Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

Naïve listeners' perception of prominence and boundary in French spontaneous speech

Guillaume Roux[1], *Roxane Bertrand*[1], *Alain Ghio*[1], *Corine Astésano*[2]

[1] Laboratoire Parole et Langage UMR7309-CNRS, Aix en Provence, France.
[2] URI Octogone-Lordat EA4156, Université Toulouse Jean Jaurès, Toulouse, France.
guillaume.roux@lpl-aix.fr, roxane.bertrand@lpl-aix.fr, alain.ghio@lpl-aix.fr, corine.astesano@univ-tlse2.fr,

Abstract

Our main goal here is to explore the link between naïve listeners' perception of prominences and boundaries in spontaneous speech and experts' annotation of prosodic hierarchy and accentuation in French. We first present the design of our corpus, which consists in 133 utterances extracted from the Corpus of Interactional Data (CID). 73 naïve listeners judged prominences and boundaries using three levels of prominence and boundary ("none", "weak" and "strong") during two separate tasks. Prominence-Scores and Boundary-Scores reveal good reliability between listeners. With a strong agreement between the two experts' annotation, we then examine the extent to which naïve judgments are in line with experts' annotations.

Index Terms: speech perception, prominences, boundaries, spontaneous speech, naïve listeners, French.

1. Introduction

This work focuses on naïve listeners' perception of prominences and boundaries in French spontaneous conversational speech and their links to experts' annotation of prosodic hierarchy and accentuation. Are naïve listeners able to perceive prominences and boundaries? Are they able to distinguish and order several levels according to experts' annotation? No studies on French have, to our knowledge, tackled the issue of prosodic perception with a three-leveled granularity.

Listeners' prominence and boundary perception in spontaneous speech has been studied in Dutch [1, 2], American English [3, 4, 5, 6, 7] and French [8]. These works showed that listeners judge efficiently boundaries and prominences in spontaneous speech with a strong agreement, but with higher agreement regarding boundaries than prominences; strong agreement was also found amongst naïve listeners and trained listeners [9, 10]. Correlations have also been established with phonetic cues such as duration and intensity [7], pre-nuclear/nuclear prominences [6] or between spontaneous and data-driven speech in French, on radiophonic data and Maptask [8]. In French, observations have been made in controlled experiments and laboratory speech for the development of automatic detection of prosody [9, 10].

Models of French prosody [11, 12, 13, 14] propose a final accent (FA) and an optional Initial Accent (IA) [14, 15], FA being post-lexical and syncretic to boundaries in French. Various studies assume that prominences and boundaries can be seen as the same underlying phenomenon [16, 17]. FA and IA are seen as right and left markers of the prosodic structures. The prosodic hierarchy is defined by two levels of phrasing: a minor prosodic phrase or accentual phrase (AP) and an intonational phrase (IP), the highest in the hierarchy. Recent studies have experimentally shown a third level of phrasing (intermediate phrase-ip) [18]. While ip requires to be refined, particularly for spontaneous speech, it has been suggested that a third level seemed necessary in spontaneous data [19].

The main goal of this research is to observe the links between naïve listeners' perception of prominences and boundaries and an experts' annotation of prosodic hierarchy and accentuation. We first wish to account for the consistency of naïve untrained listeners in perceiving prominences and boundaries in French spontaneous conversational data. We hypothesize that listeners perceive boundaries with a higher agreement than prominences as previous studies have shown [2, 5 among others]. Previous studies have investigated boundary and prominence perception according to two levels only: presence or absence. In this work, listeners were asked to identify prominences and boundaries on three-level scale: none, weak and strong.

Two experts then annotated the prosodic hierarchy (AP, ip and IP) and located prominences (IA and FA). Our goal is to investigate the link between naïve listeners' scores and experts' annotation, and to uncover whether naïve listeners are consistent in perceiving the AP, ip, IP hierarchy, and both accents.

2. Method and analysis

2.1. Corpus

133 utterances were extracted from the CID (Corpus of Interactional Data) [20], consisting of 2778 tokens and 3395 syllables. We selected short and long utterances, varying from 4 to 74 syllables and 3 to 15 seconds, from the 16 speakers of the CID.

