INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Automatic Labelling of Prosodic Prominence, Phrasing and Disfluencies in French Speech by Simulating the Perception of Naïve and Expert Listeners

George Christodoulides[1], *Mathieu Avanzi*[1], *Anne Catherine Simon*[1]

[1]Centre Valibel, IL&C, Université catholique de Louvain, Belgium

george@mycontent.gr, mathieu.avanzi@gmail.com, anne-catherine.simon@uclouvain.be

Abstract

We explore the use of machine learning techniques (notably SVM classifiers and Conditional Random Fields) to automate the prosodic labelling of French speech, based on modelling and simulating the perception of prosodic events by naïve and expert listeners. The models are based on previous work on the perception of syllabic prominence and hesitation-related disfluencies, and on an experiment on the real-time perception of prosodic boundaries. Expert and non-expert listeners annotated samples from three multi-genre corpora (CPROM, CPROM-PFC, LOCAS-F). Automatic prosodic annotation is approached as a sequence labelling problem, drawing on multiple information sources (acoustic features, lexical and shallow syntactic features) in accordance with the experimental findings showing that listeners integrate all such information in their perception of prosodic segmentation and events. We test combinations of features and machine learning methods, and we compare the automatic labelling with expert annotation. The result of this study is a tool that automatically annotates prosodic events by simulating the perception of expert and naïve listeners.

Index Terms: automatic prosodic labelling, speech perception modelling, prosodic segmentation, disfluencies, prominence

1. Introduction

As available spoken language corpora increase in size, it becomes more important to develop reliable automatic tools for enriching these corpora with multiple annotation layers, given that it is impractical to envisage manual annotation campaigns. In this contribution, we focus on the automatic labelling of prosodic events, and more specifically prominence, phrasing and hesitation-related disfluencies in French.

Prosody is central to language comprehension, by helping listeners segment incoming speech and by indicating the information status and the discourse relations between elements of utterances (for a review on the role of prosody in comprehension see [1]; more specifically on the importance of phrasing, see [2]). There is consensus that prominence is "an umbrella term encompassing various related but conceptually and functionally different phenomena, such as phonological stress, paralinguistic emphasis, lexical, syntactic, semantic or pragmatic salience" and that a linguistic unit is prominent when it stands out of its environment because of its characteristics [3]. In French, the prosodic prominence of a syllable is of crucial importance because it essentially contributes to mark the boundaries of prosodic groups (unlike variable-stress languages), i.e. prosodically prominent syllables play a central role in phrasing. While most phonological models of French (e.g. [4], [5]) admit at least three degrees of prosodic boundaries, large-scale corpora are usually limited to one or two degrees of prosodic boundaries. Similarly to the perception of prosodic prominence, there is evidence that listeners perceive prosodic boundaries as a gradual phenomenon and in relative terms, i.e. they perceive a boundary as stronger or as weaker than the previous one.

The perception of prosodic events such as prominence and phrasing is influenced by both low-level acoustic cues and top-down expectations. Acoustic correlates have been shown to include silent pauses following the event, syllable duration (lengthening), pitch movement and relative pitch, as well as spectral features (e.g. [6], [7]). Experiments have also shown that top-down linguistic expectations influence the perception of prominence and boundaries ([8], [9], [10], [11]). Hesitation-related disfluencies also affect the perception of prosodic events ([12], [13])

Finally, there are differences between the conditions under which experts annotate prosodic phenomena on a speech recording and the real-time perception of these phenomena by naïv listeners. In real-time perception experiments, non-expert listeners tend to identify fewer prosodic boundaries than those annotated by experts: this is not only due to the task constraints, but also due to the top-down influence of syntax and semantics (e.g. [14], [15], [13], [16]).

