Prosodic structure revisited: a cognitive approach

The example of French

Philippe Martin
EA 3967, UFR Linguistique, Université Paris Diderot, Paris, France
philippe.martin@linguist.jussieu.fr

Abstract
Traditionally in intonation phonology, the sentence prosodic structure is often viewed globally, taking into account all the prosodic events at once from the beginning to the end, without taking into account the sequence of events in function of time. However, from the point of view of the speaker and the listener, the situation is quite different: whereas the speaker can achieve some planning ahead in the production of the prosodic structure of read speech, it is barely the case for the listener, who has to process the linguistic information from the sequence of units perceived one by one along the time scale. In this process, prosodic events are used as signals triggering partial processing of the already perceived syllables by concatenation of strings of already stored units to form larger syllabic groups organized in stress groups. This paper looks in some details into the mechanism of this decoding process, focusing on the role of prosodic events in the specific case of French.

1. Introduction
The formal representation of the prosodic structure in intonation phonology normally extends from the beginning to the end of the sentence. Whether hypothetical-deductive or inductive, prosodic theories using the concept of prosodic structure consider prosodic events (PE) statically as part of the sentence as a whole. For instance, the autosegmental-metrical approach assigns specific tones to syntactic boundaries and stressed syllables by proceeding from a syntactic analysis [1], [2], [3], [4], whereas other theories [5], [6], [7] assign rising of falling melodic contours to lexically stressed syllables according to the declarative or interrogative terminal contour.

Applied to the encoding and decoding processes performed by the speaker and the listener, most of these formal accounts of sentence intonation assume that the entirety of the prosodic structure of each sentence is known in advance. Whereas this might be true to some extent in read speech, where punctuation marks and visual verb detection may give some indications on the prosodic structure to be associated with the text, it is less the case for spontaneous (unprepared) speech, except perhaps for short sentences containing only a few number of stress groups. From the point of view of the listener, the prediction possibilities are even more restricted (except for very short sentences).

Adopting again the points of view of the speaker and of the listener rather than the linguist, this paper proposes a (relatively) new way to describe the prosodic structure, centered on the dynamic process involved in the encoding and decoding of the prosodic events in function of time. Among other benefits, this dynamic view may lead to a better understanding of some phonological descriptions of French considered earlier as inadequate. Among other examples, the concepts of “groupe de sens” as unit of description for sentence intonation [8] and of “période intonative” [7] could be reevaluated in the light of a dynamic cognitive approach.

2. A dynamic cognitive view
From the point of view of the listener, prosodic events are perceived one after the other in function of time. Therefore their categorization (the identification of the class they belong) depends only on past and not on future events.

The prosodic structure reconsidered dynamically results from a process by which strings of syllables are hierarchically assembled thanks to the identification of each PE as belonging to classes known by the listener. This process involves 2 mechanisms: a) storage of the string of syllables perceived by the listener since the last advent of a PE and 2) concatenation of this string with all strings belonging to the same level (i.e. whose storage was triggered by same class PEs).

Among all PEs occurring in French, only those located on the last syllable of stress groups take part in this mechanism, at the exclusion of events located on the first syllable of lexical words, which are treated as secondary stress (a stress group contains either a Noun, a Verb, an Adjective or an Adverb). This means that the identification of PEs implied in the dynamic prosodic structure elaborated by the listener involves the identification of a syllable as stressed in order to qualify the associated PE as a triggering signal to process to storage-concatenation mechanism.

2.1 Storage-concatenation process
The dynamic view leads to consider that each PE, instantiated by various contrasts of variation and melody height, duration, of intensity, or vocalic quality, appears as a signal triggering the storage of syllables perceived since the appearance of the last prosodic signal belonging to the same class. This syllabic storage is accompanied by concatenation of the elements already present at the same prosodic “address”.