The CID is an audio video recording of French spontaneous face-to-face conversations. Speakers were recorded in an anechoic room and each of them was equipped with a headset microphone enabling the recording of the two speakers' voice on two different sound tracks. This results in a high quality of speech allowing a very fine-grained analysis at the different levels of speech. Based on an orthographic transcription, numerous annotations were performed at the different linguistic levels [21, 22]. For this study, we used the morpho-syntactic annotation provided by the stochastic parser Marsatag [23] which provides for each part-of-speech token an automatic annotation of its morpho-syntactic category. From this annotation, we extracted all the Noun Phrases (NP) in the

10.21437/SpeechProsody.2016-187

CID (10735) and then selected the three most frequent patterns (Determiner + Noun + Adjective, *n*= 403; Determiner + Adjective + Noun, *n*= 462; Determiner + Noun + Preposition “de” + Noun, *n*= 430). These patterns can exhibit three levels of boundaries (IP, ip and AP) [18, 24]. Also, according to [25], a long NP could favor the presence of an ip.

The syntactic functions of NP have been manually annotated by two experts. A Cohen’s kappa measured the inter-raters reliability. The kappa score reflected strong agreement (0.877). Direct object was the most frequent NP function (51.5% of the complements; *n*= 514). Choosing homogeneous morpho-syntactic patterns was meant to avoid the variability linked to other elements (spontaneous utterances, the size of the utterances, various speakers, etc.). Among 514 direct objects, we selected 133 large utterances in which these 3 types of NP occurred. Our choice was based on factors such as audibility, understandability, and no syntactic and semantic ambiguity. Moreover, we excluded the NP with disfluencies.

2.2. Experimental procedure of the perception task

73 naïve listeners without knowledge in phonetics and phonology participated to the experiment. They were between 18 and 55 years-old and did not have any auditory problems. 8 listeners could simultaneously take part in the experiment on 8 separate monitor-driven computers.

The experiment used the PERCEVAL software [26], a computer-driven system for experimentation on auditory and visual perception (developed at the LPL-CNRS lab in Aix-en-Provence). The auditory part was conducted using the LANCELOT program, run by PERCEVAL. Two scripts were created, one for each task. A three-degree evaluation scale was proposed: “rien, faible, fort” (none, weak, strong). “None” was ticked by default.

The prominence task was based on syllables, not chunks [2, 6 among others]. The utterances were cut out in syllables separated by a space, orthographically transcribed, without any punctuation nor capital letters. The judgment boxes to tick were put under each syllable. For the boundary task, the utterances were cut out in words separated by a space, without any punctuation nor capital letters. The judgment boxes were put between each word. For each task, except inside the NP, disfluencies such as repetition, filler pause, syllabic lengthening can occur in the utterances.

Listeners were divided in 4 blocks of 18, each block divided in 2 groups of 9 participants, except for block 2 which had 19 participants (10 in group 2 and 9 in group 2’). Groups of 9 were constituted to balance the number of listeners in each group. Listeners were assigned two tasks (prominences and boundaries) on the same utterances presented in a random order in each task, with 4 different sets of 33 utterances presented in 4 separate blocks. Each set of utterances was thus heard by 18 listeners (19 in block 2). Each block contains a different set of utterances. The experimenter gave complementary oral instructions to those presented on the screen before each task. For the prominence task, they were asked to focus on the musical salience of syllables. For the boundary task, they were asked to focus on a feeling of break in the utterance. Two hours were necessary to perform both tasks. They could have a break between each task and in the middle of each task. Each listener could listen to the utterance a maximum of 10 times by clicking on the sound logo. To perform their judgment, they had to click on the scaled buttons under salient syllables or between breaks, depending on the task they were doing. After their judgment, they had to validate their evaluation or they could click on “Effacer” (Clear) to start again.