2. Related Work

Previous work on the automatic annotation of prosodic events has mainly focused on using supervised machine learning models trained on hand-labelled corpora. Working on English speech, Wightman and Ostendorf [17] used decision trees and HMMs to detect and classify prosodic event sequences. Chen [18] proposed an annotator for pitch accents (presence or absence) and prosodic boundaries, based on ANN and GMM models trained on the Boston Radio Speech corpus; they report a 85% and 93% accuracy respectively, when using both acoustic and syntactical features. Rangarajan et al. [19] show that the accuracy of the automatic labelling for pitch accents and boundary tones is improved by jointly modelling simple lexical, syntactic and acoustic features, using a maximum entropy model (accuracy for accent and boundary tone detection: 86.0% and 93.1% on the Boston University Radio News corpus, and, 79.8% and 90.3% on the Boston Directions corpus, respectively). Rosenberg [20] developed an automatic system that associates ToBI labels to syllables, initially based on the expert annotation of the Boston Directions Corpus, while the AuToToBI system currently includes several models. Given that producing large training corpora is costly, [21] proposed a bootstrap process: using a large unlabelled corpus to perform unsupervised adaptation of an acoustic-prosodic model trained on a small labelled corpus. Previous work on the automatic annotation of prosodic events in French has focused on prosodic prominence and includes Analor [22], ProsoProm ([23], [24]) and Promise [25].

In this article we combine the results of our previous work ([25], [26], [27], [16]) in both expert and non-expert annotation

Copyright © 2017 ISCA

http://dx.doi.org/10.21437/Interspeech.2017-971

of prosodic prominence, phrasing and disfluencies, in order to develop an automatic annotation system that simulates the perception of these phenomena by expert or naïve listeners.

3. Data

3.1. Corpora

We used three corpora that have been compiled using comparable methodologies: CPROM [28], CPROM-PFC [29] and LOCAS-F [30]. All three corpora have been orthographically transcribed in Praat, and a phonetic transcription was automatically produced and aligned with the speech signal using the EasyAlign script; the aligned segmentation in phones, syllables and words was manually corrected by experts.

The **CPROM** corpus is a publicly available corpus of 24 recordings (70 minutes long) covering 7 speaking styles, produced by speakers from Belgium, Switzerland and France. It was one of the first publicly available corpora to contain an annotation of prosodic prominent syllables, cross validated by 3 expert annotators, according to the protocol described in [28].

The **CPROM-PFC** corpus consists of recordings extracted from the PFC database [31]. It includes speech material recorded in 14 geographical areas, spread over 3 European French-speaking countries: Metropolitan France (Béthune, Brécey, Lyon, Paris and Ogéviller); Switzerland (Fribourg, Geneva, Martigny, Neuchâtel and Nyon) and Belgium (Brussels, Gembloux, Liège and Tournai). For each of the 14 sites, 4 female and 4 male speaker, who were born and raised in the city in which they were recorded, were selected. The age of the speakers varies between 20 and 80 years. The corpus is stratified into four age groups: this parameter is controlled for each of the 14 groups of speakers ($F(13, 84) = 0.308$), between male and female speakers ($F(1, 84) = 0.110$, n.s.) and between male and female speakers across the 14 groups ($F(13, 84) = 0.114$, n.s.). For this study we used only the spontaneous speech subcorpus of CPROM-PFC, which is 11.2 hours long and consists of approximately 114.000 tokens.

The **LOCAS-F** corpus was primarily constituted to study the relationship between prosodic and syntactic boundaries, including the properties of dislocated structures. The syntactic annotation is articulated in two levels (functional sequences and dependency clauses). The prosodic boundary annotation distinguishes between major and intermediate boundaries, and was performed by trained phoneticians using a double-blind methodology. This corpus contains 14 different speaking styles, its duration is 3.5 hours and it contains approximately 43.000 tokens.

3.2. Expert Annotations

The CPROM and CPROM-PFC corpora contain an annotation of prominent syllables and syllables associated with a disfluency (fillers, lengthened syllables due to hesitations). Three different experts annotated CPROM, while CPROM-PFC was annotated by two experts, and a third one intervened in cases of disagreement to decide the final value of the syllable (+/- prominent, +/- associated to a disfluency). The LOCAS-F corpus does not contain an expert annotation of prosodically prominent syllables. However, samples from the CPROM corpus were reused in the LOCAS-F corpus. More specifically, the overlapping samples are: 3 radio news broadcasts (JPA), 3 political public addresses (POL), 3 scientific conference presentations (CNF), 2 radio interviews (INT), and 3 monologue narrations of life events (NAR). This overlap allowed us to inject the expert annotation of prosodically prominent syllables from these samples of CPROM to the corresponding subset of LOCAS-F. Inter-annotator agreement for prosodic prominence at the syllable level was assessed using Cohen's kappa statistics; for the CPROM-PFC corpus, Kappa values ranged between 0.61 and 0.88, with a mean of 0.72.

