Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition),
Traitement Automatique des Langues Naturelles (TALN, 27e édition),
Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)
Nancy, France, 8 au 19 juin 2020

Variation prosodique des styles de parole et interface syntaxe-prosodie: Étude sur corpus à grande échelle

George Christodoulides[1]
(1) Service de Métrologie et des Sciences du Langage, Université de Mons,
18 Place du Parc, 7000 Mons, Belgique
george@mycontent.gr

RÉSUMÉ

La mutualisation et diffusion des grands corpus de parole permet de réexaminer des analyses précédentes effectuées sur des corpus plus petits, afin de vérifier si les conclusions de ces analyses se généralisent aux nouvelles données. Dans cette étude, nous présentons les résultats préliminaires d'une analyse de la variation des styles de parole en français, basée sur un corpus à grande échelle (300 heures, 2500 locuteurs). Le corpus a été réaligné au niveau des phones, syllabes et mots, et une annotation morphosyntaxique et syntaxique a été ajouté en améliorant les annotations existantes. Plusieurs caractéristiques acoustiques et prosodiques sont automatiquement extraites et une analyse statistique (analyse en composantes principales, ACP) est effectuée afin d'explorer les caractéristiques des styles de parole et leur variance. Nous explorons aussi la relation entre frontières prosodique et syntaxiques comme méthode pour discriminer les styles de parole. [1]

ABSTRACT

Speaking Style Prosodic Variation and the Prosody-Syntax Interface : A Large-Scale Corpus Study

As large spoken language corpora become available, we can revisit previous analyses based on smaller datasets and verify whether the conclusions generalise to the new data. We present an analysis of speaking style variation in French, based on a large-scale corpus (300 hours, 2500 speakers). The corpus was segmented at the phonetic, syllabic and word level. Automated annotation in parts-of-speech and syntactic dependencies was performed, enhancing existing annotations, and a multitude of acoustic and prosodic features were automatically extracted. Statistical analysis (principal component analysis, PCA) is performed to explore the characteristics of speaking styles and their variance. We finally explore congruency and mismatch between prosodic and syntactic boundaries, as a method to discriminate speaking styles.

MOTS-CLÉS : style de parole, variation prosodique, classification et regroupement (clustering) de styles de parole, interface prosodie-syntaxe, linguistique de corpus.

KEYWORDS: speaking style, prosodic variation, classification and clustering of speaking styles, prosody-syntax interface, corpus linguistics.

1. Cet article est une traduction en français de la communication Christodoulides, G. (2020) Speaking Style Prosodic Variation and the Prosody-Syntax Interface : A Large-Scale Corpus Study, *10th International Conference on Speech Prosody*, 24-28 May 2020, Tokyo, Japan.

1 Introduction

La variation situationnelle, c'est-à-dire la variation linguistique liée aux différentes situations de communication, est importante pour l'étude du langage et de la parole. L'étude systématique de cette variation, à la croisée de la sociolinguistique et de la phonétique/phonologie, est du domaine de la sociophonétique. Dans ce contexte, la pertinence de la dichotomie entre « parole de laboratoire » et « parole spontanée » a été remise en question, étant insuffisante pour décrire la grande variation observée dans des corpus qui recensent plusieurs situations (Wagner *et al.*, 2015). Les caractéristiques de la parole sont déterminées à la fois par le contexte situationnel et les différences individuelles (Llisterri, 1992; Eskenazi, 1993; Léon, 1993). L'apport du contexte situationnel à un style de parole (« phonostyle ») est mieux décrit en utilisant plusieurs dimensions (cf. le modèle proposé dans (Koch & Österreicher, 2012)) plutôt que des distinctions binaires (p.ex. « discours formel » vs « discours informel »).

