TOWARDS AN INTONO-DISCURSIVE SEGMENTATION OF SPONTANEOUS UTTERANCE IN FRENCH

*C. Leleu *$, A. Lacheret-Dujour $@ and M-A. Morel **
**Paris III University, Paris, France*
$Limsi-CNRS, Orsay, France
@ELSAP, Caen, France

ABSTRACT

Our goal is to define a set of phonetic markers that can be used to segment an utterance and to predict discursive boundaries in spontaneous French speech. Our acoustic analysis of descriptive monologues is carried out on two prosodic parameters and the relationship between them: (1) melody, (2) pauses.

1 INTRODUCTION

With the hypothesis that prosody has a demarcative function in the structuring of a spoken text into discourse units, we propose a method of segmentation based on high and low points of melody and pauses. For the present, this method is not totally discursive because the prosodic status of functional words is rather ambiguous; nevertheless, we will show how the coupling of cues can be efficient.

The entity of our analysis is no longer the sentence but the paragraph which we divide into intonative and pausal constituents (refered to hereafter as **IC**s and **PC**s respectively).

Our methodology is based on the hypotheses of our research team [1, 2] on the intonative and enunciative structuring of spontaneous utterances which are given by the variation in the F0 level-points of the final syllable of prosodic constituents. We also rely on the work of some linguists, for example Mertens [3], Rossi [4], etc., on the structuring of French utterances by intonation. This linguistic analysis is however restricted in our work because we foresee formal constraints of implementation and also because our database is highly specific. Nonetheless, the scope of our analysis is made larger with pausal constituents added to intonative ones.

First, we will describe the database used to carry out our study. Next, we will describe the different stages of the segmentation allowing us to derive the discursive structure from acoustic cues. Then, we will show the formation of PCs and propose hypotheses on the relations between pausal and intonative boundaries. Finally, intono-discursive markers are interpreted on the enunciative level and corresponding boundaries are given a specific weight relative to their position in the hierarchy.

2 ANALYSIS

2.1 Description of the Database

Our data was collected from the ICY database developped in Limsi by V. Péan [5] to study inter- and intra-speaker variability. ICY was obtained from a strategy by which utterances elicited are both spontaneous, i.e. generated on the fly, and under the experimenter's control. The speakers were instructed to describe two identical pictures differing only in the colour or spatial positions of several objects chosen deliberately to make the speakers produce groups of words which contain a phonological context at word boundaries (for example: in the case of "pantalon orange", we can find a possible nasalization context). They did not know the real purpose of this experiment and were given a false pretext to persuade them to produce three different styles of speech of which we have studied the casual version.

As a result, such constraints give a corpus that consists of descriptive monologues with declarative sentences. They also explain why the resulting speech does not show as much variation as in real-life situations. However, this type of spontaneous discourse is more realistic than read-aloud, often isolated sentences generally used in automatic speech processing.

The pauses and F0 analysis were obtained using the UNICE software created in Limsi-CNRS (Orsay). In this article, we study 5 speakers among the 21 recorded (BP, AG, GM, GS, SR).

2.2 Data Analysis

Now, we will present briefly the different stages that allow the prosodic structure to be derived from acoustic data. Above all, correspondence rules between the phonetic and the structural prosodic levels are explained and illustrated with an example. As for pauses, they are used for the moment merely to assure the cohesion among certain ICs. We will therefore demonstrate that they form a structural system among pausal constituents, and that the relationship between pauses and melody is well worth taking into consideration.

2.2a Melodic Cues

Our analysis is not based on perceptual criteria, but acoustic data. The segmentation process that we propose consists of 5 stages (Cf. Figure 2) as explained below:

1. Lexical Filtering: between two different grammatical categories of words, that is functional and lexical ones, the former is not taken into account especially because of its lack of intonative autonomy. As a result, acoustic measurements concern only lexical words, i.e nouns, verbs, sentence adverbs like "sans doute" (not modifying adverbs, for example "beaucoup") and postpositional adjectives.

In agreement with Grosjean [6], we believe that lexical words have an independent status on the prosodic level because they bear most of the time high tones in our data. However, in spite of this relative autonomy, both words must be joined and belong to the same constituent.

2. Tone Labelling and Identification of Pauses: after recording, the speech was orthographically transcribed and divided into words and syllables. The value of F0 on each syllable of lexical words is then extracted manually.

Our two processes of tone labelling and prosodic regrouping are dependent on the 'relative height' notion which P. Mertens [3] used as well. In the same way, each tone label attributed to a syllable relies on the melodic interval with the height of the preceding syllable. As a result, each syllable is a potential point of reference for the syllable to its right, and for that reason we obtain only local informations (for instance, a syllable labelled L can be higher than H farther in the same paragraph).

