLISTIC VARIATION IN EP INTONATION

Stylistic variation in the intonation of European Portuguese teenagers and adults

Ana Isabel Mata¹, Helena Moniz^{1,2}, Fernando Batista^{2,3}

¹Faculdade de Letras da Universidade de Lisboa (FLUL), ²Instituto de Engenharia de Sistemas e Computadores – Investigação e Desenvolvimento em Lisboa (INESC-ID),

³Instituto Universitário de Lisboa (ISCTE-IUL)

Stylistic variation in the intonation of European Portuguese teenagers and adults

Ana Isabel Mata¹, Helena Moniz^{1,2} & Fernando Batista^{2,3}

¹Faculdade de Letras da Universidade de Lisboa (FLUL)

²Instituto de Engenharia de Sistemas e Computadores – Investigação e Desenvolvimento em Lisboa (INESC-ID)

³Instituto Universitário de Lisboa (ISCTE-IUL)

The present study aims to investigate intonation contours in phrase-final position, in a corpus of spontaneous and prepared unscripted presentations from teenagers (14-15 years old) and adults, collected in a school context. Taking into account the differences between phrasing levels (ToBI breaks 3 and 4), we show that the frequency of low/falling vs. high/rising contours – mainly (H+)L* L and (L+)H* H – varies across oral presentation types. Adults and teenagers follow distinct strategies, though cross-gender differences are also a source of variation. We interpret these changes as an adaptation effect to the speaking styles specifically required at school, which call for the speaker’s effort to speak clearly and to keep the listeners attention, and ultimately as “intelligibility-oriented” speaking style changes.

Keywords: Intonation, stylistic variation, European Portuguese, teenagers and adults

1. Introduction

Across languages, research has shown that intonation plays a major role in speaker adaptation to different situations (Swerts, 1994; Blaauw, 1995; Hirschberg, 2000; among others). Research has shown also that underlying differences between types of speech (varying along the spontaneous-read continuum) affect the distribution of final rising / falling contours, as well as the variation of global prosodic parameters – F0 height, overall range, speaking rate,

duration and frequency of pauses (for an overview and discussion, see Eskénazi, 1993 and also Barry, 1995).

It is usually recognized that rising contours are far more common in spontaneous speech than in read speech (see, for example, Silverman et al., 1992a, for American English; Blaauw, 1995, for Dutch; Serra, 2009, for Brazilian Portuguese). In monologues, rising declaratives are viewed as an effect of the more frequent expression of non-finality in spontaneous speech (Blaauw, 1995: 122). In narratives, other more specific form-function relations appear as well. High-rising declaratives seem to be exploited by speakers as a way to connect with listeners and keep their attention – for example, Wennerstrom (2001: 203) suggests that high-rising declaratives are used as “offers to clarify the scene for listeners if necessary (are you with me?)” by American English graduate students; Guy & Vonwiller (1989: 25) correlate high-rises with “the semantic complexity of the text and therefore the need for checking to see if the audience is understanding what is being said” in Australian English.

Research on speaking styles remains largely focused on adult speech. References to teenagers are sparse, since this age is quite overlooked in intonation studies. Fagyal & Stewart (2011) showed that, in conversational interactions in French, a group of male teenagers shifts from neutral falling / rising contours to a marked rising-falling contour as an effect of audience orientation. Trask (2010: 17) noted that instead of the standard falling contours, American adolescents often give declaratives a rising intonation in spontaneous speech and that this “uptalk” or high-rising pattern has spread easily. A comparable type of usage was observed for high-rises in Australian and New Zealand English, with speech task related differences – rising declaratives appear to be used more often in narratives and map-task dialogues than in non-narrative tasks (see Fletcher, Grabe & Warren, 2005, and references therein). These types of intonational differences across English varieties were

characterized as semantic by Ladd (1996) – meaning that a phonologically identical pattern is used with different functions. Alternatively, Fletcher, Grabe & Warren (2005: 399) proposed that in certain types of speech (narratives and map-task dialogues) a “neutralization of ‘phonological’ contrasts” may occur between rising and falling patterns (i.e. question rises and declarative falls) or between different rising patterns (i.e. non-final / continuation rises and question rises).

Besides speech task differences, a speaker-related factor that has received some attention in the literature is the influence of gender on the use of rising intonation in declaratives (*e.g.* Eckert & McConnell-Ginet, 2003). Recently, for Standard Southern British English, Jiang (2011) has shown that in declarative read sentences, rising boundaries are more frequent overall for teenage females, and falling boundaries for teenage males. Comparing fairy-story and “pseudo-scientific” reading passages, Clopper & Smiljanic (2011) found that young American university female students tend to use more rising phrase boundaries (L-H%), associated with continuation marking (*cf.* Pierrehumbert & Hirschberg, 1990). With regard to pitch accents, Clopper & Smiljanic (2011) have shown that L*+H is more common for females, particularly in the pseudo-scientific passage; H* is more common for males in the fairy-story passage. According to the authors, this prosodic variation may be associated not only with cross-gender differences but also with the specific characteristics and complexity of reading passages. Moreover, in other speaking styles, such as spontaneous interactions, cross-gender differences do not always appear to be significant. For instance, in (declarative) responses to teachers’ questions by 8-9 years old students, no significant gender effects were observed by Shepherd (2011). In responses in *wh*-question form on a television game show, the opposite was found by Linneman (2013): females tend to use rising intonation significantly more than males.

Cross-analyses of speech types comparing teenagers and adults are rare. With our work, we expect to help fill this gap and to gain more insight into this form of variation in European Portuguese (EP) intonation. Our corpus was built in order to investigate intonational differences between spontaneous and prepared unscripted presentations from teenagers (14-15 years old) and adults, collected in Portuguese high schools in the last year of compulsory education (the CPE-FACES corpus). Here, we focus on fluent final intonation contours in declarative utterances, selected from a subset of the corpus that was recently annotated with the ToBI prosodic system adapted to EP (Viana et al., 2007). We analyze the distribution of nuclear accents and boundary configurations per phrase break level (ToBI break indices of levels 3 and 4), as well as the final word pitch slope and standard deviation. Additionally, we take into account the mean duration of silent pauses and the proportion of delimiting punctuation marks among phrase levels.

In EP, declaratives are the most studied sentence type (e.g., Falé, 1995; Frota, 2000; Viana et al., 2007). The intonational contour generally associated with a declarative is a falling one, expressed as a prenuclear H* (in the first stressed syllable), a nuclear bitonal event H+L*, and a boundary tone L%. Although, as in most languages, a final falling contour represents the general trend associated with declaratives, studies have been reporting other final contours, as plateaus and final rising contours associated with pragmatic meanings such as continuation, and speaking style effects (Mata, 1999). These findings in Portuguese are similar to what has been reported for other languages such as Catalan (Prieto, 2014; Benet, Lleó & Cortés, 2011). In Ibero-Romance intonation in general, and for Portuguese in particular, stylistic variation based on speech outside a laboratory setting has barely been explored (see Mata, 1999, for Portuguese; Henriksen, 2013, for Spanish; Prieto, Payà & Vanrell, 2014, for Catalan). Our study provides an important and original contribution to the

development of knowledge on intonational variation in the sense that it relates speech style differences across age groups and gender in a classroom context.

Our main goal is to examine to what extent intonation patterns are affected by (i) the specific nature of the speech situation and (ii) inter-speaker differences (speaker's status/age¹ and gender). In particular, we address the following fundamental questions: is there a systematic difference in final contours between typical (prepared) school presentations and spontaneous ones? How does this difference emerge across the presentations of adults (the teachers) and teenagers (the students)? Are there general gender-related differences or do boys and girls differ in rising contour choice, for example?

Previous studies for EP (Mata, 1999, 2012; Moniz, Trancoso & Mata, 2010; Moniz, 2013) have shown cross-corpora (university lectures, high school presentations, map-task dialogues, broadcast news), intra-corpora (spontaneous vs. prepared (un)scripted) and inter-speaker (age, gender, status) variation. Building on these studies, we expect the specificity of typical school presentations to affect the distribution of final intonation patterns. We also expect the speaker's status/age and gender to play a role in final contour choices.

The chapter is organized as follows: section 2 presents the methodology followed in our study. Section 2.1 offers a brief description of the corpus CPE-FACES, section 2.2 presents the prosodically annotated set we use, and section 2.3 the method developed to create an automatic prosodic description of the data, integrating the manual prosodic information into the speech recognition output. Section 3 sets out the main patterns found in the selected data, first with an analysis of the most frequent final contours per phrase break level, followed by a cross-analysis of speech types comparing teenagers and adults. Section 4 presents a discussion of our main results and concluding remarks.

¹ In this study the terms status and age are not separated, since teenagers are directly associated to students and adults to teachers. Further research is needed in order to explore these sociolinguistic variables in a cross-comparison fashion.

2. Methodology

2.1. Corpus and Data

This study uses a subset of a corpus of European Portuguese spoken by teenagers and adults in school context, CPE-FACES (Mata, 1999). The corpus consists of spontaneous and prepared unscripted speech from 25 students (14-15 years old) and 3 teachers, with a total of 16h, collected in Portuguese high schools in the last year of compulsory education.