La variation de caractéristiques prosodiques selon les styles de parole ont été étudiés en analysant des corpus (Goldman *et al.*, 2014; Beliao *et al.*, 2013), ou parfois en vue d'une application spécifique, comme la classification automatique des échantillons sonores (Veiga *et al.*, 2012). La mutualisation et la diffusion de plus grands corpus de parole nous amène à revoir les analyses précédentes qui étaient basées sur des corpus plus petits, afin de vérifier si les résultats se généralisent, ainsi que pour effectuer des analyses plus fines. Cette contribution présente une étude sur un corpus de français parlé à large échelle (voir section 2.1) et sa méthodologie s'inspire de (Goldman *et al.*, 2014). Nous présentons également une analyse de la relation entre segmentation prosodique et syntaxique, et de sa variation selon les styles de parole, sur la base du même corpus.

2 Méthodologie

2.1 Corpus

La présente étude a été réalisée sur le sous-corpus de parole du *Corpus d'Étude pour le Français Contemporain* (CEFC) (Benzitoun *et al.*, 2016). Il s'agit d'une collection de corpus du français parlé provenant de diverses sources (Branca-Rosoff *et al.*, 2012; Cresti *et al.*, 2005; Baldauf-Quilliatre *et al.*, 2016; Avanzi *et al.*, 2016; Carruthers, 2008; Équipe Delic, 2004) qui ont été homogénéisés, transcrits, annotés automatiquement en parties du discours et en syntaxe des dépendances, et ont été approximativement alignés au niveau des tokens. La composition du corpus est présentée dans le tableau 1. Le corpus contient 900 échantillons, a une durée de 300 heures et contient plus de 2500 locuteurs. Sa taille est d'environ 3,7 millions tokens, ce qui après la phonétisation correspond à environ 4,6 millions de syllabes.

Une grande variété de situations de communication sont représentées dans le corpus CEFC, du fait qu'il est composé de plusieurs sources. Afin d'organiser les échantillons, les métadonnées du corpus CEFC proposent quatre dimensions principales : secteur (public ou privé), type/genre (fournissant une description large de l'activité ou de la situation communicative), milieu (p.ex. : amical, familial, affaires, politique, universitaire, etc.) et modalité (discours en public, face à face, radio, télévision, téléphone). Une distinction est également faite entre les monologues, les dialogues et les conversations multipartites. Nous avons combiné l'attribut secteur avec l'attribut type/genre (regroupant certaines situations de communication qui sont sous-représentées), afin d'arriver à une

catégorisation des styles de parole, qui est présentée dans le tableau 1. Dans cet article, nous présentons nos résultats principalement à travers ce regroupement en *genres* / *styles de parole* ; cependant, des analyses supplémentaires peuvent être effectuées sur les *sous-genres* plus détaillés, ou selon une autre dimension.

Style de parole	Nb	Dur	Loc	Tok	Syll
Privé					
Activité	10	2,3	14	21,4	27,4
Conversation	174	65,5	488	902,5	1075,2
Repas	12	8,0	39	102,0	120,7
Entretien	351	137,2	829	1672,3	2076,4
Narration	37	8,3	43	89,4	111,7
Public					
Activité	13	8,1	126	62,7	77,9
Conversation	81	6,2	209	67,6	86,5
Entretien	9	3,2	31	42,1	52,6
Didactique	16	4,3	63	35,3	49,2
Média	31	10,4	271	117,2	163,0
Réunion	50	28,1	319	348,6	443,6
Narration	86	15,8	96	148,2	188,3
Discours public	30	5,9	66	58,9	84,3
Total	900	303	2594	3668,4	4556,9

TABLE 1: Composition du corpus CEFC : nombre d'échantillons, durée (heures), nombre de locuteurs, syllabes (en milles) et tokens (en milles).

