Prosody in a corpus of French spontaneous speech: perception, annotation and prosody ~ syntax interaction

Irina Nesterenko[1,2], *Stephane Rauzy*[1], *Roxane Bertrand*[1]

[1]Aix-Marseille Université – Laboratoire Parole et Langage CNRS (UMR 6057)
[2]Institut National des Langues et Civilisations Orientales *(INALCO)*
irina_nesterenko@yahoo.com, stephane.rauzy@lpl-aix.fr, roxane.bertrand@lpl-aix.fr

Abstract

Our study focuses on the issue of prosodic annotation and of the prosody ~ syntax interface in conversation and is based on a large corpus of conversational speech in French. The results of inter-transcriber agreement tests show that two expert transcribers are consistent in their labeling of prosodic phrasing and the consistency is well above the chance. A qualitative analysis reveals transcribers' individual strategies, namely in reference to Intermediate Phrases sometimes found for French in specific intonation patterns.

The syntactic division of the corpus both in terms of syntactic chunks and in terms of pseudo-phrases is further analyzed in its interaction with the distribution of major prosodic breaks. In more than 60% of cases the boundaries of the pseudo-phrases co-occurs with the boundaries of major prosodic units (Intonational Phrases, *IPs*). At the same time, 50% of *IP* boundaries are aligned with smaller syntactic constituents. On the other hand, in our study beginnings of intonational phrases are more often misalign with syntactic constituent boundaries than their ends.

We discuss as well the issue of conversational corpus annotation in terms of prosodic units, given specific constraints on planning and execution in spontaneous speech.

Index Terms: prosody ~ syntax interface, prosodic phrasing, conversational speech, corpus annotation

1. Introduction

Our study focuses on the issue of prosodic phrasing in conversational speech and of the relationship between prosody and syntax. Prosodic phrasing refers to the structuring of speech material in terms of boundaries and groupings. These boundaries vary as to their relative strength thus defining a number of levels in prosodic constituency. A body of psycholinguistic research has shown that this information is of importance for effective speech processing [1] and ambiguity resolution [2, 3]. In addition, to implement a reliable algorithm for prosodic boundary placement is important for speech technologies, both in speech synthesis and in speech recognition.

Prosodic phonology [4] has proposed a universal hierarchy of prosodic constituents. Two approaches to prosodic phrasing may be distinguished: an edge alignment approach [5] and an intonation-based approach [6]. If the latter largely relies on the structure of melodic contour, the former approach acknowledges the sensitivity of prosodic structure to syntactic constituency; at the same time, it states that prosodic structure is not isomorphous to syntactic structure and cannot be predicted from syntactic information alone [7,8].

In the models of prosodic phrasing proposed for different languages it is common to distinguish two levels of phrasing above the word: the level of phonological phrases[1] (*PhP*) (to use the terminology of Selkirk [9]) and the higher level of intonational phrases (*IPs*). Our study deals with the larger prosodic units alone. Intonational phrases are primarily defined in terms of domains of distinctive pitch contours, though semantic-pragmatic information is also taken into account. At the same time, it is considered that several syntactic constructions form *IPs* of their own. This is the case for root clauses, vocatives and parentheticals; certain left-peripheral constituents (sentential adverbs) and certain right-peripheral elements are also separated by an obligatory IP boundary.

Though there seems to be a large body of research on the syntax ~ prosody interface, very few studies deal with the issue in connection to spontaneous speech. Our study aims to fill in this gap and provide evidence relating to the syntax ~ prosody interface in conversation. It should be noted that conversational speech is subject to specific constraints on planning and execution: interruption of speech flow, restarts, hesitations and pausing, for example, are typical for conversational speech and reflect planning and lexical search processes. Another characteristic property of conversational speech resides in the role of context and non-verbal information, which influence the syntactic and prosodic structure of utterances.

