Perception of prosodic boundaries by naïve listeners in three different types of subordinate syntactic constructions

Manon Lelandais, Gaëlle Ferré

LLING – Université de Nantes, France

manon.lelandais@univ-nantes.fr, gaelle.ferre@univ-nantes.fr

Abstract

This paper investigates the use of prosodic and syntactic information in the perception of boundaries in extracts of spontaneous speech. 30 naïve listeners had to measure boundary strength for 52 extracts on a 5-point scale. The stimuli all contained three tone-units, the second being a syntactic subordinate construction, which was established as a variable. The prosodic cues at the boundary between the tone-units were also established as variables, and were subject to manipulation (addition of a single cue associated with the perception of a prosodic boundary). The stimuli were also resynthesized in another set to obliterate lexical and syntactic content while keeping syllabic structure and intonation. Results show that naïve listeners are able to identify different degrees of break, and that the three syntactic types show different interactions with ratings. Although a silent pause is the strongest cue to boundary perception for all three types of subordination, the orders and levels of association with other prosodic cues are not the same across syntactic types.

Index Terms: speech perception experiment, prosodic boundaries, spontaneous speech, English, subordinate structures.

1. Introduction

During natural speech perception, listeners rely on a wide range of cues to support comprehension, from prosodic information to semantic context. Prosodic boundaries have been shown to be linked with syntactic boundaries (e. g. [1]; [2]), but the association is usually documented and recognized for large constituents [3]; [4] e.g. independent and/or complex clauses.

Production studies on the prosodic characteristics of syntactic subordination in spontaneous speech have shown that subordinate constructions display different degrees of prosodic autonomy depending on their syntactic type [5]; [6]. This study focuses on adverbial clauses, appositive clauses, and restrictive relative clauses, as illustrated in examples (1-3). In (1), the adverbial clause restricts the spatial scope in which the referential elements must be understood.

> (1)   i tried driving once in her car **when were on a little road in the countryside**

In (2), the appositive relative clause qualitatively evaluates *a place called Tropicana*, which can however be identified independently as a referent.

> (2)   and then we went into a place called Tropicana **which was horrible**

Lastly, in (3), the restrictive relative clause increases the relevance of *the Spanish girls*, creating a subcategory for this referent.

> (3)   the Spanish girls **that were there on our second lesson**

Among these three syntactic types, appositive clauses feature the biggest combination of disjunctive cues at segmental and suprasegmental levels [5]. In this context as in spontaneous speech in general, variations in duration and silent pauses are the preferred cues to mark a boundary in discourse [7]; [8]; [9], as well as initial pitch upsteps [10]. Interlocutors supposedly use these cues to segment the flow of speech and to process information [11]; [12]; [13]. However, the data collected on the production of subordinate constructions do not document the relative weight of the vocal cues compared to the (lexico-syntactic) verbal cues in the interpretation of subordinate structures. Likewise, the production data do not determine whether the preferred cues for discourse segmentation in production are also the preferred cues in perception.

Recent studies about the prosodic perception of subordinate clauses are scarce, but [1] have shown that prosodic boundaries are optional in some cases because a given syntactic structure can have a number of equally acceptable intonational phrasings [14].

Our perception test aims at answering two questions: can listeners assign degrees of perceived boundary strength between two clauses (one being syntactically subordinate), and if so, what is the relation of these different degrees to prosodic cues on the one hand, and to the different syntactic types of subordinate constructions on the other hand?

2. Related work

2.1. Theoretical vs. perceptual boundaries

Many spoken corpora include a segmentation of data into various prosodic units, partly relying on "boundaries" (e.g. [15], [16]). These units are either manually annotated by experts or automatically detected, based on acoustic features. They are useful to investigate the relations between prosodic, syntactic, and discourse phenomena. In the theories in which speech is considered as a hierarchy of prosodic domains, the largest prosodic unit to be intonationally marked is the Intonational Phrase (see [17] and [18] for Prosodic Phonology theory; [2], [19], [20] for the Autosegmental-metrical theory). Throughout the languages reviewed so far, an IP is delimited with a boundary tone (i.e. pitch movement on the last syllables of the unit) at its right or more rarely, left extremity, final syllabic lengthening, and an optional pause after the unit [20]; [21].