The subset used is comprised of 9 spontaneous presentations and 9 prepared unscripted presentations, from 6 students (3 boys and 3 girls) and 3 teachers (2 female and 1 male), all speakers of standard EP (Lisbon region). The prepared situation corresponds to typical school presentations, about a book the students must read following the Portuguese syllabus. In the spontaneous situation, students and teachers were unexpectedly asked to relate a(n) (un)pleasant personal experience. It was assumed that the involvement of speakers on topics related to their personal interests and day-to-day life would manifest in the naturalness and spontaneity of their speech (Labov, 1976).

In terms of the three-dimensional scheme proposed by Eskénazi (1993) to characterize speaking styles, the differences between these oral presentations may be classified primarily as “intelligibility-oriented” differences. Since both presentations were recorded in a classroom context (the speakers classroom of Portuguese as their L1), the remaining dimensions of variation – related with the context and background of interlocutors as well as with the familiarity between interlocutors – are stable within each group both in spontaneous and prepared data.

Basically, the data differs in the degree of planning involved, the type of information communicated, the speakers’ attention to the speech task and effort to speak clearly. In spontaneous presentations, the speakers’ freedom is not restricted: they can talk freely about any topic of their choice; they can change topic and move on to another topic whenever they

want. As far as typical (prepared) oral presentations at school are concerned, it has been argued before that “more than talking about a pre-determined theme, an oral presentation presupposes the capacity to individually produce a greater amount of utterances, organizing the information that is given to the public in a clear structured form” (Mata, 1999: 6).

2.2. Annotation procedures

Excerpts of the corpus have been prosodically annotated with the ToBI system (Silverman et al., 1992b), for experiments on automatic ToBI-labeling adapted to EP, as part of an ongoing project (see Mata, A. I., Moniz, H. & Batista, F. & Hirschberg, J., 2014; Moniz, H., Mata, A. I., Hirschberg, J., Batista, F., Rosenberg, A. & Trancoso, I., 2014). Taking into account the first proposal *Towards a P_ToBI* by Viana et al. (2007)², this task has been done by two of the authors, using Praat (Boersma & Weenink, 2013). Additionally, punctuation marks were added to the corpus by an independent annotator – a Portuguese teacher – following the guidelines provided by Duarte (2000), in which punctuation marks are described based on syntactic criteria. It is widely accepted that speech units do not always correspond to sentences, as established in the written sense. They may be quite flexible, elliptic, restructured, and even incomplete (Blaauw, 1995). Structural metadata events (Liu *et al.*, 2006; Ostendorf *et al.*, 2008), i.e. punctuation marks and disfluencies, are being added to several corpora in EP, including CPE-FACES, in order to enrich automatic speech recognition outputs, for legibility purposes and also for the empirical study of interactions among different linguistic levels of analysis.

In the 18 oral presentations, a total of 48.6 minutes (33.2 minutes of useful speech) was prosodically annotated. Since students’ presentations vary from about 1-25 minutes and teachers’ from about 2-78 minutes, the shortest presentations were fully annotated. As for the

² Downloadable from: <http://www.clul.ul.pt/english/equipa/amata/Extended-Abstract-P-ToBI.pdf> ; www.clul.ul.pt/english/equipa/amata/WkPaPI07P-ToBI.ppt

longer ones, only the initial part was included (approximately 1 minute for teenagers and 5 minutes for adults). All the pitch accents (H+L*, H*+L, L*+H, L+H*, H*, L*, H+!H*) and the final boundary tones (L%, H%, !H%, LH%, HL%) that are covered in the *Towards a P_ToBI* proposal were used in this subset (see the schematic F0 contours in Figure 1).

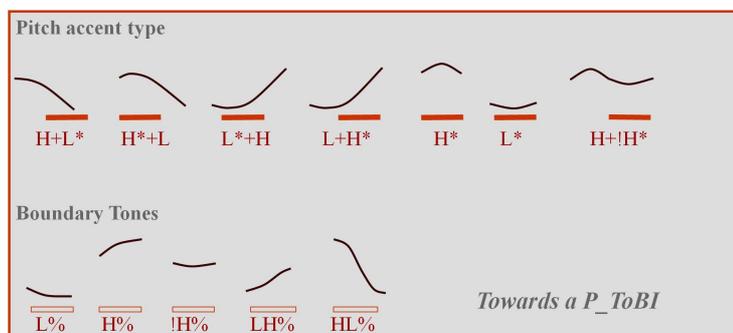


Figure 1. Schematic F0 contours for pitch accents and boundary tones. (Lines — indicate the stressed syllable.)

Prosodic and intonational studies have shown that there are two levels of intonational phrasing in EP, with boundaries of varying strength. Frota (2000) and Viana et al. (2007) equated both to the intonational phrase type – the major IP and the minor IP, in line with Ladd (1996) – and proposed to account for the different break strengths with ToBI levels 4 and 3, respectively, as well as with the corresponding tonal labels for boundaries, “%” and “-”. These coding procedures were applied to CPE-FACES, in order to allow for a comparison of final contours across intonational phrase levels. We will call them simply: “intonational phrases” and “minor phrases”.

In order to calculate the inter-annotator agreement, we compared the results of 57 files (with 900 break index marks and 729 tonal labels) from the two annotators. The agreement was measured in terms of Fleiss’ kappa (Fleiss, 1971). An agreement rate of 71.8 was achieved for both pitch accents and boundary tones³, and 93% for break indices, which compares well with ToBI inter-transcriber consistency for other languages (see Escudero *et*

³ The most salient disagreement is related to the number of tones that were annotated by annotator A and not by annotator B.

al., 2012, and references therein). According to the scale proposed by Landis and Koch (1977), there is *substantial agreement* for pitch accents and boundary tones and *almost perfect agreement* for break indices. Table 1 presents more detailed statistics on this process (<http://dfreelon.org/utis/recalfront/recal3/>).

	breaks	tones
n cases	900	729
average pairwise percent agreement	95.78%	76.13%
Fleiss' kappa	92.97%	71.78%
FK observed agreement	95.78%	76.13%
FK expected agreement	39.91%	15.41%
average pairwise Cohen's kappa	92.98%	71.95%
Krippendorff's alpha	92.98%	71.80%

Table 1. Agreement between two annotators.

Table 2 shows the overall characteristics of the annotated data set per speaker and presentation. It contains 5853 words from 9 speakers – 63.4% (22.2 minutes of useful speech) from adults; 36.6% (11 minutes of useful speech) from teenagers – with a total of 1871 phrase boundaries: 47.7% in the spontaneous subset; 52.3% in the prepared subset. The first subset has a phrasing rate of 30.7% and the second 33.2%.

		A1	A2	A3	B1	B2	B3	G1	G2	G3	Total
Spontaneous presentation	Useful speech (mins)	3.83	1.76	5.48	0.77	0.82	0.84	0.61	1.09	0.8	16
	Words (N)	621	315	930	135	141	186	119	271	185	2903
	Phrasing rate (%)	30	35.2	28.9	33.3	34	32.8	39.5	25.1	30.8	30.7
Prepared presentation	Useful speech (mins)	3.53	4.42	3.22	0.79	0.84	0.87	1.22	1.41	0.91	17.21
	Words (N)	602	659	583	127	164	171	180	279	185	2950
	Phrasing rate (%)	35	35.5	30.2	38.6	31.7	34.5	33.9	27.6	32.4	33.2

Table 2. Overall characteristics of the annotated data set (A stands for adult, B for boy and G for girl; A3 is the male adult).

It is worth noting that speech is edited online in both situations. Consequently, both presentations display disfluent phrase breaks (marked with ‘p’) – corresponding to filled pauses, prolongations, repetitions, substitutions, deletions, insertions. For the purposes of the present study, we selected all intonation contours that were realized with a fluent phrase break and that carried a pitch accent in phrase-final position (i.e. in the last word). Furthermore, we did not include interrogatives. 1552 (corresponding to 83%) of final

contours met all conditions in the annotated set: 717 in the spontaneous presentation and 835 in the prepared presentation. As our main data set, these final contours in declarative utterances – 63.9% from female speakers (67.5% adults; 32.5% teenagers) and 36.1% from male speakers (54.6% adult; 45.4% teenagers) – will be analyzed in the next sections. For the sake of comparison, disfluent phrase breaks marked with ToBI label 3p (130 in total, 70 in spontaneous presentations and 60 in prepared ones, 36.9% from females and 63.1% from males) will be analyzed for boundary configuration in section 3.2.3.

2.3. Automatic feature extraction processes

This section is concerned with the automatic prosodic feature extraction process, which involves synchronizing the manual prosodic labels with complex sets of acoustic features. It relies on information coming from the automatic speech recognition system (ASR) output, from manual transcripts, and from the signal itself. After producing the ASR transcripts, all relevant manual annotations are transferred to the ASR transcripts, including all the prosodic labels, by means of word alignments. The adopted annotation scheme allows associating relevant prosodic information to different units of analysis, including intonational units, words, syllables and phones. Duration of words, syllables and phones were derived from the ASR output. Information regarding pitch (f_0) and energy (E), not available in the ASR pipeline when this study started, has been directly extracted from the speech signal, using the Snack toolkit (Sjölander, K., Beskow, J., Gustafson, J., Lewin, E., Carlson, R., and Granström, B., 1998).

Acoustic-phonetic parameters of segmental and supra-segmental units were automatically extracted to study intonational events. Organizing such information into hierarchies, meaning into the smallest unit of analysis (phones or even sub-phone units) up to higher order constituents, was crucial to the experiments conducted, making it possible to

automatically extract a complex set of acoustic measures. At this point, the information extracted encompassed phones, syllables, words, intonational units, and speech-acts.

