



Spontaneous Speech Segmentation: Functional and Prosodic Aspects with Applications for Automatic Segmentation

A segmentação da fala espontânea: aspectos prosódicos, funcionais e aplicações para a tecnologia

Plínio A. Barbosa

Universidade Estadual de Campinas, Campinas, São Paulo / Brazil

pabarbosa.unicampbr@gmail.com

Tommaso Raso

Universidade Federal de Minas Gerais, Belo Horizonte. Minas Gerais /Brazil

tommaso.raso@gmail.com

This issue of *Revista de Estudos da Linguagem* is dedicated to a theme addressed in several other initiatives promoted by its guest editors, along with colleagues from the international community. The theme, which in recent years has played an increasingly important role in the disciplines that study speech production and perception, is the segmentation of speech into smaller units addressed from both formal and functional perspectives, fundamentally under a theoretical approach coupled with an empirical focus. Among the main initiatives, we mention:

- Two international workshops (IV Leel and X Lablita International workshop *Unit of Reference for Spontaneous Speech and their Correlation Across Language*, held in August 2015 at UFMG; and the workshop *Spoken Corpora advances: prosody as the crux of speech segmentation, annotation and multilevel linguistic studies*, organized in June 2018 at Cape Town as part of the 20th International Congress of Linguists activities);

- The book *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*, to be released soon by John Benjamins;
- A special issue of the Journal of Speech Sciences scheduled to come out in mid-2019.

All of these initiatives are dedicated to the prosodic segmentation of speech, a subject that has become increasingly central to understanding speech structuring at various levels, as well as the relationship of this structuring with the communicative functions of language. The disciplines interested in the subject, and Linguistics *in primis*, have evolved enormously from the contribution of technological advances and statistics applied to linguistic studies, and from the contribution of the advances of linguistic theories themselves. In fact, until recently, the study of speech segmentation considered almost exclusively the segmentation of the so-called lab speech. This includes read speech and speech elicited in various forms (XU, 2010) from the manipulation of external events (such as by proposing tasks with one or more participants such as map task and electronic games, conducting interviews on specific topics, *inter alia*). A few years ago, however, it became possible to approach good-quality, recorded non-scripted speech extracted from spontaneous speech corpora in varied natural communicative situations. In this introductory article to this thematic issue of RELIN, we present a partial overview of the scientific issues at stake, the results achieved so far, and the steps already announced for the future.

1. Prosodic segmentation: between form and function

Contrary to writing, which is a product that can be preserved in time and space, speech is a process whose result disappears shortly after its manifestation, if we set aside in this examination the current recording technologies. Only some cognitive consequences of discourse remain, but not speech itself (LINELL, 2005; BLANCHE-BENVENISTE; JEANJEAN, 1987). Absent from writing in its acoustic manifestation, except for mere indications inferred from punctuation marks, prosody is the essential component for speech segmentation studies. It is now possible, thanks to technology and dedicated software, to reproduce speech for as many times as necessary and to annotate the speech chain into different units by procedures of labelling and segmentation: syllables,

groups of syllables or words, prosodic units of different dimensions and theoretical status, as well as utterance sequences. This allows the systematic observation and measurement of many aspects of speech that, without technology, had to some extent only been intuited through the auditory sensitivity of the precursors of contemporary prosody research (see PIKE, 1945; LIEBERMAN, 1960; BOLINGER, 1965) without the possibility of being deepened or demonstrated. Among these aspects, a place of crucial importance is occupied by the different units in which it is possible to segment the flow of speech and by the development of a current of thought on its forms and functions. Finally, it has become possible to attempt the reconstruction of the complex prosodic structure (and not only) of human speech.

In addition, technology has made it possible to compile and investigate large amounts of speech data, treated and annotated in different ways and specifically suited to several research fronts, in a line with the view that privileges the acquisition of knowledge from huge corpora (cf. the concept of “big data” in FURHT; VILLANUSTRE, 2016). The automatic processing of the acoustic signal allows us to segment discourse into smaller units, from the utterance (or perhaps from larger units like the “paragraphs”) to the syllable and its constituents; furthermore, it allows us to investigate how human speech conveys boundaries (or their absence) at different hierarchical levels.

Depending on the interest of the study, the speech chain can be segmented into units of different sizes and types, conveying their own properties and delimited by some type of boundary. For the sake of exemplification, let us only look at the units above the word level. We can divide the speech chain into stress groups (or n-ary feet, groups of syllables up to a stressed syllable, in the case of right-hand languages), into prosodic units called intonational or tonal or prosodic groups, in sentences, or, under a syntactic perspective, in intonational phrases (IP), intermediate phrases (ip) and sentences. Each type of segmentation is directly or indirectly associated with a theoretical view, but in many cases this does not preclude an empirical investigation whose results can be analyzed in the light of different theoretical perspectives. In recent years, several corpora with prosodic annotation of the boundary have been compiled for different languages (AURAN *et al.*, 2004; DU BOIS *et al.*, 2000-2005; OSTENDORF *et al.*, 1996; CRESTI; MONEGLIA, 2005; SCHUURMAN *et al.*, 2003; IZRE’EL, 2002; RASO; MELLO, 2012; Forthcoming; METTOUCHI *et al.*, 2010; GAROFOLO *et al.*, 1993).

Any kind of segmentation implies the presence of a boundary, either actually perceived or theoretically proposed. Thus, the boundary can be understood as a physically perceived rupture, it may refer to a testable limit for the realization of linguistic phenomena, and it may further be considered as a region between two units, a region that can be auditorily perceived or not.

This thematic number seeks to study the segmentation of what can be considered as the reference unit of the speech process (IZRE'EL *et al.*, Forthcoming). The very notion of reference unit can be understood in different ways, but we can provisionally define it as a minimal unit of complete and autonomous communicative meaning that composes a spoken text (CRESTI, 2000; MONEGLIA; RASO, 2014). This definition can be challenged, but it allows us to have a point of departure.

All the aforementioned types of units, regardless of how they are defined, are separated by boundaries that are defined by highlighting greater or lesser perceptual or theoretical grounds, since hardly one of these two criteria completely excludes the other. In the articles in this thematic issue, a perceptual basis is always present, but some papers assign a greater weight to theoretical aspects, and these aspects vary from one article to another. With these differences of perspective, the concept of boundary changes as well.

Of theoretical nature are the boundaries of constituents in syntactic and informational approaches. This does not mean that they cannot be associated with prosodic boundaries, which constitute the primary interest of this thematic number. In fact, we understand that prosody guides syntactic interpretation, as in cases such as the sentence *A ovelha de raça brasileira* (The sheep of Brazilian breed; word-by-word: The-sheep-of-race-Brazilian). From this unit of writing, two utterances can be uttered in two distinct forms of grouping, where “/” represents a strong non-terminal boundary:

[A ovelha de raça] / [brasileira] vs. [A ovelha] / [de raça brasileira]

In the first case, it is a sheep born in Brazil from a non-informed breed and, in the second case, a sheep from a breed developed in Brazil. It is precisely the prosodic constituents that allow the proper scrutiny of the syntactic structure of each utterance. That is, prosody allows for disambiguation between the two possible interpretations, since the limited resources of writing do not allow deciding between the two possible

interpretations. In this example, the appropriate prosodic structure guides a single syntactic interpretation with syntactic and prosodic constituents being congruent, that is, having the same limits. Because of the prevalence of prosody, the authors of this thematic number who deal directly with the issue of speech segmentation take prosodic constituents as the only appropriate units related to the speech chain.

Furthermore, almost all contributions of this issue assume the organization of speech in units that can be considered extensive to intonational units. When we use the expression “intonational unit” in this panorama, however, we mean not only a unit organized by patterns of fundamental frequency (f_0), but also by patterns of duration and possibly voice quality. A single work (that of Ph. Martin) segments speech into accent phrases, which does not exclude the fact that a single accent phrase or a set of accent phrases coincide with an intonational unit. The segmentation in accent phrases can, therefore, be seen as an opportunity to investigate the internal structure of the intonational unit, thus enriching, and not contradicting, the perspectives that prefer to focus on the analysis of the intonational unit.

It is difficult to define the intonational unit without reference to perception or to a postulate of a theoretical nature. In general, the intonational unit is defined as a group of words (it can also be a single word and, in rare cases, where the emphasis on syllables comes into play, less than a word. In the latter case, the boundary is a perceptual consequence of the prominence of the unit) delimited between a prosodic boundary and the immediate subsequent boundary. The unit is characterized by a coherent f_0 contour separated both physically and perceptually from the preceding and following contours (DU BOIS *et al.*, 1992, p. 17; CRUTTENDEN, 1997). This definition masks some difficulties in capturing the properties of an intonational unit without reference to its boundaries, and, on the other hand, without identifying the boundary independently of the concept of intonational unit, there is a clear risk of circularity. The very definition of “coherent contour” is not completely satisfactory since we do not know clearly which parameters favour or break coherence.

