Exploring Acoustic and Syntactic Cues to Prosodic Boundaries in French A Multi-Genre Corpus Study

George Christodoulides, Anne Catherine Simon

Centre Valibel, Institute for Language & Communication, Université catholique de Louvain, Belgium
george@mycontent.gr, anne-catherine.simon@uclouvain.be

ABSTRACT

This paper investigates the way prosody and syntactic structure combine in the perception of prosodic boundaries in French. Based on a 3.5-hour balanced corpus, we first analyse the distribution of boundary types across genres, and then examine the acoustic correlates of prosodic boundaries their relationship to linguistic features (part-of-speech categories and syntactic clauses).

Keywords: prosodic boundaries, pauses, prosody-syntax interface.

1. INTRODUCTION

Prosodic segmentation is known to be fundamental for discourse comprehension ([11], [28]). However, there is no consensus on how prosodic boundaries (PB) should be defined and identified.

In order to model the prosodic segmentation of speech, several dimensions have to be taken into account. First, we need to decide on the degrees of PB strength (assuming that PBs are perceived as categorical and not gradual phenomena). Although most models on French prosody admit at least three degrees of PBs ([24], [21], [15]), most large-scale corpus annotations are limited to 1 or 2 degrees ([10]). Second, we have to decide on which acoustic features to base the modelling of prosodic segmentation. Wagner & Watson [27] reported that silent pauses, duration, f_0 movement and phonation type are the most salient cues to PB, and those cues are known to be somewhat language-specific. In French, since the primary (final) accent is located on the last syllable of the unit, it co-occurs with the PB (see [14]). However, this does not mean that French listeners cannot distinguish between prominence and PB (as evidenced experimentally by [1]).

Third, we know that prosodic and linguistic (lexical-grammatical, syntactic) structures interact with each other. There is no consensus on how these structures are related: phonological approaches posit that grammar strongly constrains the size and structure of prosodic units (see for French [19], [20], [23]), while acoustic or interactionist approaches consider that speakers make strategic use of PB to avoid ambiguity and guide listeners' interpretation (e.g. [26], [2]).

More recent studies tried to investigate to what extent signal-related properties and expectation-based (linguistic) information are responsible for the perception of PB by listeners, under natural conditions of speech production. In English, [9] demonstrated that the contribution of syntax to the perceived PB slightly overcomes the contribution of (prosodic) duration, both factors being partially dependent. Modelling the contribution of acoustic features and linguistic information to the prediction of PB in Swedish, [18] found that the combination of silent pause duration and word-final syllable rhyme duration yielded 86.2% correctly classified PBs, while the combination of content/function word distinction and POS gave 66.3% correct results.

Our contribution seeks to investigate strong and intermediate perceived PB, as well as their acoustic and linguistic correlates in a multi-genre corpus of French. The acoustic features studied include silent pauses (position and length), word-final syllable lengthening, pitch movement and relative pitch of the last syllables before a PB, while the linguistic features include syntactic units and POS categories.

2. METHOD

2.1. Material

This study is based on the LOCAS-F Corpus [12], which is a balanced corpus of spoken French with samples of 14 different speaking styles (discourse genres). Its duration is 3.5 hours and it contains 43.000 tokens. The discourse genres covered are the following: official speech in an academic setting [aca], scientific conference presentation [conf], face-to-face dialogue in a formal [conv-f] or informal [conv-i] setting, political debate [deb], sermon / homily [hom], radio interview [int-rad], socio-linguistic interview [int-soc], informal TV interview [int-tv], monologue narration of life event [mono-n], semi-prepared radio monologue [mono-r], radio news bulletin [news], political public address [pol], and neutral reading of a newspaper text [read]. The corpus composition is detailed in Table 1.

Table 1: Corpus description: number of samples, duration (min:sec), tokens, prosodic boundaries (strength 2 and 3), and syntactic units.