The speech recognizer (Neto et al., 2008), trained for the broadcast news domain, is unsuitable for the oral presentations domain. The scarcity of text materials in Portuguese to train language models for this domain has motivated the decision of using the ASR in a forced alignment mode, in order to avoid a bias of the bad results obtained with an out-of-domain recognizer. Therefore, current experiments rely on forced aligned transcripts that still contain about 1.5% of unaligned words (mainly due to overlapping speech).

The features were extracted for intonational units, either break levels 3 or 4, involving the final word itself and the adjacent contiguous words or the syllables within those units. Features involving a single word/syllable include: pitch and energy slopes; ASR confidence score; word/syllable duration; number of syllables and phones. Features involving two consecutive words/syllables include: pitch and energy slopes and shapes; pitch and energy differences; word durations, pitch medians, and energy medians. Pitch slopes were calculated based on semitones rather than Hz. Slopes in general were calculated using linear regression. Pitch and energy shapes are based on slope values and expanded to 9 binary features, assuming one of the following values {RR, R-, RF, -R, --, -F, FR, F-, FF}, where F=Fall (if the slope is negative), -=plateau (if the slope is near zero), R=Rise (otherwise). For a more detailed analysis vide Batista et al. (2012a).

3. Results

3.1. Overall results

The majority of the final contours analyzed in both presentations are marked with break index 4, corresponding to the intonational phrase level (434/60.5% in spontaneous and 540/64.7% in prepared presentations). Final contours marked with break index 3, which codes a minor

level of intonational phrasing, are less frequent overall (283/39.5% in spontaneous and 295/35.3% in prepared presentations).

Results show a clear difference in terms of nuclear accent type and boundary configuration between the two levels of intonational phrasing ($p < .001$; $U = -12.177, -5.574$, respectively). Low/falling accents are significantly more frequent in final contours marked with break index 4 and high/rising accents are significantly more frequent in final contours marked with break index 3, in both oral presentations. An identical contrast is established by the overall distribution of low vs. high boundary configurations, resorting above all to L and H. HL and !H are residual. LH is relatively rare and it is basically used with break index 4 – most often by female speakers, as we will see in section 3.2.3.

As Table 3 shows, two low/falling accents ($H+L^*$, L^*) and two high/rising accents (H^* , $L+H^*$) are the most frequent nuclear accents in the corpus. Together, they account for 86.9% of all accents in phrase-final position – 87.5% in the spontaneous and 86.4% in the prepared subset. $H+L^*$ (a sharp fall reaching a low target within the accented syllable) and L^* are mainly used in intonational phrases – 41.1% and 17.4%, respectively. H^* is mainly used in minor phrases (44.5%). $L+H^*$ (a sharp rise reaching a high target within the accented syllable) displays a more regular distribution – 16.6% in minor phrases and 14.7% in intonational phrases.

Nuclear Accent	Break index	Boundary configuration							N %		Total N %	
		(x)H				(x)L						
		LH	!H	H	all	L	HL	all				
$H+L^*$	3	8.8		2.7	11.5	88.5		88.5	113	19.6%	513	33.1%
	4	14.2	0.2	2.2	16.8	83.2		83.2	400	41.1%		
H^*+L	3					100.0		100.0	32	5.5%	107	6.9%
	4	4.0			4.0	96.0		96.0	75	7.7%		
L^*	3	9.4		9.4	18.9	81.1		81.1	53	9.2%	222	14.3%
	4	14.2		26.6	40.8	59.2		59.2	169	17.4%		
H^*	3		4.3	78.2	82.5	16.7	0.8	17.5	257	44.5%	374	24.1%
	4		0.9	76.1	76.9	20.5	2.6	23.1	117	12.0%		
$L+H^*$	3		5.2	75.0	80.2	11.5	8.3	19.8	96	16.6%	239	15.4%
	4		7.7	82.5	90.2	4.9	4.9	9.8	143	14.7%		

L*+H	3			100.0	100.0				11	1.9%	60	3.9%
	4	2.0	2.0	95.9	100.0				49	5.0%		
H+!H*	3			93.8	93.8	6.2		6.2	16	2.8%	37	2.4%
	4		9.5	85.7	95.2		4.8	4.8	21	2.2%		
N	3	15	16	307	338	230	10	240	578		1552	
	4	85	16	326	427	536	11	547	974			
%	3	2.6%	2.8%	53.1%	58.5%	39.8%	1.7%	41.5%		37.2%		100%
	4	8.7%	1.6%	33.5%	43.8%	55.0%	1.1%	56.2%		62.8%		

Table 3. Overall distribution of nuclear accents and boundary configurations per break index

H+L* is common in EP: H+L* L% is the canonical contour of declaratives, H+L* LH% the contour of information-seeking yes-no questions (Frota, 2002, 2014; Viana et al., 2007; Mata & Santos, 2010). As for the perception of these intonational contrasts, Falé (2005) and Falé & Faria (2006) have proven that the perception of a declarative is mainly based on the final fall in the last stressed syllable; the perception of an interrogative depends mostly on the pitch range of the phrase-final rise (rising boundaries must be higher than 2 semitones).

L* and H* are unusual in nuclear position in lab speech, in standard EP (see Viana et al., 2007). In reading materials, non-final phrases within complex sentences can take L* as their nuclear accent (Falé, 1995). In map-task dialogues, (H+)L* L% appear associated with the expression of agreement or completeness (Cabarrão & Mata, 2012); L* H% and H* H% with the suggestion of continuation (Viana et al., 2007; Cabarrão & Mata, 2012).

L+H*, virtually absent from lab speech, is a frequent nuclear accent in both professional reading and spontaneous speech, as Viana et al. (2007) pointed out. Furthermore, the authors suggested that L+H* nuclei may be used by speakers “to signal new information or to highlight given information that needs to be (re)activated”. In fact, in spontaneous adult-child dialogues, L+H* and H* followed by a non-falling tonal boundary are the most frequent contours in questions seeking confirmation of perception (Mata & Santos, 2010).

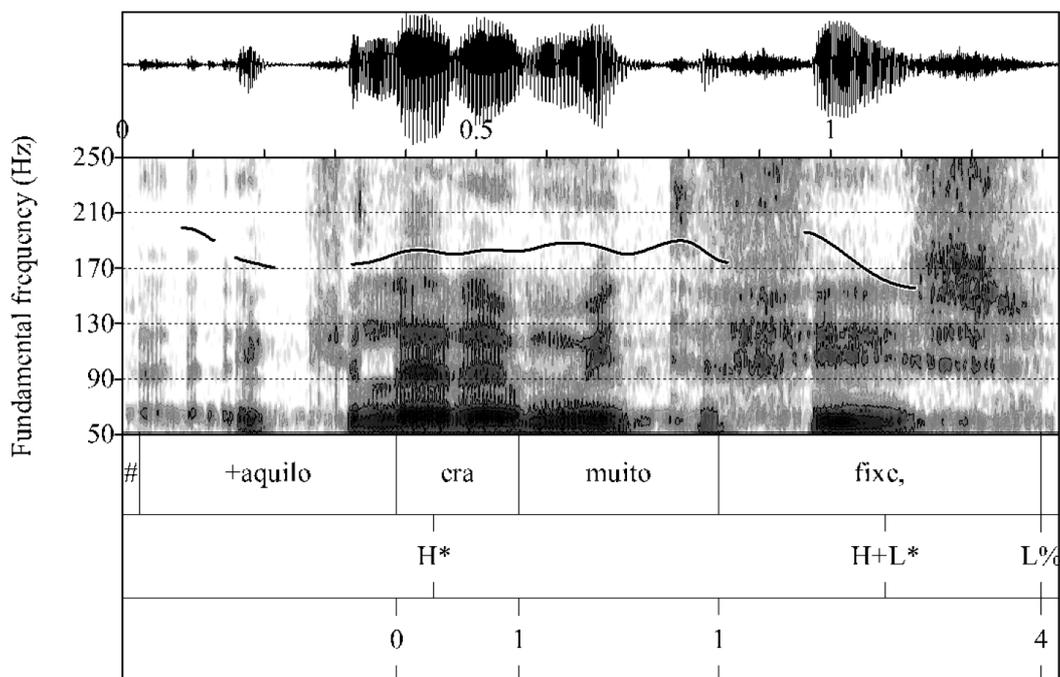


Figure 2. Example of H+L* L% in the excerpt: *aquilo era muito fixe,* that was pretty cool (extracted from a male teenager spontaneous presentation).

