PROSODIC CHARACTERISTICS OF READ AND SPONTANEOUS SPEECH IN FRENCH

Philippe Martin

CLILLAC-ARP, Univ. Paris Diderot Paris Cité, Paris, France
philippe.martin@linguist.jussieu.fr

ABSTRACT

This study compares the read and spontaneous speech prosodic characteristics of two relatively small corpora in French (about 3 minute's length). Acoustic data such as syllabic rate, number of effective stressed syllable vs. theoretical prediction, prosodic hierarchy and realization of melodic contours are compared for both styles. The predicting power of two theoretical approaches, autosegmental-metrical and functional-cognitive, is then assessed when confronted with the experimental data.

Keywords: read speech, spontaneous speech, intonation, autosegmental-metrical, functional-cognitive

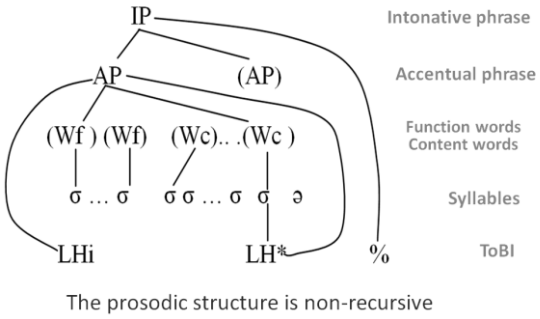
1. INTRODUCTION

Although the Autosegmental-Metrical (AM) framework for intonational analysis became the *de facto* standard, one of its limitations pertains to the difficulty to distinguish clearly between abstract categorical phonological elements and gradients of phonetic implementation. Indeed the generalized use of the ToBI transcription system to label pitch accents and boundary tones highlights the recurrent difficulty to choose between one label and another whereas this choice is of paramount importance to establish phonological categories. An obvious problem resides in the possibility given to a *bona fide* transcriber to select the best label that would later best suit some preferred underlying theoretical phonological model.

Given the large variance of data found in spontaneous speech (pertaining in particular to melodic curves), things may get worse when the AM approach is used to analyze this type of data. It was even suggested that no transcription should occur before their analysis to prevent some uncontrolled pre-analysis filtering of the data. It is not surprising that most if not all studies in French using the AM framework limit themselves to the analysis of laboratory type read sentences hardly more complex that *le colléreux garcon ment à sa mère* (the angry boy lies to his mother) [3] or *la mamie de Rémy demadait l'institutrice* (Remy's grandmother asked for the teacher) [2].

Another drawback of the AM approach in our view is the lack of explanatory power: the prosodic structure (PS) as defined in Fig. 1 appears as the only valid model valid for all languages [7], although numerous claims have been made to allow for an intermediate IP in the prosodic structure [4].

Figure 1: The Autosegmental-Metrical prosodic structure.



This prosodic structure is presented as universal, but does not propose any explanatory principle than itself, i.e. to be well designed to accommodate all prosodic events observations.

2. THE FUNCTIONAL-COGNITIVE APPROACH

To analyze spontaneous data, we will rather use another – largely ignored – approach called Functional-Cognitive (FC) based on a principle allowing data NOT to have to be transcribed beforehand by a ToBI or other notation system. The FC theory derives from observations made on short term memory capabilities of the listener, who has trouble to recollect strings larger than some 7 or 8 syllables without converting them into a higher rank linguistic unit. These converted syllabic sequences form a minimal prosodic unit assembling lexical and functional words. The conversion is triggered by a specific prosodic event (PE) which in French is (almost) always located on

the last syllable of the sequence (which is therefore stressed). Although appearing superficially very similar to AM definition of prosodic words, this view induces a set of properties and constrains which actually define the FC prosodic structure. This structure is a priori not congruent to any other structure that would organize the sentence, and in particular the syntactic structure [5].

In the Functional-Cognitive approach, the prosodic structure is recursive: minimal prosodic units are assembled into prosodic groups which are themselves grouped into larger groups until the last level corresponding to the whole sentence is obtained. Furthermore, the constraints pertaining to the prosodic structure are not defined by the Strict layer Hypothesis as in AM [3] but by a set of rules such as:

The 7 syllables rule: in a sequence of 7 syllables, one at least must be stressed [1]. For example, a long (and rare) word such as *paraske'videkatripho'bie* (the fear of Friday 13th) requires two stressed syllables;

Stress clash rule: no consecutive stressed syllables, unless separated by a pause of a sequence of different modes consonants (for example stop + fricative);

Syntactic clash rule: no prosodic grouping of two stress groups dominated by two distinct nodes in the syntactic structure. For example [*le grand-père*] [*de Marie adore*] [*les chocolats*] (*Mary's grandfather loves chocolates*) is not allowed;

Eurhythmicity rule: if more than one prosodic structure can be associated with a given syntactic structure, the eurhythmic one would be preferred. For example [*Marie adore*] [*les chocolats belges*] (*Mary loves Belgian chocolates*) (4 syllabes vs. 5 syllables) is preferred to [*Marie*] [*adore les chocolats belges*] (2 syllables vs. 5 syllables) although the latter is congruent to syntax.

