Statistical and temporal properties of prosodic units in French conversational speech

Irina Nesterenko

INALCO, France

irina.nesterenko@inalco.fr

Abstract

Our study investigates statistical and temporal properties of prosodic units, which were previously identified within laboratory phonology paradigm, in a corpus of French conversational speech. Prosodic annotation of our corpus implements two-level hierarchical model distinguishing major prosodic units (Intonational Phrases, *IPs*) and minor prosodic units (Accentual Phrases, *AP*). Both temporal data and distribution of the number of APs in an IP evidence the global tendency to produce shorter units in conversation. Moreover, Intonational phrases containing no more than two Accentual phrases cover 80% of the data. We discuss the implication of these results for both phonological studies on the constraints on prosodic phrasing and oral document tagging.

Index Terms: French, prosodic phrasing, conversational speech, temporal organization

1. Introduction

Our study focuses on the issue of prosodic phrasing in conversational speech. 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. Recently, prosodic phrasing has been the subject of advanced formal modeling within the framework of prosodic phonology [1-2] though the extensive phonetic and phonological studies dealt with laboratory speech.

In the models of prosodic phrasing proposed for French it is common to distinguish two levels of phrasing above the word (cf. [3]): Intonational phrases (IP) and Accentual phrases (AP, though the label can differ between the authors, cf. rhythmic units of [4], prosodic words of [5] and phonological phrases of [6]). Both units receive in speech a specific phonetic marking. Thus, there is an obligatory F0 rise on the last syllable of the accentual phrase (labeled LH* in autosegmental-metrical model of Jun & Fougeron) accompanied by pre-boundary lengthening [7] and, optionally, by an initial rise (Hi) on the first syllable. On the other hand, the intonational phrase is the domain of a major intonation contour and is characterized by greater degree of pre-boundary lengthening [3, 8]. There is as well semantic-pragmatic and syntactic constraints on Intonational phrase boundaries distributions: for example, several syntactic constructions, such as root clauses, vocatives and parenthetical expressions, form *IPs* of their own; cf. as well Sense unit constraint as defined in [9] and tested in [10] .

In our study we rely on the axiom that prosodic units previously identified within laboratory phonology paradigm could be identified in conversational speech [11]. We consider that such corpus studies are about how the language is spoken and how available prosodic means are exploited by the speakers, prosodic units in conversation being endowed with organizing function. We addressed the issue of interaction between two levels of prosodic hierarchy in conversational speech and we focused on temporal properties of prosodic constituents.

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 found themselves involved. From the interactional point of view, the corpus is not homogenous, combining negotiation sequences, question-answer exchanges and substantial monologues. 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 IP and AP boundaries. This annotation was guided by perception, based on a distinction between strong and weak prosodic breaks. Acoustic cues to prosodic phrasing, salient in perception, were also taken into account. Thus 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. Acoustic and perceptual cues to an AP boundary are: i) specific pitch movement (final rise); ii) relative scaling of adjacent APs; iii) slight degree of preboundary lengthening.

10.21437/SpeechProsody.2014-154

Before proceeding with the phonetic and statistical analyses, we ascertained that our annotations were sufficiently reliable (cf. [13, 14] attesting mean kappa Cohen values of 0.79).

3. Results

3.1. Two levels of hierarchy

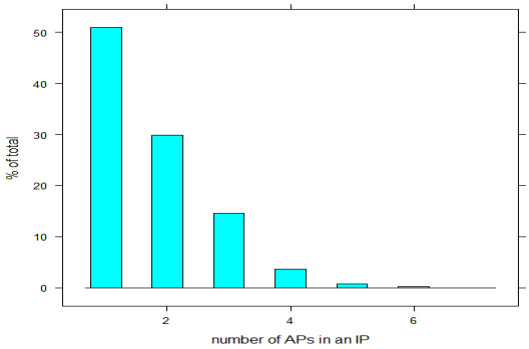

Figure 1: Distribution of the number of Accentual Phrases per Intonational Phrase