In the present study, we undertake the division of speech into pseudo-phrases (analogues of phrases in the written text) and further focus our attention on the relationship between predicted punctuation marks and perceived prosodic boundaries. In punctuation, we distinguish between strong punctuation marks, which separate sentences, and weak punctuation marks, which separate clauses, sentential adverbs and parentheticals. From the discussion above, it follows that both strong and weak punctuation marks are quite likely to coincide with IP boundaries. The rest of the paper is organised as follows: first, in section 2 we present the corpus our study is based on and we detail the undertaken prosodic and syntactic annotations. Section 3 presents the results of inter-transcribers agreement tests and the results of our analyses concerning the co-occurrence of prosodic and pseudo-orthographic boundaries, the prosody ~ syntax interface and the influence of planning and execution constraints in conversation. Finally, we discuss the impact of the adopted approach as well as future work that remains.

2. Corpus and methodology

Our study is based on an excerpt from the *Corpus of Interactional Data* [12] (http://crdo.up.univaix.fr/corpus.php?langue=fr). We focused on one dialogue between two familiar female speakers who conversed on humorous situations in which they may have

[1] Related units are Intermediate phrases of [10] and for French – Accentual phrases described in [11]

found themselves involved. The total size of the corpus was 12681 words.

The corpus was manually transcribed using an enriched orthography: in order to facilitate further processing of the corpus, our transcription conventions include special notations to signal a number of reduction phenomena (i.e. elisions, word truncations). Next, this transcription was automatically converted to a phonemic transcription of speech material and then automatically aligned to the speech signal. Subsequently, the corpus was enriched with various linguistic annotations (manual or (semi-)automatic) as a means to study interfaces between phonetics, phonology, prosody, morphology, syntax, pragmatics, discourse and gesture as they operate in conversational speech. In the following paragraphs we detail the syntactic and prosodic annotation underlying our study.

2.1. Prosodic annotation

The general prosodic annotation scheme for the corpus includes

- metrical structure in terms of perceived prominences;
- tonal structure: we distinguish the level of underlying tones and the level of surface tones (INTSINT);
- prosodic constituency.

The corpus was manually annotated in terms of Intonational phrase boundaries by two of the authors. This annotation was guided by perception, based on a distinction between strong and weak prosodic breaks. Other acoustic and perceptual cues to an IP boundary are: i) an intonation unit is associated with a specific melodic contour; ii) there is a high (H) or a low (L) boundary; iii) there is *pitch reset*; and iv) there is pre-boundary lengthening.

As this work deals with spontaneous speech, we introduced one more category at the level of *IP*s: an uncompleted *IP* (*ipa*) corresponding to a stretch of speech larger than *AP* which wasn't completed due, among others to planning constraints. In reference to the typology of acoustic/phonological cues to *IP* boundaries, these units were not full *IP*s as there was no distinctive pitch contour associated with them; at the same time, there was a perceived pitch reset at the beginning of the following *IP*. Note that this category was introduced to satisfy the Exhaustivity constraint on prosodic phrasing as formulated in [5]. Note that prosodic structuring of speech flow in conversation could by masked by performance phenomena: we sought not to melt in one category of major prosodic units both structural units (*IPs*) and performance units (marked by speech flow interruption or a filled pause).

2.2. Syntactic annotation

On the basis of enriched orthographic transcription and phonetic transcription larger units such as words were recovered and automatically aligned with speech signal; they form the basic input to the syntactic analyser. Non-syntactic objects such as laughter, and dysfluencies were removed from the input. After the first filtering stage, a modified version of the syntactic parser StP1 was applied on the data.

The syntactic parser StP1 [13] is a stochastic parser for written French text developed at the Laboratoire Parole et Langage. In the first step, for each word token it provides an automatic annotation of its morphosyntactic category. In the second step, the tokens are grouped in larger units (chunks) following the EASY flat grammar [14] described in the PEAS guidelines [15]: in such an approach, chunks represent minimal syntactic phrases of the tree structure. The EASY grammar comprises six constituents: GN (Noun Phrase), GP (Prepositional Phrase), GR (Adverbial Phrase), GA (Adjective Phrase), NV (Verbal Nucleus), PV (Verbal group introduced by a preposition), organized in a sentence with a flat structure.