A large body of work shows that naïve listeners are able to perceive boundaries consistently in spontaneous speech, in a variety of languages without previous syntactic or prosodic knowledge ([22]; [23]; [24] for English). Some studies have compared the ratings of naïve listeners with those of experts and show strong agreement rates (e.g. [25]; [4]).

2.2. Prosodic cues associated with boundary perception

One of the strongest cues to boundary perception throughout languages is the presence of a silent pause [26]; [27]. Duration is often presented as the decisive predictor, since the influence of other cues increases as soon as silent pause duration decreases [9]. A silent pause longer than 200 milliseconds becomes a decisive cue for spontaneous speech [7]. Although subject to variations, this threshold is widely accepted in the literature [8].

Final syllable lengthening [28] also plays an important role in boundary perception. Stressed vowels positioned immediately before boundaries are significantly longer than those positioned elsewhere in speech segments [20]; [21]. However, final syllabic lengthening often appears in correlation with other prosodic cues, such as silent pauses or initial pitch upsteps [27].

The association between falling vs. rising pitch contours and boundary perception is less stable [4]. In French however, rising contours tend to be associated with the perception of weak boundaries while falling contours tend to be associated with strong boundaries [29]. [30] reports that pitch movement amplitude is a better correlate for perceived boundary strength than pitch movement direction.

Throughout a vocal paragraph (i.e. a group of tone-units forming a global intonation movement [31]), pitch height naturally declines in a progressive manner. The reset of this declination line is also a cue to the presence of a boundary [10].

Although intensity plays a role in boundary perception, it is not measured in this study. Disfluencies are also excluded both from our stimuli and the analysis, given the shortness of our stimuli and their important number of variables. [32] found that the presence or absence of some of these phonetic parameters in conjunction with each other participates in the perception of stronger or weaker boundaries in conversational speech.

2.3. Assigning prosodic boundary strength

In perception experiments, the point scales presented to listeners for boundary strength assignment vary in size depending on the theoretical principles about prosodic boundaries. Scales generally include four [25] or five [33] degrees although some studies have presented listeners with a larger point scale [11]. The influence of the Autosegmental-metrical theory and its notation system ToBI (Tone and Break Indices [34]; [35]) leads to the adoption of a five-point scale based on ToBI's Break Index.

Other studies investigating the degree of consensus in boundary perception among various speech units choose to identify consensus boundaries at locations where a certain proportion of the participants have identified a boundary [25]; [30]. Boundary strength can also be calculated as the proportion of subjects indicating that they perceived a boundary at a given location, and expressed as a value between 0 and 1 [4]; [36].

3. Method

3.1. Perceptual experiment design and hypotheses

In our perceptual experiment, participants had to listen to short samples of spontaneous speech on an online platform. Extracts in which prosodic and lexico-syntactic information was available were mixed with stimuli with delexicalized speech (the delexicalization process is described below). They were instructed to rate the presence or absence of a "boundary in speech" at a specific location, specified by a star in the on-screen transcription of each extract. The delexicalized stimuli read as follows: "X*X". Participants had to rate each stimulus on a 5-point scale, from "no boundary" to "uncertain", weak boundary", "boundary" and "strong boundary". A boundary was defined as "any cue which acts as a separator between parts of speech".

Our hypotheses were as follows: (1) perceived boundary strength varies in function of syntactic type; (2) boundary perception does not rely on the same prosodic cues depending on the syntactic type.