Genre	Nb	Dur.	Tokens	PB //	PB ///	S.Seq
aca	3	15:16	2332	161	401	361
conf	4	16:43	2939	318	321	556
conv-f	3	12:51	2714	354	269	883
conv-i	3	12:24	2945	280	377	1251
deb	4	19:17	5216	883	529	1463
hom	3	13:21	1759	120	344	428
int-rad	4	20:28	4313	746	476	1032
int-soc	3	15:23	2958	453	335	766
int-tv	3	15:31	4003	517	482	1333
mono-n	3	10:20	2367	288	135	862
mono-r	3	13:22	2591	293	278	564
news	4	14:44	2902	463	207	564
pol	5	20:23	2889	307	475	610
read	3	15:17	3151	445	414	485
Total	**48**	**3:35:20**	**43079**	**5628**	**5043**	**11158**

2.2. Corpus processing

The corpus was orthographically transcribed within Praat [6], and automatically phonetised and segmented into phones, syllables and words using the *EasyAlign* software [16]. All aligned transcriptions were manually corrected and several annotation layers were added to the corpus. In order to process this multi-level annotation, we used *Praaline* [7], a toolkit for corpus management, annotation, querying and visualization. Using its interface, we applied *Prosogram* [22] for pitch stylisation on the entire corpus. *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 the database, and used to calculate the acoustic correlates (see section 3). The morphosyntactic annotation (see section 2.4) was also added to the database.

2.3. Prosodic boundaries annotation

The speech material was manually annotated for perceived prosodic boundaries by two experienced transcribers. 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 perceive the speaker was 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.

2.4. POS and syntactic annotation

The corpus was annotated using the *DisMo* [8] morphosyntactic and disfluency tagger. *DisMo* attributes detailed part-of-speech tags on minimal tokens and multi-word expressions (on separate annotation levels), and performs an automatic detection of common types of disfluencies, using a combination of language resources and statistical models based on Conditional Random Fields. For this study we used the output of the automatic annotation without manual correction (*DisMo* typically attains 95-97% precision in POS-tagging spoken French corpora). We have based our analyses on the level of annotation that takes into account multi-word expressions.

Regarding syntactic structures, a manual annotation has been carried out by two expert linguists, based on the principles of dependency syntax for spoken French (outlined in [3], [4], [13]). The annotation results in two tiers: the encompassing level delimits dependency clauses, the maximal syntactic unit that is any governing element (most often a verb) and its governed elements; dependency clauses are analysed into functional sequences: Verb Sequence (SV), Subjet Sequence (SS), Dependent Sequence (SR), Incomplete Sequences (I), Adjuncts (A), Other (mostly non-verbal sequences), Discourse Markers (DM).

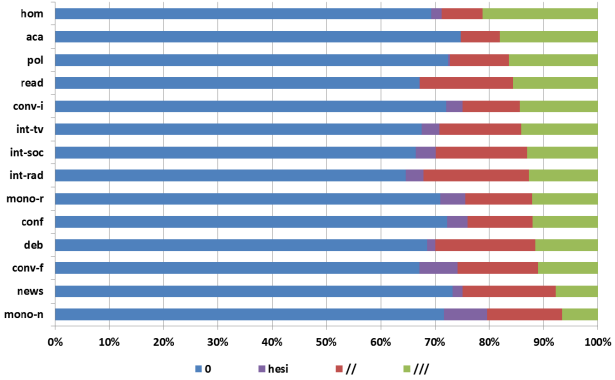
3. RESULTS AND DISCUSSION

3.1. Genres and prosodic boundaries

On average, there is 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, as can be seen in Figure 1. This variation corresponds mainly to the extent the speaker has had the opportunity to prepare. We observe that there is a positive correlation between the degree of preparation and the number of strong PBs (///), with the sole exception of radio news bulletins (this can be explained, as the pressure to transmit a lot of information in a limited amount of time leads to a particular speaking style, with a high speech rate and few pauses; see [25]). On the other hand, the number of perceived hesitations is greater in more spontaneous, non-prepared speaking genres. In only

3 cases does the number of strong PBs (///) exceed the number of intermediate PBs (//): ACA, HOM and POL, leading to what can be perceived as an over-segmented speech flow.

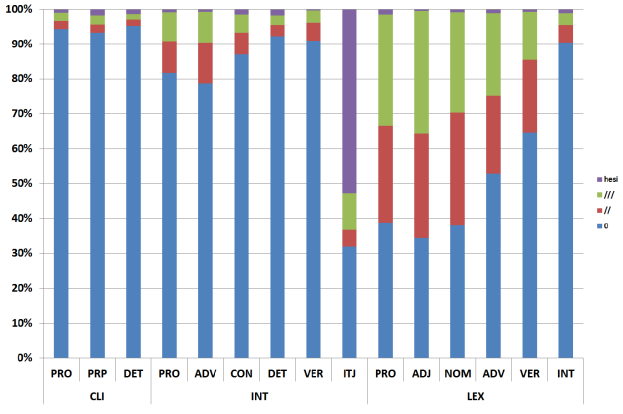
Figure 1: Distribution of PBs per discourse genre.



