Brain waves and prosodic structure

Philippe Martin
LLF, UFRL, Université Paris Diderot, France
https://doi.org/10.36505/ExLing-2018/09/0002/000335

Abstract

There is a general agreement among prosodists about the definition of the sentence prosodic structure (PS) as a hierarchical organization of minimal prosodic units called accent phrases (AP). However, disagreements appear among models about the characteristics of the accent phrases, the role of their pitch accents, and the function of the prosodic structure itself in the linguistic system. This paper describes some properties of the Incremental Dependency (ID) model and their relations with brain waves, compared to the dominant Autosegmental-Metrical (AM) model.

Instead of being described as containing only one content word (a noun, an adjective, a verb or an adverb), accent phrases in the Incremental Dependency model are defined by the time taken to pronounce them orally or process them in silent reading. In non-lexically stressed languages such as French or Korean, AP duration varies from 250 ms to about 1250 ms, values corresponding to the range of delta brain oscillations. This suggests that delta waves are involved in the coding and decoding of accent phrases. A eurhythmic process aiming to balance the duration of successive AP determines the actual its duration in the limits of delta wave variations.

Whereas the AM approach does not consider interactions between pitch accents, the ID model describes melodic movements of pitch accents as markers of dependency relations existing between accent phrases, relations which determine the sentence prosodic structure.

Furthermore, while the AM model views the prosodic structure as an emanation of the sentence syntactic structure, the ID model considers the prosodic structure as independent from grammar and actually generated in the speech process chunk by chunk before syntax and lexical selection. The prosodic structure appears then as a syntactic preselection device to process quickly in real time the flow of information, given the limits of short-term memory for speech (about 3 seconds).

Key words: prosodic structure, delta brain waves, autosegmental-metrical, Incremental-Dependency, theta brain waves.

Introduction

The dominant Autosegmental-Metrical model in prosodic phonology envisions the prosodic structure as a hierarchy of accent phrases (AP). Accent phrases are prosodic minimal units containing one single content word and their associate grammatical words. As content words, verbs, adjectives, adverbs and nouns, are normally stressed, accent phrases contain one single stressed syllable (excluding eventual emphatic stress). This hierarchy has been for a long time limited to one level (cf. the Strict Layer Hypothesis, Selkirk, 1978), but more recent studies admit the existence of two levels in the prosodic structure (Michelas, 2011). These levels are respectively called intermediate (intonation) phrase (ip) (lower level) and Intonation Phrase (IP) (highest level). Therefore, in a two-levels prosodic structure, ip groups accent phrases, IP groups ip's and the PS groups IP's.

In this view, pitch accents, located on stressed syllables and more precisely on stressed vowels of accent phrases, do not play any role in the indication of the prosodic structure. This role is devoted to boundary tones, which are located on the last syllable of intermediate phrases and Intonation Phrases.

The relation between the prosodic structure, which defines how accent phrases are grouped together in relation (or not) with syntactic units, can be viewed in the AM approach as an emanation of syntax, i.e. a hierarchy inferred from syntax with some adaptation rules. The PS is therefore more or less congruent with syntax as perfect congruence cannot be usually realized given among other reasons the limitation of the prosodic structure to two levels (except for short sentences), whereas the syntactic structure may need more levels.

In practice, descriptions proceed form a transcription of acoustic (or perceived) data using ToBI notations, in order to establish a set of well-formed sequences of tones levels, infer a prosodic grammar from the set of these sequences and finally determine adaptation rules that would predict well-formed tone sequences from a given sentence and its syntactic and possibly semantic structures.

In the AM view, the prosodic structure just as the syntactic structure is static, meaning that no temporal effect is involved in the definition of its description. Indeed, the general use of ToBI notation eludes the transcription of any duration factor of accent phrases, ip or IP, except for the perceived pauses noted (rarely) on a perception scale from 1 to 5.

The Incremental Dependency model (IP), developed since 1975 (Martin, 1975), takes almost all possible opposite views on the prosodic characteristics found in the AM approach. The only point of agreement pertains to the definition of the prosodic structure itself as a hierarchy of accent phrases (called earlier prosodic words), whereas the main differences are as follows:

Accent phrases are not defined by the category of words they contain but by the time it takes to pronounce them, either orally or silently.

The prosodic structure is not generated from syntax through some deep structure, but the other way around where the PS precedes the encoding and decoding of the syntactic structure.

