Automatic detection of accent phrases in French

Philippe Martin
LLF, UFRL, Université de Paris
https://doi.org/10.36505/ExLing-2020/11/0030/000445

Abstract

In lexically-stressed languages such as English or Greek, accent phrases usually include one lexical word (noun, verb, adverb or adjective), together with some syntactically bound grammatical words (conjunction, pronoun or preposition). In non-lexically languages such as French or Korean, accent phrases are delimited by a final syllabic stress and may contain more than one lexical word, depending on the speech rate and limited to a 250 ms to 1250-1350 ms duration range. As perception of syllabic stress is strongly influenced by the listeners current own speech rate making perception agreement between annotators elusive, an interactive software program has been implemented imbedding constrains external to acoustic data to better investigate the actual distribution of stressed syllables in oral recordings of French.

Keywords: accent phrase, French, speech rate, WinPitch

Introduction

When we speak or read a text, we process speech not word by word, but accent phrase by accent phrase. Each accent phrase may include a single word, a group of words, or even a single syllable, but does contain only one stressed syllable (excluding emphatic stress). In languages such as English, Italian or Greek, this unique stressed syllable is located on a so-called lexical or content word, i.e. a noun, a verb, an adverb or an adjective, and its position is defined by the lexicon. Pronouns, conjunctions and articles belong to the category of grammatical words and are not normally stressed. An accent phrase therefore consists of one content word together with syntactically bound grammatical words.

The segmentation into accent phrases is the first step to recover the prosodic structure intended by the speaker (or the writer) in order to bootstraps the discovery of the syntactic structure. Even when we read a text silently ("in our head"), we do generate mentally a prosodic structure organizing the accent phrases in a hierarchy to access the meaning of the text, although in this case any acoustic input is of course absent.

French, however, does not have lexical stress but a group stress located on the last syllable of accent phrases. Rather than systematically constitute *groupes de sens*, accent phrases ended by syllabic stress are constrained by external rhythmic constrains, among which:

a. The minimal duration between consecutive stressed syllables, about 250 ms (Martin 2014).
b. The maximum duration between consecutive stressed syllables in continuous speech, about 1250-1350 ms (Martin 2014).
c. The eurhythmic tendence to balance the duration of successive accent phrases (Wioland 1983).
d. The perception of any word final syllable followed by at least 250 ms of silence as stressed (Martin 2018).

The only acoustic condition for a syllable to be identified as stressed pertains to the change of its vowel fundamental frequency perceived as a change in frequency rather than a static tone, i.e. above the glissando threshold (Rossi 1971).

The simple fact that we can restore stress locations when we read aloud or silently suggests that we may not really need any acoustical input to perceive stressed syllables when we listen to speech (again non-emphatic). The predictability of lexical stress leads to consider the perception of stressed syllables as a process which compares the actual acoustic features of syllables with a position predicted from the listener knowledge of accent phrases. Furthermore, eurhythmicity makes the perception of stressed syllables and the segmentation into accent phrases sensitive to listeners speaking or reading rate.

Automatic annotation of stressed syllables

The annotation of stressed syllable in speech corpora analysis is an essential step especially for macrosyntactic description of spontaneous speech based on accent phrase chunks. The perception of stress will be influenced by the annotator own prediction process, and stressed syllables located where they would have placed by reading or speaking at the annotator own pace. The problem for an annotator is therefore to adapt to the speech rate of the recording when stressed syllables are annotated (Martin 2020).

Automatic detection of stressed syllables in French operates usually in a bottom-up fashion from the speech acoustic data, looking for significant variations between consecutive syllables in duration, fundamental frequency and intensity. Vowel quality does not appear as a significant parameter for stress detection in French.

With the advent of relatively large corpora of both read and spontaneous speech, more systematic experimental studies were carried out, only the highlight the confusion in the domain. A paper from Avanzi (2013) for instance, faced with the uncertainty in annotating stressed syllables in French, describes in detail a complex procedure involving two experts. Even with this protocol, agreement between annotators varies between 60 % and 80 %.

Later, Christodoulides and Avanzi (2014) implemented an automatic detector of prominence (i.e. not just accent phrases stressed syllable) by

machine learning methods applied to a large corpus which included two different styles. Although they use a comprehensive set of acoustic parameters, their best results, evaluated against manual placement by experts in syllabic prominence, reaches a 90% correct identification level.

Considering these difficulties, it appears that stress detection should proceed not only from speech wave acoustic analysis, but also from the implementation of the rhythmic constrains evoked above.

Interactive computer annotation of stressed syllables

Rather than proceed by annotating stressed syllables by experts, or develop another algorithm proceeding from the speech signal, an interactive software program has been implemented in the WinPitch package. The idea was to automatically locate stressed syllables from the rules given above, each rule being adjustable by the user in order to evaluate their effect. It follows that missing or superfluous cases are quickly detected visually, thanks to a specific color code.

The implemented analysis proceeds as follows:

- Loading of a recording to analyze, with orthographic transcribed sections (each section of about 1 to 5 seconds).
- Automatic segmentation into vowels and consonants based on forced alignment with a text-to-speech segment generated from the annotated text.
- Automatic scanning and detection of stressed syllables (actually stressed vowels) based on the criteria enounced above:
  a. Any syllable followed by more than 250 ms silence is stressed
  b. Any final syllable of a noun, adjective, verb or adverb is stressable (from original accent phrase definition)
  c. If 2 consecutive stressed syllables are separated by less than 250 ms, the first one is unstressed (accent phrase minimum duration)
  d. Any stressable syllable with F0 change over the glissando threshold is stressed
  e. If 2 consecutive stressed syllables are separated by more than 1250 ms in continuous speech, at least one stressable syllable in this interval is stressed (accent phrase maximum duration). Make stressed the one with the highest glissando value
  f. One stressable syllable must exist in any time window duration equals to the accent phrase average duration (eurhythmy)

All parameters, glissando threshold, minimum and maximum accent phrase syllable… are user adjustable. The eurhythmic aspect is implemented by evaluating the first accent phrases realizations and the number of syllables they contain. This starting accent phrase duration will them be used to define a

sliding time window, in which most prominent syllables with a glissando value above the threshold are retained as stressed. The size of this sliding window defines a speech rate assumed to be constant for the whole recording.

Conclusions

Rather than being simply another stressed syllable detection algorithm, the automatic accent phrase detection process implemented in the speech analysis software WinPitch appears as an exploration tool allowing to quickly and efficiently test the rhythmic and acoustic parameters of stress in French. Both large read and spontaneous corpora have been processed, establishing the validity of these parameters, and more important, highlighting a perception process of accent phrases stressed syllables operating in both bottom-up and top-down mode.

References

Avanzi, M. 2013. « Note de recherche sur l'accentuation et le phrasé à la lumière des corpus du français », Tranel, vol. 58, 5-24.

Christodoulides, G., Avanzi, M. 2014. An Evaluation of Machine Learning Methods for Prominence Detection in French, Proc. Interspeech 2014, 116-119.

Martin, Ph. 2014. Spontaneous speech corpus data validates prosodic constraints, Proc. Speech Prosody 2014, 525-529.

Martin, Ph. 2018. Intonation, structure prosodique et ondes cérébrales, London: ISTE, 322 p.

Martin, Ph. 2020. L'annotation prosodique dans Orfeo, Langages 2020/3 (N° 219), 103-115.

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

Wioland, F. 1983. La rythmique du français parlé, Strasbourg : Institut international d'études françaises.

WinPitch. 1995-2020. Speech analysis software, www.winpitch.com.