2.3. Experts’ annotation

The 133 utterances were originally transcribed on Praat [27]. For this study, IP, ip and AP were aligned with tokens, and IA and FA were aligned with syllables by the two experts, separately. The experts annotated four levels of boundaries (IP, ip, AP and “none”) and three kinds of accentuation (IA, FA and “none”). Once we had estimated the agreement between the two experts, their annotation was compared to naïve listeners’ scores.

3. Results and discussion

3.1. Reliability measurements

3.1.1. Kappa scores

Inter-raters reliability was first measured on all sites (3395 syllables and 2617 intervals), with Fleiss’ kappa for each block of 18 participants (19 for the second block), and with Cohen’s kappas on pairs of raters in each block (590 pairs for the prominence and boundary tasks). Fleiss’ kappa scores and most of pair-wised Cohen’s kappa scores were low, probably due to the inherent difficulty of the tasks. Complex factors are indeed at play: the number of sites to be judged, the variability of utterances’ lengths (4 to 74 syllables and 3 to 15 seconds), speakers’ variability (16 different speakers), and the three-level judgment for each site (none, weak, strong).

3.1.2. B-scores and p-scores

We thus chose to measure inter-reliability agreement with prominence-scores (P-scores) and boundary-scores (B-scores) [5]. The new scores were obtained for each site by adding the number of listeners that perceived a prominence or a boundary on this site. This score can vary from 0 to 18 for blocks 1, 3 and 4, and 0 to 19 for block 2. For block 2, we had to convert the scale to 18 raters. As a first step, scores were brought down to presence/absence (weak+strong vs. none), in order to measure prominences and boundaries similarly to previous studies [5]. Then, we measured each level (none, weak, strong), keeping the weak and strong distinction in order to determine the relevance of our three-level granularity, and to have a finer-grained analysis of prominences and boundaries perception. The level was determined by the majority threshold ($n \geq 10$ listeners). Figures 1 and 2 illustrate the P-Score and B-Score calculations, respectively. A score of 10 on the blue line is when most of the participants chose “weak”, a score of 10 on the red line is when most chose “strong” and a score of 9 or lower on the blue and the red lines mixed is when most chose “none”. For example, in Figure 1, the first syllable “et” has a score of 0 on the red line and of 5 on the blue line: the total score of 5 shows that most participants chose “none”. The syllable “bref” has a score of 16 on the red line, which indicates the presence of a strong prominence. The last syllable “rrons” has a score of 10 on the blue line which shows a weak prominence.

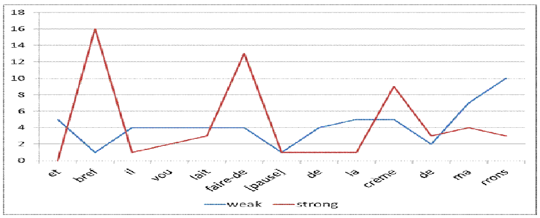

Figure 1: *P-Scores on utterance n°3 on the 4th group.*

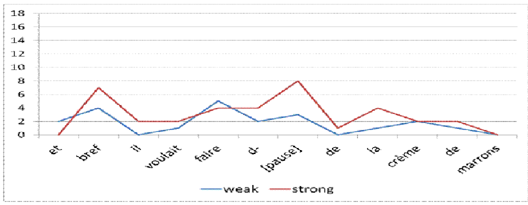

Figure 2: *B-Scores on utterance n°3 on the 4th group.*

3.2. Naïve listeners' B-scores results and discussion

96% of B-scores were well identified. Category "none" constitutes the largest category (84%). The majority of judges scored 12% of the remaining sites as a weak or strong or weak+strong boundary. Among those, only 1.3% were scored by a majority of judges as strong boundaries and only 0.9% predominantly as weak boundaries. It means that only a small percentage of boundaries are recognized by a majority of judges as clearly strong or weak. Most were judged weak+strong without reaching a majority of the scores for weak or strong only (9.8%). The remaining 4% of sites are ambiguous, that is 50% of the listeners chose "none", 50% "weak" or "strong" or "weak"+"strong".