From a functional point of view, the intonational unit can be studied and linguistically defined based on different perspectives. The main ones are the syntactic perspective, the informational perspective (CHAFE, 1994; RASO; MELLO, 2014) and the conversational

perspective (BARTH-WEINGARTEN, 2016). However, the very individualization of the intonational unit is problematic. In fact, the recognition of a coherent prosodic profile or a prosodic boundary is not always obvious. As regards the identification of a boundary, studies are usually based on the statistical agreement between annotators. In this kind of task, a certain chunk of speech is segmented into smaller units by a set of annotators. The agreement between them is used to identify a particular kind of boundary. Other approaches consider the perception of a boundary as associated to a particular f_0 movement visible by using a dedicated software, such as the so-called boundary tone, a movement of f_0 aligned to the end of the unit, in the framework of the Autosegmental-Metrical Theory (LADD, 1996; PIERREHUMBERT, 1980).

Statistical tests of inter-rater reliability show that the agreement among annotators for the identification of boundaries, and consequently of units, is very high (more than 80%, especially in the case of the terminal boundaries; MELLO *et al.*, 2012; MONEGLIA *et al.*, 2005; YOON *et al.*, 2004; BUHMANN *et al.*, 2002). It is therefore consensual that the intonational unit constitutes an important level of speech organization, although the reasons for this organization remain controversial. According to some authors, this segmentation of the speech chain is due to the limits of memory (cf. COWAN, 1998), which impose groupings of a limited number of syllables for linguistic processing. According to others, the units would have cognitive motivations (CHAFE, 1994; CROFT, 1995; BYBEE, 2010). As for yet a third view, the segmentation corresponds to units of a syntactic nature and therefore prosodic boundaries and syntactic boundaries would be correlated, especially in the phonological approaches of prosody that presuppose a mapping between syntactic constituents and the limits of prosodic units (NESPOR; VOGEL, 1986; SELKIRK, 1995). A fourth proposal, dominant in this thematic issue, attributes to the prosodic boundary the role of delimiting units of informational nature, independently of its syntactic organization. Others still see a correspondence between prosody and units of another discursive domain (COUPER-KUHLEN, 2004; SCHEGLOFF, 1998). Those who study prosody as correlated to linguistic domains of a non-syntactic nature also tend to consider prosody as a structural element implemented before the segmental elements (see the Frame/Content theory by MacNEILAGE, 1998). An interesting view within prosodic studies (HIRST; DI CRISTO, 1998; BARBOSA, 2006) attempts a

compromise between syntactic and prosodic constituents by proposing that the syntactic structure imposes some restrictions, but would not determine the position of the realized boundaries. In this proposal, the prosodic boundaries would only appear in positions compatible with the syntactic structuring without necessarily establishing constituents of this nature. After all, given a certain sentence, there are several positions compatible with the syntactic structuring where a boundary could be placed, with each position signalling a different cognitive-informational interpretation. On the other hand, many syntacticians have realized how prosody is essential for explaining particular structures that resist simple explanations in the framework of traditional syntactic theories. This is the case for the so-called insubordination phenomenon (EVANS; WATANABE, 2016, BOSSAGLIA *et al.*, Forthcoming). In such cases, the interpretability of the structure depends decisively on its prosodic coding.

2. The main theoretical questions

Previous research has also shown that the study of prosodic boundaries depends on speaking style and partially on the typology of the spoken text as well. In fact, until recently, research had focused on the study of prosodic segmentation in read texts or limited sequences performed in laboratory with interesting results, but that does not seem to be comparable with what happens in spontaneous speech, a priority objective of this issue. In prosody studies linked to syntax and phonology, laboratory speech is often used to test relations between prosody and syntax (as in the case of disambiguation and in the investigation of the relation between prosodic and phonological/syntactic constituents delimited by theoretical boundaries). Read texts present a much smaller number of variables than spontaneous speech, in addition to greater predictability (PRICE *et al.*, 1991). What is more, read speech is the sonorous realization of a written text, therefore being structured based on principles distinct from those of spontaneous speech.

Recently, some works on spontaneous speech have obtained promising results in the investigation of segmentation mechanisms. This has been done either by observing a high agreement (greater than 80%) among human annotators (MELLO *et al.*, 2012; MONEGLIA *et al.*, 2005; TEIXEIRA FALCÃO, 2017) or by developing software able to segment spontaneous speech automatically, achieving results that are

highly comparable with the tasks performed by humans (AVANZI *et al.*, 2008; NI *et al.*, 2012; MITTMAN; BARBOSA, 2016).

The development of software capable of automating prosodic segmentation in intonation units (cf. MITTMAN; BARBOSA, 2016) is only possible because the investigation of the acoustic parameters responsible for boundary perception has greatly advanced, thanks to the work done with read speech and speech sequences performed in the laboratory, which allowed a first understanding of the highly complex phenomena at play. From that, it came up that the parameters responsible for our perception boundaries are diverse; they are not always all co-present; their weight may vary depending on the languages and the circumstances of a particular speech style. This leads to the question of whether it is possible to speak of boundaries as a homogenous category at all, and points in the direction in favour of speaking of different types of boundaries.

In the literature, the parameters that are most mentioned as boundaries markers are fundamental frequency (f_0), duration and intensity, as well as parameters that refer to voice quality (BARTH-WEINGARTEN, 2016; MO *et al.*, 2008; WAGNER; WATSON, 2010), especially creaky voice (DILLEY *et al.* 1996; GORDON; LADEFOGED. 2001; REDI; SHATTUCK-HUFNAGEL, 2001; HANSON *et al.*, 2001; CARLSON *et al.*, 2005). From them, the main boundary cues that emerge are: the silent pause, which we will simply call “pause” (later on we will discuss the role of the filled pause), whose presence automatically seems to convey the perception of a boundary (MARTIN, 1973; SWERTS, 1997; SHRIBERG *et al.* 2000; TSENG; CHANG 2008; MO; COLE 2010; TYLER, 2013); the lengthening of the final syllables of the unit, that is, a decreasing of speech rate during the last syllables before a boundary (WIGHTMAN *et al.*, 1992; BARBOSA, 2008; MO *et al.*, 2008; FUCHS *et al.*, 2010; FON *et al.*, 2011; TYLER, 2013); the shortening of the first syllables of the unit, that is, speech rate increases just after a boundary (AMIR *et al.* 2004; TYLER, 2013), correlated with phenomena of anacrusis; the reset of the f_0 curve; the abrupt change of direction of the f_0 curve; the change of intensity at the beginning of the prosodic unit (SWERTS *et al.*, 1994; TSENG; FU, 2005; MO, 2008); creaky voice and perhaps other non-modal voice qualities. To these parameters, at least for some languages, some phenomena of a segmental nature must be added. For example, for English, final stop release and creakiness or glottal closure in the vicinity of final segments may be cues of a boundary.

Each of these cues brings some issues for the researcher. For example, the pause, which intuitively seems an obvious notion, is not identified consensually: what is the minimum amount of silence considered as a pause? How does the presence of a pause affect the other parameters that contribute to boundary perception? Is the pause a clue of boundary type or not? As for the f_0 curve, what is the relative contribution of f_0 level difference, f_0 excursion, the direction of f_0 movement, and of f_0 variation rate? When considering syllabic duration, what is the extent of the region affected by the boundary, measured in number of syllables? Additionally, if the change in duration involves more than the syllable just before and after the boundary, does the change occur in the same proportion for each syllable involved or not? Furthermore, previous experimental work has shown that, in order to reliably evaluate duration measures, some form of normalization that sets aside the intrinsic properties of the segments is necessary, which, in this case, decisively influences the duration (BARBOSA, 2012). It should also be noted that the measure of duration appropriate for prosodic analysis should consider phonological and phonetic syllables. The former is important for the perception of speech, because it involves syllable perception through the cognitive system, while the latter is the basis for the production of the speech chain and the structural organization of the corresponding consonants and vowels.

Research on the acoustic parameters that, together, convey the perception of a boundary should consider the weight or relative contribution of each acoustic cue. For this, it is important to consider not only that each cue is perceptible only if it surpasses a certain threshold, but that this threshold varies by varying the other cues (t'HART *et al.*, 1990). This means, first, that we are not able to perceive just any change in f_0 or any change in duration or intensity, but only changes that exceed a certain threshold. Although for each parameter or cue in isolation we can know its *Just Noticeable Difference* (JND), that is, the minimum variation of this parameter that we can perceive (see HUGGINS, 1972; KLATT; COOPER, 1975 for segmental duration, t'HART, 1981; and RIETVELD; GUSSENHOVEN, 1985, for f_0 as well as KOFFI, 2018, for intensity), as well as the way in which the JND varies with the modification of another parameter (for example how we perceive intensity variation at different frequencies), we do not know yet how these complex combinations of parameters vary with respect to the ability to convey boundary perception.

It is not simple to model the boundary phenomenon given the possibility of combining so many parameters in the speech flow. In fact, it would not be surprising if the weight of a cue changes by changing the combinations of the other cues, or by changing speaking style - reading or spontaneous conversation, or other styles of spontaneous speech, or different linguistic functions of the units delimited by the boundaries, without considering variations related to the characteristics of the speakers.