3.2. Corpus and stimuli preparation

The speech stimuli were extracted from the ENVID corpus [5], gathering 2 hours and 10 minutes of conversational British English. Two groups of 3 authentic stimuli were selected, with an average duration of 4 seconds (min: 2.1, max 7.2). Each stimulus contained three tone-units: one corresponding to the subordinate construction SC (1 appositive relative clause, 1 adverbial clause, 1 restrictive relative clause), one corresponding to its left co-text L (i.e. the tone-unit immediately preceding SC), one corresponding to its right co-text R (i.e. the tone-unit immediately following SC). Each extract could then be described as an L-SC-R sequence. One group of stimuli (2 extracts from a female speaker, 1 extract from a male speaker) did not feature any prosodic cue associated with a boundary, and one group (1 extract from a female speaker, 2 extracts from a male speaker) contained stimuli featuring a clear prosodic break in between L and SC, with a combination of three boundary cues (silent pause between L and SC, final lengthening on L's last stressed vowel, pitch upstep on SC's first syllable). These two sets of stimuli were validated by a pre-test on 6 naïve listeners: no boundary was annotated on the first set (inter-rater agreement 100%), while a boundary was annotated for each stimulus of the second set (inter-rater agreement 100%). 10 distractors were also selected as L-SC-R sequences, but their prosodic cues were not controlled.

3.3. Stimuli manipulation

3.3.1. Synthesis

While the set of authentic stimuli containing boundaries was left intact, the set of authentic stimuli without any prosodic break was duplicated and subject to five different manipulations, realised in Praat [37] by two experts.

A 500-ms silent pause was added between L's final syllable and SC's first syllable. The silent pause was taken from the original dialogue. Its duration was based on the average length of silent pauses in the corpus (0.56 seconds). Given the short duration of our extracts, we made sure the silent pause remained shorter than speech duration.

L's final stressed syllable was lengthened by 50% [8] with the addition of duration points in Praat. We relied on the lengthening threshold established in [5], depending on each syllable's phonemic weight, to make sure each syllable reached the last threshold, corresponding to a "very significant" lengthening.

A falling tone was added to L's nuclear contour by modifying its pitch trajectory with the curve stylization function. The excursion starts at the onset of the target syllable

and ends after its coda. Its amplitude varies from 1.5 to 4 semitones [3]; [38] depending on the extracts: the movement had to be clear but the extract had to stay natural. Pitch height adjustments were performed if the modification affected other melodic parameters (i.e. pitch resets, octave jumps).

The exact same process was applied to add a rising tone on L's nuclear contour.

Finally, a pitch reset was created between L's final syllable and SC's first syllable. The gap varies between 4 and 6 semitones [38] depending on the extracts. Similar adjustments to those made for pitch contours were performed if the reset affected other melodic parameters.

Our set of manipulated stimuli then includes:

- 3 extracts with a silent pause
- 3 extracts with final syllable lengthening
- 3 extracts with a falling tone
- 3 extracts with a rising tone
- 3 extracts with a pitch reset.

3.3.2. Filtering

A delexicalized version of the total 21 (authentic and manipulated) stimuli was created with the Pass Hann band filter, only giving access to the prosodic content of speech. Given the fact that intensity is not measured or controlled in this study, we varied the filtered frequencies from 0 to 350/400 Hz depending on the segments, so as to limit the phonotactic cues showing through.

3.4. Participants and procedure

Our 52 stimuli (12 authentic and 30 manipulated stimuli + 10 distractors) were randomized and presented sequentially to 30 British listeners (aged 19 to 45, mean = 24) via a specifically designed web interface (eSurv; [39]), permitting to run perception experiments through the internet using a standard web browser. The test lasted approximately 40 minutes.

Half of the participants were recruited online, via mailing lists and social media, while the other half were recruited in person in Bristol, UK. For the latter half, the test was realised in presence of one of the experimenters. Preliminary questions on the web interface secured the fact that all participants were native speakers of English, had no hearing deficiency, had no experience in prosodic annotation, and had headphones plugged in. In total, 30 participants completed the task. However, 8 (all from the first half of participants) were excluded from the analysis, for not fulfilling one or more of the mentioned criteria.

Participants were first presented with a short description of the study, then had to answer the aforementioned questions. They were then presented with 10 sound extracts per page, each extract featuring its orthographic transcription (except the delexicalized stimuli). The transcript did not feature any punctuation, but featured a star to indicate the location of the potential boundary to be identified. Participants had to tick one of the five boxes of the point scale to rate boundary strength. They could play the sound file any number of times they wanted.

3.5. Data analysis

We computed successive ratios respecting the five-point scale according to each group of stimuli. A series of Generalized Linear Mixed Models fit by maximum likelihood estimation using the R 3.4.0 statistical programming language [40] and the lme4 package [41] was performed on the data to account for the degree of influence of each separate variable on boundary perception.

