Neurophysiological research explains prosodic structures constraints

Pesquisa neurofisiológica explica restrições nas estruturas prosódicas

Philippe Martin
UFR Linguistique, Université Paris Diderot

Resumo

Na abordagem cognitivo-funcional (CF), a estrutura prosódica associada a um determinado texto é condicionada por um conjunto de regras que limitam 1) o número máximo de sílabas em cada palavra prosódica (stress group), 2) a presença de duas sílabas tônicas sucessivas (stress clash), 3) o agrupamento no mesmo grupo tonal de palavras dominadas por diferentes ligações na estrutura sintática (syntatic clash). Essa abordagem também prevê a preferência dos falantes em escolher estruturas prosódicas eurrítmicas dentre todas as estruturas prosódicas possíveis. Essas regras explicam adequadamente um número considerável de dados provenientes de discurso lido e de discurso espontâneo. Além disso, pesquisas recentes no campo da eletroencefalografia e da imagética nuclear oferecem informações interessantes sobre os mecanismos da atividade cerebral ligada à percepção da entonação sentencial. Quando confrontados com as propriedades da estrutura prosódia, esses mecanismos conduzem a um conjunto de explicações convincentes, relativas às restrições na estrutura prosódica, que, por conseguinte, têm sua origem na universalidade das características do cérebro humano.

Palavras-Chave

Estrutura prosódica, Choque de acento, Choque sintático, Eurritmicidade, EEG.

Abstract

In the Functional-Cognitive (FC) approach, the prosodic structure associated with a given text is constrained by a set of rules limiting 1) the maximum number of syllables in each prosodic word (stress group); 2) the presence of two successive stressed syllables (stress clash); 3) the grouping in the same stress group of words dominated by distinct nodes in the syntactic structure (syntactic clash). It also predicts the observed preference of speakers to choose eurhythmic prosodic structure among all possible prosodic structures. These rules give a proper account of numerous data made on both read and spontaneous speech.
Besides, recent research in the domain of electroencephalography and nuclear imagery gives interesting details about the mechanisms of brain activity linked to the perception of sentence intonation. When confronted with the prosodic structure properties, these mechanisms lead to a set of convincing explanations pertaining to the prosodic structure constraints, which therefore find their origin in the universality of human brain characteristics.

Keywords

Models for the prosodic structure

For two decades at least, the so-called Autosegmental-Metrical (AM) model has been dominant in intonation phonology. In this model, the prosodic structure organizes hierarchically prosodic events (PE) in three non-recursive levels: a first level assembles syllables σ, content words Wc (verbs, nouns adjectives and adverbs) and function words Wf (conjunctions, pronouns,…) into accentual phrases (AP); a second level groups AP into intonation phrases (IP); finally a phonological utterance (PU) eventually groups sequences of IP.

The prosodic events PE are aligned on accentual phrases' specific syllables and are described as sequences of tones belonging to the ToBI notational system (tones and break indices). This system uses High (H) and Low (L) symbols to transcribe melodic targets as perceived or observed on fundamental frequency curves obtained from the speech signal acoustic analysis.

However, other approaches have been proposed to model sentence intonation. One of the developed models for Romance languages, called Functional-Cognitive (FC), considers the prosodic structure as an a priori independent hierarchical organization of stress groups (equivalent to prosodic words PW) associated with syntax and constrained by a set of specific rules.

These rules pertain to the properties of prosodic words and the way they can form a prosodic structure (PS). More specifically (MARTIN, 2009), they are:

a. Stress clash rule: no PS can have consecutive stressed syllables, unless sufficiently separated in time. *Du vin 'rouge* "some red wine" with a stressed syllable on *'vin* is only possible with a pause between the two stressed syllables *du 'vin # 'rouge*;

b. Syntactic clash rule: no prosodic word syntagm (i.e. a group of stress groups) is allowed if any of its components is dominated by distinct nodes in the syntactic structure. *Le frère de Max aime le café* "Max's brother likes coffee" cannot be prosodically parsed into [*Le frère de*] [*Max aime le*] [*café*] stress groups;