3.2. Linguistic and syntactic cues to prosodic boundaries

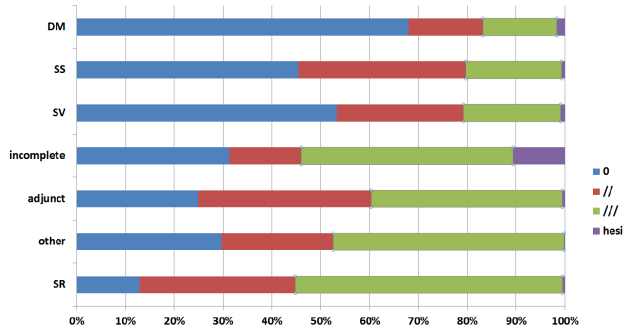
Figure 2 shows the relationship between the types of PBs and the part-of-speech tag of the associated token. We observe that PBs occur mostly on lexical words (LEX: adjectives ADJ, nouns NOM, adverbs ADV, verbs VER and clitic pronouns), while less than 3% of PBs would occur on a clitic word (CLI). An intermediate category (INT, including interrogative or relative pronouns PRO, negation particles ADV, conjunctions CON, determiners DET and auxiliary verbs VER) co-occurs with a PB in < 10% of potential positions. These findings seem to confirm the analysis in [17].

Figure 2: Relationship between PBs and POS tag of the token on which it occurs.



Sequences may be rather long and complex syntactic units. As can be seen in Figure 3, in approximately 50% of the cases, verb sequences (SV) and subject sequences (SS) do not bear a PB, because they are followed by another sequence that is prosodically grouped with them (cf. word order in French, [4]). Conversely, the majority of dependent sequences (SR), Adjuncts and Others (sequences without verb) are followed by strong (///) prosodic boundary. Incomplete sequences are found in-between these two tendencies.

Figure 3: Relationship between PBs and syntactic structures.



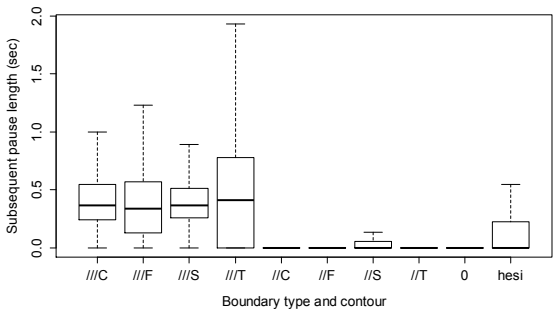
3.3. Acoustic correlates to prosodic boundaries

For each potential PB position (last syllable of each token), several acoustic features were extracted:

- the duration of a subsequent silent pause, excluding the pauses at turn-taking;
- relative duration: the duration of the last syllable divided by the average duration of 2…5 previous syllables;
- relative pitch: the difference between the pitch (in semitones) of the last syllable and the average pitch of 2…5 previous syllables;
- intra-syllabic pitch movement (in semitones)

We observed that a local analysis (window of 2 syllables) is sufficient: the results presented here remain valid for windows of 3…5 syllables.

Figure 4: Subsequent silent pause duration per PB type and contour.



A first analysis indicates that the presence or absence of a silent pause is the main feature that distinguishes between strong (///) and intermediate (//) PBs (see Figure 4). 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 (Figure 5). 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 (Figure 6). Focus (F) PBs have the most dynamic (pronounced) intrasyllabic pitch movement, followed by C (continuation). Strong continuation (///C) PBs are clearly marked with both inter- and intra-syllabic pitch movement, whereas intermediate continuation (//C) PBs are only marked with relative pitch differences. The interaction of the 3 most important parameters in marking different types and contours of prosodic boundaries can be readily seen in Figure 8.

4. CONCLUSION AND PERSPECTIVES

Our main findings confirm that silent pauses are a decisive cue to strong PBs in French, and that the use of other acoustic cues differs according to the strength (intermediate or strong) and function (continuation, finality, focus or suspense) of the PB. To some extent, hesitation markers behave like "suspense" PBs (lengthening and relative pitch). Linguistic cues are also decisive, as PBs co-occur with certain grammatical categories. Types of PB differ significantly across speech genres; this incites us to extend our modelling to macro-genres (e.g. interaction vs. monologue).

Figure 5: Syllable lengthening (relative syllable duration) by PB type and contour.