3.3. Naïve listeners' P-scores results and discussion

94% of P-scores were well identified, with "none" constituting the largest category (70%). The majority of judges scored 24% of the remaining syllables as prominent, whether weak or strong or weak+strong. Among them, 6.3% are judged as strong prominences by a majority of judges and 0.6% as weak prominences. It means that only a small percentage of prominences are clearly identified as strong or weak by a majority of judges: 17% of prominences are weak+strong and do not reach consensus as to their weak or strong status. The remaining 6% of syllables are ambiguous.

There are differences between perception of prominences and boundaries ($X^2 = 309.3481$, df = 4, $p < .001$). Weak prominences and boundaries are the less predominantly well-identified events. Altogether, prominences are more likely to be judged as strong than boundaries ($X^2 = 38.9188$, df = 1, $p < .001$). They indeed hardly reach a majority of judgments. Also, prominences have more ambiguous judgments than boundaries ($X^2 = 20.2088$, df = 1, $p < .001$).

3.4. Experts' annotation

Two experts annotated all the utterances. Cohen's kappa scores between the pair of experts were 0.94 for IP/ip, 0.90 for AP and 0.89 for AI/AF. According to the experts' annotation, a third expert annotation based on the sum of the two annotations was done, to avoid ambiguous annotations. Only annotations with perfect agreement were kept. In the end, we selected 2592 tokens' sites and 3495 syllables with perfect agreement between the two experts.

3.5. Naïves' and Experts' links on boundaries

We wished to test for the correspondence between the sites annotated by the experts and naïve listeners' judgments. Experts' annotation provides categorical predictor variables for ANOVA tests.

As stated before, B-scores and P-scores for predominantly weak or predominantly strong judgments were very low. Moreover, the scale distinguishing weak and strong makes it difficult to precisely categorize a value on a continuous scale. It also overestimates the weight of "strong" judgments: for example, how should a prominence or a boundary with 6 weak and 4 strong or 9 weak and 1 strong be categorized on a continuous scale? For those reasons, a scale from 0 to 18 was chosen to account only for "presence"/"absence". We postulated that the number of answers determine the hierarchical categories of the experts' annotation.

First, to measure the reliability between b-scores and experts' annotation of prosodic hierarchy, we used a one-way ANOVA with repeated measures between IP, ip, AP and "none". Results show significant difference (F=177.9; df=3; *p*<.001). IP, ip and "none" are also significantly distinguished from one another (F= 290.5; df= 2; *p*<.001). Post-hoc tests indicate that when listeners indeed perceive a boundary, they can distinguish between an *ip* and a stronger *IP* (IP mean = 8.84; ip mean = 6.9 and "none" mean = 4.29; F=26.91, df=1, p<.001). The same goes for the distinction between AP and "none" (AP mean = 6.58 and "none" mean = 4.29; F=338.1; df=1; p<.001), for IP and "none" (F=452.3; df=1; p<.001), for ip and "none" (F=104.8; df=1; p<.001), and for IP and AP (F=61.48; df=1; p<.001). These results indicate that naïve listeners perceive the difference between a strong boundary and a weak one. There is however no significant difference between ip and AP (F=1.876; df=1; *p*=0.171). To summarize on B-scores, our results indicate that naïve listeners perceive boundaries, which can be linked to our IP and ip/AP. They make the difference between weak and strong boundaries to a certain extent, but not between *ip* and AP. Experts' annotation is on a four-level hierarchy, which was not asked to the naïve listeners who only had three levels. Yet, they did perceive three levels, which is really good. According to the fact that *ip* is not precisely defined, its distinction from AP would not be expected by naïve listeners. However, the standard deviation of IP is higher than the one's for *ip* and AP (3.67>3.25 for *ip* and 3.62 for AP). "None" has the lowest *sd* (2.83). It indicates that "none" B-scores are the most consistent and that *ip* boundaries are the most consistent boundaries.