c. <u>Maximum number of syllables rule</u>: a prosodic word cannot contain more than 7 syllables. If it does, as in the word *paraskevidekatriapho'bie* "fear of Friday 13", it would require at least two stressed syllables: *paraske'videkatriapho'bie*;

d. <u>Eurhythmicity rule</u>: among all possible PS that can be associated with a given syntactic structure, speakers will favor the most eurhythmic one (i.e. the one that balances the number of syllables in prosodic words syntagms in equivalent levels in the PS). So the prosodic phrasing [*Marie 'aime*] [*les choco'lats*] "Mary likes chocolates" is preferred to the phrasing congruent to syntax [*Ma'rie*] [*aime les choco'lats*]. Alternatively, this latter realization would involve the insertion of a pause in order to balance the duration of both stress groups [*Ma'rie*] # [*aime les choco'lats*];

e. <u>Planarity rule</u>: all PS are planar; no tangled structures are admitted. This rule derives from the limited capabilities for prosodic markers instantiated by prosodic events (PE) to indicate a dependency relation towards another marker not immediately adjacent to the group considered. Forming a hierarchy with the stress groups A, B and C in the sequence A B C where A and C form a first prosodic syntagm, which then forms a group with B at a later stage is not allowed: *[A[B]C].

The main differences of this approach with the AM model are as follows:

a. Classically in the literature, stress clash can be resolved either by the cancellation of the first stress involved (as in *vin 'rouge*, "red wine") or by shifting the first stress to the left (as in *'café 'noir* "black coffee"). However, in French and in other Romance languages, depending on the syntactic structure, stress clash is possible. In *Max aime le ca'fé 'noir* "Max likes his coffee black" for example, as an answer to the question *Comment Max aime-t-il son café?* "How does Max likes his coffee?" whereas the realization of a stress shift as in *Max aime le 'café 'noir* "Max likes black coffee" would answer the question *Qu'est ce que Max aime boire ?* "What does Max like to drink?";

b. In French, prosodic words do not necessarily contain a unique lexical word (Adjective, Noun, Verb or Adverb). Actually, they can contain just one syllable (as in *po-li-ment* as an emphasis to the word *poliment* "politely" where all syllables are pronounced separately for emphasis, or more than one lexical word (as in *Max aime le ca'fé* "Max likes coffee"), as long as the number of syllables does not exceed 7 syllables;

c. The AM prosodic structure is non-recursive, whereas the FC is. This difference may stem from the fact that very short sentences were used as experimental justifications in AM driven data analysis. In contrast, the FC approach was designed from an extensive set of both read and spontaneous speech data;

d. The AM model uses the ToBI transcription system, which does not take duration parameters into account. The ToBI system has no explicit provision to describe temporal aspects of sentence intonation other that the perceived break durations (which is seldom used);

e. While other transcription systems are either available or could be more or less easily adapted to fit specific properties of a given language, the quasi exclusive use of the ToBI system involves an oversimplification of the description of melodic events. Oversimplification sometimes compensated at a later stage by complex tone alignment rules aimed at better taking the phonetic details of melodic movements into account;

f. Contextual properties of prosodic events are almost always ignored, whereas the FC model uses contextual acoustic features to describe prosodic events correlated to the PS;

g. In early versions of the AM framework, the prosodic structure was assumed to be congruent with the sentence's syntactic structure. This implies that only one prosodic with structure could be associated to given sentence. Even if congruence with syntax is not necessarily retained today as obligatory, it is rare to find an author considering the possibility of associating more than one prosodic structure with a given syntactic structure.

h. As other less known theoretical approaches, AM ignores a basic property of sentence intonation, i.e. to be encoded by prosodic events encoded and decoded sequentially by the speaker and the listener along the time axis. Therefore, it may be misleading to consider prosodic events on a piece of paper as emerging at once to represent the prosodic structure, as they appear in reality in a timely fashion one after the other. This dynamic time domain aspect may modify the way we envision sentence intonation and the prosodic structure;

About this last point, the FC approach envisions the perception of the sequences of syllables by the listener as follows: in the process, the flow of syllables is stored in a short time listener memory that can only accumulate a

limited number of syllables, in the order of 7 +/- 2 (actually this limit depends on the speech rate, as discussed below). See Miller (1956) for short term memory limitations for objects belonging to the same class. To avoid overflow, this sequence of syllables has to be converted into some higher order linguistic unit, the stress group (which rarely correspond to orthographic words). This conversion is triggered by specific processes, involving the presence of a stressed syllable (in final position of the syllabic sequence in French), the direct identification in known patterns in the sequence, and by rhythmic properties identified in the syllabic sequence.