4. Results and Discussion

The first result to be noted is that the difference in ratings between the authentic stimuli and the manipulated stimuli is not significant for any variable. This means that our naïve listeners rate boundaries in both types of stimuli in an equal way.

We first explored possible interactions among the three syntactic types (fixed factor = Type; values = appositive, adverbial, restrictive) and boundary weight assignment (fixed factor = Ratings; values = Strong boundary, Boundary, Weak boundary, Uncertain, No boundary). Interactions were found for "strong" and "weak" boundaries. The main effect of strong boundaries (fixed factor = Strong; values = yes, no) was significant for appositive clauses ($ß = 1.72$, SE = .27, $p = .0001$) only. However, the differences with adverbial and restrictive relative clauses are not significant. Nonetheless, as seen in Table 1, this same main effect is not significant for the same stimuli containing filtered speech. Likewise, the main effect of weak ratings (fixed factor = Weak; values = yes, no) was significant for adverbial clauses ($ß = 0.36$, SE = .17, $p = .05$) only. Although the difference is only significant with appositive clauses ($ß = -1.46$, SE = .27, $p < .0001$), this same main effect is not significant for the same stimuli containing filtered speech. No significant interaction was found for restrictive relative clauses, while in the same stimuli containing filtered speech, the main effect of strong ratings was significant for restrictive relative clauses ($ß = 0.65$, SE = .31, $p < .05$).

Table 1. *Significant interactions between syntactic types and boundary ratings in non-filtered and filtered speech (*** = p < .0001; * = p < .05).*

	no filter	filter
appositive clauses	strong boundary***	n.s
adverbial clauses	weak boundary*	n.s
restrictive relative clauses	n.s	strong boundary*

While appositive clauses are significantly associated with the perception of strong boundaries and adverbial clauses are significantly associated with the perception of weak boundaries, restrictive relative clauses are not associated with boundary perception. These results cannot be explained in terms of constituent length, sequential position, and discourse status, as these parameters were controlled in the study. However, constituent complexity and weight remain potential factors. Although syntax and prosody do not necessarily correspond, our results are in accordance with macro-syntactic results, as appositive clauses are considered to be macro-syntactically detached from the matrix clause [42].

The fact that adverbial clauses are associated with a weak boundary also finds an echo on macro-syntactic grounds, since adverbial clauses have been shown to feature a flexible macro-syntax [43]. They often show partial autonomy, in which syntactic freedom does not necessarily correspond to pragmatic freedom [44].

Restrictive relative clauses are not associated with prosodic boundary perception, while their filtered counterpart is

associated with the perception of strong boundaries. Given the fact that subjects labelling prosodic boundaries in their own language usually perceive more boundaries than subjects annotating delexicalized speech [45], this result suggests that restrictive relative clauses are negatively correlated with boundary perception. In terms of macro-syntax, they have been shown to be mostly fully integrated to their left co-text, as they are in majority syntactically governed [42].

We then explored possible interactions between isolated prosodic boundary cues and boundary weight assignment, in our three syntactic types. As shown in Table 2, appositive and adverbial clauses are quite similar in terms of interactions, but restrictive relative clauses show differences in orders and levels of interactions. Silent pauses are the strongest cue for boundary perception in all three syntactic types. However, the main effect of silent pauses is significant for the "strong" rating in the case of appositive ($ß = 2$, SE $= .40$, $p = .0001$) and adverbial clauses ($ß = 1.67$, SE $= .39$, $p = .0001$), while it is significant for the "boundary" rating in the case of restrictive relative clauses ($ß = 1.04$, SE $= .32$, $p = .005$). This is in accordance with our previous findings: boundaries are less perceived in restrictive relative clauses, even with the strongest cue. In filtered speech, the main effect of silent pauses is diminished for appositive and adverbial clauses (only significant for "no boundary" ratings; appositive clauses; $ß = -1.38$, SE $= .48$, $p = .005$; adverbial clauses: $ß = -2.53$, SE $= .67$, $p = .0001$), while it remains similar for restrictive relative clauses (significant for "boundary" ratings; $ß = 0.79$, SE $= .40$, $p = .05$).