Another important difference with AM theory pertains to the grammar of intonation defined for a given language. Although in both theories prosodic markers are attached to stressed syllables and prosodic structure boundaries, in FC, prosodic markers are defined contextually, i.e. their characteristics in terms of phonetic features such as syllabic duration, melodic level and movement depend directly of the complexity of the prosodic structure, according to the number of contrasts to be realized the ensure the differentiation of prosodic markers in their function of encoding the PS and thus allowing the listener to retrieve it. Whereas for instance only one prosodic feature is necessary and sufficient to contrast the PE occurring on the stressed syllables of *le frère de **Max** adore les choco**lats*** (Max brother loves chocolates), two of those features are needed to contrast the 3 PE located on the stressed syllables of *les grand-pa**rents** de Marie-Char**lott**e adorent le choco**lat*** (Marie-Charlotte grandparents adore chocolates). This simple minimal contrast principle giving a proper account of the data is totally absent in the AM framework (although it could of course be introduced with complex adjustment rules).

3. DATA

About sixty seconds taken from two examples of read and spontaneous speech were analyzed in terms of syllabic duration and fundamental frequency movements and level. These recordings were taken 1) AM from the CFPP 2000 corpus [1] and 2) NS from the Rhapsody corpus [6].

AM a female speaker, age 46 at the time of the recording, was born and lives in Paris. She presently assists handicapped children in the primary school system;

NS a male speaker, age 53 at the time of the recording, was born and lives in Paris. He is presently president of the French Republic.

The NS corpus consists of a read address delivered in an official occasion in front of a military audience. It is not clear if the text was written by the speaker himself on not.

3.1. General characteristics of spontaneous speech

As with most of non-prepared speech data, characteristic occurrences of hesitations (filled by *euh* 16 cases), repetitions (3 cases), reformulations (6 cases) and abandons (9 cases) are found in the AM corpus. These types of occurrences are normally not found in read speech, and indeed none are found in the NS recording.

Table 1: Compared number of hesitations, repetitions, reformulations and abandons between speakers NS and AM.

	NS Read	AM Spontaneous
Hesitation	0	16
Repetition	0	3
Reformulation	0	6
Abandon	0	9

Detailed differences of both corpora are given in tables 2 to 6.

3.2. Pauses and syllabic rate

As expected, the syllabic rate for read speech is higher compared to read speech. The resulting syllabic rate is similar, but the large number of long pauses by NS implies a faster syllabic tempo as found normally in read speech.

Table 2: Compared duration, pause duration, number of syllables and syllabic rate (excluding pauses) between NS and AM. The NS syllabic rate equals 1.34 the AM syllabic rate.

	NS Read	AM Spontaneous
Total duration	190 s	149.5 s
Total pause	60.9 s	7.3 s
Nb. syllables	772	632
Syllabic rate	5.97 syl/s	4.44 syl/s

3.3. Stressed syllables

The rate of effectively perceived stresses syllables over the theoretical number of stressed syllables (corresponding to the number of content words) is larger for read speech, which can be easily explained by the context and the nature of a read discourse delivery. Effectively stressed syllables were acoustically characterized by a glissando (in Semitone/sec) higher than the glissando threshold).

Table 3: Compared effective stressed syllables related to stressable syllables (predicted phonological stress).

	NS Read	AM Spontaneous
effective stress / predicted stress	203 / 219 = 0.92	157 / 204 = 0.77

Distressing of a stressable syllable occurs if the number of syllables in the stress group is inferior or equal to 7 (depending of the tempo, which is faster for speaker NS).

3.4. Secondary stress

Secondary stress occurrences (located on the first syllable on content words (i.e. verbs, nouns, adverbs and adjectives, vs. non grammatical words such as conjunctions, auxiliaries, clitics…).

Table 4: Compared number of secondary stress occurrences.

	NS Read	AM Spontaneous
Secondary stress	23	6

Although clearly describes by some autosegmental-metrical description of French intonation [2], the "arc accentuel" scheme is found only once in our data. The rare cases observed in both recordings have the last syllable of the first word involve stressed as well.

3.5. Hierarchy and distribution of melodic contours

The observation of melodic contours on both recordings (acoustical data were obtained with the WinPitch software) leads to the following hierarchy:

NS read corpus

Table 5: Types of melodic contours ending stress groups (prosodic words) for speaker NS.

