ANALOR. A Tool for Semi-Automatic Annotation of French Prosodic Structure
Mathieu Avanzi, Anne Lacheret-Dujour, Bernard Victorri

To cite this version:

HAL Id: hal-00334656
https://hal.archives-ouvertes.fr/hal-00334656
Submitted on 27 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
ANALOR
A Tool for Semi-Automatic Annotation of French Prosodic Structure

Mathieu Avanzi1; Anne Lacheret-Dujour2; Bernard Victorri3

1U. Neuchâtel, Switzerland & U. Paris X Nanterre, France; 2U. Paris X Nanterre & IUF, France; 3Lattice, ENS, Paris, France

Abstract
In the area of large speech corpora, there is a definite need for common prosodic notation system based on efficient (semi-)automating tools of prosodic segmentation and labelling. In this context, we present the software program ANALOR, developed in order to process semi-automatically prosodic data. From a text-sound alignment, this computer tool detects major prosodic units, on the basis of global and local melodic variations. That leads to the segmentation of an utterance in prosodic periods. Inside those prosodic periods, prominent syllables are then automatically detected.

1. Introduction
Linguistics and speech technology have dealt with prosody from various points of view, which make a precise definition of the scope of research on prosody difficult. Nevertheless, a complete analysis is very useful as part of a linguistic analysis in order to determine the optimum number of functional prosodic units and to determine their nature according to precise acoustic correlates. In this context, most of the existing transcription systems, whether they engender a phonologic interpretation, like ToBI system [2], or not (see for example IVIE [18] or IVTS [9] systems), necessarily share the point of view that prominence processing represents the cornerstone of the prosodic annotation. Actually, most of prosodic annotation systems are based on manual processing. This situation remains problematic for at least two reasons.

First, it is a well-known fact that manual prosodic annotation varies greatly from an expert to another. See for example [16]’s experiment on the SEC. The authors report that the disagreement between the two experts who annotated the same sub-part of the whole corpus (nearly 4190 syllables) was of about 27%. Regarding the assignment of tonetic stress marks, [5]’s calculations reveal that consensus was of about 55% at best. See also, for an example on spontaneous French, [17]. The author asked seven prosodic experts to annotate prominent syllables in a small stretch corpus (165 syllables). The proportion of syllables marked as prominent varied from 19% to 49%.

Others studies, like [4], showed that better results could be obtained if the annotators followed a strict protocol (set of symbols reduced, common training, etc.). Thus, they respectively got a very satisfactory inter-transcribers agreement on a 45-minute long spoken Dutch corpus. However, such a manual procedure is extremely time-consuming. The authors concluded that a non-expert annotator would need about 40 times the duration of the corpus to annotate minimal prosodic phenomena as strong and weak breaks, segmental lengthening and prominent syllables.

As a consequence, automating the procedure of prosodic annotation in spoken corpora is of great importance. In this paper, we present what the two different steps that punctuate the procedure of prosodic annotation proposed consist of. The division in major prosodic units (called prosodic periods) and the methodology which leads to the development of this first algorithm is described in section 2. Then, we present the algorithm used for the detection of prominent syllables (section 3). We finally conclude with the specificity of our software, compared with other quite similar tools (section 4).

2. Segmentation in major prosodic units (prosodic periods)
The concept of prosodic period stems from an inductive approach, which rests on the comings and goings between a manual observation of the data and a computer modelling. This method, introduced by [11] and [12], is structured around three fundamental steps.

2.1. Methodology
We first conducted a manual analysis on a small corpus of French radio talk (about two hours, with male and female speakers, see [12] for details). The goal was to isolate a set of phonetic cues (silent pauses of a certain minimal duration, major contour of specific amplitude) associated to what is commonly perceived of as a strong prosodic break.

The computer modeling, which comes after this first stage, rests on the processing of local and global pitch variations in a given time interval. The implementation of an automatic segmentation mechanism was thought up in order to systematically test the principles issued from manual processing.

Comparing the manual analysis to the automatic data processing permits the highlighting of differences between the two treatments. In fact, two tracks were then studied to report these differences: (i) questioned by the manual decision-making and/or (ii) redefining the criteria used for the automatic processing. In the first case, the computer modelling intervenes as a control mechanism, i.e. it allows for adjusting the intuition of the experimenter by pointing to incoherencies in the analysis, thus rationalizing the initial intuitions. In the second, the observation leads to refine the criteria used for the automatic data processing. It is after all these comings and goings between manual observation and automatic data processing that we developed a stable algorithm of automatic segmentation of the statements that we decided to call prosodic periods.
The automatic algorithm relies on basic relative acoustic parameters (see also [19] for a review of automatic modeling of the internal periodic structure. In order to appreciate it, we invoke the notion of prominence. In practice, accentual labelling does not rely on a structural feature of the word or word group such as lexical stress, but on a neutral phonetic definition of prominence, as a perceptual salience of a word or word group such as lexical stress, but on a neutral accentual labelling does not rely on a structural feature of the word or word group such as lexical stress, but on a neutral phonetic definition of prominence, as a perceptual salience within background speech ([13], [14], [18], [19] and [20]). The main advantage of this approach is to be independent of any theoretical framework, and of any morphologic or syntactic considerations.