2.2 Traitement de données

Dans un premier temps et afin d'améliorer la performance de l'alignement automatique texte-parole, une procédure de restauration et d'amélioration a été appliquée à tous les échantillons audio du corpus, en utilisant le logicieel *iZotope RX 6 Audio Editor*. Les filtres suivants ont été appliqués en séquence : de-clip (restaurer les échantillons écrêtés à haute qualité), de-click (supprimer les clics aléatoires), de-hum (supprimer le bruit parasite et ses harmoniques), de-noise (réduction adaptative du bruit), equaliser (en mode dialogue) et leveller (normalisation des niveaux audio, respectant la dynamique du dialogue). Les métadonnées du corpus (fichiers XML encodés en TEI) et les annotations (fichiers CoNLL-U avec les informations sur les locuteurs et l'alignement approximatif de chaque token) ont tous été importés dans une base de données SQL à l'aide du logiciel de gestion de corpus *Praaline* (Christodoulides, 2014) pour le traitement ultérieur.

Une transcription phonétique avec des variantes de prononciation a été produite à partir de la transcription orthographique, et a été alignée au niveau du phone, de la syllabe, du token et de l'énoncé, en utilisant le système d'alignement forcé de *Praaline*, qui pour le français, utilise le système de reconnaissance vocale *Kaldi* (Povey *et al.*, 2011) et un lexique de prononciation basé sur *GLÀFF* (Hathout *et al.*, 2014).

Le corpus a été ré-annoté en utilisant *DisMo* (Christodoulides & Barreca, 2017) qui fournit une annotation morphosyntaxique détaillée, une détection des unités polylexicales ainsi qu'une annotation automatique des disfluences. Ces annotations ont été combinées avec l'annotation en syntaxe de dépendance originale. Le corpus aligné a été analysé à l'aide de *ProsoGram* (Mertens, 2004), qui détecte le noyau de chaque syllabe en fonction de la fréquence fondamentale (F0) et l'intensité ; la courbe F0 est ensuite stylisée selon un système perceptif, produisant une annotation en segments

tonales. Nous avons ensuite appliqué le plug-in *Promise* afin d'effectuer une détection automatique des syllabes proéminentes (Christodoulides & Avanzi, 2014) et une détection automatique des frontières prosodiques majeures et mineures (Christodoulides, 2018) ; les algorithmes statistiques de *Promise* ont été entraînés sur des corpus du français annotés manuellement, comme détaillé dans les références de l'outil. Une segmentation automatique en unités prosodiques (phrases intonatives et groupes accentués) a été effectuée sur la base de ces annotations, et est corrélée à l'annotation syntaxique (voir section 3.3).

Nous avons finalement appliqué les outils d'analyse statistique de *Praaline* (plug-ins d'analyse temporelle, profil prosodique et extraction d'unités) afin d'extraire plusieurs caractéristiques (mesures) acoustiques et prosodiques pour chaque échantillon du corpus et pour la participation de chaque locuteur à chaque échantillon. Ces mesures peuvent être regroupées comme suit : mesures temporelles (par exemple durée de pauses silencieuses et pleines, débit de parole) ; mesures de dynamique conversationnelle (longueur des tours de parole, pauses interlocuteurs et chevauchements) ; mesures prosodiques (ex. registre et mouvements dynamiques) ; mesures de proéminence ; et les mesures de segmentation (unités prosodiques, unités syntaxiques et leur corrélation). La base de données SQL de *Praaline* a été liée au logiciel *R* (R Core Team, 2017) pour l'analyse statistique.

3 Résultats

Dans ce qui suit, nous présentons des résultats statistiquement significatifs, selon la classification en *styles de parole*, pour certains groupes de mesures.

3.1 Mesures temporelles et prosodiques

Le taux d'articulation (pourcentage du temps d'enregistrement pendant lequel un locuteur articule) par situation (style de parole) est présenté dans la figure 1. Le style "narration privée" (narration spontanée d'expériences personnelles), ainsi que le style "narration prof" (contes de fées récités par des professionnels) ont un taux plus faible, indiquant que les pauses silencieuses sont plus nombreuses et/ou plus longues. Une observation similaire peut être faite pour le style "prof-leçon" (conférences académiques et leçons scolaires), ainsi que le style "prof-public speech" (où les pauses sont principalement utilisées pour l'effet rhétorique).