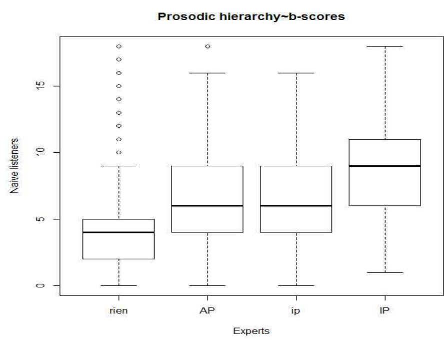

Figure 3: *Boxplots of B-scores for Experts' annotation of prosodic hierarchy ("rien" stands for "none").*

The different ANOVAs between experts' categories and naïve listeners' B-scores show that the means difference between boundaries and sites with no boundaries were significant. So, naïve listeners perceive boundaries and those boundaries are quite similar to what experts coded as IP, ip and AP. Naïve listeners differentiate between weak and strong prominence well enough to determine a difference between IP and ip/AP but not between a weak one and an intermediate one. This tendency follows our expectation (Figure 3). The absence of difference between ip and AP is in line with Tobi for French [25], suggesting that ip predominantly characterize syntactic structures. Because syntactic structure can be retrieved somewhat independently from prosodic cues, our participants may not have felt the necessity to over-specify prosodic structure's depth. This echoes the idea that prosodic cues can be quite variable or 'flexible' in the marking of syntactic structure [3].

3.6. Naïves' and Experts' links on accentuation

To measure the reliability between naïve listeners' P-scores and experts' annotation of accentuation, a one-way ANOVA was performed on P-scores, with prominences (IA, FA and "none") as the independent variable. Results indicate a significant difference between both accents' and "none" categories: more specifically, FA scores are significantly higher than those of "none" (mean scores for "None" = 5.74, and for FA = 9.98; F=979.4; df=1; p<.001); IA scores are significantly higher than those of "none" (mean score for IA = 9.95; F=227.2; df=1p<.001); IA and FA scores are however not significantly distinguished (F=0.005; df=1; p= 0.942).

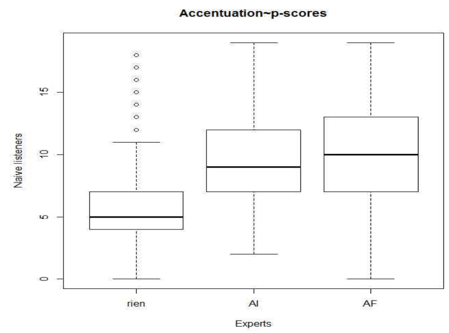

Figure 4: *Boxplots of p-scores and Experts' annotation of accentuation ('rien' stands for 'none')*

Our results indicate that naïve listeners distinguish between prominences and non-prominent syllables. They also indicate that both accents (IA and FA) are judged as equally weak or strong prominences. Hence, weak or strong prominences are not characteristic of one or the other type of accent: FA is not (metrically or perceptually) stronger than IA. However, the standard deviation is larger for FA than IA (4.02>3.68), and "none" one's is the lowest (3.15). It shows that "none" P-scores are the most consistent, and that, as far as accentual categories are concerned, IA is perceived in a slightly more consistent way than FA (Figure 4).

Globally, listeners' perception of prominence is consistent with experts' annotation of IA and FA. They do not perceive prominence on otherwise unaccented syllables. Their perception however does not help functionally distinguishing IA from FA, in terms of their perceived strength, but the more consistent *sd* for IA could be indicative that the initial accent initiate less variable strength perception.