Final conclusive contours + pause	**17**
Major continuation rising contours + pause	**57**
Continuation rising contour + no pause	**28**
Major continuation falling contours	**37**
Minor flat, rising and falling contours	**64**
Falling-rising contrast of melodic slope	**39**
Rising-falling contrast of melodic slope	**2**
Rising contour secondary stress	**18**
Falling contour secondary stress	**5**
Total number of contours	**203**

AM spontaneous speech corpus

Table 6: Types of melodic contours ending stress groups (prosodic words) for speaker AM.

Final conclusive contours + pause	**21**
Major continuation rising contours + pause	**0**
Continuation rising contour + no pause	**55**
Major continuation falling contours	**15**
Minor flat, rising and falling contours	**66**
Falling-rising contrast of melodic slope	**11**
Rising-falling contrast of melodic slope	**8**
Rising contour secondary stress	**6**
Falling contour secondary stress	**0**
Total number of contours	**157**

Abandons with rising melodic contour: 4

Besides falling conclusive melodic contours ending sentences, speaker AM uses ponctuants such as *bon, ben, ah* at the beginning of sentences, and *voila* at the end. At three occasions, the same speaker abandons the construction of the syntactic structure and its prosodic counterpart which ends with a rising melodic contour.

3.6. Contrast of melodic slope

Tables 5 and 6 show how this contrast of slope is more present in read speech (61 occurrences) whereas spontaneous data contain only 19 cases.

A basic differentiating mechanism between prosodic markers in French involves a contrast of melodic slope, where the dependency to the right indicated by prosodic markers placed on the last pronounced syllable of the prosodic word is instantiated by a melodic variation of opposite slope (i.e. rising vs. falling or falling vs. rising)

relatively to the contour ending a lager unit located on its right.

Contrary to languages such as English, stressed syllables are always in final position in prosodic words and prosodic groups, which imply that prosodic markers located on stressed syllables, are boundary markers as well. In this sense, prosodic contours, excluding those indicating an emphasis and located on the first syllable of content words, are all indicate a continuation (An rather old terminology still used today qualify those contours as major and minor continuation [8]).

Using the ToBI transcription, since high and low tones indicate targets, the contrast in melodic slope appears indirectly in the alignment of the H* tone, at the beginning of the stressed syllable for falling contour, and at its end for a rising contour.

4. COMPARED AM-FC PREDICTING POWER

Analysis of the data indicates that, due to the use of different transcription systems (Tone targets vs. contours), the AM approach has some difficulty to give a proper account of the prosodic events pertaining to these two recordings.

The contrast in melodic slope, more frequent in read speech, could be easily handled by some appropriate rule such as H*L -> L*H / --H*L and L*H -> H*L / --L*H, or perhaps with an appropriate tone alignment rule.

On the other hand, differentiating with the ToBI transcription between large and restricted melodic excursion as observed in the data for both rising and falling contour may be difficult to achieve with a reasonable intuitive understandable mechanism.

Let us add that the AM model lacks, at least when applied to French data, an explanation principle giving a proper account of the contextual realization of melodic contours.

5. CONCLUSION

The predicting power and appropriateness of both AM and FC theoretical approaches have been (somewhat briefly) assessed for two short corpora of opposite nature, involving read speech and spontaneous non prepared speech. The rather sketchy results seem to indicate that the Autosegmental-Metrical model somewhat lacks of explanatory power to give a proper account of prosodic events observed in both recordings.

6. REFERENCES

[1] CFPP 2000 Corpus de français parlé parisien, *http://ed268.univ-paris3.fr/syled/ressources/Corpus-Parole-Paris-PIII/CFPP2000.pdf*

[2] d'Imperio, M., Michelas, A. 2010. Embedded register levels and prosodic phrasing in French, *Proc. 5th Speech Prosody Conference* Chicago.

[3] Jun, S.-A., Fougeron, C. 2002. The realizations of the accentual phrase in French intonation, *Probus* 14, 147-172.

[4] Ladd, R. 1996. Intonational phonology. *Cambridge Studies in Linguistics* Cambridge: Cambridge University Press, 79, 334.

[5] Martin, P. 2009. *Intonation du français*, Paris: Armand Colin, 256.

[6] Rhapsodie 2010. Corpus prosodique de référence en français parlé *http://rhapsodie.risc.cnrs.fr/en/archives.html*

[7] Selkirk, E.O. 2005. Comments on intonational phrasing in English. In Frota, S., Vigario, M., Freitas, M.J. (eds.), *Prosodies: Selected Papers from the Phonetics and Phonology in Iberia Conference, 2003*. *Phonetics and Phonology Series 2005*. Berlin: Mouton de Gruyter.