Stress groups parsed from the flow of syllables are not simply concatenated in a flat prosodic structure. Instead, they are hierarchically grouped into prosodic syntagms until the whole structure is formed. This hierarchy is reconstituted by the listener thanks to the differentiations existing between prosodic events, differentiations encoded by the speaker and instantiated mainly (at least in French) by melodic contours located on stressed syllables. The identification of these melodic contours into classes allows the listener to assemble strings of prosodic words belonging to the same level, and to concatenate at various levels the prosodic syntagms formed by this process. This Storage-Concatenation mechanism suggests that the prosodic events acting as markers of the prosodic structure should not be considered globally, but locally, and that their realizations should be analyzed in context relatively to the necessary and sufficient contrasts to be maintained by the speaker.

Delta and Theta EEG waves

The formal constraints cited above limit the number of possible prosodic structures that can be associated with a given syntactic structure, thus with a given text (the AM approach rarely considers more than one prosodic structure for a given text). Several recent neurophysiological studies lead to consider hypotheses giving convincing explanations to the PS rules, making them not only the result of constraints established from detailed data observations, but also rooting them in mechanisms specific to the human brain. These studies essentially use the evoked potentials techniques (EEG, electroencephalography) to establish possible correlations between some brain activity and the processing of speech perception of intonation by listeners. In some instances, magnetic resonance imagery is also used by some authors,

With these techniques, Steinhauer & al. (1999) for example showed that the processing of the prosodic structure by listeners preceded syntactic parsing. Later, Gilbert & Boucher, (2007) and Obrig & al. (2010) demonstrated that segmentation of the syllabic flow into stress groups was realized by two concurrent process, involving prosodic tags and direct pattern matching of sequences already stored in listener memory (the latter process being more time consuming). For adults, these processes lead to a preferred right brain lateralization for prosodic information, and a preferred left brain lateralization for information already stored in memory (lexical access). According to Wartenburger & al. (2007), this hemispheric specialization appears in children above 4 years old in the language acquisition process.

Isel & al. (2005) have also highlighted the differences in processing time between prosodic information and lexical information, access to which obviously requires more time, particularly when the sequence presents syntactic "errors" or is less frequently used.

Following Friederici's (2002) proposals, all these observations lead to the following hypothesis, allowing the drafting of a coherent set of explanations to give an appropriate account of the constraints rules quoted above. We know that cortex waves Delta and Theta, among others, govern the flow of information from neuronal sets to other neuronal sets. The functions of these waves have been particularly described in sleep studies, and their frequencies vary for Delta waves from 1 to 4 Hz, and for Theta waves from 4 to 10 Hz (values vary slightly among authors). If considered in terms of periods rather than frequencies, the variations are 250 ms to 1000 ms for Delta, and 100 ms to 250 ms for Theta waves. The average observed values pertaining to syllabic durations, say 100 ms to 250 ms, and for stress group duration, about 250 ms (including pauses in the case of consecutive stressed syllables) to 1000 ms (for longest stress groups) suggest that 1) Theta waves may synchronize syllabic perception by listeners, and 2) that Delta waves may synchronize the transfer of sequences of syllables into another part of memory storing larger linguistic units.