The StP1 chunker obtains relatively good results on written texts (see [13] for more details), an F-measure score being of 0.94 for the tagging stage and of 0.92 for the chunking stage. These performances are reduced for conversation speech corpora, an F-measure score being of 0.79 for chunks formation, but still remain interesting for providing us with an automatic annotation of the syntactic information.

The StP1 chunker has been modified in order to account for the specificities of conversational speech. Two levels of hierarchy were introduced in the syntactic treatment, corresponding to the strong punctuation marks (final point, exclamation mark) and weak or soft punctuation marks (comma) that can be found in written text. The modified stochastic parser automatically inserts these two kinds of frontiers on the basis of the syntactic context.

This symbolic annotation underlines probabilistic modelling of speech division into phrase-like units (4 levels) undertaken in our study. We would like to emphasize that all syntactic analyses in this study are automatic and there was no manual correction applied.

3. Results

We present first the results on the distribution of *IP* boundaries and we evaluate the reliability of prosodic annotation. The data on the most probable division of our corpus in terms of pseudo-phrases are then analysed in their interaction with the distribution of perceived *IP* boundaries. Finally, we analyse the distribution of *ipa* boundaries in relation to syntactic structure.

3.1. Distribution of prosodic boundaries

In Table 1 we present the data on the distribution of prosodic boundaries independently for each annotator and each speaker. It follows that 24-30% of word boundaries in the corpus were annotated as boundaries of higher constituents in prosodic hierarchy: there was a prosodic boundary perceived every 3-4 words.

Another finding deals with the use of *ipa* labels: it was used in 8-13% of cases (mean 10,9%), the differences between transcribers being non-significant (for AB: $\chi^2 = 3.7877$, df = 1, p-value = 0.052; for CM $\chi^2 = 2.1008$, df = 1, p-value = 0.15). At the same time, RB used significantly more *ipa* labels for CM compared to AB (8% *versus* 11,3%; $\chi^2 = 10.2708$, df = 1, p-value = 0.001).

Table 1: *Distribution of perceived prosodic boundaries*

Transcriber	*Speaker*	*ipa*	*IP*	*# of words*	*Mean interval*
IN	AB	162	1309	6162	4,19
IN	CM	230	1546	6519	3,67
RB	AB	132	1438	6162	3,92
RB	CM	225	1757	6519	3,29

A qualitative analysis reveals that many of disagreements between transcribers could be imputed to the existence of a third level of phrasing in French, namely the level of intermediate phrases: several authors provide evidence for the existence of Intermediate phrase units in French [11, 16], which occur in a restricted set of marked constructions. In our annotation study, transcribers use different strategies when there seems to be an Intermediate phrase boundary: while one of the annotators signalled the presence of an IP boundary (adding a special note for the unit to be an intermediate phrase, *ip*), the other transcriber restricted the use of *IP* labels for the full *IP*s, cf.

Bon je vois qu'tu es tellement à court d'idées
(Well I see you really lack the ideas)
1st transcriber [[ ]IP (=ip)]IP
2nd transcriber [ ]IP

Note that relative clauses and sententional adverbs represented the most frequent disagreement contexts.

3.2. Inter-Transcribers' reliability test

Traditionally, one resorts to Cohen's kappa statistics when the question of measuring inter-annotator agreement arises. Both pair-wise agreement and kappa coefficients allow accessing the consistency in annotators' performance, though only the latter proceeds by comparison of the observed agreement with the probability of the two transcribers agreeing by chance.