In fact, the studies in different languages confirm the importance of the aforementioned cues for the perception of a boundary, while revealing that each one of these cues acts with a distinct weight to mark this same boundary (TEIXEIRA FALCÃO, 2017). This varying hierarchy of acoustic cues seems to be linked to the functions that a certain parameter has in the language. For example, in tonal languages, f_0 has the role of conveying linguistic functions that in non-tonal languages are conveyed by other parameters. In these languages, f_0 differences implement tone distinctions that serve to contrast lexical items. In addition, the weight of f_0 is affected when this parameter is used to mark the boundary, with duration and f_0 reset being the most relevant parameters for signalling boundaries (YANG; WANG, 2002). This is likely to be the case with other parameters, which would behave differently to signal the prosodic boundary depending on how important they are to convey other functions in a given language. Very little is known about how the weight of a given parameter changes within a large combination of other parameters for marking boundaries of functionally different units.

While some studies focus on investigating the opposition between presence vs. absence of a boundary (MO *et al.*, 2008; BARBOSA, 2010), other studies investigate a potential diversity among the boundaries. In the latter case, some authors propose the existence of a certain number of boundaries, while others propose a continuum between presence and absence of boundaries. In this second case, there is always a risk of finding some degree of boundary, no matter how small, and losing the boundary vs. non-boundary contrast, making any consideration of a functional nature attributable to a boundary extremely difficult, if not impossible.

On the other hand, the researchers who consider that the boundary is a gradient phenomenon, although categorical, propose a gradation of strength for the different boundaries, which occur in a limited number. Among these authors there is disagreement about the amount of different strengths that can be recognized and perceived (see BARBOSA, 2006,

for a discussion). Some studies distinguish between strong and weak boundaries, while others consider it possible to individualize more than two degrees of strength (see WIGHTMAN *et al.*, 1992, for English, BARBOSA, 2006, for Brazilian Portuguese, and BARBOSA, 1994, for French) with some of them reaching up to seven degrees, which is in line with the phonological theories for prosody such as those by Nespor and Vogel (1986) and Selkirk (1995).

Another possibility to infer degrees of boundary strength is the use of local maxima of the acoustic parameters that convey a prosodic boundary as indices of the strength of this boundary (TEIXEIRA FALCÃO, 2017). Even if local maxima vary continuously, it is possible to use clustering techniques to infer a limited number of boundary strengths that do not exceed four (see BARBOSA, 2006, for BP, and BARBOSA, 1994, for French). In the work for BP, Barbosa (2006) used z-score-normalized syllable duration maxima to obtain 3 to 4 distinct levels, partially correlated with syntactic boundaries obtained by the projection of a dependency tree in line with Tesnière's (1965). The different degrees of strength allowed establishing a hierarchy of prosodic constituents that open the possibility of inferring the prosodic structure of an utterance. This procedure had already been proposed by Grosjean and colleagues (GROSJEAN; GROSJEAN; LANE, 1979; GROSJEAN; DOMMERGUES, 1983; GEE; GROSJEAN, 1983) by asking people to read at increasingly slow rates and subsequently analysing vowel durations associated with silent pauses when applicable and from segmentation indices for utterances obtained from perception tests. This procedure reveals what they called a *structure de performance*, a prosodic structure with the following properties: constituents of similar size, hierarchical organization and symmetric structure (GROSJEAN; DOMMERGUES, 1983). These properties emerged from two competing constraints: the speaker's tendency to respect the linguistic structure of the sentence and the tendency to balance the extension of the constituents it produces (MONNIN; GROSJEAN, 1993, p. 28; MARTIN, 1987). The tendency to equilibrate the extension of prosodic constituents would explain why subjects do not systematically group the verb with the object noun phrase when pronouncing English phrases, as would be predicted by syntax, but prefer groupings of type (SV)O (GROSJEAN; GROSJEAN; LANE 1979, p. 59).

The discussion about boundary types, however, is not just quantitative in nature. Many authors distinguish between boundaries that convey perception of prosodic and linguistic completion (with distinct interpretations of the nature of the completed linguistic unit) and boundaries that convey the perception of discourse continuity. The latter signals that the discourse segment in progress cannot be considered complete even if the boundary signals the end of a constituent, this one having distinct types, depending on the theoretical approach (MONEGLIA; CRESTI 1997; CRYSTAL, 1969; SWERTS, 1994; SWERTS *et al.*, 1994). For several authors these two types of boundaries are called terminal and nonterminal, respectively.

But some authors who consider the distinction between terminal and non-terminal boundaries argue for a fine-grained difference. For these authors, there would not be a single type of terminal boundary, nor a single type of non-terminal boundary. According to this proposal, we can observe that some terminal boundaries are “more terminal” than others. For example, the boundaries of utterances would be less terminal compared to the boundary between larger discursive blocks, called paragraphs by some (van DONZEL 1999). Similarly, there would be several types of non-terminal boundaries, some more prominent than others, or perceptually closer to the terminal boundaries, or announcing the fact that the conclusion is close. These proposals should not be considered as mutually exclusive, since they are able of capturing different aspects of the complexity of the phenomenon (SWERTS *et al.*, 1994; TEIXEIRA FALCÃO 2017).

In fact, if we examine the phonetic-acoustic parameters correlated to boundary perception, in particular the non-terminal boundary, we observe varied combinations within the same language and text (see TEIXEIRA FALCÃO, 2017). We have, for example, boundaries clearly marked by a movement of increasing f_0 , an acoustic cue of continuity, which, along with other prosodic cues like duration, conveys the perception that the discourse will continue. On the other hand, this increasing movement of f_0 or final lengthening may be lacking in other boundaries that are also perceived as non-terminal (WAGNER, 2010).

As for conclusive boundaries, it is often observed that they are characterized by a downward movement of f_0 to the lowest level, followed by a reset of f_0 at the beginning of the next unit, which would start with an f_0 value at a clearly-defined distinct height. However, it

is commonly recognized that not all utterances conclude with a low f_0 value. Although the most obvious and studied case is that of the yes/no questions in languages such as English and Peninsular Spanish, there are other illocutions, according to the terminology and categorization we adopt, which are marked, among other parameters, by a higher f_0 at the end (CRESTI 2000; Forthcoming; MORAES; RILLIARD, 2014 *inter alia*).

The variability of the physical realization of the boundaries can be correlated with different functional values on the linguistic plane. We would then have not only a correlation between types of boundaries conveying completion and types of boundaries conveying continuation, but also between different conclusive types, in the case of different illocutions, and between different non-conclusive types, which, by hypothesis, would mark different constituent types (syntactic or other kinds). In this perspective, the specific realization of a prosodic boundary would not only have a demarcating value, but would depend heavily on the linguistic function of the unit delimited by the boundaries, the associated cues would also point to these same linguistic functions.

Thus, in this perspective, studying how boundaries are physically realized and studying the nature of the units delimited by these very boundaries (one on the left and the other on the right) would no longer belong to distinct scopes. The former having been of a prior interest to Phonetics and the latter to those who are interested in higher linguistic levels or in cognitive mechanisms would, therefore, become much more integrated. The perspective that unites the functions of the units to the concrete manifestation of the boundaries that delimit them is still incipient and can give us interesting answers about the nature of the units that are delimited by these boundaries.

Before moving on to the different theoretical approaches to units, it is worth making an observation about some kinds of boundaries (and units) that are much less frequent in laboratory speech, at least in the case of read speech, but which are extremely common in spontaneous speech: the different types of disfluencies. In spontaneous speech, the phenomena of interruption, retractings and hesitation are very frequent. Many units come to an end not because the speaker planned their completion, but because some unforeseen internal (improper word retrieval, change of mind, or any problem in the articulation or elaboration of content) or external cause (interruption by another speaker or any environmental

event) leads to the momentary interruption of the utterance before it is completed semantically and prosodically. As for retraction, the statement is not interrupted, but is fragmented by repetitions of words or parts of words, which the speaker then ideally cancels and corrects, continuing to produce the utterance as if they had not been pronounced. This is the result of difficulties in the realization of the utterance that do not lead to interruption of the statement and are more or less present in all speakers, but especially in those who have less mastery of speech, or because they are very young, or because they are from a lower diastrophic category, or for other reasons. In the case of hesitation, difficulties in speech are manifested under different guises, such as vowel stretching or time taking by producing filled pauses (e.g., anh, ehh). One or two boundaries (one in the case of the interruption and usually two in the other two cases) always or nearly always occur when one of these three phenomena takes place. However, in principle, these boundaries are not planned by the speaker and do not mark units with a linguistic function. In the analysis of the prosodic boundary cues, they constitute an element of noise, and cannot be compared to the boundaries that the speaker makes to build the meaning of the utterance.

A last type of boundary we have to consider is the one that delimits the units that, in the model of the *Language into Act Theory* (L-AcT; CRESTI, 2000; MONEGLIA; RASO, 2014; MONEGLIA; CRESTI, 1997), are called *Scanning Units*. A *Scanning Unit*, according to L-AcT, is an informationally non-autonomous unit constituting one part of a bigger information unit (e.g. a Topic divided into two or more intonation units). In this case, the units before the last one are *Scanning Units*, and the prosodic profile conveying the information unit function always appears in the last intonation unit. For L-AcT, boundaries that delimit these types of units are due to different possible reasons: emphasis (in order to make parts of an information unit text prominent, its content is segmented into more intonation units); lack of skill in speech (such as small hesitations or retractings without any added segmental material); articulatory necessity (when an information unit features too many syllables for them to fit comfortably in one intonation unit). These kinds of boundaries that, as we have seen, do not constitute a homogeneous group, constitute a problematic typology with regard to the other kinds of boundaries, since the individualization of a *Scanning Unit* is possible

only after a text has been informationally annotated, and this annotation follows text segmentation and cannot be automatized.