4. Conclusion

This study was designed to observe the consistency of naïve listeners' judgment of French spontaneous speech on prominences and boundaries in a difficult experiment based on two tasks, using three levels of judgment scores. It is a first step towards a methodological assessment of prosodic events on spontaneous speech. This research allowed us to confirm that French naïve listeners can perceive prominences and boundaries with a consistent agreement, despite the three-level judgment. A task using three levels of judgment seems to be relatively difficult for naïve listeners: when binary judging for presence or absence, scores are much higher, confirming previous studies [2, 5 among others]. It also reveals that, in line with results on other languages, boundaries elicit less ambiguous judgments than prominences. Stronger prominences are also easier to perceive, thus more numerous than weak ones. The link between the experiment on naïve listeners and the experts' annotation shows a positive trend that needs to be further investigated. Our results indeed show consistent links within naïve listeners' judgments and within experts' judgments. But precise correspondence between listeners and experts for prominence judgments, and predominantly for finer-grained boundary judgment (AP *vs* ip) is more difficult to assess. It is possible that the two populations do not judge comparable events. It is thus necessary to qualitatively evaluate what influences naïve listeners' judgment. If *ip* indeed characterize syntactic structuring [25], a certain redundancy may take place between syntactic and prosodic cues, lowering listeners' ability to distinguish between the two boundary types. On the contrary, experts' annotations respect the proposed three-layered prosodic hierarchy, even on spontaneous speech, which is a noticeable result. Another task may be needed to help bring to light a comparable three-layered hierarchy in naïve listeners' judgment. Also, more work needs to be done to refine the characterization of the *ip* before linking it to perceptual results. More generally, B-scores and P-scores standard deviations indicate that the perception of boundaries is more consistently annotated by the experts than prominences. Finally, disfluences could hinder naïve listeners' judgment more than experts' [3]. Future investigations are planned to test the relationship between Part-of-Speech and boundary/prominence, and to more precisely focus on the comparison between naïve listeners' judgments and experts annotators *within* the sole NP.

5. Acknowledgments

This study is supported by the Agence Nationale de la Recherche grant ANR-12-BSH2-0001 (PI: Corine Astésano).

6. References

[1] B. M. Streefkert, L. C. W. Pols and L. F. M. ten Bosch, "Prominence in read aloud sentences as marked by listeners and classified automatically," *Proceedings of IFA* (Amsterdam), pp. 101-116, 1997.

[2] J. Buhmann, J. Casper, V. J. Heuven van, H. Hoekstra, J. P. Martens and M. Swerts, "Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus," in *Proceedings of LREC 2002* (Las Palmas), pp. 779-785, 2002.