First, we looked at the distribution of number of APs in an IP (Fig.1). On average, an IP groups together 1,74 APs. Note that almost half of the identified IP units contain only one AP. In [15] studying radio speech in English, the author communicates that the IPs with no more than two accents cover 80% of the corpus, though almost equal proportions of IPs with one (39%) and two (41%) prosodic prominences was observed in this study. In our data, IPs containing only one AP, i.e. with only one intonational prominence, dominate. We consider such a distribution as quite typical for conversational data: in fact, conversation abounds in short replies and interaction words, speakers need to provide context settings for their interlocutors; as a consequence, quite often, but not always, such units tend to form a prosodic unit on their own and are quite often cumulated at the beginning of a phrase (cf. Fig 2 which illustrate the waveform, the F0 contour and phrasing annotation of the utterance $[(Demain)_{AP}]_{IP}[(c'est)_{AP}(le\ repas)_{AP}(de\ Noël)_{AP}]_{IP}$, 'Tomorrow, there is a Christmas meal': time setting adverbe *demain* 'tomorrow' being set apart in a separate IP (of 0.517s); *we should specify that there is no error in Praat's pitch detection at the end of first IP, the gap being induced by creaky voice*).

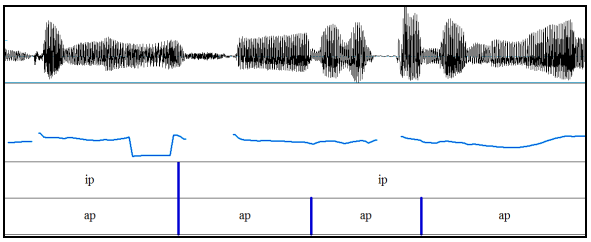

Figure 2: Waveform, F0 contour and phrasing annotation of the utterance $[(Demain)_{AP}]_{IP}[(c'est)_{AP}(le\ repas)_{AP}(de\ Noël)_{AP}]_{IP}$, 'Tomorrow, there is a Christmas meal'

Our data attest that there is a relatively small number of units with more than 3 APs: such a finding has its implications for studies of prosody-syntaxe mapping and theoretical issue of prosodic levels in French. In fact, laboratory studies showed that IPs with more than three APs tend to be restructured and such a restructuring is accompanied by the emergence of an additional level of phrasing, that of Intermediate Phrases (*ips*) (such restructuring applies to specially identified syntactic structures [17,18],; cf. as well the constraint on prosodic phrasing authorising only 2 prosodic prominences per unit). Our data indicate that such a restructuring is probable in only 20% of units, all the syntactic structures put together. We suggest that Intermediate phrase level in French should not be described not within Strict Layer Hypothesis on prosodic phrasing [9], but within a probabilistic framework for which corpus studies are an important source of data: the level of Intermediate Phrases emerges whenever the required conditions are met, as to the distribution of prosodic prominences, focusing and speech rate.

3.2. Temporal organisation of perceived units

Next we looked at temporal organisation of prosodic units in spontaneous speech and analysed the distributions of durations.

3.2.1. Intonational phrases

Mean duration of IPs in our data is of 0.841s. (75% of units are shorter than 1.14s., though they exhibit a rather important variation, coefficient of variation $C_v = 0.589$). In our previous study based on Russian spontaneous speech [16] we obtain similar results for major prosodic units (mean duration of 0.86s.). Both contrast with the data on reading in which the authors communicate the mean durations of prosodic units as of 1.2-1.6 s. Consequently, mean IP duration seems to be a property, which allows distinguishing conversation and reading and consequently, tagging oral documents from these two styles.