The data on inter-transcribers agreement and kappa statistics are presented in Table 2. The kappa values in Table 2 show that there is a good agreement between the annotators well above chance. At the same time, there is a significant difference in the % of inter-transcriber agreement between speakers (χ^2 = 33.1609, df = 1, p-value < 0.001): further acoustic, phonetic and phonological analyses will show whether there is a difference in salience of acoustic cues used by each of the speakers to signal prosodic structure as well as in the frequency of use of different boundary cues (melodic contour, pitch resetting, presence of high or low boundary tone, pre-boundary lengthening etc.)

Table 2. *Inter-transcriber agreement and Cohen's Kappa scores*

Speaker	AB	CM
% of agreement	92.84	89.95
Cohen's kappa	0.81	0.76

3.3. The Prosody ~ Syntax interface

In this section we present the results on the relationship between prosodic annotation and syntactic structure. Left and right boundaries of *IP* were dealt with separately.

Figure 1 plots the mean number of IP boundaries, which co-occur with strong and weak punctuation marks and with syntactic constituents. In table 3 we present evaluation statistics of the algorithm for pseudo-phrases only. It appears that left IP boundaries co-occur with a punctuation mark in 46% and right boundaries in 48% of cases. At the same time, if there is a punctuation mark inserted by the algorithm, there are around 65 % of chances that there would be a prosodic boundary aligned with it. F-measure statistics are 0.54 and 0.55 respectively. Note though that the underlying grammar was built from the corpus of written texts; consequently, we could expect better performance if the model is trained on an annotated corpus of spontaneous speech.

At the same time the data on the distribution of prosodic boundaries with respect to syntactic constituency show a significant asymmetry between the left and the right IP boundaries for both speakers (χ^2 = 37.7547, df = 3, p-value <<0.001 for AB; χ^2 = 22.5927, df = 3, p-value << 0.001 for CM).

Another important result is the proportion of IP boundaries which are located within syntactic chunks (24% of left IP boundaries and 17% of right IP boundaries; χ^2 = 24.1606, df = 1, p-value << 0.001). This means that beginnings of intonational phrases are more often misalign with syntactic constituent boundaries than their ends. At the same time, this result quantifies the non-isomorphism between prosodic and syntactic structure: for while most theories assume that prosodic phrase breaks do not always coincide with syntactic phrase boundaries, the magnitude of the effect was not evaluated.

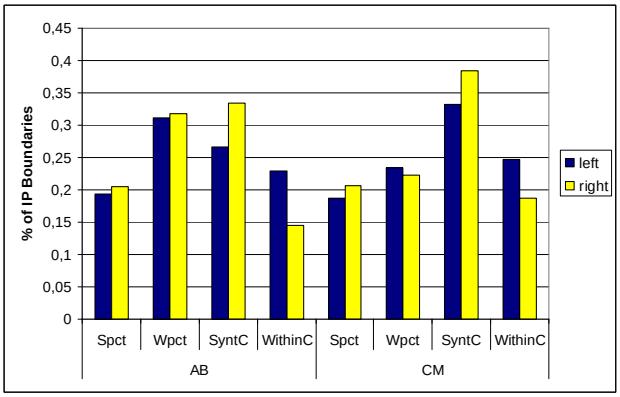

Figure 1. Mean percentage of co-occurrence of *IP* boundaries with punctuation mark or syntactic constituent

Table 3. *Evaluation statistics.*

	Left boundary	Right Boundary
Recall	0.46	0.48
Precision	0.65	0.66
F-measure	0.54	0.55

3.4. Uncompleted *IPs*

Uncompleted *IP* (*ipa*) account for suspensions in speech delivery. They correspond to approximately 11 % of all annotated units: note though that this annotation doesn't take into account all the perceived dysfluencies and interruptions within intonational phrases. Our further analyses will reveal whether or not the boundaries of *ipa* coincide with *AP* boundaries.