Table 2. *Significant interactions between boundary cues and syntactic types (*** = p < .0001; ** = p < .005; * = p < .05).*

	appositive clauses	adverbial clauses	restrictive relative clauses
(1)	pause*** "strong"	pause*** "strong"	pause** "boundary"
(2)	final lengthening* "strong"	final lengthening* "strong"	pitch reset* "boundary"
(3)	n/a	falling tone* "weak"	final lengthening* "no boundary"

The second main effect also differs for restrictive relative clauses only. In the case of appositive and adverbial clauses, the main effect of final syllable lengthening is significant for "strong" ratings (appositive clauses: $ß = 0.78$, SE $= .38$, $p = .05$; adverbial clauses: $ß = 1.13$, SE $= .42$, $p = .05$). In the case of restrictive relative clauses however, the second strongest main effect is that of pitch reset, to a lesser extent since it is only significant for "boundary" ratings ($ß = 0.8$, SE $= .37$, $p = .05$). In conversational speech, pitch reset is uncommon in restrictive relative clauses [5]. While they are in majority produced in a common tone-unit with L, the others are typically preceded with a continuation pitch contour [5]. The end of L and the beginning of SC are then usually uttered at the same pitch height. The interaction is then not based on any production-based learning effect involving a strict correspondence between a syntactic type and a prosodic feature.

No other main effect between boundary cues and ratings were found in the case of appositive relative clauses. This means that boundary cues used in appositive clauses only play a role in the identification of a "strong" boundary. Production studies have shown that silent pauses and final syllable lengthening are typical features of sequences containing appositive clauses [5]; [6], contrary to adverbial and restrictive relative clauses. However, although appositive clauses are also typically preceded with falling tones on L's nuclear syllable [5] and that they are commonly produced with a change in pitch height [5], or with a higher onset [6], falling or rising tones do not play any significant role in boundary perception. Pitch reset as an isolated cue is not used either.

Adverbial clauses and restrictive relative clauses show a third interaction. The fact that these two types show an important number of interactions with prosodic cues despite their weaker association with prosodic boundaries in general are evidence of naïve listeners' sensitivity to prosodic cues [24]. In the case of adverbial clauses, the main effect of falling tones is significant for "weak" ratings ($ß = 1.5$, SE $= .48$, $p = .05$). A falling tone is less reliable in boundary perception than a silent pause or final syllable lengthening, but is still identified as a boundary cue in this syntactic type. In conversational speech, adverbial clauses are commonly preceded with a falling tone on L [5], usually linked with the orientation of the information flow [46].

In the case of restrictive relative clauses, the main effect of final syllable lengthening is significant for "no boundary" ratings ($ß = -1.09$, SE $= .39$, $p = .01$), as there are less "no boundary" ratings for stimuli with final syllable lengthening than for those without. Final syllable lengthening then does not play a role in boundary weight assignment, but is still useful in boundary identification. However, in conversational speech, final syllable lengthening rarely occurs on L in sequences containing restrictive relative clauses [5].

5. Conclusion

We have presented an experiment to study the perception of prosodic boundaries by naïve listeners, in extracts of spontaneous speech containing syntactic subordinate structures We have confirmed the following of our hypotheses: (1) prosodic boundary perception varies across different syntactic types of subordination; (2) the different syntactic types are not distinguished with the same prosodic boundary cues. Future developments of the study will aim at establishing a stronger statistical model, which will focus on the role of individual differences and attentional fatigue in perception, and on determining the nature of the association between each variable and prosodic boundary strength in a more specific way.

6. References

[1] L. A. Fromont, S. Soto-Faraco and E. Biau, "Searching high and low: prosodic breaks disambiguate relative clauses", *Frontiers in Psychology*, vol. 8, 2017. [Accessed 6 February 2017 from http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00096/full].

[2] J. Pierrehumbert and M. E. Beckman, "Intonational structure in Japanese and English", *Phonology*, vol. 3, no. 1, pp. 255–309, 1986.

[3] A. C. Ritveld and C. Gussenhoven, "On the relation between pitch excursion size and prominence", *Journal of Phonetics*, vol. 13, pp. 299–308, 1985.