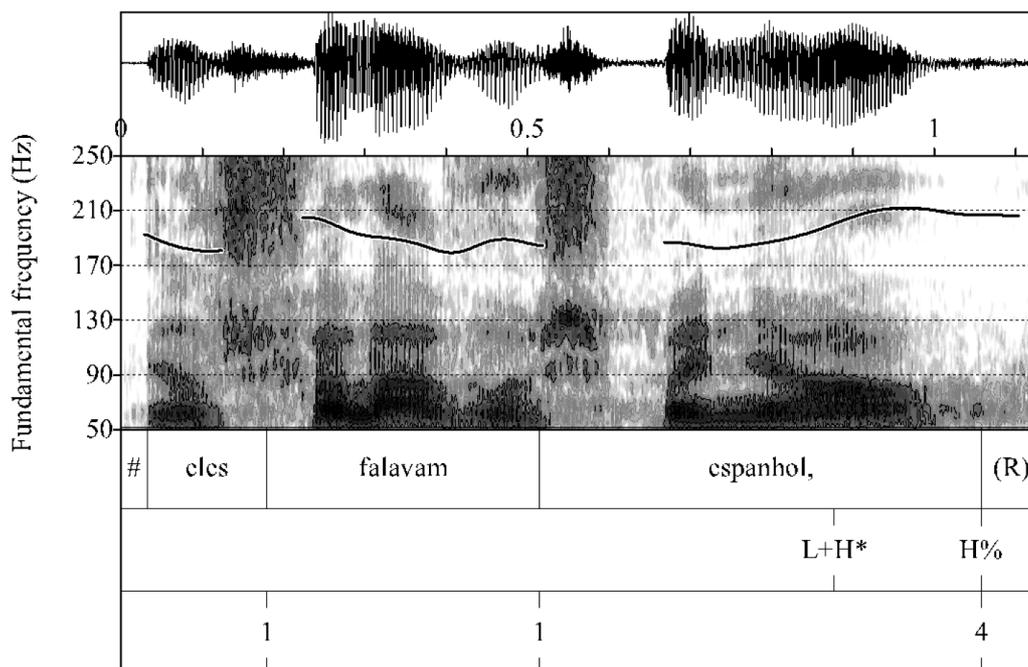


Figure 3. Example of L+H* H% in the excerpt: *eles falavam espanhol,* they spoke Spanish (extracted from a male teenager prepared presentation).

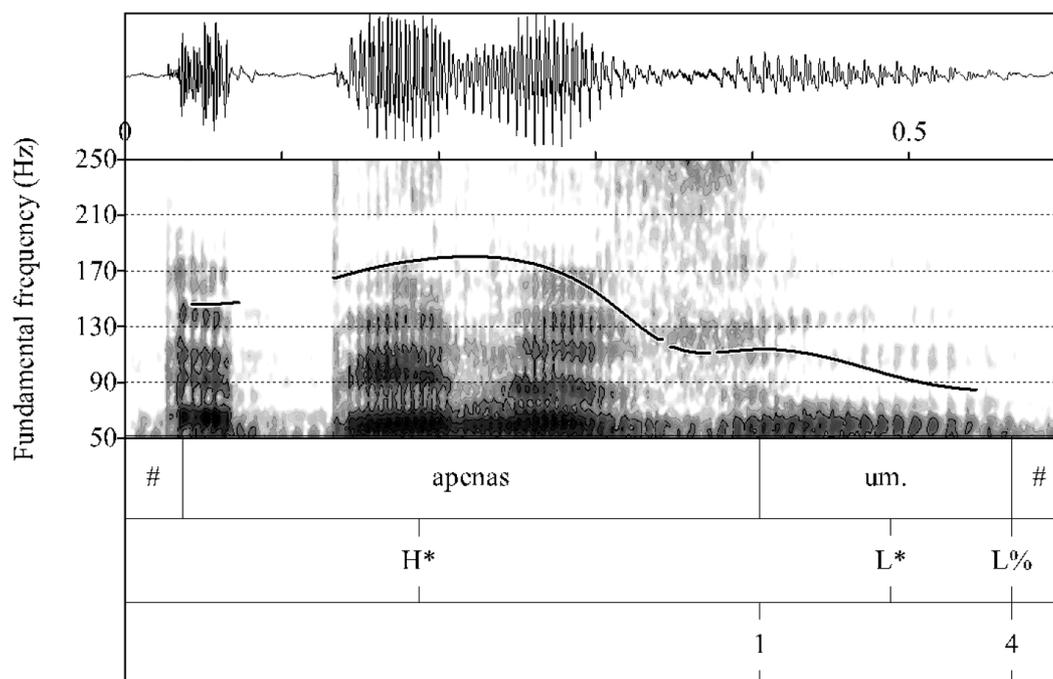


Figure 4. Example of L* L% in the excerpt: *apenas um./* just one (extracted from a male adult prepared presentation).

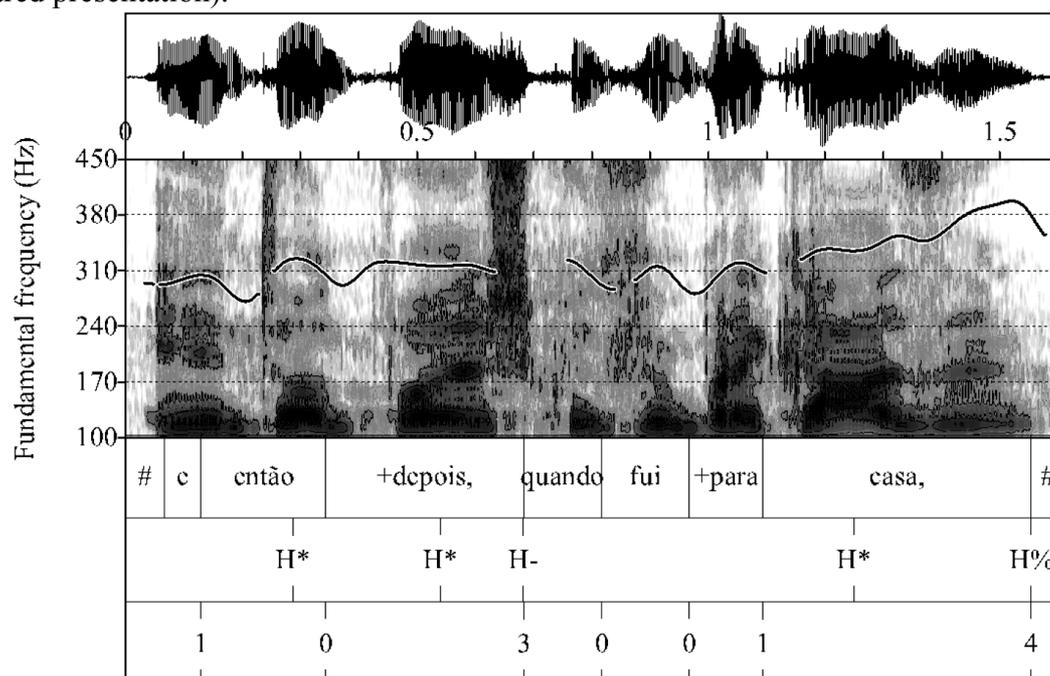


Figure 5. Example of H* H-/H% in the excerpt: *e então depois, quando fui para casa,/* and then later when I went home (extracted from a female teenager spontaneous presentation).

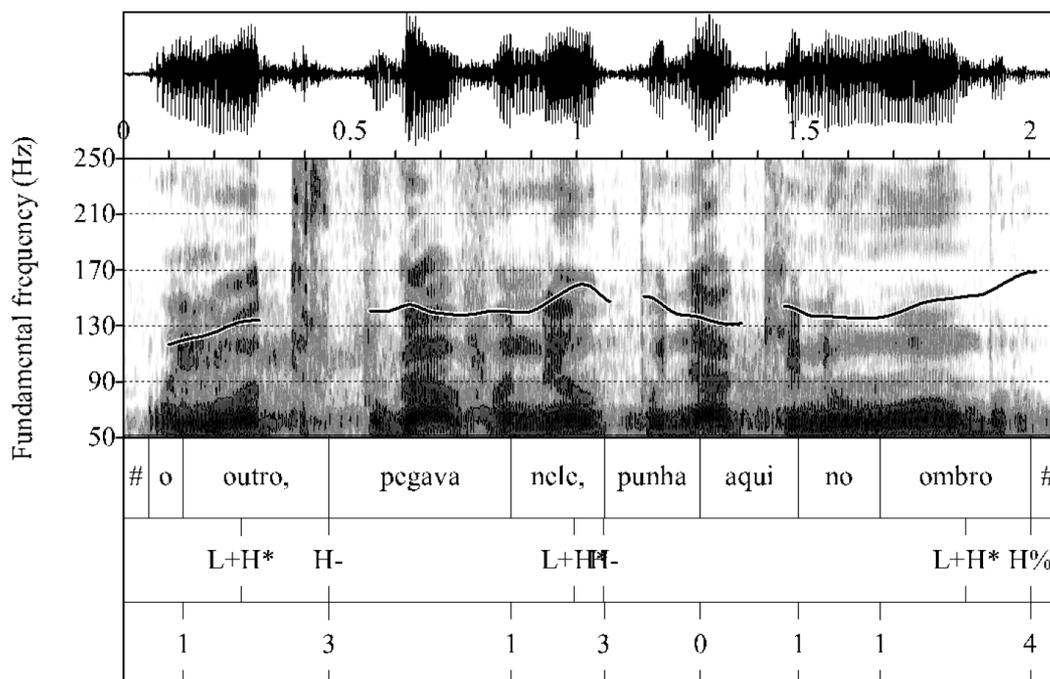


Figure 6. Example of L+H* H-/H% in the excerpt: *o outro, pegava nele, punha aqui no ombro/* the other, I would take it and put it here on the shoulder (extracted from a male teenager spontaneous presentation).

H+L* L%, L* L%, H* H- and L+H* H%/H- (illustrated in figures 2-6 above) are the main nuclear contours in our data. We did not include interrogatives in the analyzed data, as previously mentioned. Measurements of final word pitch slopes and standard deviations (cf. section 2.3) show that these are distinct contours. A statistical analysis confirms that these results are highly significant ($p < .001$, $U = -10.414$, -8.183 , respectively for normalized slopes and standard deviations across L+H* H and H* H; $p < .001$, $U = 5.789$, -9.437 , respectively for normalized slopes and standard deviations across H+L* L and L* L). As Figure 7 shows, although with varying degrees across phrase levels, H* H and L* L differ mainly in pitch slope ($p < .001$); L+H* H and H+L* L contrast both in pitch slope and standard deviation ($p < .001$; $p < .01$). Significantly higher values in both pitch features discriminate H+L* L and L+H* H from L* L and H* H, respectively.