3.1. Algorithm

The automatic algorithm relies on basic relative acoustic parameters (see also [19] for a review of automatic modeling of the internal periodic structure. In order to appreciate it, we invoke the notion of prominence. In practice, accentual labelling does not rely on a structural feature of the word or word group such as lexical stress, but on a neutral phonetic definition of prominence, as a perceptual salience within background speech ([13], [14], [18], [19] and [20]). The main advantage of this approach is to be independent of any theoretical framework, and of any morphologic or syntactic considerations.

3.1. Algorithm

The automatic algorithm relies on basic relative acoustic parameters (see also [19] for a review of automatic modeling of the internal periodic structure. In order to appreciate it, we invoke the notion of prominence. In practice, accentual labelling does not rely on a structural feature of the word or word group such as lexical stress, but on a neutral phonetic definition of prominence, as a perceptual salience within background speech ([13], [14], [18], [19] and [20]). The main advantage of this approach is to be independent of any theoretical framework, and of any morphologic or syntactic considerations.
prominence detection). Today, we exploit the **pitch parameters** (pitch range, melodic movement) and **duration** of the syllabic segments composing a given period. Those parameters are calculated relatively, that is to say a syllable will be identified as prominent by the software if it stands out from its environment according to a certain threshold.

For the moment, the formula implemented is this one: Let $M_0(\alpha)$ be the average and $E_0(\alpha)$ the distance-type of fundamental frequency on a given period $P$. A syllable is said prominent for height (and it is labelled “H”, cf. figure 1) if it contains a local maximum pitch, marked $h(s)$, verifying the condition

$$h(s) > M_0(\alpha) + K_E E_0(\alpha)$$

where $K_E$ is an adjustable parameter (called “threshold height distance”: its value is 1.5, by default).

In others words, the algorithm is based on a Gaussian distribution of F0 centred on the median axis (average of all the points of F0 for the given period), from which we calculate a standard deviation which allows the researchers to quantify the distribution of points around the average of F0, and beyond which we can detect a salient acoustic event. With a variable threshold, the interest of this kind of arithmetic is that it is strong concerning the inter-variability of different speakers. As a consequence, it does not matter much if the speakers modulate a lot or a little: the software will detect an event whatever happens to the significant variations of F0.

The formula is the same for duration.

**3.2. Evaluation phase**

Because we thought that these two parameters were not sufficient to track down all prominent syllables (like boundary tones which does not manifest a significant pitch or duration variation, see for example the last syllable of the period in figure 2), we decided to use **silent pauses** to hone the detection. Consequently, each syllable followed by a silent pause, whatever the pause duration, will be considered as prominent.

**3.3. Visualization**

The user can consult the results of the automatic detection of the prominences in the bottom window (cf. figure 2), and thus compare those results with the manual coding (labelled prominent syllables ‘P’ or ‘P’ in the median tier, top window). If necessary (for example if the results are clashing), it is possible to correct the manual note or change the thresholds of the automatic detection. The software thus allows making permanent comings and goings between the model and the phonetic analysis, driven by the empirical data: like this we can, at the same time, check the coherence of the analyses and improve the adequacy of the model to the analyzed data.

All the analyses and the results obtained with ANALOR can then be repatriated in TextGrid files (Praat format).
4. Discussion and conclusion

In this article, we presented a software program for prosodic analysis which constitutes a very useful tool assistant to facilitate the prosodic annotation of spontaneous spoken French corpora. A first formula leads to an automatic division of the prosodic continuum in major prosodic units, or prosodic periods, by basing itself on the interaction of melodic and temporal parameters. Within the identified segments, a second operation proceeds in a prominent syllables detection, according to a certain threshold of pitch and duration change.

The specificity of the elaboration of our tool is summed up in two points: (i) it is about an emergent approach of a bottom-up type, (ii) even though the approach is driven by strong hypotheses, it is on no account forced by a predetermined theoretical frame (it does not rest on a phonological categorization of the detected accents, as ToBI system [2] for example). Besides, contrary to other tools of the same type, ANALOR does not establish its measures on a stylised signal (like MOMEL [10] for example). Moreover, it is not reserved for the treatment of read sentences, stemming from laboratories in which they were recorded, contrary to the plug-in WaveSurfer developed by [19]. Finally, we would like to emphasize that ANALOR offers dynamic results, not static ones such as the plots allowed by ProsoProm [8], derived from [15]’s Prosogram.

In future works, we will develop a procedure for automatic detection of segmental lengthening resulting from a hesitance. For the moment, it must be pre-identified manually. The robustness of the tool on corpora of more varied genres also remains to be checked, and its performances must be compared to the competing tools mentioned supra.

ANALOR can be downloaded from:
http://www.lattice.cnrs.fr/Analor.html.
Sources are in free access.

5. Acknowledgments

This publication was supported by the FNS, subsidy n°100012-113726/1, “La structure interne des périod es”, hosted in Neuchâtel University, Switzerland.

6. References