Besides these open issues, it would also be interesting to consider some other non-linguistic ones: do male and female voices use the acoustic parameters that convey perception of boundary in the same way? What happens in the different speech pathologies, in which articulatory or cognitive functions are endangered? How do skills that deal with this functional goal develop along ontogenesis?

Along the past decades, research has greatly improved its investigation and understanding of the complex combinations of factors that affect boundary expression; more recently many works have begun the investigation of this phenomenon in spontaneous speech. However, there remains a long way to be covered. Finally, to face the parameter problem is still not sufficient. It is necessary also to look carefully at each parameter in their different combinations and at their weight (hierarchy) in each combination. Of course, this increases the variables responsible for signaling prosodic boundaries, and imposes the use of computational and statistical tools in order for them to be satisfactorily captured.

More recently, prosodic boundaries have been the object of psycholinguistic investigations in an attempt to better understand how their perception is processed (DRURY *et al.*, 2016; GLUSHKO, *et al.*, 2016; NICKELS *et al.*, 2013; HWANG; STEINHAUER, 2011; PAUKER *et al.*, 2011; STEINHAUSER, 2003; STEINHAUER; FRIEDERICI, 2001), especially through the *Event-Related Potential* (ERP) technique. Steinhauser *et al.* (1999) were the first ones who used this technique to show that perceived prosodic boundaries are associated to intervals of increased amplitude in electric activity (evoked potential), named CPS (*Closure Positive Shift*). This peak occurs between 400 and 800 ms. after a defined moment, which, in the most successful tests, was considered in the last stressed syllable before the boundary. The experiments were performed with and without the presence of pause and of other parameters considered responsible for conveying the perception of boundary, but the electric activity peak was always detected. It seems that syllabic lengthening and the presence of a boundary tone are sufficient to trigger the hearer's encephalic reaction. Currently, researchers are trying to refine further the observation of human reaction to isolated parameters, or to their combinations, for the perception of boundaries.

The fact that segmentation (*phrasing*) seems to be sensible to cues of different modalities is especially interesting: not only acoustic cues, but also graphic ones, such as commas in reading, seem to cause an increase of electric activity when there is a boundary. Besides this, the phenomenon also occurs for musical segmentation, but with a greater latency (may be due to the lack of linguistic information, like syntax or lexicon). It also seems that CPS can be encountered only after a certain age (more or less three years of age), and this could be explained if we consider that it depends on a minimal capacity for structuring, either syntactically or prosodically, *stricto sensu*. This result is compatible with data about language acquisition (THORNTON, 2016; HYAMS; ORFITELLI, 2015 *inter alia*). Finally, CPS seems to be more evident when the boundary is less expected, that is, when it is not or is minimally predictable based on information of different natures; but it also seems clear that prosody, as a vehicle for boundaries, prevails when it is in conflict with syntactic expectations (BÖGELS; TORREIRA, 2015; BÖGELS *et al.*, 2013; 2010).

Because boundaries are marked by the combination of all the prosodic parameters, mainly syllabic duration, f0 and intensity, it is important to add that dextral individuals have a predominant temporal processing in the left hemisphere, while spectral processes mainly activate areas of the right hemisphere (ROBIN *et al.*, 1990; ZATORRE, 1997). This is confirmed by studies on impaired individuals, either on the left or on the right hemisphere, the former losing capacity of temporal processing (SHAH *et al.*, 2006). As far as the neuronal areas involved in speech perception, both temporal cortical areas and parietal ones are bilaterally activated (HICKOK; POEPPPEL, 2000).

3. Segmentation and linguistic meaning

Speech segmentation is essential to build linguistic meaning (cf. FERY, 2017, for a review). Prosody is used to mentor the hearer in reconstructing the different functional units and their hierarchy and function, in order to decode the message. This is the main reason that motivates researchers to study the physical nature of boundaries and its relation with the different linguistic levels. Let us look at some examples in different languages.

In English, a sequence as *People give John the book I promised him* can be parsed at least in the four following ways, giving rise to very different meanings, from both illocutive and syntactic points of view:

- (a) *People* (Calling)! *Give John the book I promised him* (Order)!
- (b) *People give John the book I promised him* (Assertion).
- (c) *People give John the book* (Question)? *I promised him* (Assertion).
- (d) *People* (Calling)! *Give John the book* (Order)! *I promised him* (Assertion).

In (a), (c) and (d) we find two terminal boundaries, while in (b) we find just one, which is terminal, too. However, when we look at the acoustic parameters, terminal boundaries associated to the different possible segmentations vary, at least as far as f_0 movements are concerned. If the second boundary in (a), (c) and (d) is preceded by a falling movement, the first boundary features a rising one. These rising movements are different, as much as the different falling movements of the other cases. A similar distinction could be made for the values of duration and intensity.

In Portuguese, a sequence such as *João vai pro Rio até amanhã* (*João will go* (or *go*) *to Rio until tomorrow* (or *see you tomorrow*) can be parsed at least in three different ways:

- a) *João* (calling)! *Vai pro Rio até amanhã* (order)! (*João! Go to Rio until tomorrow*)
- b) *João vai pro Rio até amanhã* (assertion). (*João will go to Rio until tomorrow*)
- c) *João* (calling)! *Vai pro Rio* (order)! *Até amanhã* (greeting)! (*João! Go to Rio! See you tomorrow*)

In these three sentential organizations, it is evident that segmentation affects the syntactic and the semantic-pragmatic interpretation of the sequence.

Finally, the following example shows how segmentation can decide syntactic and semantic interpretation in Italian:

- (a) *Claudia* (calling)! *Guarda* (deixis)! *Quanto è bello* (expressive)!
(*Claudia! Look! How beautiful it is!*)
- (b) *Claudia* (calling)! *Guarda quanto è bello* (deixis)! (*Claudia! Look how beautiful it is!*)
- (c) *Claudia guarda quanto è bello* (assertion). (*Claudia looks how beautiful it is.*)

The series of examples could easily be more complex, considering different interpretations and other types of units. It could also easily be extended to other languages. However, what is relevant for us is to make the importance of the role of prosodic parsing in the construction of linguistic meaning evident, both at the syntactic and at the semantic level. The presence of a boundary certainly affects the phono-morphological level too, for instance, inhibiting sandhi phenomena.

In the previous examples, we have observed some cases of terminal boundaries; they isolate pragmatically and prosodically autonomous linguistic sequences that can be uttered in isolation. However, meaning is also affected in the case of non-terminal boundaries, that is, when the (syntactic or informational) relationship between two units separated by a boundary must be maintained. For example, the sequence *the film I like it* can be analyzed as a noun phrase modified by a relative clause. However, if we insert a boundary, the analysis can change: in *the film, I like it* the analysis can show a Topic-Comment relationship that can be interpreted like: *as for the film* (TOP), *I like it* (COM).

Let us go back to the notion of *unit of reference for speech*, as the minimal unit of the text that carries an autonomous communicative (in the actional sense) meaning. If we consider the prosodic dimension, it is hard to define this unit only through the syntactic criteria used to characterize traditional categories like clause or sentence. Prosody has a communicative dimension that leads researchers to rather pay attention to production and perception of speech, even if we do not lack more abstract perspectives (but possible only outside a communicative context).

Many of the linguists who incorporate prosody as one of the main elements of their models consider prosodic perception of terminality of a communicative sequence as the main cue of the unit of reference for speech (CRESTI, 2000; MONEGLIA; RASO, 2014; IZRE'EL, 2002). Others prefer to consider the intonation unit as unit of reference, no

matter if its prosodic profile is perceived as conclusive or non-conclusive (METTOUCHI *et al.*, 2010). In both these perspectives, the main cue that defines a unit of reference corresponds to the boundary of an intonation unit. The difference consists on whether any kind of boundary determines a reference unit or only boundaries with a specific quality can do it. This discussion goes along with that concerned with the linguistic relations that occur within an intonation unit, those that occur among different intonation units pertaining to the same terminated sequence, and also those across the boundary between different terminated sequences (for some aspects of this discussion in a different but similar framework, see Izre'el in this volume; CRESTI, 2014; PIETRANDREA *et al.*, 2014).

4. The papers in this volume and their contribution to the debate

The nine papers presented in this thematic volume deal with different aspects of prosodic segmentation of spontaneous speech. A first group of papers focuses on the development of software that allow the extraction of data and information useful to clarify some of the many questions related to prosodic segmentation. Of course, also behind these works there is a theoretical hypothesis, either about the function or the number of different boundaries to be identified.

The paper by Xu and Gao presents the FormantPro script, which uses the software Praat as its platform for the automatic extraction of formant trajectories. Although the theme of this article does not directly focus on the problem of prosodic segmentation, the tool and the examples that the authors bring open a discussion about the isomorphism between acoustic and articulatory events that delimit the boundaries of consonants and vowels. These boundaries are discussed with relation to the issue of the alignment of these segmental landmarks with trajectories of f_0 that eventually might have implications to delimitate prosodic boundaries. The software also generates values of duration and intensity and allows the presentation of the mean trajectories in terms of temporal normalization, which helps observing the equivalencies among instances of different utterances with words in contrast. The values of duration can be used to investigate cues of prosodic boundaries in case of important changes with respect to context.