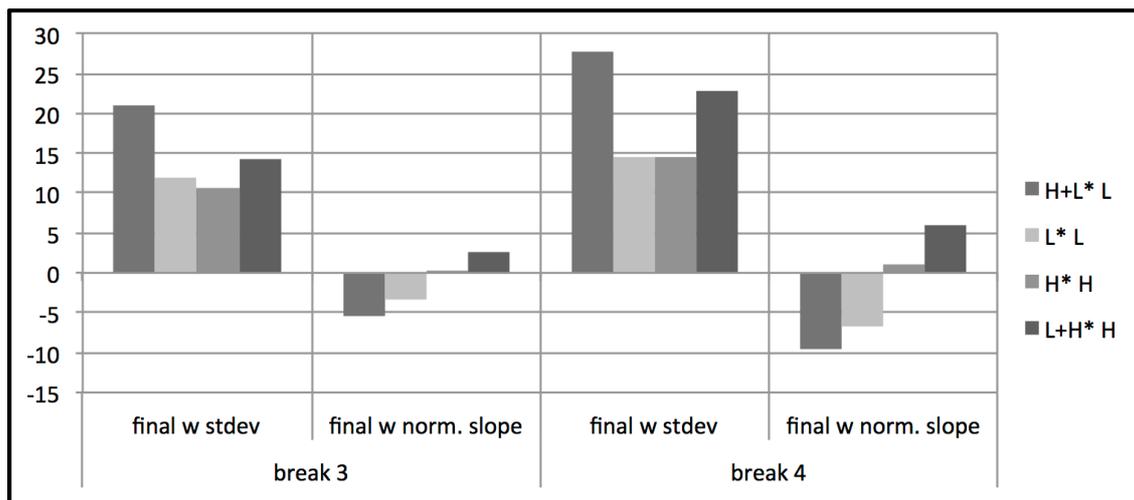


Figure 7. H+L* L, L* L, H* H and L+H* H contours: final word pitch slopes and standard deviations

Besides nuclear accent type and boundary configuration contrasts, overall differences between the higher and lower level of intonational phrasing include the mean duration of silent pauses and the proportion of punctuation marks delimiting each level.

Intonational phrases take longer subsequent silent pauses than minor phrases ($U=27.359$, $p<.001$) – 1.02 seconds is the mean duration of silent pauses after break index 4 and 0.07 seconds the mean duration of silent pauses after break index 3; this correlation was confirmed for all speakers in both presentations. Intonational phrases also tend to be longer (mean number of words, syllables and phones in intonational phrases: 3.43, 6.40, 13.53; mean number of words, syllables and phones in minor phrases: 3.11, 5.73, 12.13). Furthermore, intonational phrases tend to be characterized by higher mean values in pitch; minor phrases by higher mean values in energy. As expected, when the final word of a phrase is compared with the first word of the following phrase, the data shows a prosodic reset at the start of the adjacent context, though with varying degrees across phrase levels.

Finally, looking into punctuation differences among phrase levels, we note that intonational phrases boundaries display a higher proportion of punctuation marks than minor phrases boundaries ($\chi^2(4) = 286.886$, $p <.001$): 69.5% and 35.3%, respectively. Periods are almost exclusively associated with intonational phrases (98.5%). Commas, an abundant

punctuation mark in Portuguese speech transcripts (Batista et al., 2012b; Moniz et al., 2013), may delimit both phrase levels in our data (intonational phrases: 34.8%; minor phrases: 34.1%). However, the large majority of the punctuation marks inserted in minor phrases are commas (96.6%).

In this section we have shown that there are significant differences among phrase levels regarding the frequency of low/falling contours, the mean duration of subsequent silent pauses and the proportion of delimiting punctuation marks (higher for intonational phrases than for minor phrases, in every case). In the next sections we turn our attention to the variation of nuclear accent and edge tone types within phrase levels, comparing adults vs. teenagers' spontaneous and prepared presentations.

3.2. Spontaneous vs. prepared presentations

Figure 8 illustrates the percentage differences of nuclear accent type (low/falling vs. high/rising) and edge tone type (low vs. high), across adults and teenagers per presentation subset.

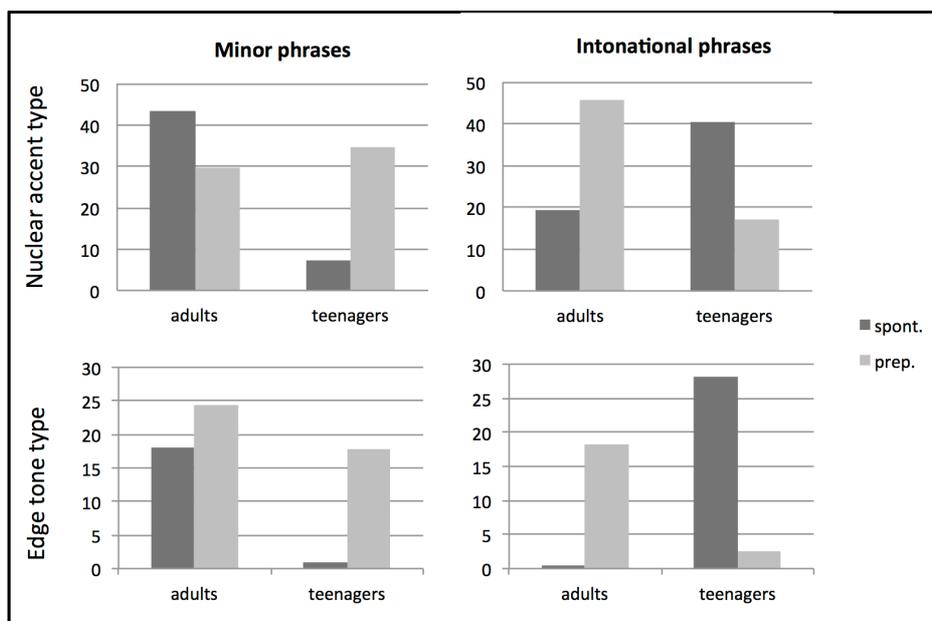


Figure 8. Percentage differences of nuclear accent type (top row) and edge tone type (bottom row) for adults and teenagers in spontaneous and prepared presentations

Spontaneous vs. prepared differences are visible in edge tone type distribution. However, the most striking differences concern nuclear accent types, within and across groups of speakers. For presentation type, the contrast between minor phrases and intonational phrases frequency did not reach statistical significance ($p = .093$, $U = -1.682$). In our data, variation in the balance of phrase levels is an additional evidence of gender-related differences across presentations.

3.2.1. Phrase levels

The distribution of phrase levels per presentation subset shows that break index 4 is predominant in all the prepared presentations, independently of inter-speaker differences (Figure 9).

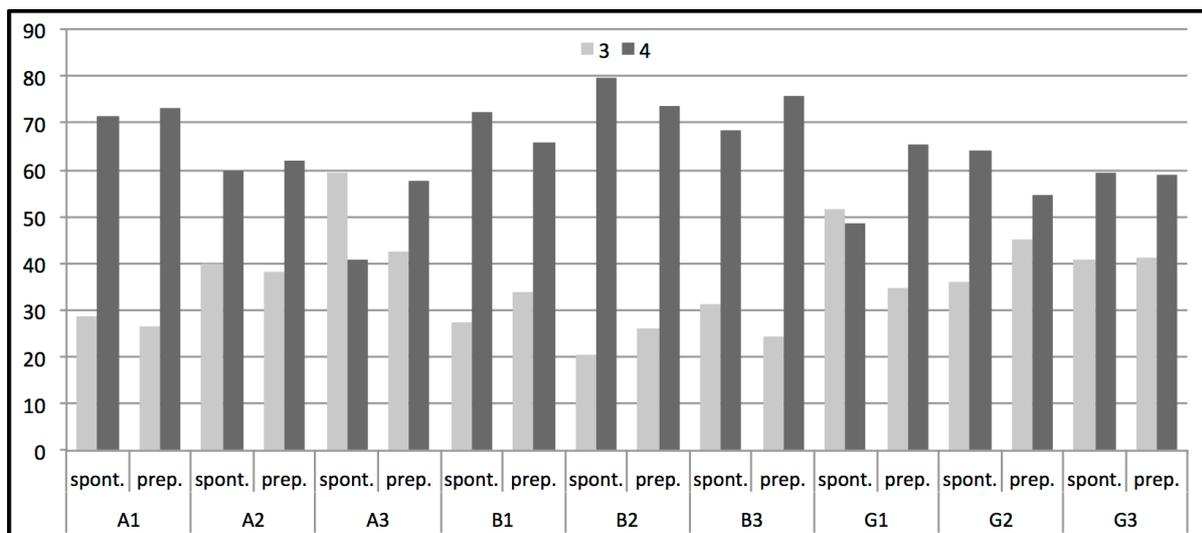


Figure 9. Distribution of phrase levels per speaker and presentation subset

In the spontaneous presentation, this effect is not equally consistent across speakers: break index 3 is the most frequent for one of the adults (speaker A3, the only male adult in our data) and for one of the female teenagers (speaker G1). Removing the respective data from our corpus analysis, the results do not display a clear distinction between phrase level

distribution and type of presentation or age group. For all the adults, intonational phrases increase and minor phrases decrease from the spontaneous to the prepared presentation. In the second group, half of the teenagers (B1, B2, G2) follows a strategy that is opposite to the one used by adults – increasing the frequency of minor phrase breaks and decreasing the frequency of intonational phrase breaks from the spontaneous to the prepared presentation, and thus reducing the differences between phrase level distribution – and one teenager (G3) does not change the relative frequency of minor and intonational phrases between presentations. Only two of the teenagers (B3 and G1) behave in a way similar to adults. Therefore, comparing adults versus teenagers, no significant correlation was found between phrase level distribution and type of presentation ($p = .754$, $U = .313$, for teenagers; $p = .980$, $U = .026$, for adults).