[4] A-C. Simon and G. Christodoulides, "Perception of Prosodic Boundaries by Naïve Listeners in French", in *Proceedings of Speech Prosody 2016*, Boston, USA, 2016, pp. 1158–1162.

[5] M. Lelandais and G. Ferré, "Prosodic boundaries in subordinate syntactic constructions", in *Proceedings of Speech Prosody 2016*, Boston, USA, 2016, pp. 183–187.

[6] C. Auran and R. Loock; "The prosody of discourse functions: the case of Appositive Relative Clauses in spoken British English", *Corpus Linguistics and Linguistic Theory*, vol. 7, no. 2, pp. 181–201, 2011.

[7] P. Fraisse, *La Psychologie du Rythme*. Paris: PUF, 1974.

[8] D. Duez, "Perception of silent pauses in continuous speech", *Language and Speech*, vol. 28, pp. 377–389, 1985.

[9] Y. Mo and J. Cole, "Perception of prosodic boundaries in spontaneous speech with and without silent pauses", *Journal of the Acoustical Society of America*, vol. 127, no. 3, p. 1956, 2010.

[10] J. t' Hart, R. Collier, and A. Cohen, *A Perceptual Study of Intonation.* Cambridge: CUP, 1990.

[11] R. Collier, J. R. de Pijper, and A. Sanderman, "Perceived prosodic boundaries and their phonetic correlates", in *Proceedings of the workshop on Human Language Technology*, Stroudsburg, USA, 1993, pp. 341–345.

[12] M. Ostendorf, P. J. Price, J. Bear, and C. W. Wightman; "The use of relative duration in syntactic disambiguation", *Proceedings of the workshop on Speech and Natural Language*, Hidden Valley, USA, 1990, pp. 26–31.

[13] J. P. Gee and F. Grosjean, "Performance structures: A psycholinguistic and linguistic appraisal", *Cognitive Psychology*, vol. 15, no. 4, pp. 411–458, 1983.

[14] D. Watson and E. Gibson, "Intonational Phrasing and constituency in language production and comprehension", *Studia Linguistica*, vol. 59, no. 2–3, pp. 279–300, 2005.

[15] J. Svartvik and R. Quirk, *A Corpus of English Conversation.* Lund, Sweden: Lund University Press, 1980.

[16] C. Auran, C. Bouzon, and D. Hirst, "The AixMARSEC project: an evolutive database of spoken English", in *Proceedings of Speech Prosody 2004*, Nara, Japan, 2004, pp. 561–564.

[17] C. Gussenhoven and A. Rietveld, "Intonation contours, prosodic structure and preboundary lengthening", *Journal of Phonetics*, vol. 20, no. 3, pp. 283–303, 1992.

[18] D. R. Ladd and N. Campbell, "Theories of prosodic structure: evidence from syllable duration", in *Proceedings of the 12th International Congress of Phonetic Sciences*, Aix-en-Provence, France, 1991, pp. 290–293.

[19] R. D. Ladd, *Intonational Phonology*. Cambridge: CUP, 2008.

[20] S. A. Jun and J. Fletcher, "Methodology of studying intonation: From data collection to data analysis," in *Prosodic Typology II.* New York: Oxford University Press, 2014, pp. 493–519.

[21] A. Katsika, J. Krivokapic, C. Mooshammer, M. K. Tiede, and L. Goldstein, "The coordination of boundary tones and its interaction with prominence", *Journal of Phonetics*, vol. 44, pp. 62–82, 2014.

[22] T. J. Yoon, S. Chavarria, J. Cole, and M. Hasegawa-Johnson, "Intertranscriber Reliability of Prosodic Labeling on Telephone Conversation using ToBI", in *Proceedings of Interspeech 2004,* Jeju, South Korea, 2004, pp. 2722–2732.

[23] Y. Mo, J. Cole, and E. K. Lee, "Naïve listeners' prominence and boundary perception", in *Proceedings of Speech Prosody 2008*, Campinas, Brazil, 2008, pp. 735–738.