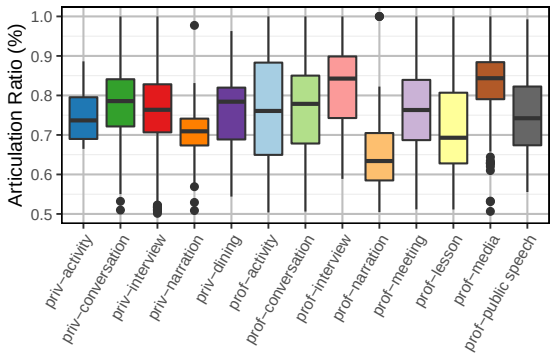

FIGURE 1 – Taux d'articulation (%) par situation.

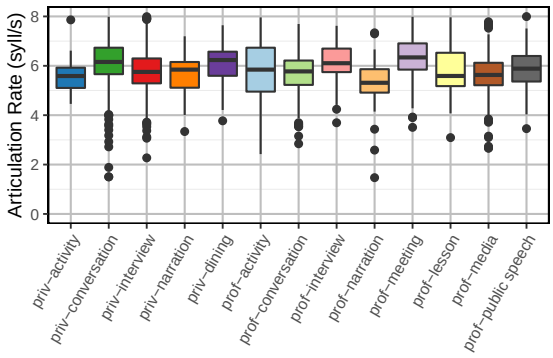

FIGURE 2 – Débit d'articulation (syll/s) par situation.

Nous observons également que la narration professionnelle a un débit d'articulation (syllabes par seconde) plus faible, et que la variabilité est plus élevée dans les conversations spontanées, tandis que les différences entre les autres styles de parole restent mineures. Le nombre de pauses pleines (normalisé au nombre de tokens) est présenté dans la figure 3 et est un descripteur prosodique qui discrimine les styles de parole avec un faible degré de planification, ou spontanés (p.ex. conversation, que ce soit dans un cadre privé ou professionnel, entretiens). Cependant, nous observons que les hésitations ainsi que les autres types de disfluences sont présentes dans tous les styles de parole (à l'exception de la narration professionnelle).

La figure 4 montre la distribution des trajectoires mélodiques (mouvements de F0 stylisée en demi-tons par seconde) par modalité de communication. Nous observons que les deux activités médiatiques (radio et télévision) ont des trajectoires mélodiques plus importantes, ce qui peut s'expliquer par l'adoption d'un style plus expressif par ces locuteurs professionnels.

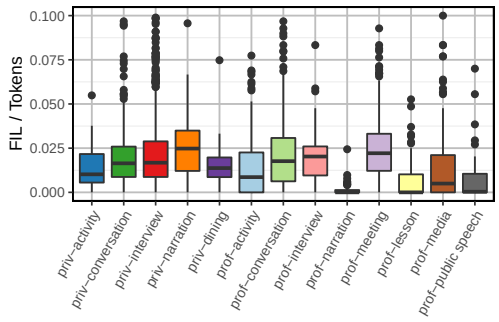

FIGURE 3 – Pauses pleines (normalisé au nombre de tokens) par situation.

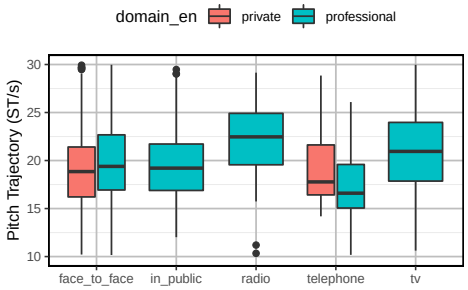

FIGURE 4 – Trajectoire mélodique (ST/s) par médium.

3.2 Dynamique conversationnelle

En ce qui concerne la dynamique conversationnelle, la figure 5 montre la durée moyenne (en secondes) des tours de parole.