The work by Teixeira Falcão and Mittmann presents an interesting procedure to extract models of acoustic parameters for different types

of boundaries in stretches of spontaneous speech corpora previously segmented by 14 segmentators. The data from corpora were treated to make them readable by the script in Praat. After this, a very high number of measurements is extracted in a window of ten V-V units to the left and 10 V-V units to the right of each position that is a candidate to be a phonological word boundary. The V-V segmentation (BARBOSA, 2006) shows how other levels of speech segmentation necessarily interact with the level of the intonation unit. A statistical procedure, after human refinement, reveals the combinations of parameters that better explain the boundaries and their weight. The whole work was planned considering that prosodic boundaries can be distributed into two big groups: terminal and non-terminal. The work about non-terminal boundaries suggests that it would be necessary to consider these boundaries as at least three different sub-groups, with three different models to account for non-terminal boundaries. These findings encourage the hypothesis that we should differentiate between terminal and non-terminal boundaries, and that we need more subtle distinctions. It would be very important to investigate what accounts for the latter.

The paper by Bigi and Meunier evaluates the SPPAS software, which allows the automatic segmentation of read and spontaneous speech, placing main focus on disfluencies found in spontaneous speech. The tool presupposes the existence of an orthographic transcription and a lexicon pronunciation dictionary. It uses an acoustic model of the sounds of French speech, which allows the alignment of phonetic symbols with the speech signal. The errors in the alignment are approximately 11% in read speech and 15% in spontaneous speech, but they can be reduced using an enriched orthographic transcription that identifies disfluency types. The tool has been tested in nine corpora, including read speech, spontaneous conversation and political debates, for the cases with disfluencies, laughter, filled pauses and noises. The authors show that, when preceded by a pre-processing that segments the speech flow into inter-pausal units, it is possible to achieve a precision level of about 20 ms in the segmentation task.

The article by G. Christodoulides uses two French spoken corpora with the annotation of boundaries of different strengths, in order to verify: (a) degree of agreement between prosodic annotations originated from two different theoretical perspectives, the autosegmental-metrical theory (PIERREHUMBERT, 1980) and the distinction between micro

and macro-syntax (BLANCHE-BENVENISTE, 2002; 2003) referring to two comparable levels of annotation; (b) which acoustic parameters are more important to convey the two types of boundaries and what their hierarchy is. The use of corpora depending on such different theoretical perspectives is an important test for research about prosodic boundaries. This is even more true considering that one corpus is segmented based on theoretical criteria and the other based on perceptual ones. The investigated parameters are: presence and duration of pause, pre-boundary lengthening and two measurements of f_0 associated with boundary. The analysis shows a very high agreement between the two corpora as far as the prosodic parameters in the positions marked as boundary and the distinction between the two types of comparable boundaries are concerned. The conclusion is that the most important parameters associated with boundary and boundary strength is pause, followed by syllabic lengthening. f_0 seems to be important to distinguish between presence or absence of boundaries, but not to signal boundary strength and therefore distinguish the two types of boundaries.

Ph. Martin's work differs from the others because it analyses a different unit: the stress group. The object of the paper is therefore a smaller unit than intonation unit, even if sometimes the two units may coincide. Martin individualizes a limited number of possible f_0 movements in the stress group inside the intonation unit, and observes that there is a dependency criterion among them. This allows us to investigate the internal structure of an intonation unit, based on smaller units marked by stress. Among other consequences, the results of this analysis may bring to light some characteristics of the internal structures of different intonation units, and may show how these structures correlate with the linguistic function of a specific intonation unit. Different aspects of the unit, with the presence of some prominences in defined positions, have already been discussed in the literature, even if not conclusively in our perspective. Proposals like that by Martin lead us to consider the role played by other prosodic levels and their specific linguistic functions, that, besides other characteristics (prominences, type of boundary), may give us a better understanding of how we build a sequence with a definite linguistic function dealing with different levels of the prosodic structure.

A third group of papers investigates the boundaries clearly with linguistic goals, either syntactic or informational.

The study by A. Mettouchi on Kabyle, an Afro-Asiatic language of Algeria, shows how the presence/absence of a boundary can constitute the linguistic cue that marks a syntactic function, in this case the direct object. The boundary reveals itself as the decisive cue in order to distinguish this structure from structures that can have different functions, probably informational ones, but that appear in the utterance with the same formal cues, except for the presence (other functions) or absence (direct object) of a prosodic boundary. This study raises an important issue: the relationship between the presence of boundaries and the rupture of syntactic compositionality. Other studies (CRESTI, 2014; RASO; VIEIRA, 2016; BOSSAGLIA *et al.*, Forthcoming) treat this important aspect, which is still controversial. If on the one hand it is easy to find cases in which it seems clear that syntactic compositionality is interrupted where there is a prosodic boundary (making it possible to think that some type of boundary has the possibility of marking this interruption), on the other hand, we still have cases that are interpretable, thus saving the syntactic compositionality across a prosodic boundary.

The article by da Silva and Fonseca also presents several aspects of interest. The first one, as with the previous and the following studies, is the importance that a prosodic cue has for the identification of a linguistic unit, in this case the unit of Topic. The second reason is the experimental basis of the research, about which we will come back later. A third reason is that the work shows how results presented within a formalist framework can also be useful for the study of Topic in different perspectives, making it clear how the empirical view on data can benefit the scientific debate. The experiments idealized and implemented by da Silva and Fonseca can be of great interest for the debate among researchers about information structure in speech. The results can be used to compare a syntactic definition of Topic with definitions of a pragmatic nature, especially the one proposed by L-AcT, which assigns to prosody a crucial weight besides presenting many results investigating different languages, among which BP (cf. CRESTI, 2000; SIGNORINI 2004; FIRENZUOLI; SIGNORINI, 2003; MONEGLIA; RASO, 2014; ROCHA; RASO, 2013; CAVALCANTE, 2016; MITTMANN, 2012; RASO; CAVALCANTE; MITTMANN, Forthcoming). Actually, the non-expected results found for the third experiment reported in the paper could be easily explained assuming that Topic is a pragmatic category that does not depend on argument structure and, therefore, can

occupy the subject position, but is marked by a prosodic boundary and a functional prosodic focus that distinguish it from subject. The subject, on the contrary, does not present a prosodic boundary between itself and the rest of the utterance and does not carry any prosodic functional focus. In this case, the difference between Topic and subject would not consist in their being two different syntactic functions, but would be explained as a difference of linguistic level: the subject would be a syntactic function and an argument of the verb in the Comment unit, while the Topic would be a pragmatic function, external to the Comment unit. A more in-depth debate between these different theoretical perspectives could clarify the notion of Topic and stimulate both approaches to refine their analyses and their argumentation, using both experimental procedures, like those proposed by da Silva and Fonseca, and data extracted from spontaneous speech corpora, like those compiled taking L-AcT into account (CRESTI; MONEGLIA, 2005; RASO; MELLO, 2012; Forthcoming).

The study by Panunzi and Saccone is also clearly theory-oriented. In fact, its goal is to observe if, to which extent and how boundaries between different pairs of information units are performed in different ways. The two pairs (rarely sequences of more than two items) that are explored in the article are different combinations of illocutionary units. One type of pair is characterized by two prosodically and pragmatically patterned illocutions that build a unique interpretation. The other type, on the contrary, is constituted by two independent illocutions, even if separated by a non-terminal boundary. Therefore, in order to analyze the boundaries, the text must be informationally tagged according to a theoretical framework, in this case L-AcT (CRESTI, 2000; MONEGLIA; RASO, 2014). The first results suggest that there are clear formal differences between the two pairs of units. This is an intriguing example showing how characteristics of the boundary may correlate with the function of the units separated by it. This kind of study, which tries to correlate linguistic functions of the intonation unit and boundary cues, can be applied to different kinds of units and can be based on different theoretical frameworks.

The paper by Izre'el is the last one in this volume because, based on some considerations about the linguistic role played by prosody and especially by prosodic boundaries, it proposes a general revision of the traditional categories phrase, clause, sentence and predication, showing how the incorporation of prosody may lead to a general reformulation of

canonical categories in the study of spontaneous speech. Izre'el revisits the discussion about these categories starting with the ancient Greek tradition up to Chomsky, in order to show how some categories, as they are defined in the syntactic tradition, do not work in the analysis of speech, especially of spontaneous speech, which, in principle, should be the natural domain for the analysis of language. Considering prosody and data from spontaneous speech corpora, the importance of the illocution (which Izre'el calls *modality*) clearly emerges as a crucial category to individualize the communicative unit and as a prosodically marked category. The importance of prosodic boundaries also clearly emerges as a means to define the domain of linguistic relations in their communicative realization. Like other papers in this volume, but portraying a wider scope, this paper brings more arguments to the discussion (cf. also BIBER *et al.*, 1999; the papers in RASO; MELLO, 2014; CRESTI, 2005; RASO; MITTMANN, 2012, *inter alia*). It highlights the urgency of defining the communicative unit of speech, of revising the notion of predication (and of proposition), or those of clause and sentence, and sustains how important it is to incorporate prosody as the central element to mark the unit of reference for spoken communication. As other articles in this thematic issue, the paper by Izre'el does not leave any doubt about the necessity of incorporating prosody among the levels of linguistic analysis, and, more than this, about the crucial hierarchical weight of prosody to individualize the linguistic constituents of speech.