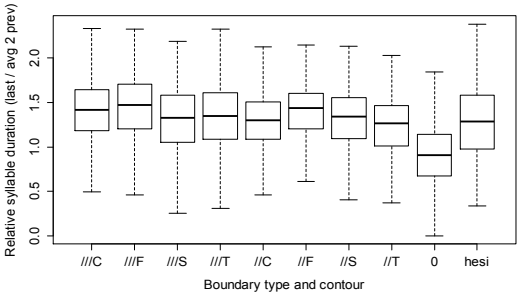


Figure 6: Relative pitch by PB type and contour.

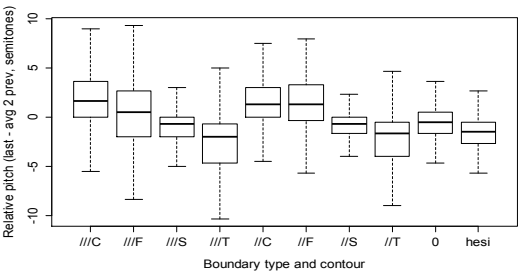


Figure 7: Intra-syllabic pitch movement by PB type and contour.

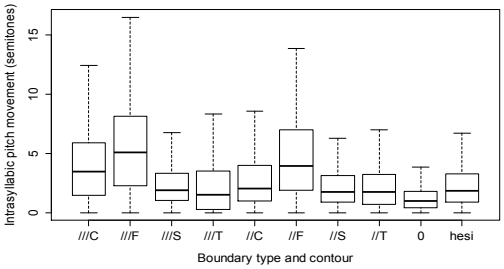
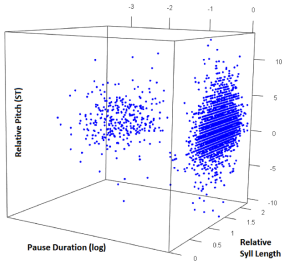
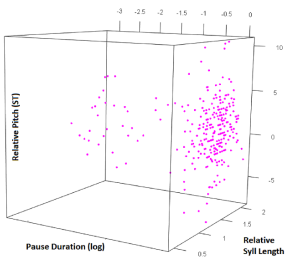


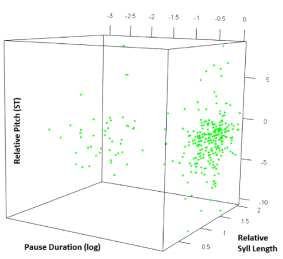
Figure 8: Scatter plots of the 8 types of prosodic boundaries over 3 dimensions of acoustic correlates: log pause duration (x axis, 0…-3.0 log sec), final syllable length relative to the average length of the previous two (y axis, 0.5…1.5), relative pitch (z axis, semitones -10…+10 semitones). Each point represents a PB of the corresponding type.



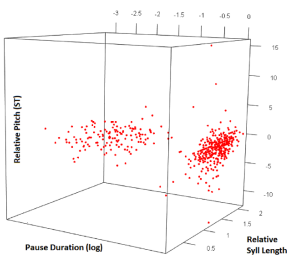
//C (strength 2, continuation)



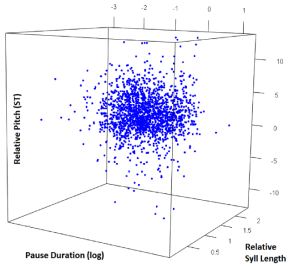
//F (strength 2, focus)



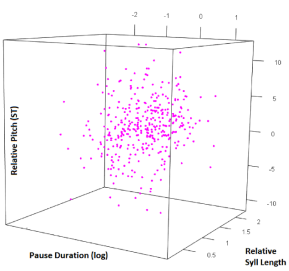
//T (strength 2, final)



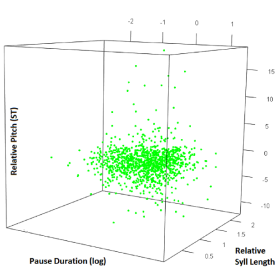
//S (strength 2, suspense)



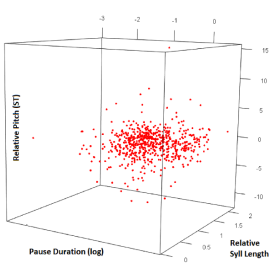
///C (strength 3, continuation)



///F (strength 3, focus)



///T (strength 3, final)



///S (strength 3, suspense)

7. REFERENCES

[1] Astésano, C., Bertrand, R., Espesser, R., Nguyen, N. 2012. Perception des frontières et des proéminences en français. *Proc. JEP-TALN-RECITAL*, Grenoble, 353-360.