In spite of that, grouping teenager speakers by gender, a clear difference is found in both spontaneous and prepared presentations ($p < .01$, $U = 3.352$). As Figure 10 shows, girls display a higher percentage of minor phrases than boys; boys display a higher percentage of intonational phrases than girls. In general, the contrast between minor and intonational phrases is stronger for boys than for girls independently of presentation type (41%/59% for girls vs. 27.6%/72.4% for boys, respectively). This contrast is also evident for the female adults, when adult speakers are grouped by gender, ($p < .001$; $U = -5.684$). Moreover, it distinguishes female adults from female teenagers ($p < .01$; $U = 2.608$).

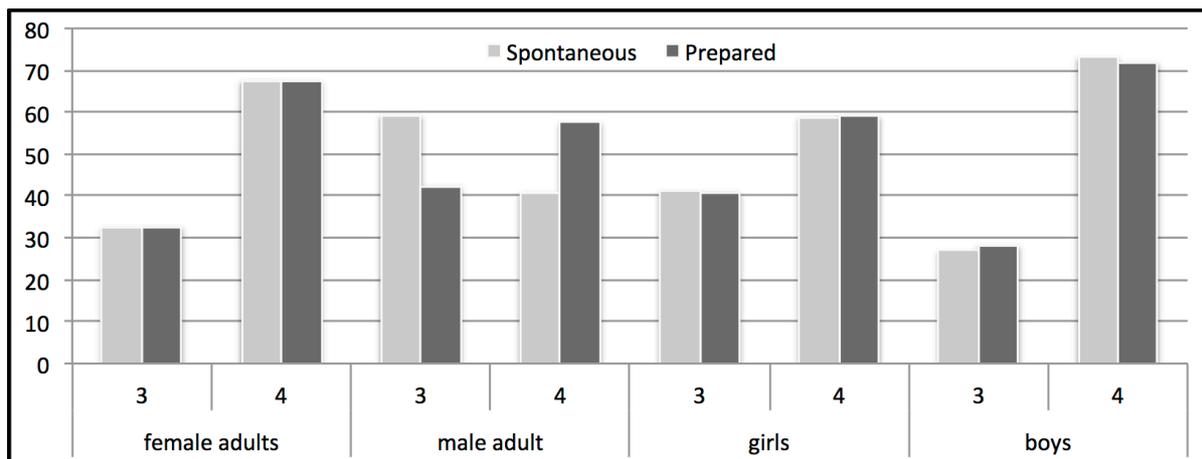


Figure 10. Distribution of phrase levels per gender

3.2.2. Nuclear accents

Comparing spontaneous vs. prepared presentations, the data shows a significant correlation between oral presentation and nuclear accent types for both the adults ($p < .001$; $U = 4.165$) and the teenagers ($p < .01$; $U = -3.042$). Each group follows a distinct strategy: for adults, low/falling accents – mainly $(H+)L^*$ – increase in the prepared presentation, and this is more evident within intonational phrases than within minor phrases, whereas the opposite is observed for teenagers, as in general there is a more balanced distribution of the main accent types used within minor phrases in the spontaneous presentation of teenagers.

Contrary to what would be expected based on the literature, high/rising accents – mainly $(L+)H^*$ – are overall less common in teenagers spontaneous presentations than in prepared ones (spontaneous: 29.8% high/rising at break 4, 53.7% at break 3; prepared: 41.5% high/rising at break 4, 67.3% at break 3). Adults differ from teenagers by using more high/rising accents in the spontaneous presentation, in both phrase levels (spontaneous: 40.3% high/rising at break 4, 71.8% at break 3; prepared: 27.1% high/rising at break 4, 64.9% at break 3).

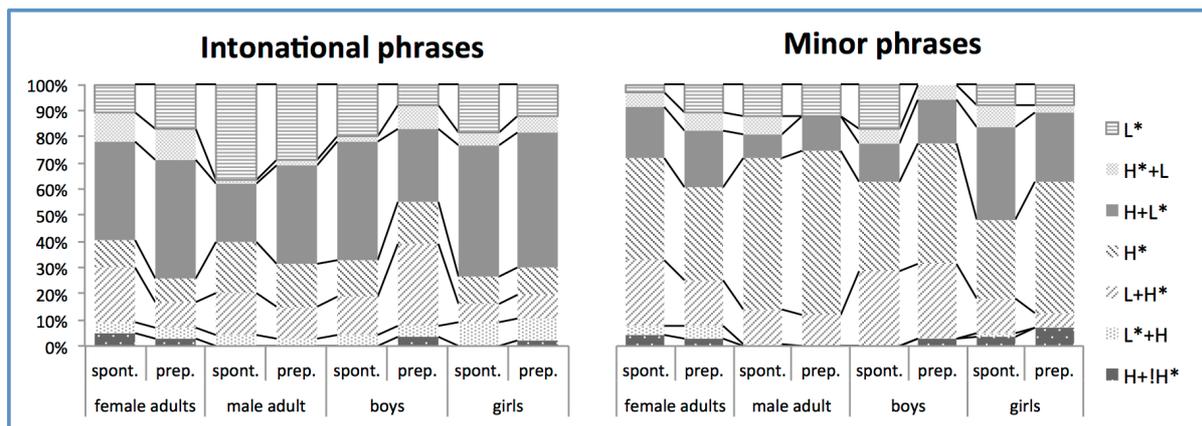


Figure 11. Distribution of nuclear accents by gender

When the data is filtered by gender (Figure 11), H+L* and L* are the most frequent nuclear accents of intonational phrases in the spontaneous presentations of teenagers, independently of gender (girls: 50% H+L*/ 18.6% L*; boys: 45.3% H+L*/ 20% L*). Although the frequency distribution of the nuclear accents most commonly used by each speaker varies, the data shows no statistically significant gender-specific patterns in the spontaneous speech of teenagers ($p = .380$; $U = .879$).

In general, the relative frequency of high/rising accents increases in the prepared presentations of teenagers (girls: 29.9% at break 4, 62.5% at break 3; boys: 55.1% at break 4, 77.2% at break 3), and this is more evident for boys in intonational phrases and for girls in minor phrases. In what concerns H* and L+H*, the most frequent high/rising accents, L+H* is used more often by boys within intonational phrases; girls display a more balanced distribution between H* and L+H* at this phrase level. In minor phrases, both genders use mostly H*; crucially boys use more L+H* than girls ($p < .01$; $U = -2.743$). Thus, the data indicates that the specificity of school presentations affects the distribution of nuclear patterns; it also offers evidence of inter-speaker differences in the adaptation process.

Comparing the female adults vs. the male adult spontaneous presentations, in general L* is used more often by the male and H+L* by the two female speakers ($p < .001$; $U =$

5.161). With regard to high/rising accents, in minor phrases both genders use mostly H*; crucially the two females use more L+H* than the male ($p < .05$; $U = 2.502$), suggesting that bitonal accents are more common for female adults, as pointed out by Clopper & Smiljanic (2011) in American English. In intonational phrases, L+H* is more frequent for females and H* for the male, but no statistically significant difference was found ($p = .100$; $U = 1.644$). In the prepared presentation, although adult speakers vary the balance between the most frequent accents, no statistically significant gender-specific patterns were found in the data (at break 3: $p = .061$, $U = 1.876$; at break 4: $p = .340$, $U = .953$).

3.2.3. Boundary configuration

Spontaneous vs. prepared differences are also evident in intonational phrase boundaries. Globally, adults increase (x)L%⁴ in the prepared presentation; teenagers increase (x)H% and tend to a more balanced distribution of (x)H% and (x)L% boundaries ($p < .05$; $U = 2.159$, -2.498 , respectively). In the spontaneous presentation adults differ from the majority of teenagers ($p < .01$; $U = 2.871$) by using more (x)H% boundary tones (spontaneous: adults 49.8%, teenagers 35.9% at break 4; prepared: adults 40.9%, teenagers 48.7% at break 4). This usage of contrasting strategies is consistent with the distributional patterns observed for nuclear accents. However, with regard to minor phrases, globally both groups tend to increase (x)H- in the prepared situation (spontaneous: adults 59%, teenagers 49.5% at break 3; prepared: adults 62.2%, teenagers 58.9% at break 3). This increase is more evident for teenagers than for adults, but differences do not reach statistical significance (teenagers: $p = .181$, $U = -1.336$; adults: $p = .527$, $U = -.633$). Interestingly, when the data is filtered by gender (Figure 12), one of the adults (the male speaker) uses H- even more often than teenagers in the prepared presentation. At this phrase level, distributional patterns are identical across the female adults' presentations.