We proceed therefore from left to right: a P_{n+1} point is labelled **'L'** (Low) when it is lower than the P_n point; otherwise, it is labelled **'H'** (High).

As for the determination of the initial tone height of each paragraph (noted **Pi** before identification), we proceed in the opposite direction.

On the other hand, we distinguish two different pauses: (1) silent pauses (silence and respiration) which seem to have a closing function (**Pc**), (2) non-silent pauses (glottal stop, buccal noises, vocal lengthening and typical French "euh") which are likely on the contrary to open a constituent (**Po**). We suppose that non-silent pauses have a cognitive function of lexical access or of lexical emphasis, except for false starts. As a result, pauses at the beginning of a paragraph are not taken into consideration. When several pauses follows one another, only the first one is kept. Furthermore, we need to extract the durations relative to such labels.

3. Delimitation of the Paragraph: we can determine the right hand boundary of the paragraph (noted L-) when the three conditions below are satisfied:

a. If $L_n < P_i$
b. If L_n is followed by a long (> 50 ms) and a silent pause (Pc category).
c. And if $L_n < L_{n+1}$

Then $L_n = \mathbf{L\text{-}}$

Let us add that we have noticed, at least in our data, that the declination line and the length of the paragraph coincide, which explains condition a.. As for c., it corresponds to the typical resetting at the beginning of a new utterance observed by many linguists.

4. Determination of the Prosodic Structure: in order to segment a paragraph into ICs, the notion of relative height is applied between labelled high tones. Some neighbouring units ending with H can thus be grouped. Therefore we can say that there is a conjunctive relationship between two units when the high tone of the first is lower than that of the second. On the contrary, when the high tone of the first unit is higher, there is a disjunction, in which case the highest

tone is relabelled **FH** (Final Height) and corresponds to the right hand boundary of an IC (noted #). The highest FH is noted **FH+**. We formalize this process as follows:

If $H_n > H_{n+1}$
Then $H_n = FH$

5. Grouping Function of Pauses: we have noticed that pauses can take over from melody by linking two ICs. We distinguish two kinds of pausal strategies assuring this cohesion: (1) the two ICs are separated by a Po, or (2) the second IC is between two pauses, i.e. Pc or Po In all these cases, the intermediate boundary is demoted and relabelled **H+** (Cf. Figure 2). Ex. GMS223: "un bouquet de fleurs[216] # **Po** rouges[211] # avec les tiges vertes[195] # **Pc**" (F0 in Hz).

Above all, this kind of cohesion seems very common inside nominal groups (whether its function is rather local is questionable) and allows morpho-syntactic rules to be dismissed such as: a noun cannot be separated from its post-positional adjective or complement.
In case of strong disjunction between a noun and its postpostional adjective, for instance the insertion of other words, the cohesion between them is restored using linking by both intonation and pauses. In this way, cues can be added to assure a strenghthened cohesive function.
Ex. GMS29: "y'a une petite lampe[186] en revanche **Po** en bas **Po** rose[250] **Pc**", GMS221: "il a un petit noeud papillon[138] **Po** à droite[145] **Po** orange[444] et **Pc**".

2.2b Pauses

Besides this latter partial use, pauses also form a structured system with minor and major contituents (Cf. Figure 1). The first stage of its formation concerns minor PCs all closed by a Pc. In other words, we proceed from left to right and note the boundary of a minor PC each time we encounter a Pc. If a Po is found, then we close the minor PC immediately after the nominal group which it contains. Next, we obtain major PCs by comparing the Pc duration: thus, if Pc1 < Pc2 then PC1 is included in PC2; otherwise, PC2 is included in PC1.

As for the relationship between cues, in our data, we noticed some cases of complementary distribution between FH and pauses; for example, in BPS225 we can observe that, independently of lexical and syntactic considerations quite similar here, two strategies are used by the same speaker to articulate his utterance: the first relies on pauses and the second on FH. In this case, as in our previous stage of grouping motivated by Po, we can note that pauses have a continuative function.

Ex. BPS225: "Sur le dessin[174] de gauche[163] **Pc** la table[178] basse[154] **Pc** en bas à droite[148] **Pc** est blanche[267] **FH+** sur le dessin[157] de droite[200] la table[167] basse[195] en bas[138] à droite[190] est bleue **L-**".