Pitch accents located on accent phrases stressed syllables are not without interaction with each other, on the contrary they are markers of dependency relations existing between accent phrases.

The dependency relations indicated by pitch accents are generated and decoded dynamically along the time scale and not generated from some deep structure.

The prosodic structure is recursive and not constrained by the Strict Layer Hypothesis.

All of these characteristics find an explanation in the brain wave features involved in speech production and perception. In particular, it has been shown the theta and delta brain waves timing properties do correspond closely respectively to the syllables (Ghitza, 2011) and to the accent phrases ranges of duration (Martin, 2018). The generation of the sentence prosodic structure itself, which cannot be avoided in either spontaneous or read speech whether produced orally or silently, finds its origin if the limited capability of our short-term memory dealing with speech.

Accent phrases

In lexically-stressed languages such as English or Italian, accent phrases usually include only one content word and therefore only one stressed syllable, as nouns, verbs, adverbs and adjectives are normally stressed at some morphological boundary, their eventual associated grammatical words being unstressed (excluding emphatic stress). In non-lexically stressed languages such as French or Korean, the position of stressed syllables is governed by rhythmic constrains whose effects are hidden in lexically-stressed languages by morphological stress rules. These rhythmic constrains are easy to observe in the pronunciation of long words (ex. *paraske**vi**dékatriapho**bie*** in French) which requires at least one extra stressed syllable on top of the final default position of the rhythmic stress. On the other hand, sequences involving one syllable word (ex. *par le **fait** **que***) require a gap of at least 250 ms to keep the final syllable of the preceding AP perceived as stressed.

These observations lead to confer to accent phrases a minimal duration of 250 ms (including the gap preceding a single syllable accent phrase), and a maximal duration of about 1250 ms. If thanks to speech synthesis manipulation, two consecutive stressed syllables are separated by more than 1250 ms, listener will tend to perceive an extra intermediate stressed syllable, even if not showing any particular acoustic characteristic of stress, such as a longer duration of a perceived pitch movement.

This means that the actual content of a given accent phrase will depend on the speech rate, the stressed syllable being always placed in final position in non-lexically and non-tonal stressed languages. While an average speaking rate corresponds to about 4 to 5 syllables per second, resulting to accent phrases containing a maximum of 6 to 7 syllables, very slow speech rate will result in only one syllable in each AP (a detached pronunciation syllable by syllable), and a very fast speech tempo will pack up to 10 or 11 syllables in a single AP (cf. *parole de jeunes* in French).

Between the 250 ms – 1250 ms extreme values, a eurhythmy principle applies in order to reduce the variation of duration between consecutive accent phrases (Wioland, 1985). As the number of syllables of every successive word cannot be changed, speakers have to choose to either adapt the speech rate for every accent phrase to achieve a relative eurhythmicity, speaking faster for accent phrase with many syllables and slower for those with few syllables, or selecting consecutive words to obtain a similar number of syllables in consecutive accent phrases, possibly at the expense of congruence with syntax. The first strategy is preferred for spontaneous speech, the second for read speech.

Prosody before syntax

The precedence of prosodic structure generation over syntax can be shown in many ways (Keating and Shattuck-Hufnage, 2002 , Martin 2018), but one of the most convincing argument comes from the reading process, where this property may seem at first the least probable.

Reading operates by saccades, where the eye focalizes on successive words by jumps which do not exceed some 10 to 20 characters (Reyner et al., 2010). This means that a reader who never read a given text before can only adapt its (partial) prosodic structure to a few consecutive words, eventually helped by some punctuation marks. In configurations such as (A) (B C) where the number of syllables of the words B and C would exceed some 10 to 20 characters, the reader could not normally anticipate the syntactic relation existing between B and C, and would instead group A and B. Discovering C at the next reading saccade, the only solution would then to put prosodically C at the same level than B, ending the sequence AB. The prosodic hierarchy will then be [A B] [C], not congruent with the syntactic (A) (B C). In the French example *deux alpinistes allemands ont trouvé le cadavre d'un homme dans un glacier* "two

German mountaineers found the corpse of a man in a glacier", readers will group prosodically *ont trouvé* with *le cadavre* and then add *d'un homme* at the next step of the prosodic structure elaboration.