We can look into this mechanism in some details with the simple example schematized on Fig. 1 which gives a diagram of the storage-concatenation mechanism relative to a syllabic sequence σ₁, σ₂,… σₙ. In this sequence, some syllables are stressed and are therefore associated with a PE in the prosodic structure. When one of these stressed syllables is perceived, the PE attached to that syllable is identified by a grid of perception acquired by the listener.
Fig. 1. Schema of the storage-concatenation process

In our examples, this grid pertains to the knowledge of the linguistic system of French, but can possibly be adapted to idiosyncratic systems used by some speakers (see below). Each one of these identifications by the grid starts the following operation:
1) Storage of the syllabic sequence appeared since the last PE;
2) Concatenation with the sequences stored at the lower levels, if they exist, and deletion of the existing strings of syllables already stored. This procedure implies a hierarchy between the PEs: EP0 > EP1 > EP2 >… > EPn.

In the example, EP1 assembles all the sequences stored in the lower memories N2, N3 taken in the order with the sequence of syllables appeared since the last PE. This operation can be generalized on all levels. We will call this general procedure Storage-Concatenation Process (SCP).

It also results from this dynamic cognitive view that prosodic marks enter a network of relative contrasts, and do not need invariant acoustic characteristics, as long as they are identified by the listener as belonging to the same class. Another argument going in this direction comes from the large variety of the styles of production of word. An emotionally depressed speaker will realize little or no melody variations. In whispered voice, contrasts between prosodic marks will have to be realized by other means than laryngeal frequency variations, with segment durations, about 50 to 70% higher than the equivalent achievements of duration in not whispered voice. There cannot thus be invariant characteristics for the prosodic marks, as opposed to what many of researchers in this field seem to admit.

2.2 A simple example

Fig. 2 gives a (very) simple example involving a sequence of syllables of the French sentence le frère de Max a mangé les tartines. This time, PEs are instantiated by various melodic contours found in many description of French intonation C0, C1, C2 and Cn in the sequence Cn C1 C; C C0 [9].

The advent of the first contour C4 involves the storage of the syllabic sequence appeared since the beginning of the sentence up to the C4 contour. From its position, C4 must be differentiated of all other contours which could appear in its place, and in particular C2, C1 and C0 and its variations. The combinations of prosodic features being relatively limited, C4 will be often realized without noticeable melody movement, by a flat contour simply ensuring enough prominence to the syllable. If it is the case, the listener can expect after C4 the appearance of another C4, C2, C1 as well as C0 announcing the end of the sentence. But if C2 is identified, which must be this time be well differentiated from C4, the listener proceed to the storage of the syllabic sequence appeared since C4, and assemble elements put previously in memory by C4 with the new sequence finished by C2.

Then C1 appears. This time a new syllabic sequence will be concatenated with the one ended by C2. The string of syllables appeared between C1 and C3 is then put in memory, but the listener has to wait for the following contour to know the level of concatenation: if C1 appears, the listener will concatenate the sequence with all the other sequences ended by C1; if C0 appears, the listener will assemble the sequence with the other sequences ended by C0 before assembling the whole of the sequences ended by C1 with the ones ended by C0. In other words, C0 ends to whole process by triggering the concatenation of all remaining syllable strings and cleaning all storages afterwards. The same applies to C1, and to other contours instantiating the PEs.

Each stressed syllable triggers a mechanism of grouping or storing depending on the class of the perceived PE, “strong” PE triggering concatenation and “weak” PE storage. In French “stronger” coding is generally carried out by a contrast of melody slopes, but of other achievements are possible. This
principle was already stated in Martin [9], in Dell [10] specifying the nodes formed well by the sequence w - s (weak-strong), and taken again later by Mertens [11].

3. Some theoretical consequences

3.1 The prosodic word

In many phonological theories, the prosodic word, minimal unit of prosody organized by the prosodic structure (in a limited number of levels in the autosegmental-metrical approach, and in an undefined number in other approaches) does correspond to a lexical word (Noun, Adjective, Verb or Adverb) around which revolve grammatical words such as conjunctions, prepositions, and the like. This unit is frequently called stress group.

Experimental data tend to contradict this statement. First in French, stress groups as defined above carry a stressable syllable which is not necessarily stressed. In our example le frère de Max a mangé les tartines the actual realization of stress on the stressable syllables ending the stress groups [le frère] [de Max] [a mangé] [les tartines] depends on the time distance between the consecutive stressable syllables: a slow rhythm would indeed involve a stress on each of the stressable syllables, whereas a very rapid rhythm would imply stress only on Max and tartines.