[3] 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*, no 25(7), pp. 1141-1177, 2010.
[4] T-J. Yoon, S. Chavarria, J. Cole and M. Hasegawa-Johnson, "Intertranscriber Reliability of Prosodic Labeling on Telephone Conversation using ToBI," *Proceedings of Interspeech 2004* (Jeju), pp. 2722-2732, 2004.
[5] J. Cole, T. Mahrt and J. I. Hualde, "Listening for sound, listening for meaning: Task effects on prosodic transcription," *Speech Prosody* 7, Dublin, Ireland, 2014.
[6] Y. Mo, J. Cole and E.-K. Lee, "Naïve listeners' prominence and boundary perception," *Speech Prosody 2008*, Campinas, Brazil, pp. 735-738, 2008.
[7] Y. Mo, "Duration and Intensity as perceptual cues for naïve listeners' prominence and boundary perception," *Speech Prosody 2008*, Campinas, Brazil, pp. 739-742, 2008.
[8] C. L. Smith, « Naïve listeners' perceptions of French prosody compared to the predictions of theoretical models," in H-Y. Yoo & E. Delais-Roussarie (Eds.), *Proceedings from IDP 2009* (Paris), pp. 335-349, 2009.
[9] A. Lacheret, S. Kahane, J. Beliao, A. Dister and K. Gerdes, "Rhapsodie: un Treebank annoté pour l'étude de l'interface syntaxe-prosodie en français parlé," *4e Congrès Mondial de Linguistique Française* (Berlin, Allemagne), pp. 2675-2689, 2014.
[10] M. Avanzi, L. Rousier-Vercruyssen, S. Schwab, S. Gonzalez and M. Fossard, "C-PROM-Task : A new annotated dataset for the study of French Speech Prosody," *TRASP'2013* (Aix-en-Provence, France), pp. 27-30, 2013.
[11] S. A. Jun and C. Fougeron, "Realizations of accentual phrase in French *intonation," Probus no 14(1), pp. 147-172*, 2002.
[12] B. Post, "Tonal and phrasal structures in French intonation," *Thesis*, The Hague, 2000.
[13] A. Di Cristo, "La prosodie au carrefour de la phonétique, de la phonologie et de l'articulation formes-fonctions," *Travaux interdisciplinaires du Laboratoire parole et langage d'Aix-en-Provence*, vol. 23, pp. 67-211, 2004.
[14] P. Welby, "French intonational structure: Evidence from tonal alignment," *Journal of Phonetics*, vol. 34, pp. 343-371, 2006.
[15] C. Astésano, E. Bard and A. Turk, "Structural influences on Initial Accent placement in French," *Language and Speech*, no 50(3), pp. 423-446, 2007.
[16] J. Vaissière, "Rhythm, accentuation and final lengthening in French," in J. Sundberg, L. Nord & R. Carlson (Eds.), *Music, language, speech and brain*, New York, MacMillan Press, pp. 108-120, 1991.
[17] M. E. Beckman, "Evidence for Speech Rhythms across Languages," in Tohkura et al. (Eds.), *Speech Perception, Production and Linguistic Structure*, Tokyo, pp. 457-463, 1992.
[18] A. Michelas and M. D'Imperio, "Uncovering the role of the intermediate phrase in the syntactic parsing of French," in *17th International Congress of Phonetic Sciences*, pp. 1374-1377, 2011.
[19] I. Nesterenko, S. Rauzy and R. Bertrand, "Prosody in a corpus of French spontaneous speech: perception, annotation and prosody syntax interaction", in *Speech Prosody 2010-Fifth International Conference*, Chicago, pp. 1-4, 2010.
[20] R. Bertrand, P. Blache, R. Espesser, G. Ferré, C. Meunier, B. Priego-Valdere and S. Rauzy, "Le CID – Corpus of Interactional Data-: protocols, conventions, annotations", *Travaux interdisciplinaires du Laboratoire parole et langage d'Aix-en-Provence (TIPA)*, vol. 25, pp. 25-55, 2006.
[21] P. Blache, R. Bertrand and G. Ferré, "Creating and exploiting multimodal annotated corpora: the toma project," in *Multimodal Corpora*, pp. 38-53, 2009.
[22] P. Blache, R. Bertrand, B. Biggi, E. Bruno, E. Cela, R. Espesser, G. Ferré, M. Guardiola, D. Hirst, R. Muriasco, J. C. Martin, C. Meunier, M. A. Morel, I. Nesterenko, P. Nocera, B. Palaud, L. Prévot, B. Priego-Valverde, J. Seinturier, N. Tan, M. Tellier and S. Rauzy, "Multimodal annotation of conversational data," in *Proceedings of Linguistic Annotation Workshop*, Uppsala, Sweden, pp. 186-191, 2010.
[23] S. Rauzy, G. Montcheuil and P. Blache, "Marsatag, a tagger for French written texts and speech transcriptions," in *Proceedings of Second Asian Pacific Corpus linguistics Conference*, Hong Kong, Hong Kong, pp.220, 2014.
[24] C. Astésano, R. Bertrand, R. Espesser & N. Nguyen, "Dissociation between prominence and boundary phenomena in French : a perception study. pS-prominenceS", *Viterbo*, Italy, Dec. 12-14, 2013.
[25] E. Delais-Roussarie, B. Post,,M. Avanzi, C. Buthke, C. Di Cristo, I. Feldhausen, S. A. Jun, Ph. Martin, T. Meisenburg, A. Rialland, R. Sichel-Bazin, H. Yoo, "Developing a ToBI system for French", in Frota, S., Prietro, P. (eds). *Intonational variation in Romance*, Oxford: OUP, pp. 63-100, 2015.
[26] C. André, A. Ghio, C. Cavé and B. Teston, "PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception," *Proceedings of XVth ICPhS*, Barcelona, Spain, pp. 1421-1424, 2003.
[27] P. Boersma & D. Weenink, "Praat: doing phonetics by computer [Computer program]." Version 6.0.14, retrieved 11 February 2016 from http://www.praat.org/