The LOCAS-F corpus was manually annotated for perceived prosodic boundaries (PBs) by two experts. Each word was marked as being followed by a strong PB (///), an intermediate PB (//), or as not followed by any boundary (0). The annotators used the code "hesi" to indicate that they perceived the speaker as hesitating: this includes filled pauses (e.g. euh) and drawls. A function was also attributed to each PB, based on the shape of the corresponding intonation contour. Four types of contours were used: C (continuation), T (final prosody), S (suspense) and F (focus). This annotation was primarily based on the annotators' perception; however they did have visual access to the pitch contour as displayed in Praat. In cases of disagreement, the annotators listened to the relevant section once again and agreed on the final PB and contour label.

3.3. Perceptual Experiments

In addition to the expert annotations described above, an on-line experiment reported in [16] allowed us to validate the perception of prosodic boundaries by naïve listeners. Participants (N = 84) listened to the short samples of speech extracted from the LOCAS-F corpus, and were instructed to press a key whenever they perceived the end of a "group of words" (this instruction was deliberately vague, in order to avoid biasing subjects towards a syntax-based analysis). Participants could only listen to each sample once and the collection of responses was done in real time, in order to be as close as possible to natural conditions of speech perception and comprehension. The role of syntax was also explored using delexicalised versions of the stimuli. We defined the measure of "boundary force" to mean the percentage of participants that perceived a PB on a specific syllable.

3.4. Perceptual Experiment Results

On average, experts annotated a prosodic boundary at the end of 27.7% of the tokens in the corpus; 14.5% are of intermediate strength (//) and 13.1% are strong (///), while 2.9% are marked as hesitations (hesi). The distribution of prosodic boundary types presents significant variation across genres, with a positive correlation between the degree of preparation and the number of strong prosodic boundaries (///), with few exceptions. The number of expert annotated hesitations (hesi) increases in the more spontaneous speaking styles. Regarding the syntactic correlates of prosodic boundaries, we observed that prosodic boundaries occur mostly on lexical words, while less than 3% of PBs would occur on a clitic word (CLI). An intermediate category co-occurs with a PB in $<$10% of potential positions. Taking the acoustic correlates of the expert annotation of prosodic boundaries, we observed that the presence or absence of a silent pause is the main feature that distinguishes between strong (///) and intermediate (//) prosodic boundaries. Strong PBs (///) are almost always followed by a silent pause, while this is rare for intermediate PBs (//); hesitations are occasionally followed by a pause. Syllable lengthening occurs on PBs regardless of their associated contour, and it is more pronounced in cases of focus-contour PBs. Taking non-PB syllables as the baseline, PBs associated with the C (continuation) and F (focus) show a rising intonation, while T (final) PBs exhibit a falling intonation, followed by S (suspense) PBs, while hesitations are more similar to T (final) PBs. Focus (F) PBs have the most dynamic (pronounced) intra-syllabic pitch movement, followed by C (continuation). Strong continuation (///C) prosodic boundaries are clearly marked with both inter- and intra-syllabic pitch movements, whereas intermediate continuation (//C) prosodic boundaries are only marked with relative pitch differences. These results were reported in [26].