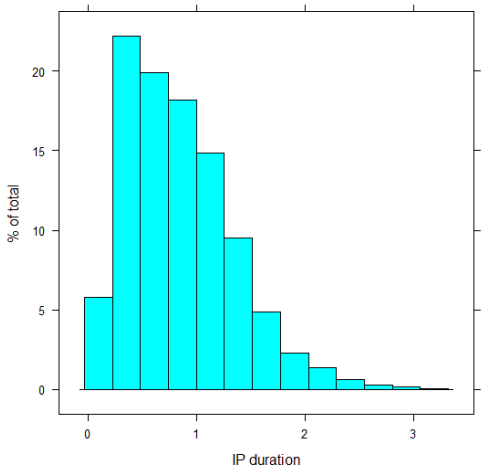

Figure 3: Distribution of Intonational Phrase durations

On the next step, we compared IP length distributions for two speakers separately (Table 1). In our study, one of the speakers tends to produce shorter units than the other, but the mean values are still under the level observed in earlier studies of reading. The differences between the speakers could be related both to speech rate differences (cf. [17] on longer units in fast speech) and to different strategies for prosodic structuring of speech (this hypothesis could be tested by comparing unit lengths in number of syllables).

Table 1. *Description of IP durations distributions for two speakers*

Speaker	AB	CM
Min	0.08	0.084
1st quartile	0.528	0.358
Median	0.845	0.625
Mean	0.905	0.709
3rd quartile	1.208	0.96
Max	3.194	2.844

3.2.2. AP durations

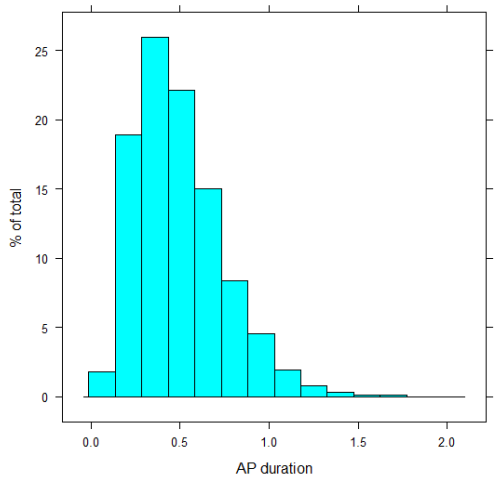

Figure 4: Distribution of Accentual Phrase durations.

The AP lengths in our corpus present the similar distribution as IP lengths with mean AP duration being of 0.496s. (75% of units being shorter than 0.636s. with substantial internal variation, $C_v = 0.501$).

3.2.3. Speech-in-interaction effect

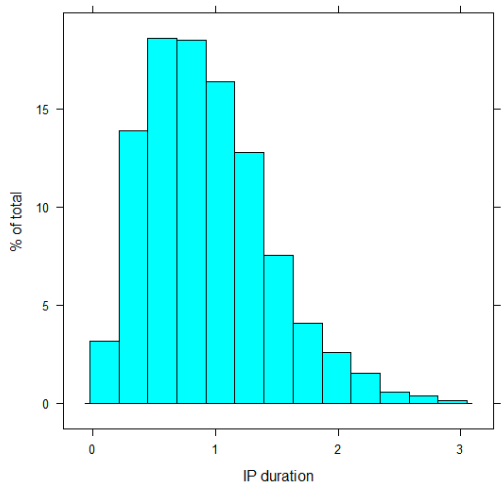

Figure 5: Distribution of Intonational Phrase durations in spontaneous monologues

As described earlier, our corpus contains both task- and strategy negotiation sequences and monologues from one and the other speakers. During these monologues, the participation of the second interlocutor is limited to backchannels and some rare questions and reactions. Thus, to sort out speech-in-interaction effect we limited ourselves to monologue sequences only and looked for IP duration distributions in this sub-corpus. IP number was thus limited by 30%. The mean IP duration for sub-corpus is slightly higher than for the totality of data: mean IP duration is of 0.937 and 75% of IPs are shorter than 1.234s, variation coefficient value (C_v=0.541) indicating great dispersion of values around the mean. This finding goes in the sense of our hypothesis

4. Discussion and Conclusions

Based on a large corpus, our study is about prosody production. Starting with the idea that prosodic constituents previously identified in laboratory speech are relevant for the analysis of conversational prosody, we investigated hierarchical relation between major (Intonational phrases) and minor (Accentual phrases) prosodic units in French. In the analysis of the results we took into account both phonological constraints on prosodic phrasing and details of phonetic (temporal) organisation of these units.