Secondly, speakers can put a stress (not an emphatic stress) on each syllable of a lexical word. In an example such as je te le demande po-li-ment, with each syllable of poliment pronounced with a pause, each syllable of poliment can receive a stress. The prosodic word in this case does not correspond to the classical stress group. It should be noted that such emphasis must be done on each syllable of the word, so that they are assembled at the same level in the storage-concatenation process.

3.2 Variations of contours

![Fig. 3. Different prosodic contours encode correspond levels in the prosodic structure](image)

Prosodic contours at the same levels must be phonologically identical; they should not contrast with each other in order to be classified in the same category in the storage-concatenation process. This implies that contours belonging to similar levels in the prosodic structure must not be identical, as each of them has to contrast with only one contour at a time, as shown on Fig. 3. In French, this can be observed in the last prosodic group. Due to the contrast of melodic slope, contours on the last stress groups tend to have a melodic slope opposite to the final contour. In declarative sentences ended by a falling contour they are raising, and in interrogative cases with a final rising contour, they are falling.

4. Compatibility issues

4.1 Sense group

The old concept of sense group used in descriptions of French intonation (e.g. Delattre, [8]) can be revisited in the light of the storage-concatenation process. Indeed, sequences of strings of syllables, although they may contain more than one lexical category, depending on the speech rate [12], do correspond to units of decoding.

4.2 Intonation period

The “période intonative” (Lacheret [7]) corresponds to the prosodic groups at the highest level in the prosodic structure. Indeed, these periods exhibit various realizations on melodic contours inside the stress groups, as explained in 3.2. The storage-concatenation process implies that the prosodic groups at the highest level must carry same class PE. In [7] these are assumed to be instantiated by rising contours, but other realizations are possible, as shown on Fig. 6 and 7.

4.3 Dominance to the right

Likewise, the “dominance to the right” by contours of various slope advocated by Mertens [11] is easily explained by the mechanism of Fig. 3. As long as the prosodic structure is limited to two levels, PEs can vary inside the same group as long as they are recognized as belonging to the same category. As structures of more than 2 levels are relatively rare in spontaneous speech (and in examples found in laboratory phonology…), the apparent free phonetic variations of contours seems valid, and justify in some sense axioms such as the Strict Layer Hypothesis [1]. However, if the structure becomes more complex with more than 2 levels, theses variations are not observed anymore, as contours have now to contrast with PE pertaining to prosodic groups situated lower in the structure [6].

4.4 Eurhythmicity

Prosodic restructuration may appear when there is a strong unbalance between the number of syllables on the left and right of the syntactic structure. In Marie adore les chocolats belges, the prosodic structure may be align with syntactic units [Marie] [adore les chocolats belges], but also be eurhythmic as to tend to equilibrate de number of syllables defined at the first level of the prosodic structure [Marie] [les chocolats belges]. On the other hand, the prosodic reorganization can help the listener in the storage-concatenation process (SCP). As it does not seems easy to predict which association prosody syntax will be used by the speaker, it may be vain to look for a constant correlation between tones and syntactic categories.

5. Spontaneous speech examples

Fig. 4 and Fig. 5 give show different realizations of PE ending stress groups located at the same level in the prosodic structure. In each case, the various groups are assembled in larger groups ("periods") ended by a rising contour, but their contours are flat in the first case (Fig.4) and falling in the second case (Fig. 5). Indeed, these prosodic markers do not have to contrast with contours other than final $C_0$ and rising $C_1$, so they can be realized with any variations of Fo, intensity,
duration as long as they will be identified as belonging to the same class of prosodic events. On the other hand, inside the period, they must be phonetically similar enough so that they can be recognized by the listener as belonging to the same class, but again only inside the larger group which assembles them.

In Fig. 6, the speaker SR consistently uses a falling melodic contour to end prosodic periods, and another falling contour but placed at a higher frequency level on the last syllable of stress groups inside the period.

In Fig. 7, the speaker NS proceeds by periods containing stress groups ended by a falling contour, contrasting with rising-falling contour ending periods. The final declarative contour is also falling, but without any rising part.

6. References