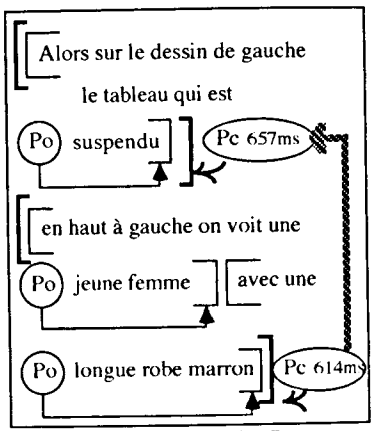

Figure 1. Organization of PCs.

However, FH and pausal effects can also be combined to mark a boundary more obviously; for instance, the major boundary in GMS21 is marked twice (FH+ and Pc). We already exploited in our rules the cohesive effect of this particular boundary on the preceding ICs, and saw that it could also restore a syntactic linking between two separated words .

2.2c Enunciative Interpretation of Prosodic Markers

According to enunciative criteria, the goal of this last part is to decide on a weight of previously determined boundaries. The paragraph, divided into blocks with indetermined statuses of utterance, frame, and rheme, is limited by Pi and L- at its edges. From an enunciative point of view, it is characterized by its thematic

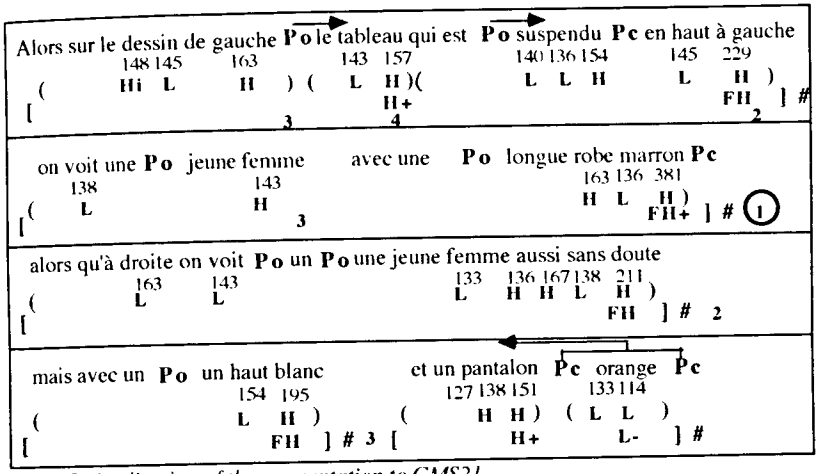

Figure 2. Application of the segmentation to GMS21.

unity due to the attention of the speaker drawn on a same object in pictures. The paragraph also contains at least two utterances, the major frame to the left and its rheme. These two words in turn are separated in the same way; in other words, utterance is recursive.

More precisely, we proceed as follows: (1) the blocks ending with FH+ and L- are considered, (2) starting from these final labels and from right to left, we divide the block into two parts at the highest point, i.e. frame and rheme, (3) the same operation is repeated up to H inside new blocks, (4) after this retroactive analysis, we attribute hierarchical values to the boundaries (1 to the strongest boundary) that correspond to the degree of inclusion of blocks (Cf. Figure 2).

ACKNOWLEDGEMENTS

This paper was supported by EA 1483 "Recherche sur le français contemporain", Paris III-Sorbonne Nouvelle.

3 CONCLUSION

Our findings, that we have recovered from several speakers, have helped to highlight the importance of the relationship between pauses and melody and to confirm that it deserves to be studied in order to improve the discursive segmentation of utterances Now, we need to give an equal weight to pauses and melody by comparing their structures more precisely. Furthermore, we have to perceptually validate the relevance of our boundaries and extend our study to spontaneous speech in different situations. As for the processing of functional words, there still exists an obstacle in this area. Further studies need to be carried out in the future in order to overcome the present problem of automatic implementation of spontaneous speech .

REFERENCES

[1] Morel, M.-A. (dir.) (1995), *Structuration linguistique du dialogue oral*, n° 1; *Langue orale et linguistique du discours*, n° 5, Documents de travail du centre de recherches sur le français contemporain.
[2] Rialland, A., Vaissière, J. (dir.) (1994), "Prosodie du français", *Rapport d'activité de l'U.R.A. 1027*, pp. 16-22.
[3] Mertens, P. (1987), "L'intonation du français. De la description linguistique à la reconnaissance automatique", *Doctorale dissertatie:*, Katholieke Universteit Leuven.
[4] Rossi, M., Di Cristo A. et alii (1981), *L'intonation, de l'Acoustique à la Sémantique*, Paris: Klincksieck.
[5] Péan, V. (1992), "Conception de la base de donnée ICY", *Notes et documents LIMSI*, n° 92-1: Orsay.
[6] Grosjean, F. & Monnin, P. (1993), "Les structures de performance en français: caractérisation et prédiction", *l'Année Psychologique*, fasc. 1, pp. 9-30.