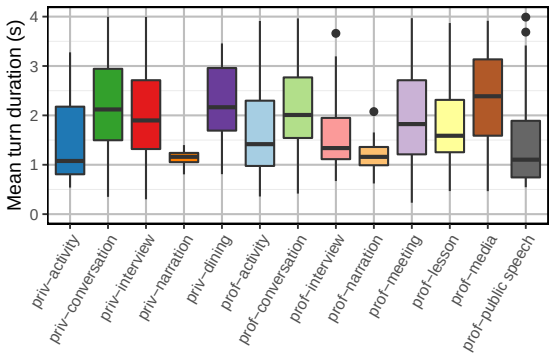

FIGURE 5 – Durée de tour de parole moyenne (s) par situation.

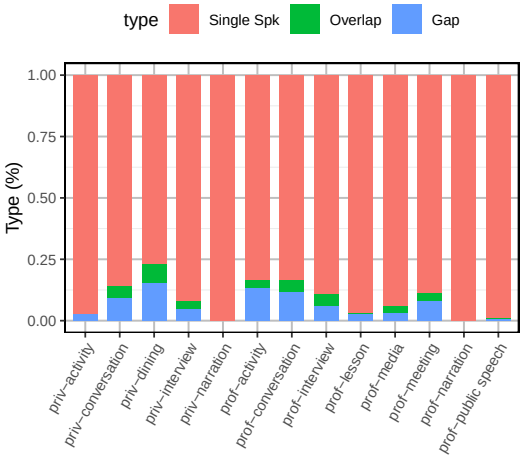

FIGURE 6 – Dynamique conversationnelle par situation.

Les différences entre les situations de communication plus ou moins interactives peuvent ainsi être observées, surtout si la durée du tour est combinée avec le pourcentage de chevauchements et de gaps (pauses interlocuteurs, entre deux tours de parole appartenant à des locuteurs différents). Cette distribution par situation est illustrée dans la figure 6.

3.3 Interface prosodie-syntaxe

Le corpus CEFC est annoté en syntaxe de dépendance en utilisant un nombre réduit de relations de dépendance, proposé dans le cadre du projet ORFEO. L'annotation syntaxique de la parole présente des difficultés et des choix doivent être faits concernant le traitement des faux départs (phrases inachevées), des parenthèses et des disfluences. Pour cette raison, dans la présente étude, nous avons utilisé l'annotation fournie et limité les analyses à la relation entre les unités syntaxiques majeures (c'est-à-dire les unités de dépendance complètes) et les unités prosodiques majeures. L'annotation en dépendance définit les unités syntaxiques (SU) et leur taille en syllabes, selon la situation de communication, est illustrée à la figure 7.

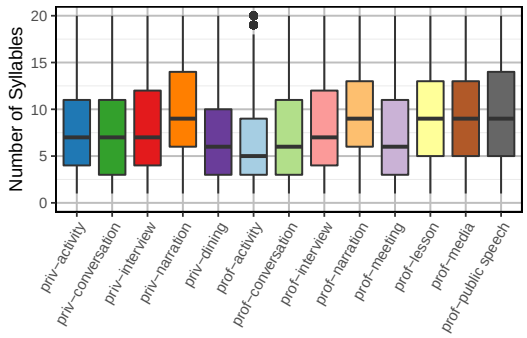

FIGURE 7 – Nombre de syllabes par unité syntaxique, par situation.

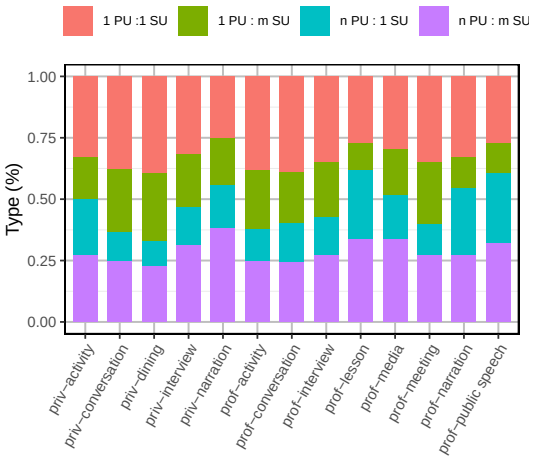

FIGURE 8 – Relation entre unités syntaxiques et prosodiques, par situation.