The data on the co-occurrence of punctuation marks and syntactic constituents and *ipa* prosodic boundaries are presented on Figure 2. These data show that in 52 % of cases for the speaker AB and in 46% of cases for the speaker CM the inserted punctuation marks are aligned with left boundaries of uncompleted *IP*s (*ipa*), though this is the case for only for 25% and 27 % of right boundaries respectively. On the other hand, 50 % and 36 % of right *ipa* boundaries occur within syntactic constituents; for the left *ipa* boundaries we observed such a pattern in only 23% and 28 % of cases. In this respect the *ipa* units differ from full *IP*s, though their special status should be further investigated. We have mentioned previously that the main criterion used for annotating an uncompleted *IP* was a perceived pitch reset: we plan to address the issue of pitch reset across boundaries in our future work. At the same time, melodic contours associated with the *ipa* units need also to be investigated first, with the reference to the typology of distinctive pitch contours for French and secondly, in their interaction with discourse structure.

A preliminary analysis of the morphosyntactic category of words occurring at *ipa* boundaries indicates that uncompleted prosodic units tend to end in syntactically unmotivated positions. At the same time, these interruptions tend to occur rather at the beginning of a syntactic constituent (after initial function words, such as determiner, preposition, auxiliary verb or conjunction, 49.62% of cases). This result suggests that speakers start their utterances before planning it completed. Alternatively, in an interactive setting, it may reflect a strategy whereby the speaker seeks to keep his turn in conversation and non to be interrupted.

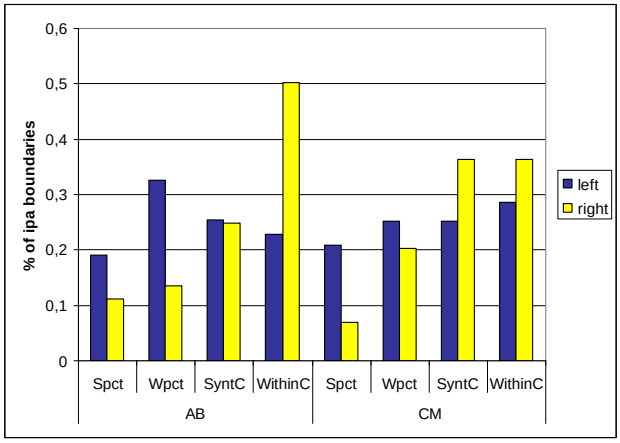

Figure 2 : Co-occurrence of ipa boundaries and syntactic constituent boundaries

4. Discussion and Conclusions

In this study we investigated the dependency relationships between syntactic and prosodic organisation in a large annotated corpus of French conversational speech. The results of inter-transcriber agreement test show that two annotators are consistent in their labelling of prosodic phrasing at the level well above the chance. At the same time, a qualitative analysis reveals annotators' individual strategies, namely in respect with potential Intermediate Phrases in French: the issue of phonetics and phonology of Intermediate phrase boundaries needs thus further research based both on attested corpora and on specially designed stimuli.

One of the central syntactic units in our study was the pseudo-phrase, as defined by strong and soft punctuation marks. In more than 60% of cases the boundaries of these pseudo-phrases are the boundaries of *IPs*. At the same time, 50% of *IP* boundaries are aligned with smaller syntactic constituents or fall within these constituents. It should be taken into account that statistical models underlying syntactic parsing were first developed for the analysis of written texts. On this basis we expect that syntactic parsers specially designed for and trained on conversational speech corpora would perform better.

We also found that 24% of left IP boundaries and 17% of right IP boundaries occur within syntactic constituents. These data provide a quantitative measure of the non-isomorphism between syntactic and prosodic structure. In future studies we plan to compare the misalignment effect in spontaneous and read speech as well as to investigate the impact of eurhythmic constraints in their interaction with syntactic structure in conversation.

In this study we have introduced a new unit, the uncompleted Intonational phrase (*ipa*). This unit appears to account for cases where a continuous delivery of speech is interrupted. We assume that such interruptions could not be treated as major prosodic breaks, or *IP* boundaries, since they do not serve any structural function in speech. They do however intervene in turn-taking organisation in conversation. In fact, *ipa* boundaries, more often than structural *IP* boundaries, are found within syntactic constituents. Specifically, right *ipa* boundaries are internal to syntactic constituents in 43% of cases, though this is the case for only 17% of full IP boundaries. We observed that the *ipa* units were associated with at least 3 different conversational events: word search, turn holding or turn yielding. We assume (and our preliminary informal observations seem to confirm it) that *ipa* units differ in their prosodic properties according to the category of conversational event.