References

- AMIR, N.; SILBER-VAROD, V.; IZRE'EL, S. Characteristics of Intonation Unit Boundaries in Spontaneous Spoken Hebrew: Perception and Acoustic Correlates. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 2004. Nara. *Proceedings...* Nara: ISCA, 2004. p. 677-680.
- AURAN, C.; BOUZON, C.; HIRST, D. The Aix-MARSEC Project: an Evolutive Database of Spoken British English. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 2004. Nara, Japan. *Proceedings...* Nara: ISCA, 2004.
- AVANZI, M.; LACHERET-DUJOUR, A.; VICTORRI, B. ANALOR. Tool for Semi-Automatic Annotation of French Prosodic Structure. In: ANALOR. A Tool for Semi-Automatic Annotation of French Prosodic Structure. Campinas, Brazil, May 2008. p. 119-122.

BARBOSA, P. A. *Caractérisation et génération automatique de la structuration rythmique du français*. 1994. Tese (Doutorado) – Institut National Polytechnique de Grenoble, França, 1994.

BARBOSA, P. A. *Incursões em torno do ritmo da fala*. Campinas: Pontes, 2006.

BARBOSA, P. A. Prominence-and Boundary-Related Acoustic Correlations in Brazilian Portuguese Read and Spontaneous Speech. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 4., 2008, Campinas. *Proceedings...* Campinas: ISCA, 2008. p. 257-260.

BARBOSA, P. A. Automatic Duration-Related Saliency Detection in Brazilian Portuguese Read and Spontaneous Speech. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 5., 2010, Chicago. *Proceedings...* Chicago: ISCA, 2010.

BARBOSA, P. A. Panorama of Experimental Prosody Research. In: GSCP INTERNATIONAL CONFERENCE – SPEECH AND CORPORA, VIIth., Belo Horizonte. *Proceedings...* Florence: Firenze University Press, 2012. p. 33-42.

BARTH-WEINGARTEN, D. *Intonation Units Revisited*. Cesura in Talk-In-Interaction. Amsterdam: John Benjamins, 2016.

BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited, 1999.

BLANCHE-BENVENISTE, C. Macro-syntaxe et micro-syntaxe: les dispositifs de la rection verbale. In: ANDERSEN, H. L.; NØLKE, H. (Éd.). *Macro-Syntaxe e Macro-Sémantique*. Bern: Peter Lang, 2002. p. 95-115.

BLANCHE-BENVENISTE, C. Le recouvrement de la syntaxe et de la macro-syntaxe. In: SCARANO, A. (Ed.). *Macro-syntaxe et pragmatique*. L'analyse linguistique de l'oral. Roma: Bulzoni, 2003. p. 53-75.

BLANCHE-BENVENISTE, C.; JEANJEAN, C. *Le français parlé*. Transcription et édition. Paris: Didier Érudition; Institut National de la Langue Française, 1987.

BLANCHE-BENVENISTE C. Macro-syntaxe et micro-syntaxe: les dispositifs de la rection verbale. In: ANDERSEN, H. L.; NØLKE, H. (Éd.). *Macro-syntaxe et macro-sémantique*. Actes du Colloque International d'Århus [17-19 mai 2001]. Bern: Peter Lang, 2002. p. 95-118.

BÖGELS, S.; SCHRIEFERS, H.; VONK, W.; CHWILLA, D. J.; KERKHOFS, R. The Interplay Between Prosody and Syntax in Sentence Processing: The Case of Subject- and Object-Control Verbs. *Journal of Cognitive Neuroscience*, Cambridge, v. 22, n. 5, p. 1036-1053, 2010.

BÖGELS, S.; SCHRIEFERS, H.; VONK, W.; CHWILLA, D.; KERKHOFS, R. Processing Consequences of Superfluous and Missing Prosodic Breaks in Auditory Sentence Comprehension. *Neuropsychologia*, Oxford, v. 51, p. 2715-2728, 2013.

BÖGELS, S.; TORREIRA, F. Listeners Use Intonational Phrase Boundaries to Project Turn Ends in Spoken Interaction. *Journal of Phonetics*, [s.l.], v. 52, p. 46-57, 2015.

BOLINGER, D. Pitch Accent and Sentence Rhythm. In: ABE, I.; KANEKIYO, T. (Ed.). *Forms of English: Accent, Morpheme, Order*. Cambridge, Mass: Harvard University Press, 1965. p. 139-180.

BOSSAGLIA, G.; MELLO, H.; RASO, T. Insubordination and the Syntax/Prosody Interface in Spoken Brazilian Portuguese: Data on Adverbial Clauses. In: IZRE'EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*. Amsterdam: John Benjamins. (Forthcoming).

BUHMANN, J.; CASPERS, J.; HEUVEN, V. J. van; HOEKSTRA, H.; MARTENS, J-P.; SWERTS, M. Annotation of Prominent Words, Prosodic Boundaries and Segmental Lengthening by Non-Expert Transcribers in the Spoken Dutch Corpus. In: LREC, 3rd, 2002, Las Palmas. *Proceedings...* Las Palmas: ELRA, 2002. p. 779-785.

BYBEE, J. *Language, Usage and Cognition*. Cambridge: CUP, 2010.

CARLSON, R.; HIRSCHBERG, J.; SWERTS, M. Cues to Upcoming Swedish Prosodic Boundaries: Subjective Judgment Studies and Acoustic Correlates. *Speech Communication*, [s.l.], v. 46, p. 326-333, 2005. Doi: <https://doi.org/10.1016/j.specom.2005.02.013>

CAVALCANTE, F. *The Topic Unit in Spontaneous American English: a Corpus-Based Study*. 2016. Master (Thesis) – Universidade Federal de Minas Gerais, Belo Horizonte, 2016.

CHAFE, W. *Discourse, Consciousness and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: Chicago University Press, 1994.

COUPER-KUHLEN, E. Prosody and sequence organization in English conversation. In: COUPER-KUHLEN, E.; FORD, C. E. (Ed.). *Sound Patterns in Interaction: Cross-Linguistic Studies from Conversation*. Amsterdam: John Benjamins, 2004. p. 335-376.

COWAN, N. *Attention and Memory: An Integrated Framework*. Oxford: Oxford University Press, 1998.

CRESTI, E. *Corpus di italiano parlato*. Firenze: Accademia della Crusca, 2000. 2 v.

CRESTI, E. Notes on Lexical Strategy, Structural Strategies and Surface Clause Indexes in the C-ORAL-ROM Spoken Corpora. In: CRESTI, E.; MONEGLIA, M. (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam; Philadelphia: John Benjamins, 2005. p. 209-256.

CRESTI, E. Syntactic Properties of Spontaneous Speech in the Language into Act Theory: Data on Italian Complements and Relative Clauses. In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014.

CRESTI, E. The Pragmatic Analysis of Speech and Its Illocutionary Classification According to Language into Act Theory. In: IZRE'EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*. Amsterdam: John Benjamins. (Forthcoming).

CRESTI, E.; MONEGLIA, M. (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, 2005.

CROFT, W. Intonational Units and Grammatical Structure. *Linguistics*, [s.l.], v. 33, n. 5, p. 839-882, 1995.

CRUTTENDEN, Alan. *Intonation*. 2nd edition. New York: Cambridge University Press, 1997.

CRYSTAL, D. *Prosodic Systems and Intonation in English*. Cambridge: CUP, 1969.

DILLEY, L.; SHATTUCK-HUFNAGEL, S.; OSTENDORF, M. Glottalization of Word-Initial Vowel as a Function of Prosodic Structure. *Journal of Phonetics*, [s.l.], v. 24, p. 423- 444, 1996.

DRURY, J. E.; BAUM, Sh. R.; VALERIOTE, H.; STEINHAUSER, K. Punctuation and Implicit Prosody in Silent Reading: An ERP Study Investigating English Garden-Path Sentences. *Frontiers in Psychology*, [s.l.], Sept. 2016. Doi: <https://doi.org/10.3389/fpsyg.2016.01375>

DU BOIS, J. W.; CHAFE, W. L.; MEYER, Ch.; THOMPSON, S. *Santa Barbara Corpus of Spoken American English*. Washington DC: Linguistic Data Consortium, 2000-2005.

DU BOIS, J. W.; CUMMING, S.; SCHUETZE-COBURN, S.; PAOLINO, D. Discourse transcription. *Santa Barbara Papers in Linguistics*, Santa Barbara, v. 4, n. 1, p. 225, 1992.

EVANS, N.; WATANABE, H. The dynamics of insubordination: An overview. In: _____. (Ed.). *Insubordination*. Amsterdam: John Benjamins, 2016.

FERY, C. *Intonation and Prosodic Structure*. Cambridge: CUP, 2017.

FIRENZUOLI, V.; SIGNORINI, S. L'unità informativa di topic: correlati intonativi. In: MAROTTA, G. (Ed.). *La coarticolazione: atti delle XIII Giornate di Studio del Gruppo di Fonetica Sperimentale*, [28-30 nov. 2002]. Pisa: ETS, 2003. p. 177-184.