⁴ See heading for Table 3

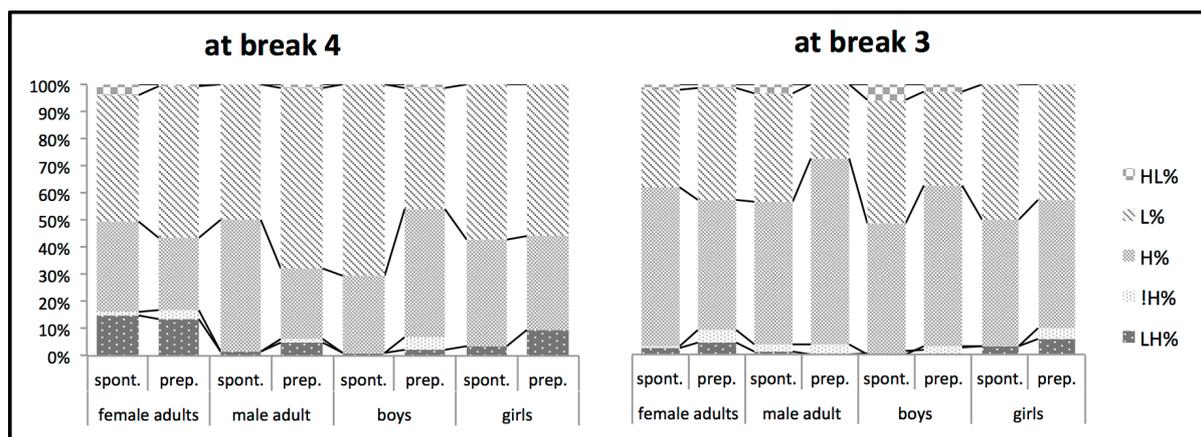


Figure 12. Distribution of edge tones by gender

L and H are the most frequent edge tones across phrase levels, independently of gender. !H and HL are residual in the data. In intonational phrases, H% and LH%, together, are used more often by females than by males. For girls, this is clearer in prepared presentations and for female adults in spontaneous presentations – $p < .05$, $U = -2.385$; $p < .01$, $U = -3.227$, respectively.

With regard to punctuation differences among final contour types, it is worth noting that in intonational phrases, commas are significantly correlated with a high boundary configuration and periods with a low boundary configuration ($\chi^2(4) = 116.805$, $p < .001$). In minor phrases, commas are largely independent of final contour type.

When fluent minor phrase breaks are compared with disfluent ones, the data offers additional evidence for an increase of H- in the prepared situation. We analyzed the boundary configuration for disfluent phrase breaks (ToBI label 3p) produced by teenagers and adults in spontaneous and prepared presentations (53.8% and 46.2%, respectively). Unlike fluent breaks, a low boundary is predominant in disfluent phrase breaks in both presentation types (74.3% in spontaneous; 58.3% in prepared). However, as observed in fluent breaks, a high

boundary is more frequent in teenagers and adults' prepared presentations than in spontaneous ones – $p < .05$; $U = -2.071$.

4. Discussion and conclusions

In this chapter we have presented a brief account of the differential characteristics of oral presentations at high school, showing that the distribution of low/falling and high/rising pitch accents and boundary tones varies in Lisbon Portuguese across styles (spontaneous vs. prepared) and across age (teenagers vs. adults) in a classroom context.

Focusing on final contours associated with declarative utterances, we have examined the distribution of nuclear accents, boundary configuration and phrase break levels in spontaneous and prepared (unscripted) presentations from adults and teenagers. Overall, both nuclear accent type and boundary configuration vary with phrase break level: final low/falling tones are more frequent with ToBI level 4, the predominant level of intonational phrasing across presentations; final high/rising tones are more frequent with break level 3, which encodes a minor level of intonational phrasing. This difference is consistent with findings on phrasing cues in spontaneous speech across languages (see, for example, Blaauw (1995), for “minor boundaries”; Benet, Lleó & Cortés (2011), for “intermediate phrases”). Furthermore, it is also particularly relevant for the analysis of intonational variation across types of speech. Considering this difference, the data shows that the frequency of low/falling vs. high/rising contours – mainly (H+)L* L and (L+)H* H – varies depending on the nature of the oral presentation, e.g. prepared vs. spontaneous.

Taking the definition of style proposed by Eskénazi (1993) as a point of departure, we interpret these changes as an adaptation effect to the speaking styles specifically required at school.

Style reflects the action of the environment upon the individual and the individual upon the environment. It is his perception of the various “status” levels of his listener and of the type of situation in which he finds himself. It is also a projection of himself, his background, and is a setting of the type and tone of conversation he wishes to have. All of this is a mixture of conscious and unconscious (voluntary/involuntary) effort on his part and is not always perceived as it was intended. (Eskénazi, 1993: 502).

Prepared oral presentations at school differ from spontaneous presentations of personal experiences in the degree of planning involved, the type of information communicated, the speakers’ attention to the speech task, and effort to organize the information that is given to the public in a clear and appealing way. Ultimately, these differences are interpretable as differences in intelligibility between styles, with regard to the dimensions of variation proposed by Eskénazi (1993). Additionally, the speaker’s status/age and gender are also a source of the observed variation.

In both spontaneous and prepared presentations, the balance of phrase levels is significantly affected by gender differences in each age group. The contrast between minor phrases and intonational phrases is stronger for the female adults than for the male adult and stronger for boys than for girls.

In the spontaneous presentations of adults (the teachers), minor phrases display a majority of high/rising contours; intonational phrases, as expected, a more balanced distribution of low/falling and high/rising contours. The differences between the two levels of phrasing are even stronger in their prepared presentations, due mainly to an increase of low/falling contours in intonational phrases. This variation of (x)H% vs. (x)L% may be associated with the more frequent expression of non-finality in spontaneous monologues, as suggested by Blaauw (1995), and in general with the interpretation of final highs and lows as a signal of continuation and finality, respectively.

Teenagers (the students) differ from adults by showing the exact opposite behavior: in the spontaneous presentations, intonational phrases display a majority of low/falling contours

while minor phrases show a more balanced distribution of low/falling and high/rising contours. Consistently with our expectations, teenagers maintain a distinction between intonational phrases and minor phrases in their prepared presentations. However, high/rising contours increase in both levels of phrasing in this type of situation – and this is even more evident for boys than for girls. Thus, considering final lows and highs as a suggestion of finality/non-finality, it appears that for teenagers non-finality is overall more frequent in prepared presentations and finality in spontaneous presentations – nevertheless, other prosodic cues to non-finality seem to be exploited as well, namely the variation of final word pitch slopes and standard deviations between phrase levels. On the other hand, the more regular use of non-low patterns is generally associated with a higher degree of expressiveness, and may be interpreted as a way to appeal to listeners' attention, a way to involve and to integrate listeners as participants.

Speculating about this result, the increase of final high/rising tones may be due to an approximation to a “model” of oral presentation common at school – a type of speech that can be organized in a more harmonious manner and can display greater regularity in contour type variation among levels of phrasing. It can be considered that, in typical school presentations, speakers try to adapt to a “standard” style that was learned, although it was not explicitly taught. This would explain the different strategies adopted by speakers in an attempt to reach an identical final effect.

The present work, focused on phrase-final intonation contours, points to intonational changes that can differentiate spontaneous and prepared (unscripted) speech from teenagers and adults. The results reported here as well as those we intend to report in the near future, including timing patterns, will also be of interest for the study of prosodic patterns in spontaneous speech and stylistic variation in EP, and ultimately for building prosodic models based on speech outside a laboratory setting. Stylistic variation in the intonation of

Portuguese, and of Ibero-Romance in general, has barely been explored hitherto, especially when variables such as age and gender are considered. By relating speech style differences across age groups and gender in a classroom context, this study provides a significant contribution to the development of knowledge on intonational variation.

Acknowledgments

This work was supported by FCT – Fundação para a Ciência e a Tecnologia – under project PTDC/CLE-LIN/120017/2010.

References

- Barry, W. (1995). Phonetics and phonology of speaking styles. In K. Elenius & P. Branderud (Eds.) *Proceedings of ICPhS 1995*, Stockholm, Sweden, vol.2, pp. 4-10.
- Batista, F., Moniz, H., Trancoso, I., Mamede, N., & Mata, A.I. (2012a). Extending automatic transcripts in a unified data representation towards a prosodic-based metadata annotation and evaluation. *Journal of Speech Sciences*, 2(2), 113-136.
- Batista, F., Moniz, H., Trancoso, I., & Mamede, N. (2012b). Bilingual Experiments on Automatic Recovery of Capitalization and Punctuation of Automatic Speech Transcripts. *IEEE Transactions on Audio, Speech, and Language Processing*, 20 (2), 474-485.
- Benet, A., Lleó, C. & Cortés, S. (2011). Phrase boundary distribution in Catalan: Applying the prosodic hierarchy to spontaneous speech. In G. Christoph and C. Lleó (Eds.), *Intonational phrasing in Romance and Germanic: Cross-linguistic and bilingual studies*, (pp. 97-126). Amsterdam/Philadelphia: John Benjamins.
- Blaauw, E. (1995). *On the perceptual classification of spontaneous and read speech*. Utrecht: LEd.
- Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.56, retrieved 15 September 2013 from <http://www.praat.org/>
- Cabarrão, V. & Mata, A. I. (2012). Prosodic and pragmatic properties of affirmative words in European Portuguese. In A. Botinis (Ed.) *Proceedings of the 5th ISEL Conference on Experimental Linguistics, ExLing 2012*, Athens, Greece, pp. 33-36.
- Clopper, C. G. & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics* 39, 237–245.
- Duarte, I. (2000). *Língua Portuguesa. Instrumentos de Análise*. Lisboa: Universidade Aberta.
- Eckert, P. & S. McConnell-Ginet. 2003. *Language and Gender*. Cambridge: Cambridge University Press.
- Escudero, D., Aguilar, L., Vanrell, M. del Mar, & Prieto, P. (2012). Analysis of inter-transcriber consistency in the Cat_ToBI prosodic labeling system. *Speech Communication*, 54 (4), 566-582.