The results of the perceptual experiment confirmed our hypothesis that due to the tasks constraints, naïve listeners would perceive a smaller number of prosodic boundaries than those annotated by the experts. In the natural speech condition (NS), 85% (434 out of 508) of PBs annotated as strong (///) by the experts were perceived by the naïve listeners, while only 17% (89 out of 533) PBs annotated as intermediate (//) by the experts were perceived by the naïve listeners in real time. The corresponding figures for the delexicalised speech condition were 93% (470 out of 508) and 9% (49 out of 533). Subjects perceived 61 PPBs in NS and 21 PPBs in MS on syllables where the experts had not annotated a prosodic boundary. The results also validated the experts' annotation of prosodic boundary strength. Boundaries annotated as strong (///) by the experts were systematically perceived by more participants (i.e. the boundary force is greater) than boundaries annotated as intermediate (//) by the experts. Focusing on PBs both annotated by experts and perceived by naïve listeners, the mean boundary force was 61% for strong PBs and 42% for intermediate PBs. Examining all PBs annotated by the experts, irrespective of whether these were perceived by the naïve listeners, in the NS condition, the mean boundary force of the 508 strong PBs is 52%, and the mean boundary force of the 533 intermediate boundaries is 7% (the difference of means is significant; Cohen d = 1,94); in the MS condition, the mean boundary force of strong PBs is 57% and the mean boundary force of intermediate PBs is 4% (the difference of means is significant; Cohen d = 2,74). The falling-contour (T) PBs had the highest mean perceived force; while level-contour (C) PPBs had a similar distribution of perceived force, regardless of whether the experts annotated them as weak or strong. Finally, an analysis of the acoustic and syntactic correlates of PPBs using linear models and linear regression trees showed that in the NS condition, the most important cue for the perception of a prosodic boundary is the presence of a subsequent silent pause, followed by the strength of a syntactic boundary (end of clause, followed by the end of a sequence), and final lengthening. These results were reported in [16].

4. Method

Our objective was to combine all the data and findings outlined in Section 3, in order to develop an automatic annotation system for prosodic events in French speech. The first step was **corpus homogenisation**: the three corpora were imported into *Praaline* [32], which is a toolkit for corpus management, annotation, querying and visualization. Using its interface, we applied Prosogram [33] for pitch stylisation on all data. Prosogram operates in two phases; for each syllable, vocalic nuclei are detected based on intensity and voicing. The f_0 curve on the nucleus is then stylised into a static or dynamic tone, using an algorithm that takes into account the perception of tones. The features extracted (duration, pitch, pitch movement etc.) were added to *Praaline*'s SQL database and used to calculate the acoustic correlates. We also applied an automatic morphosyntactic (part-of-speech and multi-word unit) annotation to the CPROM and LOCAS-F corpora, using the *DisMo* annotator; note that the morphosyntactic annotation of CPROM-PFC has been manually corrected and is used as part of the training corpus for the *DisMo* tool, as outlined in [34].

Automatic prosodic annotation is approached as a **sequence labelling** problem, drawing on multiple information sources (acoustic features, lexical and shallow syntactic features) in accordance with the experimental findings showing that listeners integrate all such information in their perception of prosodic segmentation and events. Tests of different machine learning methods for the annotation of prosodic prominence [25] had indicated that a labeller based on Conditional Random Fields [35] outperforms classification methods; a previous study by Cutungo et al. [36] compared CRF and Latent-Dynamic CRF models with equally promising results. We therefore chose to train CRF models to annotate sequences of syllables.

The **features** associated with each syllable were the following: syllable duration (ms); minimum, maximum and mean pitch (stylised f0, in semitones); pitch movement (within syllable and between successive syllables); peak intensity within the syllabic nucleus; spectral balance; token (word) to which the syllable belongs, and part-of-Speech tag of the token; presence and duration of subsequent pause; syllabic structure (C/V, whether the syllable ends with a schwa); position of the syllable relative to the token (word) in which it belongs: initial, final, penultimate, mono-syllabic word.

We trained CRF models for prosodic prominence with two output labels: prominent or non-prominent syllable (a binary model). A first model was trained on the CPROM corpus only; we then trained two CRF models for prosodic prominence (binary) on the CPROM-PFC corpus, one based on the acoustic features only, and one that also takes into consideration the gold-standard part-of-speech tags of words. Subsequently, we cross-annotated the two corpora, using the CPROM model to add an automatic annotation of prominent syllables to the CPROM-PFC corpus and vice versa. We then proceeded to train a CRF model to predict the presence of intermediate and strong prosodic boundaries (output labels: // or ///) based on the same set of features and trained on the annotation of the LOCAS-F corpus. A further CRF model was trained to predict boundary contours.