FON, J.; JOHNSON, K.; CHEN, S. Durational Patterning at Syntactic and Discourse Boundaries in Mandarin Spontaneous Speech. *Language and Speech*, [s.l.], v. 54, n. 1, p. 5-32, 2011.

FUCHS, S.; KRIVOKAPIĆ, J.; JANNEDY, S. Prosodic Boundaries in German: Final Lengthening in Spontaneous Speech. *The Journal of the Acoustical Society of America*, [s.l.], v. 127, n. 3, p. 1851, 2010. Doi: 10.1121/1.3384378

FURHT, B.; VILLANUSTRE, F. *Big Data Technologies and Applications*. [s.l.]: Springer, 2016.

GAROFOLO, J. S.; LAMEL, L. F.; FISHER, W. M.; FISCUS, J. G.; PALLETT, D. S.; DAHLGREN, N. L.; ZUE, V. TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Web Download. Philadelphia: Linguistic Data Consortium, 1993.

GEE, J. P.; GROSJEAN, F. Performance Structures: A Psycholinguistic and Linguistic Appraisal. *Cognitive Psychology*, [s.l.], v. 15, n. 4, p. 411-458, 1983.

GLUSHKO, A.; STEINHAEUER, K.; De PRIEST, J.; KOELSCH, S. Neurophysiological Correlates of Musical and Prosodic Phrasing: Shared Processing Mechanisms and Effects of Musical Expertise. *PLoS ONE*, San Francisco, v. 11, n. 5, 2016.

GORDON, M.; LADEFOGED, P. Phonation Types: a Cross-Linguistic Overview. *Journal of Phonetics*, [s.l.], v. 29, p. 383-406, 2001.

GROSJEAN, F.; DOMMERGUES, J. Y. Les structures de performance en psycholinguistique. *L'Année Psychologique*, [s.l.], v. 83, n. 2, p. 513-536, 1983.

GROSJEAN, F.; GROSJEAN, L.; LANE, H. The Patterns of Silence: Performance Structures in Sentence Production. *Cognitive Psychology*, [s.l.], v. 11, n. 1, p. 58-81, 1979.

HANSON, H. M.; CHUANG, E. S. Glottal Characteristics of Male Speakers: Acoustic Correlates and Comparison with Female Data. *Journal of the Acoustical Society of America*, [s.l.], v. 106, n. 2, p. 1064-1077, 2001.

t'HART, J. Differential Sensitivity to Pitch Distance, Particularly in Speech. *The Journal of the Acoustical Society of America*, [s.l.], v. 69, n. 3, p. 811-821, 1981.

t'HART, J.; COLLIER, R.; COHEN, A. *A Perceptual Study on Intonation: An Experimental Approach to Speech Melody*. Cambridge: CUP, 1990.

HICKOK, G.; POEPEL, D. Towards a Functional Neuroanatomy of Speech Perception. *Trends in Cognitive Sciences*, [s.l.], v. 4, n. 4, p. 131-138, 2000.

HIRST, D.; Di CRISTO, A. A Survey of Intonation Systems. In: _____. (Ed.). *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 1998.

HUGGINS, A. W. F. Just Noticeable Differences for Segment Duration in Natural Speech. *The Journal of the Acoustical Society of America*, [s.l.], v. 51, n. 4B, p. 1270-1278, 1972.

HWANG, H.; STEINHAUER, K. Phrase Length Matters: The Interplay Between Implicit Prosody and Syntax in Korean ‘Garden Path’ Sentences. *Journal of Cognitive Neuroscience*, Cambridge, v. 23, n. 11, p. 3555-3575, 2011.

HYAMS, N.; ORFITELLI, R. The Acquisition of Syntax. In: CAIRNS, H.; FERNANDEZ, E. (Ed.). *Handbook of Psycholinguistics*. [s.l.]: Wiley; Blackwell Publishers, 2015.

IZRE’EL, S. The Corpus of Spoken Israeli Hebrew: Textual Samples. *Lesson Enu*, v. 64, p. 289-314, 2002.

IZRE’EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. (Ed.). *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*. Amsterdam: John Benjamins. (Forthcoming)

KLATT, D. H.; COOPER, W. E. Perception of Segment Duration in Sentence Contexts. In: COHEN, A.; NOOTEBOOM, S. G. (Ed.). *Structure and Process in Speech Perception*. Berlin; Heidelberg: Springer, 1975. p. 69-89.

KOFFI, E. A Just Noticeable Difference (JND) Reanalysis of Fry’s Original Acoustic Correlates of Stress in American English. *Linguistic Portfolios*, St. Cloud, v. 7, 2018.

LADD, D. R. *Intonational phonology*. Cambridge: Cambridge University Press, 1996.

LIEBERMAN, P. Some Acoustic Correlates of Word Stress in American English. *Journal of the Acoustical Society of America*, [s.l.], v. 32, p. 451-454, 1960.

LINELL, P. *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations*. London; New York: Routledge, 2005.

MacNEILAGE, P. F. The Frame/Content Theory of Evolution of Speech Production. *Behavioral and Brain Sciences*, Cambridge, v. 21, n. 4, p. 499-511, 1998.

MARTIN, P. Les problèmes de l'intonation: recherches et applications. *Langue Française*, [s.l.], v. 19, p. 4-32, 1973.

MARTIN, P. Prosodic and Rhythmic Structures in French. *Linguistics*, [s.l.], v. 25, n. 5, p. 925-950, 1987.

MELLO, H.; RASO, T.; MITTMANN, M. VALE, H.; CÔRTEZ, P. Transcrição e segmentação prosódica do *corpus* c-oral-brasil: critérios de implementação e validação. In: RASO, T.; MELLO, H. (Ed.). C-ORAL-BRASIL I. *Corpus* de referência do português brasileiro falado informal. Belo Horizonte: UFMG, 2012. p. 125-174.

METTOUCHI, A.; CAUBET, D.; VANHOVE, M.; TOSCO, M.; COMRIE, B.; IZRE'EL, S. CORPAFROAS, A Corpus for Spoken Afroasiatic Languages: Morphosyntactic and Prosodic Analysis. In: ITALIAN MEETING OF AFRO-ASIATIC LINGUISTICS, 13th, 2010, Padova. *Proceedings...* Padova: Sargon, 2010. p.177-180.

MITTMANN, M. M. *O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o Tópico no Português Brasileiro*. 2012. Dissertation (Ph.D.) – Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

MITTMANN, M. M.; BARBOSA, P. A. An Automatic Speech Segmentation Tool Based on Multiple Acoustic Parameters. *CHIMERA. Romance Corpora and Linguistic Studies*, Madrid, v. 3, p.133-147, 2016.

MO, Y. Duration and Intensity as Perceptual Cues for Naïve Listeners' Prominence and Boundary Perception. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 4th., 2008, Campinas. *Proceedings...* Campinas: ISCA, 2008. p. 739-742.

MO, Y.; COLE, J.; LEE, E. K. Naïve Listeners' Prominence and Boundary Perception. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 4th., 2008, Campinas. *Proceedings...* Campinas: ISCA, 2008. p. 735-738.

MO, Y.; COLE, J. Perception of Prosodic Boundaries in Spontaneous Speech with and Without Silent Pauses. *The Journal of the Acoustical Society of America* [s.l.], v. 127, n. 3, p. 1956, 2010.

MONEGLIA, M.; CRESTI, E. L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In: BORTOLINI, U.; PIZZUTO, E. (Ed.). *Il Progetto CHILDES Italia*. Pisa: Del Cerro, 1997. p. 57-90.

MONEGLIA, M.; FABBRI, M.; QUAZZA, S.; ANDREA; PANIZZA, A.; DANIELI, M.; GARRIDO, J. M.; SWERTS, M. Evaluation of Consensus on the Annotation of Terminal and Non-Terminal Prosodic Breaks in the C-ORAL-ROM corpus. In: CRESTI, E.; MONEGLIA, M. (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, 2005. p. 257-276.

MONEGLIA, M.; RASO, T. Notes Language into Act Theory (L-AcT). In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014. p. 468-495.

MONNIN, P.; GROSJEAN, F. Les structures de performance en français: caractérisation et prédiction. *L'Année Psychologique*, [s.l.], v. 93, p. 9-30, 1993.

MORAES, J. A.; RILLIARD, A. Illocution, Attitude and Prosody: A Multimodal Analysis. In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014. p. 233-270.

NESPOR, Marina; VOGEL, Irene. *Prosodic Phonology*. Dordrecht: Foris Publications, 1986.

NI, C. J.; ZHANG, A. Y.; LIU, W. J.; XU, B. Automatic Prosodic Break Detection and Feature Analysis. *Journal of Computer Science and Technology*, [s.l.], v. 27, n. 6, p. 1184-1196, 2012.

NICKELS, S.; OPITZ, B.; STEINHAEUER, K. ERPs Show that Classroom-Instructed Late Second Language Learners Rely on the Same Prosodic Cues in Syntactic Parsing as Native Speakers. *Neuroscience Letters*, [s.l.], v. 557, p. 107-111, 2013.

OSTENDORF, Mari; PRICE, Patti; SHATTUCK-HUFNAGEL, Stefanie. Boston University Radio Speech Corpus LDC96S36. DVD. Philadelphia: Linguistic Data Consortium, 1996.

