
Analyse acoustique du corpus CFPB

Philippe Martin^{*1}

¹Université Paris-Cité – Université Paris Cité, Université Paris Cité, Université Paris-Cité – France

Résumé

À la suite de C-Oral-Rom (1), plusieurs corpus oraux de français spontané, tels que CFFP2000 (2) pour le français parisien, CFPB français parlé à Bruxelles (3) ou OFROM (4) parlé en Suisse romande, ont été élaborés ces dernières années, et partiellement rassemblés dans Orfeo (5). La plupart des études linguistiques de ces corpus ont porté sur leur caractéristiques syntaxiques et sémantiques, telles qu'observées à partir de leur transcription orthographique. Les aspects spécifiques à l'oral sont rarement abordés, face aux difficultés techniques de l'analyse, dues à des facteurs tels que enregistrements bruités, superposition de voix, écho, niveau sonore trop faible, etc., rendant l'exploitation des données difficile avec les outils habituellement utilisés (e.g. : Praat (6))

Une version récente du logiciel WinPitch (7) s'adresse à ces difficultés, par un ensemble de fonctions dédiées à l'analyse linguistique des corpus oraux, comme par exemple : extraction automatique des "scories" caractéristiques de la parole spontanée (répétitions, abandons, reformulations...) par comparaison avec la transcription manuelle de la transcription automatique grammaticalement corrigée fournie par le logiciel intégré Whisper; annotation automatique ou par commandes graphiques des courbes mélodiques et des variations de formants même dans des cas d'enregistrements dégradés ; affichage automatique des structures prosodiques à partir des catégories phonologiques des contours mélodiques des voyelles accentuées définies par l'utilisateur.

L'utilisation de cet outil sera illustrée par l'analyse de quelques exemples relatifs à certains traits connus ou moins connus du français régional parlé à Bruxelles (8), tels que les réalisations du —r—, la substitution de () par (w), la diphtongaison des voyelles accentuées entraînant des variations des contours mélodiques différentes de celles du français parisien, des réalisations idiomatiques (par ex. —six— en position accentuée finale de groupe prononcé (siz)). On montrera que la fréquence de réalisation de ces variantes est liée aux caractéristiques sociogéographiques des locuteurs.

Références

(1) C-ORAL-ROM French, Italian, Spanish, European Portuguese,

<http://www.elda.org/en/proj/coral/fr/coralrom.html>

(2) CFFP2000 <https://www.ortolang.fr/market/corpora/cfpp2000>

(3) CFPB https://www3.usaintlouis.be/4DACTION/rechw_detail_projet/13/2332/F

*Intervenant

(4) OFROM <https://ofrom.unine.ch/>

(5) Orfeo, Le Corpus d'Etude du Français Contemporain - Le projet ORFEO www.projet-orfeo.fr/corpus/lecorpus-d-etude-du-francais-contemporain/14-orfeo

(6) Praat <https://www.fon.hum.uva.nl/praat/>

(7) WinPitch <https://www.winpitch.com>

(8) Hugo Baetens Beardsmore, 1971. *Le français régional de Bruxelles*, Bruxelles : Presses Universitaires de Bruxelles.

Mots-Clés: Français régional, Bruxelles, phonétique, intonation