Accent Phrases dependency relations

When we start a sentence, either orally or silently "in our head", we have already decided about its modality, declarative or interrogative, and their variants, imperative or implicative for the declarative, and surprise or doubt for the interrogative. The terminal conclusive melodic contour on the last pitch accent correlated with this modality is therefore planned in the future of the process, and will mark also the end of the sentence (except for deferred complements or theme-rheme constructions not discussed here). Therefore, the melodic realization as non-final of the other pitch accents that precede depend on the final prosodic event planned by the speaker in the future instantiating a dependency relation with another accent phrase in the future. This can be generalized to all non-terminal pitch accents, which may define a dependency relation with some other pitch accent located later in the prosodic structure.

In French, following the terminology of Delattre (1966) but with a somewhat different interpretation, at least five phonological categories of melodic contours located on pitch accents can be considered: the *major continuation*, which indicates a dependency relation towards the *terminal conclusive contour declarative* or *interrogative*; the *minor continuation*, which indicates a dependency relation towards the major continuation; the *neutralized* contour, which indicates a dependency towards either the minor continuation, the major continuation or the terminal conclusive contour.

The symbols attached to these categories are respectively C0 ↓terminal conclusive declarative, Ci ↑terminal conclusive interrogative, C1 ↗ major continuation, C2 ↘ minor continuation, Cn → neutralized contour. The falling and rising arrows for C2 and C1 implement the contrast of melodic slope used in French to indicate the dependency "to the right" between accent phrases, where a falling contour C2 is depending of a rising contour C1 located somewhere later in the sentence, and the rising contour C1 a dependency towards the final conclusive declarative C0. C1 and C2 melodic variation are above the glissando threshold (Rossi, 1971), which means their melodic change is effectively perceived as such by listeners, contrary to the neutralized contour Cn, and possibly C0 which usually reaches the lowest frequency level in the sentence.

Accent phrases local congruence with syntax

In non-lexically non-tonal stressed languages, accent phrases contain one or more content words with their syntactically associated grammatical words. These micro syntactic constructions constitute the elementary building blocks stored as such in speakers and listeners long-term memory. In the speech decoding process, listeners retrieve these stored elements already partially syntactically organized, and assemble them according to the hierarchy defined by the incremental prosodic structure. Therefore, referring to the example given above, accent phrases containing sequences such as **allemands ont* "Germans have" will be excluded as unlikely be part of the listener lexicon.

Given this local congruence property, the function of the prosodic structure in the linguistic system appears more clearly: to allow the speaker to quickly assemble hierarchically partial syntactic structures in the very short allowed time window before the speech sound vanishes in the short-term memory.

Prosodic structure phonetics and phonology

The prosodic annotation in the Autosegmental-Metrical framework is usually done with the ToBI notation system, basically transcribing high and low melodic targets (and not contours) with H and L symbols. Numerous variants allow to represent complex melodic shapes, as well as their position relative to pitch accents (i.e. H* or L*) or syntactic boundaries (i.e. %H or %L "left", H% or L% "right".

Besides the often-questionable validity of actual transcription made from sometimes elusive fundamental frequency curves, this annotation process will basically mix phonological and phonetic data, to be sorted at a later stage. Unfortunately, this is not always easy to do in practice, as the specific function of the prosodic structure being inferred from syntax in the AM approach is not clearly defined.

By contrast, the incremental dependency model poses the prosodic structure as a frame for micro syntactic patterns, which is actually generated before the overall syntactic structure in the speech planning process operated by the speaker, in spontaneous as well as in read speech, oral or silent. From this view of the prosodic structure function and the assumed dependency system of relations between accent phrases, it is relatively easy to discover features relative to the necessary contrast between categories of contours and to sort them from the phonetic details such as those relative to regional or idiosyncratic variations.

Brain waves

Brain waves are generated by constant electrical exchanges of the order of microvolts that take place between neuronal regions through long chains of synapses. Brain waves are usually classified by their frequency range. In the speech domain, brain oscillations of interest are delta, oscillating in the 0.8 Hz – 4 Hz range, theta 4 Hz – 10 Hz and gamma 30 Hz - 80 Hz (these approximate values may vary depending on the subject). Beta brain waves, which are endogenous (i.e. not trigger directly by an external stimulus), oscillates between 14 Hz and 21 Hz are also involved in speech processing.