Overall, our results provide a better understanding of the syntax ~ prosody interface as it is realised in conversational speech. It thereby provides an insight into the factors governing the structuring of speech. Our future work aims at developing acoustic and linguistic models of prosodic phrasing in conversation. Our present and future results are expected to be useful both for text-to-speech synthesis, for speech recognition applications and for the development of tools for corpus annotation.

5. Acknowledgement

This research was supported by the ANR French agency under grant no ANR BLAN08-2_349062 (Tools for Multimodal Annotation / Outils de traitement de l'information multimodale project).

6. References

[1] Clifton, C. Jr., Carlson, K. and Frazier, L., "Informative prosodic boundaries", Language and Speech, 45:87-114, 2002.

[2] Price, P., Ostendorf, M., S. Shattuck-Hufnagel, and C. Fong, "The use of prosody in syntactic disambiguation", Journal of the Acoustical Society of America, 90:2956-2970, 1991.

[3] Pynte, J. and Prieur, B., "Prosodic breaks and attachment decisions in sentence parsing", Language and Cognitive Processes, 11:165-191, 1996.

[4] Nespor, M. and Vogel, I., "Prosodic Phonology", Foris Publication, Dordrecht, 1986.

[5] Selkirk, E., "The prosodic structure of function words", in In J. Beckman, L. W. Dickey and S. Urbanczyk [Eds.], Papers in Optimality Theory. University of Massachusetts Occasional Papers. Amherst, Mass: GLSA, 18: 439-469, 1995.

[6] Beckman, M. E. and J. B. Pierrehumbert, "Intonational structure in Japanese and English", Phonology Yearbook 3, 255-309, 1986

[7] Gee, J. and Grosjean, F., "Performance structures: A psycholinguistic and linguistic appraisal", Cognitive Psychology, 15:411-458, 1983.

[8] Bachenko, J. and E. Fitzpatrick, "A computational grammar of discourse-neutral prosodic phrasing in English", Computational Linguistics, 16(3):155-170, 1990.

[9] Selkirk, E., "Phonology and Syntax: The Relation Between Sound and Structure", Coll. Current Studies in Linguistics Series, 10, Cambridge, MA, USA: The MIT Press, 1984.

[10] Pierrehumbert, J. and Beckman, M.E., "Japanese Tone Structure", Coll. Linguistic Inquiry Monographs, 15. Cambridge, MA, USA: The MIT Press.

[11] Jun, S.-A. and C. Fougeron, "Realizations of accentual phrase in French intonation", Probus 14:147-172, 2002.

[12] Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B. and S. Rauzy, "Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle", Traitement automatique des langues (TAL), 49(3):105-134, 2008.

[13] Blache P. and Rauzy S., « Influence de la qualité de l'étiquetage sur le chunking : une corrélation dépendant de la taille des chunks », Proceedings of the TALN conference, 290-299, 2008, Avignon, France.

[14] Paroubek P., Robba I., Vilnat A. and Ayache C., "Data Annotations and Measures in EASY, the Evaluation Campaign for Parsers in French", Proccedings of the 5th international Conference on Language Resources and Evaluation", 314-320, 2006, Genoa, Italy.

[15] Gendner V., Illouz G., Jardino M., Monceaux L., Paroubek P., Robba I. and Vilnat A., "PEAS, the first instantiation of a comparative framework for evaluating parsers of Frenc", Research Notes of EACL2003, 2003, Budapest, Hungaria.

[16] Di Cristo, A. and Hirst, D., "Vers une typologie des unités intonatives du français", XXIème JEP, 219-222, 1996, Avignon, France.