This interpretation would provide an explanation about the variation of duration of stress groups, from 250 ms to 1000 ms, which corresponds to the Delta period variations, as well as the variations of syllabic duration, from 100 ms to 250 ms also, constrained by mechanical properties of human articulators. The Theta waves would then synchronize the perception of syllables, and the

Delta waves the perception of stress groups, or more precisely the conversion of sequences of syllables stored in short term memory into larger linguistic units (FIG. 1).
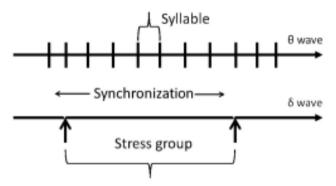

FIGURE 1 - Synchronization of syllabic perception by Theta waves and of stress groups by Delta waves.

Furthermore, this interpretation leads to specific explanations perfaining to the constraints given above limiting the number of possible prosodic structures that can be associated with a given text:

a. Stress clash is indeed allowed, but requires a minimal time interval of about 250 ms between successive stressed syllables. This amount of time corresponds to the minimal period value for Delta waves;

b. The 7 syllables rule defining the maximum number of syllables in any stress group is in fact determined by the maximum value of Delta waves period, i.e. about 1000 ms. The maximum number of syllables is thus defined by the speech rate inside a 1000 ms stress group. In spontaneous speech in French, for example, the analysis of "*parler jeune*", speech style of the young generation in large city suburbs, can reach 12 to 15 syllables;

c. The syntactic clash constraint is explained by the identification time necessary to recognize wrong or unknown syllabic sequences. The violation of this constraint involves a time consuming revision of the initial phrasing realized by the listener on the base of prosodic information, i.e. stress groups stressed syllables. This explains as well why abandoned stress groups are repeated or reformulated with a complete, and not a partial, stress group (MARTIN, 2009);

d. Eurhythmy, i.e. the preference among all possible constrained prosodic structures for balanced number of syllables at every level of the prosodic structure or for modulations of speech rate in order to balance the duration of groups at the same level, is explained by the difficulty in modifying successive periods of synchronization performed by Delta waves with extreme values, varying for example from 250 ms to 1000 ms for the next period.

Conclusion

The Functional-Cognitive approach involves a set of constrains that allow one to define well-formed prosodic structures that can be associated with a given text. When considered in the light of recent neurocognitive research pertaining to the perception of sentence intonation, an explanatory hypothesis can be proposed, linking the prosodic constrains to properties of human brain activity: the syllabic durations can be linked to Theta EEG waves, and the conversion of sequences of syllables into stress groups (prosodic words) can be connected to Delta EEG waves, as the range of period variations of these EEG waves correspond to those of syllables and stress groups, respectively. It can then be assumed that Theta and Delta waves synchronize the perception of syllables and the conversion of strings of syllables into stress groups in the listener's short time memory. This hypothesis leads to a complete explanation of the origin of the prosodic structure constraints and a better comprehension of the mechanisms underlying the association of intonation and syntax in the sentence.

References

FRIEDERICI, A. D. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, v. 6, n. 2, p. 78-84, 2002.

GILBERT, A.; BOUCHER, V. What do listeners attend to in hearing prosodic structures? investigating the human speech-parser using short-term recall. *Proc...* p. 430-433, 2007.

ISEL, F.; ALTER, K.; FRIEDERICI, A. D. Influence of prosodic information on the processing of split particles: ERP evidence from spoken German. *Journal of Cognitive Neuroscience*, v. 17, n. 1, p. 154-167, January 2005.

MARTIN, Ph. *Intonation du français*. Paris: Armand Colin, 2009. 256 p.

MILLER, G. A. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Trends in Cognitive Sciences*, v. 7, n. 3, p. 141-144, 1956.

STEINHAUER, K.; ALTER, K.; FRIEDERICI, A. D. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, v. 2, n. 2, p. 191-196, 1999.

OBRIG, H.; ROSSI, S.; TELKEMEYER, S.; WARTENBURGER, I. From acoustic segmentation to language processing: evidence from optical imaging. *Front Neuroenergics*, 2: 13, 2010.

WARTENBURGER, I.; STEINBRINK, J.; TELKEMEYER, S.; FRIEDRICH, M.; FRIEDERICI, AD.; OBRIG, H. (2007). The processing of prosody: Evidence of interhemispheric specialization at the age of four. *Neuroimage*, v. 34, n. 1, p. 416-425, Jan. 2007.