PAUKER, E.; ITZHAK, I.; BAUM, S. R.; STEINHAEUER, K. Effects of Cooperating and Conflicting Prosody in Spoken English Garden Path Sentences: ERP Evidence for the Boundary Deletion Hypothesis. *Journal of Cognitive Neuroscience*, Cambridge, v. 23, n. 10, p. 2731-2751, 2011.

PIERREHUMBERT, J. B. *The Phonology and Phonetics of English Intonation*. 1980. Dissertation (Doctoral) – Massachusetts Institute of Technology, 1980.

PIETRANDREA, P.; KAHANE, S.; LACHERET, A.; SABIO, F. In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014.

PIKE, K. *The Intonation of American English*. Ann Arbor: University of Michigan Press, 1945.

PRICE, P. J.; OSTENDORF, M.; SHATTUCK-Hufnagel, S.; FONG, C. The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, [s.l.], v. 90, n. 6, p. 2956-2970, 1991.

RASO, T.; CAVALCANTE, F.; MITTMANN, M. M. Prosodic Forms of the Topic Information Unit in a Cross-Linguistic Perspective: a First Survey. In: GSCP INTERNATIONAL CONFERENCE, 2016, Napole. *Proceedings...* Napole. (Forthcoming).

RASO, T.; MELLO, H. (Ed.). C-ORAL-BRASIL I. *Corpus de referência do português brasileiro falado informal*. Belo Horizonte: UFMG, 2012.

RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014.

RASO, T.; MELLO, H. (Ed.). The C-ORAL-BRASIL II. *Corpus de referência do português falado (formal em contexto natural, mídia e telefone)*. (Forthcoming).

RASO, T.; MITTMANN, M. As principais medidas da fala. In: RASO, T.; MELLO, H. (Ed.). C-ORAL-BRASIL I. *Corpus de referência do português brasileiro falado informal*. Belo Horizonte: UFMG, 2012. p. 177-220.

RASO, T.; VIEIRA, M. A Description of Dialogic Units/Discourse Markers in Spontaneous Speech Corpora Based on Phonetic Parameters. CHIMERA. *Romance Corpora and Linguistic Studies*, Madrid, v. 3, p. 221-249, 2016.

REDI, L.; SHATTUCK-HUFNAGEL, S. Variation in the Realization of Glottalization in Normal Speakers. *Journal of Phonetics*, [s.l.], v. 29, p. 407-29, 2001.

RIETVELD, A.; GUSSENHOVEN, C. On the Relation Between Pitch Excursion Size and Prominence. *Journal of Phonetics*, [s.l.], v. 13, p. 299-308, 1985.

ROBIN, D. A. *et al.* Auditory Perception of Temporal and Spectral Events in Patients with Focal Left and Right Cerebral Lesions. *Brain and Language*, [s.l.], v. 39, p. 539-555, 1990.

ROCHA, B.; RASO, T. O pronome lembrete e a Teoria da Língua em Ato: uma análise baseada em corpora. *Veredas*, Juiz de Fora, v. 17, n. 2, p. 39-59, 2013.

SCHEGLOFF, E. A. Reflections on Studying Prosody in Talk-in-Interaction. *Language and Speech*, [s.l.], v. 41, n. 3-4, p. 235-263, 1998.

SELKIRK, E. Sentence prosody: Intonation, Stress and Phrasing. In: GOLDSMITH, J. A. (Ed.). *The Handbook of Phonological Theory*. Oxford: Blackwell, 1995. p. 550-569.

SHAH, A. P.; BAUM, S. R.; DWIVEDI, V. D. Neural Substrates of Linguistic Prosody: Evidence from Syntactic Disambiguation in the Productions of Brain-Damaged Patients. *Brain and Language*, [s.l.], v. 96, p. 78-89, 2006.

SHRIBERG, E.; STOLCKE, A.; HAKKANI-TÜR, D.; TÜR, G. Prosody-Based Automatic Segmentation of Speech into Sentences and Topics. *Speech Communication*, [s.l.], v. 32, n. 1-2, p. 127-154, 2000.

SCHUURMAN, I.; SCHOUPPE, M.; HOEKSTRA, H.; Van der Wouden, T. CGN, an Annotated Corpus of Spoken Dutch. In: INTERNATIONAL WORKSHOP ON LINGUISTICALLY INTERPRETED CORPORA (LINC-03), 4TH, 2003, Budapest. *Proceedings...* Budapest: EACL, 2003.

SIGNORINI, S. *Topic e soggetto in corpora di italiano parlato spontaneo*. 2004. Dissertation (Ph.D.) – Università Firenze, Firenze, 2004.

STEINHAEUER, K.; ALTER, K.; FRIEDERICI, A. D. Brain Potentials Indicate Immediate Use of Prosodic Cues in Natural Speech Processing. *Nature Neuroscience*, [s.l.], v. 2, n. 2, p. 191-196, 1999.

STEINHAEUER, K. Electrophysiological Correlates of Prosody and Punctuation. *Brain and Language*, [s.l.], v. 86, n. 1, Special issue on the Neuronal Basis of Language, p. 142-164, 2003.

STEINHAUER, K.; FRIEDERICI, A. D. Prosodic Boundaries, Comma Rules, and Brain Responses: The Closure Positive Shift in ERPs as a Universal Marker for Prosodic Phrasing in Listeners and Readers. *Journal of Psycholinguistic Research*, [s.l.], v. 30, n. 3, p. 267-295, 2001.

SWERTS, M. Prosodic Features of Discourse Units. 1994. Thesis (PhD) – Technische Universiteit Eindhoven, 1994.

SWERTS, M. Prosodic Features at Discourse Boundaries of Different Strength. *The Journal of the Acoustical Society of America*, [s.l.], v. 101, n. 1, p. 514-521, 1997.

SWERTS, M.; COLLIER, R.; TERKEN, J. Prosodic Predictors of Discourse Finality in Spontaneous Monologues. *Speech Communication*, [s.l.], v. 15, n. 1-2, p. 79-90, 1994.

TEIXEIRA FALCÃO, B. H. *Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea*, 2017. Thesis (Master) – Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

TESNIÈRE, L. *Éléments de Syntaxe Structurale*. Paris: C. Klincksieck, 1965.

THORNTON, R. Children's Acquisition of Syntactic Knowledge. In: ARONOFF, M. (Ed.). *Oxford Research Encyclopedia of Linguistics*, 2016.

TSENG, C. Y.; CHANG, C. H. Pause or No Pause? Prosodic Phrase Boundaries Revisited. *Tsinghua Science and Technology*, [s.l.], v. 13, n. 4, p. 500-509, 2008.

TSENG, C. Y.; FU, B. Duration, Intensity and Pause Predictions in Relation to Prosody Organization. In: EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY, INTERSPEECH, 9th, 2005. Lisboa. *Proceedings...* Lisboa: SISCA, 2005. p. 1405-1408. Available at: <<http://www.ling.sinica.edu.tw/eip/FILES/publish/2007.4.12.99500673.0143164.pdf>>. Retrieved on: Dec. 1, 2016.

TYLER, J. Prosodic Correlates of Discourse Boundaries and Hierarchy in Discourse Production. *Lingua*, [s.l.], v. 133, p. 101-126, 2013.

Van DONZEL, M. E. *Prosodic Aspects of Information Structure in Discourse*. Den Haag: Holland Academic Graphics/IFOTT, 1999.

WAGNER, A. Acoustic Cues for Automatic Determination of Phrasing. In: SPEECH PROSODY 2010 – INTERNATIONAL CONFERENCE, 5th., 2010, Chicago. *Proceedings...* Chicago: ISCA, 2010. Available at: <https://www.isca-speech.org/archive/sp2010/papers/sp10_196.pdf>. Retrieved on: Sept. 2018.

WAGNER, M.; WATSON, D. G. Experimental and Theoretical Advances in Prosody: A Review. *Language and Cognitive Processes*, [s.l.], v. 25, n. 7-9, p. 905-945, 2010.

WIGHTMAN, C. W.; SHATTUCK-HUFNAGEL, S.; OSTENDORF, M.; PRICE, P. J. Segmental Durations in the Vicinity of Prosodic Phrase Boundaries. *The Journal of the Acoustical Society of America* [s.l.], v. 91, n. 3, p. 1707–1717, 1992.

XU, Y. In Defense of Lab Speech. *Journal of Phonetics*, [s.l.], v. 38, n. 3, p. 329-336, 2010.

YANG, Y.; WANG, B. Acoustic Correlates of Hierarchical Prosodic Boundary in Mandarin. In: SPEECH PROSODY, 2002, Aix-en-Provence. *Proceedings...* Aix-en-Provence: Laboratoire Parole et Langage, 2002.

YOON, T-J.; CHAVARRÍA, S.; COLE, J.; HASEGAWA-JOHNSON, M. Intertranscriber Reliability of Prosodic Labeling on Telephone Conversation Using ToBI. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 2004. Nara, Japan. *Proceedings...* Nara: ISCA, 2004. p. 2722-2732.

ZATORRE, R. J. Cerebral Correlates of Human Auditory Processing: Perception of Speech and Musical Sounds. In: SYKA, J. (Ed.). *Acoustical Signal Processing in the Central Auditory System*. Prague: Plenum Press, 1997. p. 453-468.