[2] Barth-Weingarten, D., 2011. The fuzziness of intonation units: Some theoretical considerations and a practical solution. *InLiSt - Interaction and Linguistic Structures*, 51.

[3] Bilger, M., Campione, C. 2002. Propositions pour un étiquetage en 'séquences fonctionnelles'. *Recherches sur le français parlé* 17, 117-136.

[4] Blanche-Benveniste, C., 1996. Trois remarques sur l'ordre des mots dans la langue parlée. *Langue française* 111, 109-117.

[5] Blanche-Benveniste, C., Bilger, M., Rouget, C., van den Eynde, K., Mertens, P. 1990. *Le français parlé: études grammaticales*. Paris: Éditions du CNRS.

[6] Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer [Computer program]. Version 5.4.04, retrieved 28 December 2014 from http://www.praat.org/

[7] Christodoulides, G. 2014. Praaline: Integrating tools for speech corpus research. *Proc. 9th LREC*, Reykjavik, Iceland, 31-34.

[8] Christodoulides, G., Avanzi, M., Goldman, J-Ph. 2014. DisMo: A morphosyntactic, disfluency and multi-word unit annotator. An evaluation on a corpus of French spontaneous and read speech. *Proc. 9th LREC*, Reykjavik, Iceland, 3902-3907.

[9] Cole, J., Mo, Y., Hasegawa-Johnson, M., 2010. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1(2), 425‑452.

[10] Cresti, E., Moneglia, M. (ed.). 2005. *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam: John Benjamins

[11] Cutler, A., Dahan, D., van Donselaar, W. 1997. Prosody in the comprehension of spontaneous spoken discourse. *Studia Linguistica* 40(2), 141-201.

[12] Degand, L., Martin, L.J., Simon, A.C. 2014. Unités discursives de base et leur périphérie gauche dans LOCAS-F, un corpus oral multigenres annoté. *Actes du 4ème Congrès Mondial de Linguistique Française 2014*, Berlin, Germany: EDP Sciences.

[13] Degand, L., Simon, A.C. 2009. Mapping prosody and syntax as discourse strategies: How Basic Discourse Units vary across genres. In: Wichmann, A., Barth-Weingarten, D., Dehé N., *Where prosody meets pragmatics: research at the interface*. Emerald: Bingley, 79-105.

[14] Di Cristo, A. 2011. Une approche intégrative des relations de l'accentuation au phrasé prosodique du français. *Journal of French Language Studies*, 21(01), 73‑95.

[15] Di Cristo, A., 1999. Vers une modélisation de l'accentuation du français: première partie. *Journal of French Language Studies* 9(2), 143‑79.

[16] Goldman, J.-P. 2011. EasyAlign: an automatic phonetic alignment tool under Praat *Proc. InterSpeech*, Firenze, Italy, 3233-3236.

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

[18] Heldner, M., Megyesi, B. 2003. Exploring the Prosody-Syntax Interface in Conversations. *Proc. 15th ICPhS* Barcelona, 2501-2504.

[19] Jun, S.-A., Fougeron, C., 2000. A Phonological Model of French Intonation. In A. Boltinis, (ed.) I*ntonation: Analysis, Modelling and Technology*. Amsterdam: Kluwer Academic Publishers, 185-208.

[20] Kentner, G., Féry, C., 2013. A new approach to prosodic grouping. *The Linguistic Review*, 30(2), 277-311.

[21] Mertens, P. 1993. Intonational grouping, boundaries and syntactic structure in French. *Proc. ESCA Workshop on Prosody*, Lund, 156-159.

[22] Mertens, P. 2004. The ProsoGram: Semi-automatic transcription of prosody based on a tonal perception model. *Proc. Speech Prosody 2004*, Nara, Japan.

[23] Post, B., 2000 *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.

[24] Rossi, M., 1999. *L'intonation, le système du français: description et modélisation*, Paris-Gap: Ophrys.

[25] Simon, A.C., Auchlin, A., Goldman, J.-Ph., Christodoulides, G. 2014. Tendances prosodiques de la parole radiophonique. In Valery, P. (Ed.) *Cahiers de Praxématique*, Université de Montpellier III, ISSN: 0765-4944. Accepted.

[26] Swerts, M. & Geluykens, R., 1993. The Prosody of Information Units in Spontaneous Monologue. *Phonetica* 50(3), 189-196.

[27] Wagner, M., Watson, D.G. 2010. Experimental and theoretical advances in prosody: A review. *Language and cognitive processes*, 25(7-9), 905-945.

[28] Watson, D., Gibson, E. 2005. Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica* 59(2-3), 279-300.