It has been showed recently that theta waves are directly related to the perception of syllables, delta to the processing of accent phrases (Martin, 2018), and gamma to the perception of higher frequency acoustic consonants characteristics, such as fricatives. Syllables need a minimum of 100 ms duration to be processed, although the perception of their acoustic features themselves can take much less time (Ghitza, 2011). On the other hand, syllables separated by gaps of more than 250 ms will be perceived as stressed.

As mentioned above, accent phrases have a minimal duration of 250 ms (one stressed syllable preceded by the necessary separation from an eventual previous stressed syllable), and a maximal duration of 1250 ms (if longer, an extra stressed syllable will be perceived in the interval).

All these observations can be explained by looking at the temporal dynamics of theta and delta brain oscillations. Their frequency range converted into periods, 4 Hz – 10 Hz as 100 ms – 250 ms for theta, and 0.8 Hz – 4 Hz as 250 ms – 1250 ms for delta, correspond respectively to the range of syllabic duration and to the inter stressed syllable duration, closely linked to the accent phrases duration. As the delta and theta oscillations are phase locked, i.e. they synchronize each other, the mechanism of syllable and stressed syllable can be interpreted as follows.

In the absence of sound and in particular of speech sound, the delta and theta waves are idle, the oscillate freely inside their duration range while being constantly phased locked. When the first stressed or unstressed syllable of a speech sequence appears, both the theta and delta spikes align on this new temporal event (Gilbert and Boucher, 2007). The upcoming syllables then synchronize the next theta oscillations until a stressed syllable appears, characterized by some perceivable acoustic difference with the preceding syllables, such as a longer duration, a difference in intensity, a marked pitch change. This event triggers a delta spike, which is in phase with the theta spike triggered by the syllable. The eurhythmy effect would prevent variations of successive delta periods to be too large.

Although in lexically-stressed languages the stressed syllable is not necessarily in final position, it has been suggested (Martin, 2018) that the delta spikes initialize a retrieving process of stored accent phrases, containing a micro syntactic structure in the listener long-term memory. This seems also to be the case and trigger a delta spike even in silent speech where stressed syllables are present with no external acoustic stimulus (Magrassi et al., 2015).

This bottom-up mechanism is concurrent with a top-down process where from a first syllabic segmentation the listener can retrieve from long-term memory the content of the incoming segmented accent phrase as well as its stressed syllable position. The switch between bottom-up and top-down processes appears to be instantiated by beta oscillations (Pefkou and al., 2017).

References

Delattre, P. 1966. Les dix intonations de base du français, *French Review* 40, 1-14.

Ghitza, O. 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm, Frontiers in Psychol. 2, 130.

Gilbert, A., Boucher, V. 2007. What do listeners attend to in hearing prosodic structures? Investigating the human speech-parser using short-term recall, Proc. Interspeech 2007, 430-433.

Keating, P., Shattuck-Hufnage, S. 2002. A Prosodic View of Word Form Encoding for Speech Production, UCLA Working Papers in Phonetics 101, 112-156.

Magrassi, L, Aromataris, G., Cabrini, A., Annovazzi-Lodi, V., Moro, A. 2015.

Sound representation in higher language areas during language generation, PNAS 112 (6) 1868-1873.

Martin, Ph. 1975. Analyse phonologique de la phrase française. Linguistics 146, 35-68.

Martin, Ph. 2018. Intonation, structure prosodique et ondes cérébrales. London: ISTE.

Michelas, A. 2011. Caractérisation phonétique et phonologique du syntagme intermédiaire en français de la production à la perception. Thèse de doctorat, université de Provence.

Pefkou, M., Arnal, L.H., Fontolan. L., Giraud, A.-L. 2017. Theta- and beta-band neural activity reflect independent syllable tracking and comprehension of time-compressed speech. Journal of Neuroscience 16, 37 (33), 7930-7938.

Reyner, K., Slattery, T.J., Béanger, N.N. 2010. Eye movements, the perceptual span, and reading speed. Psychonomic Bulletin 17(6), 834-839.

Rossi, M. 1971. Le seuil de glissando ou seuil de perception des variations tonales pour la parole. Phonetica 23, 1-33.

Selkirk, E.O. 1978. On prosodic structure and its relation with syntactic structure. In Fretheim, T. (ed.) 1978, Nordic Prosody, 111-140. Trondheim: TAPIR.

Wioland, F. 1985. Les structures rythmiques du français. Paris: Slaktine-Champion.