5. Results and Discussion

With respect to the **syllabic prominence** annotation, and as reported in [25], a 5-fold cross-validation of the CRF model using the entire feature set on the CPROM-PFC corpus yields a precision of 83.7% and a recall of 82.5%. However, when using the statistical model trained on CPROM-PFC to annotate the CPROM corpus, the result was a precision of 61.7% and a recall of 86%. The reason for that is that the expert annotation of prominent syllables in the CPROM-PFC corpus is more dense than the one in the CPROM corpus. In other words, annotators had a tendency to annotate more syllables as being prominent in CPROM-PFC than in CPROM. In order to take a middle road between the annotation choices of the experts who have annotated these two corpora, we followed the following procedure. The CRF model trained on the CPROM corpus (17.7k syllables) was used to automatically annotate the CPROM-PFC corpus (60k syllables). Then, a new CRF model, called Promise-Cross, was trained based on the resulting automatic annotation of CPROM-PFC. Table 1 shows the evaluation measures of three models by comparing the automatic annotation (using hte CPROM-PFC and Promise Cross models) with the manual annotation of the CPROM corpus. A further test and comparison was performed, using the prominence annotator PromGrad [24].

Table 1: *Precision, Recall and F-measure for three prosodic prominence models*

Model	Precision	Recall	F-Measure
Promise CPROM-PFC	63.3	83.5	72.0
Promise Cross	95.8	84.4	89.7
PromGrad	70.3	81.0	75.3

With respect to the **prosodic boundary strength** annotation, the model achieved a 92% precision in identifying strong prosodic boundaries and a 65% precision in identifying intermediate prosodic boundaries (based on the results of a 5-fold cross-validation, on the LOCAS-F corpus). Each syllable is described by a feature vector including: the duration of the subsequent silent pause (if any), the relative duration of the syllable compared to the previous 2, 3 and 4 syllables, the relative mean pitch of the syllable compared to the previous 2, 3 and 4 syllables, the inter-syllabic and intra-syllabic pitch movements, part-of-speech information, the position of the syllable within its token, and information on the presence or absence of disfluencies. The evaluation of the automatic annotation of boundary strength is show in Table 2. The overall accuracy is 93.5% and the Kappa score is 0.785.

Table 2: *Precision, Recall and F-measure for the prosodic boundary strength model*

Class	Precision	Recall	F-Measure
Major boundaries (///)	91.6	92.4	92.0
Minor boundaries (//)	65.4	67.1	66.2
No boundary	96.7	96.4	96.5

With respect to the annotation of functional **contours of boundary syllables**, we trained a separate CRF model, with the same features as the CRF model for prosodic boundary strength. The evaluation results for each of the four contours identified in the LOCAS-F corpus is shown in Table 3. The overall accuracy is 90.5% and the Kappa score is 0.685.

Table 3: *Precision, Recall and F-measure for the prosodic boundary contour model*

Class	Precision	Recall	F-Measure
No contour	97.0	96.4	96.7
Continuation (C)	71.6	65.7	68.5
Focus (F)	10.3	23.9	14.4
Suspense (S)	30.4	38.8	34.1
Final (T)	64.3	68.4	66.3

6. Conclusion

We presented a method to use expert annotation, validated by non-expert listeners, of prosodic prominence, phrasing and disfluencies, in order to create an automatic annotation system for French spoken corpora. By combining three different corpora, and taking advantage of the different annotations available to each corpus, we have developed an automatic annotation system, based on Conditional Random Fields and using both acoustic and lexical-syntactic features. This work has led to the release of a second version of the *Promise* automatic annotation, which works either as a plug-in for *Praaline* (e.g. as part of a corpus processing workflow), or as a stand-alone tool, and is distributed under a GPL3 open source license.

7. References

[1] A. Cutler, D. Dahan, and W. van Donselaar, "Prosody in the comprehension of spoken language: A literature review," *Language and Speech*, vol. 40, no. 2, pp. 141–201, January 1997.