[24] J. Cole, T. Mahrt, and J. I. Hualde, "Listening for sound, listening for meaning: Task effects on prosodic transcription", in *Proceedings of Speech Prosody 2014*, Dublin, Ireland, 2014, pp. 859–863.

[25] C. Auran, A. Colas, C. Portes and M. Vion, "Perception of breaks and discourse boundaries in spontaneous speech: developping an on-line technique", in *Proceedings of IDP05,* Aix-en-Provence, France, 2005, pp. 1–7.

[26] T. J. Yoon, J. Cole, and M. Hasegawa-Johnson, "On the edge: acoustic cues to layered prosodic domains" in *Proceedings of ICPhS XVI*, Saarbrücken, Germany, 2007, pp. 1264–1267.

[27] M. Wagner and D. G. Watson, *Language and Cognitive Processes,* vol. 25, no. 7-9, pp. 905-945, 2010.

[28] Y. Mo, "Duration and intensity as perceptual cues for naïve listeners' prominence and boundary perception", in *Proceedings of Speech Prosody 2008*, Campinas, Brazil, 2008, pp. 739–742.

[29] C. Portes, "Approche instrumentale et cognitive de la prosodie du discours en français", *TIPA*, vol. 21, pp. 101–119, 2002.

[30] C. Smith, "Naïve listeners' perceptions of French prosody compared to the predictions of theoretical models", in *Proceedings from IDP09*, Paris, France, 2009, pp. 335–349.

[31] A. Cruttenden, *Intonation.* Cambridge: Cambridge University Press, 1986.

[32] D. Barth-Weingarten, *Intonation Units Revisited. Cesuras in talk-in-interaction*. Amsterdam: John Benjamins, 2016.

[33] K. Carlson, C. Clifton, and L. Frazier, "Prosodic boundaries in adjunct attachment", *Journal of Memory and Language*, vol. 45, no. 1, pp. 58–81, 2001.

[34] J. B. Pierrehumbert, *The phonology and phonetics of English intonation*. Boston, USA: MIT, 1980.

[35] K. E. A. Silverman, M. B. Beckman, J. F. Pitrelli, M. Ostendorf, C. W. Wightman, P. Price, J. B. Pierrehumbert, and J. Hirschberg, "TOBI: a standard for labeling English prosody", in *Proceedings of ICSLP 92*, Banff, Canada, 1992, pp. 867–870.

[36] J. Cole, Y. Mo, and S. Baek, "The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech", *Language and Cognitive Processes*, vol. 25, no. 7-9, pp. 1141–1177, 2010.

[37] P. Boersma and D. Weeninck, "Praat: doing Phonetics by Computer". http://www.fon.hum.uva.nl/praat/, [Accessed 30 January 2013].

[38] S. Kakouros and O. Räsänen, "Statistical Learning of Prosodic Patterns and Reversal of Perceptual Cues for Sentence Prominence", in *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, Philadelphia, USA, 2016, pp. 1–6.

[39] "eSurv", http://esurv.org/, [Accessed 10 May 2017].

[40] R Core Team, "A language and environment for statistical computing". R foundation for statistical computing. http://www.r-project.org/, [Accessed 29 June 2017].

[41] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Linear mixed-effects models using eigen and s4". http://cran.r-project.org/, [Accessed 29 June 2017].

[42] M. De Vries, "The syntax of appositive relativization", *Linguistic Inquiry*, vol. 37, no. 2, pp. 229-270, 2006.

[43] O. Ehmer, "Adverbial patters in interaction", *Language Sciences,* vol. 58, pp 1-7, 2016.

[44] N. Evans, "Insubordination and its uses", in *Finiteness: Theoretical and Empirical Foundations*. Oxford: Oxford University Press, 2007.

[45] A. Mettouchi, A. Lacheret-Dujour, V. Silber-Varod, and I. 'el Shlomo, "Only prosody? Perception of speech segmentation in Kabyle and Hebrew", *Nouveaux Cahiers de Linguistique Française*, vol. 28, pp. 207-2018, 2007.

[46] W. Chafe, "How people use adverbial clauses", in *Proceedings of the Tenth Annual Meetings of the Berkeley Linguistics Society,* Berkeley, USA, 1984, pp. 437-449.