- Eskénazi, M. (1993). Trends in Speaking Styles Research. In *Proceedings of Eurospeech '93*, Berlin, pp. 501-505.
- Fagyal, Z. & Stewart, C. M. (2011). Prosodic style-shifting in preadolescent peer-group interactions in a working-class suburb of Paris. In F. Kern & M. Selting (Eds.), *Ethnic styles of speaking in European metropolitan areas*, (pp. 75-99). Amsterdam/Philadelphia: John Benjamins.
- Falé, I. (1995). *Fragmento da prosódia do Português Europeu: as estruturas coordenadas*. Master's thesis, University of Lisbon.
- Falé, I. (2005). *Percepção e reconhecimento da informação entoacional em Português Europeu* (Doctoral dissertation). University of Lisbon, Lisbon, Portugal.
- Falé, I. & Faria, I. (2006). Categorical perception of intonational contrasts in European Portuguese. In R. Hoffmann, H. Mixdorff & O. Jokisch, (Eds.) *Proceedings of Speech Prosody 2006*, Dresden, Germany, pp. 69-72.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76 (5), 378-382.
- Fletcher, J., Grabe, E. & Warren, P. (2005). Intonational variation in four dialects of English: the high rising tone. In Sun-Ah Jun (Ed.) *Prosodic typology: The phonology of intonation and phrasing*, (pp. 390-409). Oxford: Oxford University Press.
- Frota, S. (2000). *Prosody and focus in European Portuguese. Phonological phrasing and intonation*. New York: Garland Publishing.
- Frota, S. (2002). Nuclear falls and rises in European Portuguese: a phonological analysis of declarative and question intonation. *Probus*, 14, 113-146.
- Frota, S. (2014). The intonational phonology of European Portuguese. In Sun-Ah Jun (Ed.), *Prosodic typology II*, (pp. 6-42). Oxford: Oxford University Press.
- Guy, G. & Vonwiller, J. (1989). The high rising tone in Australian English. In P. Collins & D. Blair (Eds.), *Australian English: The language of a new society*, (pp. 21-34). St. Lucia: University of Queensland Press.
- Henriksen, N. (2013). Style, prosodic variation, and the social meaning of intonation. *Journal of the International Phonetic Association*, 43, 153-193.
- Hirschberg, J. (2000). A corpus-based approach to the study of speaking style". In M. Horne (Ed.), *Prosody: theory and experiment, Festschrift in honor of Gosta Bruce*, (pp. 335-350). The Netherlands: Kluwer.
- Jiang, H. (2011). Gender difference in English intonation. In W.-S. Lee & E. Zee (Eds.) *Proceedings of ICPHS 2011*, Hong Kong, pp. 974-977.
- Labov, W. (1976). *Sociolinguistique*. Paris: Minuit.
- Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Linneman, T. J. (2013). Intonation variation on a television game show. *Gender & Society*, 27(1), 82-105.

- Liu, Y., Shriberg, E., Stolcke, A., Hillard, D., Ostendorf, M., & Harper, M. (2006). Enriching speech recognition with automatic detection of sentence boundaries and disfluencies. *IEEE Transaction on Audio, Speech and Language Processing*, 14(5), 1526–1540.
- Mata, A. I. (1999). Para o estudo da entoação em fala espontânea e preparada no Português Europeu: metodologia, resultados e implicações didáticas (doctoral dissertation). University of Lisbon, Lisbon, Portugal.
- Mata, A. I. (2012). Intonational differences between speech tasks in school context. In A. Botinis (Ed.) *Proceedings of the 5th ISEL Conference on Experimental Linguistics, ExLing 2012*, Athens, Greece, pp. 81-84.
- Mata, A. I. & Santos, A. L. (2010). On the intonation of confirmation seeking requests in child-directed speech”. In M. Hasegawa-Johnson (Ed.) *Proceedings of Speech Prosody 2010*, Chicago, U.S.A..
- Mata, A. I., Moniz, H., Batista, F. & Hirschberg, J. (2014). Teenage and adult speech in school context: building and processing a corpus of European Portuguese. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk & S. Piperidis (Eds.) *Proceedings of 9th International Conference on Language Resources and Evaluation, LREC 2014*, Reykjavik, Iceland, pp. 3914-3919.
- Moniz, H. (2013). *Processing Disfluencies in European Portuguese* (doctoral dissertation). University of Lisbon, Lisbon, Portugal.
- Moniz, H., Trancoso, I., Mata, A. I. (2010). Disfluencies and the perspective of prosodic fluency. In A. Esposito, N. Campbell, C. Vogel, A. Hussain & A. Nijholt (Eds.), *Development of multimodal interfaces: active listening and synchrony*, (pp. 382-396). Berlin / Heidelberg: Springer.
- Moniz, H., Batista, F., Trancoso, I. & Mata, A.I. (2013). Automatic structural metadata identification based on multilayer prosodic information. In R. Eklund (Ed.) *Proceedings of DISS 2013*, Stockholm, Sweden, pp. 49-52.
- Moniz, H., Mata, A. I., Hirschberg, J., Batista, F., Rosenberg, A. & Trancoso, I. (2014). Extending AuToBI to prominence detection in European Portuguese. In N. Campbell, D. Gibbon & D. Hirst (Eds.) *Proc. of Speech Prosody 2014*, Dublin, Ireland, pp. 280-284.
- Neto, J., Meinedo, H., Viveiros, M., Cassaca, R., Martins C., & Caseiro, D. (2008). Broadcast news subtitling system in Portuguese. In A. Sayed (Ed.) *Proceedings of ICASSP '08*, pp. 1561–1564.
- Ostendorf, M., Favre, B., Grishman, R., Hakkani-Tür, D., Harper, M., Hillard, D., Hirschberg, J., Ji, H., Kahn, J. G., Liu, Y., Maskey, S., Matusov, E., Ney, H., Rosenberg, A., Shriberg, E., Wang, W., & Wooters, C. (2008). Speech segmentation and spoken document processing. *IEEE Signal Processing Magazine*, 25(3), 59–69.
- Pierrehumbert, J. & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan & M. E. Pollack (Eds.), *Intentions in Communication*, (pp 271-311). Cambridge, MA: MIT Press.
- Prieto, P. (2014). The intonational phonology of Catalan. In S. Jun (Ed.), *Prosodic typology II*, (pp. 43-80). Oxford: Oxford University Press.
- Prieto, P., Payà, M. & Vanrell, M. M. (2014). Estils prosòdics en el llenguatge publicitari”. In D. Casals & J. Cornudella (Eds.), *Publicitat en català. Estudis de llenguatge*

- publicitari* (Zeitschrift für Katalanistik, special volume 12), (pp. 5-22). Aachen: Shaker.
- Serra, C. R. (2009). *Realização e percepção de fronteiras prosódicas no Português do Brasil: fala espontânea e leitura* (doctoral dissertation). Rio de Janeiro, Universidade Federal do Rio de Janeiro.
- Shepherd, M. A. (2011). Functional significance of rising-intonation declaratives in settings with special discursive norms. *LSA Annual Meeting Extended Abstracts 2011*. Pittsburgh.
- Silverman, K., E. Blaauw, J. Spitz & J. F. Pitrelli. (1992a). A prosodic comparison of spontaneous speech and read speech. In J. J. Ohala, T. M. Nearey, B. L. Derwing, M. Hodge & G. E. Wiebe (Eds.) *Proceedings of ICSLP 92*, Banff, Canada, vol. 2, pp. 1299-1302.
- Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert and Hirschberg, J. (1992b). "TOBI: a standard for labeling English prosody". In J. J. Ohala, T. M. Nearey, B. L. Derwing, M. Hodge & G. E. Wiebe (Eds.) *Proceedings of ICSLP 92*, Banff, Canada, vol. 2, pp. 867-870.
- Sjölander, K., Beskow, J., Gustafson, J., Lewin, E., Carlson, R., and Granström, B. (1998). Webbased educational tools for speech technology. In R. H. Mannell & J. Robert-Ribes (Eds.) *Proceedings of ICSLP'98*, Sydney, Australia, pp. 3217-3220.
- Swerts, M. (1994). *Prosodic features of discourse units* (doctoral dissertation). Eindhoven, University of Eindhoven.
- Trask, R. L. (2010). *Why do languages change?* Cambridge: Cambridge University Press.
- Viana, C., Frota, S., Falé, I., Fernandes, F., Mascarenhas, I., Mata, A. I., Moniz, H. & Vigário, M. (2007). Towards a P_ToBI. Paper presented at the Workshop on the Transcription of Intonation in Ibero-Romance. University of Minho, Portugal.
- Wennerstrom, A. (2001). *The music of everyday speech: prosody and discourse analysis*. New York: Oxford University Press.