[2] L. Frazier, K. Carlson, and J. Clifton, "Prosodic phrasing is central to language comprehension," *Trends in Cognitive Sciences*, vol. 10, no. 6, pp. 244–249, 2006.

[3] P. Wagner, A. Origlia, C. Avezani, G. Christodoulides, F. Cutugno, M. D'Imperio, D. Escudero Mancebo, B. Gili Fivela, A. Lacheret, B. Ludusan, H. Moniz, A. Ní Chasaide, O. Niebhur, L. Rousier-Vercruyssen, A. C. Simon, J. Simko, F. Tesser, and M. Vainio, "Different parts of the same elephant: a roadmap to disentangle and connect different perspectives on prosodic prominence," in *Proc. of the 18th International Congress of Phonetic Sciences, August 10–14, Glasgow, UK*, 2015.

[4] S.-A. Jun and C. Fougeron, "A phonological model of french intonation," in *Intonation: Analysis, Modelling and Technology*, A. Boltinis, Ed. Amsterdam: Kluwer Academic Publishers, 2000, pp. 185–208.

[5] P. Mertens, "Intonational grouping, boundaries and syntactic structure in french," in *Proc. ESCA Workshop on Prosody, Lund*, 1993, pp. 156–159.

[6] M. Swerts, "Prosodic features at discourse boundaries of different strength," *Journal of the Acoustical Society of America*, no. 101, pp. 514–521, 1997.

[7] Y. Mo, "Duration and intensity as perceptual cues for nave listeners prominence and boundary perception," in *Proc. of the 4th Speech Prosody Conference, May 6–9, Campinas, Brazil*, 2008, pp. 739–742.

[8] J. Cole, J. 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.

[9] J. Cole, J. Mo, and M. Hasegawa-Johnson, "Signal-based and expectation-based factors in the perception of prosodic prominence," *Laboratory Phonology*, vol. 1, pp. 425–452, 2010.

[10] D. Arnold and P. Wagner, "The influence of top-down expectations on the perception of syllable prominence," in *Proc. ISCA Workshop on Experimental Linguistics*, 2008, pp. 25–28.

[11] A. Lacheret, A. Simon, J.-P. Goldman, and M. Avanzi, "Prominence perception and accent detection in French: from phonetic processing to grammatical analysis," *Language Sciences*, vol. 39, pp. 95–106, 2013.

[12] J. G. Martin and W. Strange, "The perception of hesitation in spontaneous speech," *Perception & Psychophysics*, vol. 3, no. 6, pp. 427–438, 1968.

[13] C. Auran, A. Colas, C. Portes, and M. Vion, "Perception of breaks and discourse boundaries in spontaneous speech: developping an on-line technique," in *Proc. of IDP*, 2005, p. 7 pp.

[14] Y. Mo, J. Cole, and E. Lee, "Naive listeners prominence and boundary perception," in *Proc. of the 4th Speech Prosody Conference, May 6–9, Campinas, Brazil*, 2008, pp. 735–738.

[15] C. Smith, "Perception of prominence and boundaries by naïve French listeners," in *Proc. of the 17th International Congress of Phonetic Sciences, August 17–21, Hong Kong, China*, 2011, pp. 1874–1877.

[16] A. C. Simon and G. Christodoulides, "Perception of prosodic boundaries by naïve listeners in French," in *Proc. of the 8th Speech Prosody Conference, May 31–June 3, Boston, USA*, 2016.

[17] C. W. Wightman and M. Ostendorf, "Automatic labeling of prosodic patterns," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, Oct 1994.

[18] K. Chen, M. Hasegawa-Johnson, and A. Cohen, "An automatic prosody labeling system using ann-based syntactic-prosodic model and gmm-based acoustic-prosodic model," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, May 2004, pp. I–509–12 vol.1.

[19] V. Rangarajan, S. Bangalore, and S. S. Narayanan, "Exploiting acoustic and syntactic features for automatic prosody labeling in a maximum entropy framework," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 797–811, May 2008.