We found that Intonational phrases containing only one Accentual Phrase are dominating in our corpus, quite in agreement with its style. In this aspect, our corpus contrasts with the read speech and news broadcasts [15], though we do not possess directly comparable data in French. Moreover, Intonational units containing no more than two Accentual Units cover 80% of the data. This finding has its implications for the issue of number of levels in prosodic structuring in French: in fact, it means that there are the appropriate conditions for the emergence of an intermediate level of phrasing in only 20% of contexts. We need to further test the hypothesis if there really are any phonetic indices (greater pre-boundary lengthening and/or tonal cues) indicating the presence of an intermediate phrase boundary in these potential locations. We propose that further modeling of prosodic phrasing in spontaneous speech integrates this probabilistic data of the number of phrasing levels.

As to the temporal dimension of prosodic units, we obtained that spontaneous speech is structured in shorter units than read speech (mean IP duration of 0.841s in our study and that of 1.2 -1.6 s. communicated in reading). For the restricted corpus of monologue sequences, mean IP length slightly augments up to 0.937s. Restricted corpus contained less short units which function as backchannels, though there is still a number of context setting phrases promulgated to the level of IPs in conversation. Both the data on IP duration and that on AP duration exhibit a great dispersion around the mean value. At the same time, our data on conversational speech show that mean IP duration could potentially be one of the parameters allowing for tagging apart conversations from other oral documents.

We observed as well speaker's effect on phrase duration, our data attesting the tendency of one of the speaker to produce shorter units than the other. This difference could be imputed to speech rate differences and subsequent tendency to produce larger units (cf. [17] uncovering at least two mechanisms available to the speaker in fast speech rate which have an impact on the number and the strength of prosodic boundaries) or individual strategies in speech structuring. Both hypotheses need further investigation.

5. References

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

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

[3] Jun, S.A. and C. Fougeron.,"A Phonological Model of French Intonation", Probus, 14, 147-172, 2000.

[4] Di Cristo, A., "Intonation in French". In D. Hirst, & A. Di Cristo (Ed), Intonations systems: A survey of twenty languages, Cambridge: Cambridge University Press, 195-218, 1998.

[5] Vaissière, J., "Rhythm, Accentuation and Final Lengthening in French". In J. Sundberg, L. Nord and R. Carlson [Ed], Music, language, speech and brain, 59,108-120, Stockholm : Wenner-Gren, International Symposium Series, 108-120, 1992.

[6] Post, B. "Tonal and Phrasal Structures in French Intonation". The Hague: Holland Academic Graphics, 2000.

[7] Pasdeloup, V., "Modèle de règles rythmiques du français appliqué à la synthèse de parole". Doctoral thesis, Université de Provence, 1990.

[8] Di Cristo, A., "De la microprosodie à l'intonosyntaxe". Aix-Marseille : Université de Provence, 1985.

[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.

[10] Frazier, L., Clifton, Ch. Jr., and Carlson, K., "Don't break, or do: prosodic boundary preferences", Lingua, Vol. 114(1), 2004, 3-27.

[11] Portes, C., Betrand, R., " Permanence et variation des unités prosodiques dans le discours et l'interaction", Journal of French Language Studies, 2011.

[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] Cohen, J., "A coefficient of agreement for nominal scales", Educational and Psychological Measurement, Vol. 20, 1960, 37-46.

[14] Nesterenko, I., Bertrand, R., Rauzy, S., "Prosody in a corpus of French spontaneous speech: perception, annotation and prosody ~ syntax interaction", Speech Prosody, Chicago, USA, 2010.

[15] Dainora, A., "An Empirically based Probabilistic model of Intonation in Americain English", PhD Dissertation, University of Chicago, IL, 2001.

[16] Nesterenko, I., Skreline, P., "Some evidence on the phonetics and phonology of prosodic phrasing in Russian". Interspeech 2007 : 438-441

[17] Jun, S.A. and Fougeron, C., "The Realizations of the Accentual Phrase in French Intonation". Probus, 14, 147-172, 2002.

[18] D'Imperio, M., Michelas, A., « Embedded register levels and prosodic phrasing in French ». Speech Prosody, Chicago, USA, 2010, 100879:1-4