Automatiser l'analyse prosodique des corpus oraux

Flora Badin, Emmanuel Schang, François Nemo, Camille Leroux

Automatiser l’analyse prosodique des corpus oraux

Flora Badin; Emmanuel Schang; François Nemo; Camille Leroux

LLL UMR7270

1. Introduction

Le travail sémantique sur la prosodie est confronté au problème de l’hétérogénéité des données authentiques issues de conditions de recueil diverses. Mais partir de données produites dans des conditions contrôlées pose à l’inverse le triple problème de ne pas permettre de disposer de la diversité effective des formes de réalisation prosodique des séquences étudiées, de conduire le cas échéant à des artefacts prosodiques (en produisant des séquences qui ne seraient en réalité jamais réalisées) ou encore d’associer sans le savoir aux séquences des réalisations de fait très spécifiques. Une solution à ce problème consiste à dépasser le problème de la diversité des situations et conditions d’enregistrement et de la variabilité des paramètres que celles-ci induisent en recourant à des techniques de classification automatique opérant sur une masse importante de données (Gharbi & alii, 2015). Disposer d’une masse de données suffisantes, en particulier quand il s’agit de tester parallèlement un ensemble d’hypothèses, ne peut se faire qu’à condition d’automatiser l’extraction des segments à partir du corpus d’étude puis d’automatiser leur caractérisation prosodique. Nous présenterons donc la chaîne de traitement mise au point pour le traitement automatique du corpus ANCOR-Centre (Muzerelle & alii 2014), qui constitue la plus importante ressource annotée en relations anaphoriques pour le français oral. En effet, bien que l’ensemble des travaux portant sur la résolution des phénomènes anaphoriques et de coréférence font référence à la notion de saillance (Landragin (2004) par exemple), celle-ci n’est principalement abordée que d’un point de vue cognitif (v. Huang (2000), Walker & alii (1998), Gundel (2010) notamment), tout en postulant l’existence d’une saillance acoustique visible par la prosodie. Nous proposons ici une méthode de recherche des traits prosodiques révélant (éventuellement) cette saillance acoustique dans le corpus ANCOR. La question que nous posons est donc : peut-on améliorer les scores des classifieurs pour la résolution anaphorique par l’ajout de traits prosodiques identifiables dans le signal ?

2. Extracteur, caractérisateur prosodique et création de banque d’emplois : une application en devenir

La multiplicité des formats, la conservation de l’alignement parole/son et l’interopérabilité entre les logiciels font parties des problématiques de notre application. L’avantage et l’inconvénient de notre étude réside dans le fait que nous travaillons avec un ensemble volumineux de données, ce qui nous permet de pouvoir faire abstraction de certaines données sonores inexploitables par les logiciels mais aussi ce qui nous pousse à trouver des solutions informatiques pour le traitement en masse de ces données.

Nos données sources (format .xml) sont alignées au son (format .wav) en tour de paroles à l'aide du logiciel Transcriber (Boudahmane, Manta, Antoine, Galliano & Barras 1998-2008). Elles seront amenées à être converties selon le traitement sélectionné dans différents formats dont .textgrid, format issu du logiciel PRAAT (Boersmal & Weenink, 2016) voir en simple texte.

La chaine de traitement a été pensée de la façon suivante :

- Découpage en tour de parole de la donnée sonore (ffmpeg ou shnsplit) et récupération du texte associé (xslt).
- Alignement de la donnée sonore au mot via le logiciel Jtrans (Cerisara, Mella & Fohr, 2009) disponible en open source sur la plateforme GitHub. Un avantage à son utilisation est l'absence de langage de script, ce qui permet une adaptation presque immédiate du composant.
- Détection de la prosodie des séquences sonores avec le logiciel SLAM (Obin, Beliao, Veaux, & Lacheret, 2014), logiciel de stylisation et labellisation automatique de la prosodie qui se lance en ligne de commande sous distribution LINUX uniquement et disponible sur la plateforme GitHub également. L'utilisation de ce logiciel est d'autant plus intéressante qu'il se base sur des données chiffrées et non sur une perception de locuteur pour déterminer la prosodie d'un segment. Ceci nous permet alors de pouvoir comparer sur un même niveau tout un corpus d'emploi.

Les résultats obtenus (csv) permettent la consultation d'une banque de données contenant les informations liées à la prosodie de la majorité des mots d'un corpus d'entrée.

3. Application au corpus ANCOR.

Le projet ANCOR a été réalisé en grande partie à partir des enregistrements du corpus ESLO. Il contient des fichiers d'annotations au format XML, produits par le logiciel de transcription GLOZZ (Widlöcher & Mathet 2014). Ce corpus de 450 000 mots environ (27, 5 heures d'enregistrement) nous a permis de tester notre application.

A partir des segments annotés dans la transcription, le segment sonore correspondant a été repéré dans le fichier son. A la suite de quoi, ces segments de signal ont été annotés automatiquement en traits prosodiques pour chaque mot du segment. Ces traits prosodiques provenant du logiciel SLAM utilisés sont représentés par la variation de F0 sur le signal. Cette variation est calculée selon la première position de F0, la dernière et le point de saillance de F0. Elle est représentée sur 5 niveaux de hauteur d'extrêmement basse à extrêmement haute.

Un ou plusieurs traits prosodiques ayant ainsi été assignés à chaque segment anaphorique de notre corpus, il est devenu possible de rechercher des patterns prosodiques corrélés à certains types d'entités anaphoriques et d'évaluer leur poids (s'il en est un) dans la perspective de la résolution automatisée de la coréférence.

4. Perspectives.

La chaîne de traitement décrite vise donc à mettre en lien des segments de parole avec des annotations sémantiques riches. Elle est libre de droits et disponible pour des projets similaires qui visent à mettre en correspondance des annotations sémantiques et de la parole spontanée.

Cette application vise à être utilisée par la suite sur différents corpus de langues (somali, créole guadeloupéen, anglais, français, …)

Mots-clés : discrimination prosodique, classification automatique, chaîne de traitement, contours prosodiques

Références bibliographiques

Corpus Ancor : http://tln.li.univ-tours.fr/Tln_Ancor.html

Corpus Eslo : http://eslo.huma-num.fr/

Application Prosodorus : https://gitlab.com/trugliadophe/SEMORAL

Elordieta, G & Prieto, P. (eds.) (2012). *Prosody and Meaning*. Berlin: De Gruyter Mouton.

Gundel, Jeanette K. (2010). Reference and accessibility from a Givenness Hierarchy perspective. ***Intern****ational Review of Pragmatics*, 2(2):148–168.

Landragin, F. (2004). Dialogue homme-machine multimodal. Modélisation cognitive de la référence aux objets. *Hermès-Lavoisier, Paris*.

Muzerelle, J. Lefeuvre, A. Schang, E. Antoine, J.-Y., Pelletier, A. Maurel, D., Eshkol, I. Villaneau, J (2014). ANCOR_Centre, a Large Free Spoken French Coreference Corpus: Description of the Resource and Reliability Measures. *LREC2014 Proceedings*, Reykjavik.

Petit, M,. Nemo, F & Létang, C. (2016). “Prosodic constraints on pragmatic interpretation: a new chapter in linguistic pragmatics”. *Lodz Papers in Pragmatics*. Berlin : Mouton de Gruyter. Volume 12, Issue 1. 53-64. (DOI 10.1515/lpp-2016-0005)