[20] A. Rosenberg, "Automatic detection and classification of prosodic events," PhD dissertation, Columbia University, 2009.

[21] S. Ananthakrishnan and S. S. Narayanan, "Automatic prosodic event detection using acoustic, lexical, and syntactic evidence," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 216–228, Jan 2008.

[22] M. Avanzi, A. Lacheret-Dujour, and B. Victorri, "A corpus-based learning method for prominence detection in spontaneous speech," in *Proc. of Speech Prosody 2010, Prosodic Prominence Workshop, May 11–14, Chicago, USA*, 2010.

[23] J.-P. Goldman, A. Auchlin, S. Roekhaut, A.-C. Simon, and M. Avanzi, "Prominence perception and accent detection in French. a corpus-based account," in *Proc. Speech Prosody 2010, Chicago, May 11-14*, 2010.

[24] J.-P. Goldman, M. Avanzi, and A. Simon, A.C.and Auchlin, "A continuous prominence score based on acoustic features," in *Interspeech 2012 – 13th Annual Conference of the International Speech Communication Association, September 9-13, Portland, Oregon, USA, Proceedings*, 2012, pp. 2454–2457.

[25] G. Christodoulides and M. Avanzi, "An evaluation of machine learning methods for prominence detection in French," in *Interspeech 2014 – 15th Annual Conference of the International Speech Communication Association, September 14–18, Singapore, Proceedings*, 2014, pp. 116–119.

[26] G. Christodoulides and A. C. Simon, "Exploring acoustic and syntactic cues to prosodic boundaries in French. a multi-genre corpus study," in *Proc. of the 18th International Congress of Phonetic Sciences, August 10–14, Glasgow, UK*, 2015.

[27] G. Christodoulides and M. Avanzi, "Automatic detection and annotation of disfluencies in spoken French corpora," in *Interspeech 2015 – 16th Annual Conference of the International Speech Communication Association, September 6–10, Dresden, Germany, Proceedings*, 2015, pp. 1849–1853.

[28] M. Avanzi, A. C. Simon, J.-P. Goldman, and A. Auchlin, "C–PROM: An annotated corpus for French prominence study," in *Proc. of Speech Prosody 2010, Prosodic Prominence Workshop, May 11–14, Chicago, USA*, 2010.

[29] M. Avanzi, "A corpus-based approach to French regional prosodic variation," *Nouveaux Cahiers de Linguistique Française (Proceedings of the 3rd SWIP)*, vol. 31, pp. 309–332, 2014.

[30] L. Degand and A. C. Simon, "Unités discursives de base et leur priphérie gauche dans LOCAS-F, un corpus oral multigenres annoté," in *Actes du 4ème Congrès Mondial de Linguistique Française*. Berlin, Germany: EDP SciencesC, 2014.

[31] J. Durand, B. Laks, and C. Lyche, *Phonologie, variation et accents du français*. Paris: Hermes, 2009.

[32] G. Christodoulides, "Praaline: integrating tools for speech corpus research," in *LREC 2014 – 9th International Conference on Language Resources and Evaluation, May 26–31, Reykjavik, Iceland, Proceedings*, 2014, pp. 31–34.

[33] P. Mertens, "The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model," in *Proc. of Speech Prosody 2004, March 23–26, Nara, Japan*, 2004, pp. 549–552.

[34] G. Christodoulides, M. Avanzi, and J. P. Goldman, "DisMo: A morphosyntactic, disfluency and multi-word unit annotator," in *LREC 2014 – 9th International Conference on Language Resources and Evaluation, May 26–31, Reykjavik, Iceland, Proceedings*, 2014, pp. 3902–3907.

[35] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. of the 18th International Conference on Machine Learning*. Morgan Kaufmann, 2001, pp. 282–289.

[36] F. Cutugno, E. Leone, B. Ludusan, and A. Origlia, "Investigating syllabic prominence with conditional random fields and latent-dynamic conditional random fields," in *Interspeech 2012 – 13th Annual Conference of the International Speech Communication Association, September 9-13, Portland, Oregon, USA, Proceedings*, 2012, pp. 2402–2405.