Sur la base de la détection automatique des frontières prosodiques et des syllabes proéminentes, nous avons annoté le corpus en unités prosodiques (PU) majeures (groupes intonatifs) et mineures (groupes accentués). Il existe des congruences et des décalages entre ces unités, créant quatre configurations possibles : un PU correspond à un SU (1 :1), un PU regroupe plusieurs SU (1 :m), plusieurs PU sont regroupés dans un SU (n :1), et une série de décalages entre les frontières prosodiques et syntaxiques qui conduisent à plusieurs PU correspondant à plusieurs SU (n :m). Cette méthode d'analyse est similaire à celles précédemment présentées par (Beliao *et al.*, 2013) et (Martin *et al.*, 2014) pour des corpus multi-genres du français parlé, de taille plus petite. Les résultats de l'analyse de la relation entre les unités syntaxiques et prosodiques (congruence/décalage), selon la situation, sont présentés dans la figure 8. Les situations de communication sont caractérisées par des différences dans la planification de la parole, et celles-ci affectent la longueur des structures syntaxiques, ainsi que le nombre et la durée des pauses silencieuses (qui sont le principal corrélat acoustique des frontières prosodiques majeures) ; on peut donc utiliser ces mesures de congruence / décalage entre les unités syntaxiques et prosodiques majeures afin de différencier (certains) styles de parole.

3.4 Analyse en composantes principales

Nous avons constaté qu'aucun paramètre prosodique unique n'est suffisant pour différencier les styles de parole. Étant donné que plusieurs mesures sont fortement corrélées entre elles, nous avons procédé en appliquant une analyse en composantes principales (ACP), qui réduit l'ensemble de mesures à un petit ensemble de *composantes principales* non corrélés linéairement (chaque composante principale est une combinaison linéaire des mesures initiales). Les résultats de l'ACP indiquent que les 2 premières composantes principales (PC) expliquent 25,1 % de la variance, les 4 premiers PC expliquent 44,6 % de la variance et 8 PC expliquent 71,0 % de la variance. On peut comparer ces résultats à ceux de (Goldman *et al.*, 2014) (avec 9 styles de parole et 105 échantillons), où les 2 premiers PC expliquaient seulement 43 % de la variance et les 8 premiers expliquaient 78,2 %.

Dans la figure 9, chaque point représente la participation d'un locuteur dans un échantillon de corpus, codé par couleur selon la situation / le style de parole, et est tracé sur un plan défini par les 2 premières composantes principales (PC1 sur l'axe x et PC2 sur l'axe y). Les ellipses de confiance indiquent la variabilité de chaque style de parole sont présentées dans la figure 10 : on peut ainsi observer que certains styles de parole sont très homogènes (par exemple les présentations médiatiques), tandis que les conversations et les entretiens sociolinguistiques ont une variabilité plus élevée.

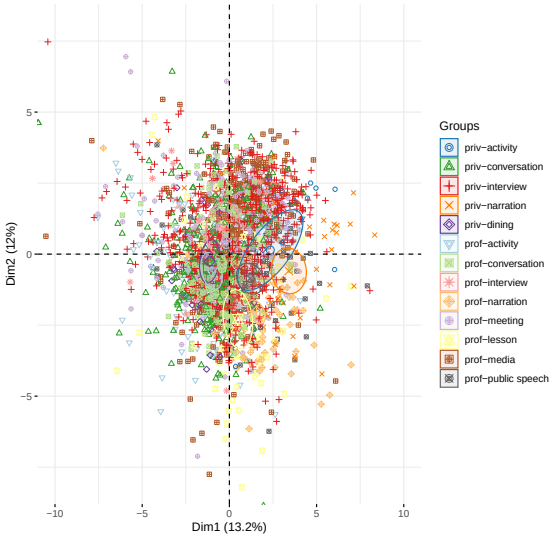

FIGURE 9 – Deux premières composantes principales, tous les échantillons.

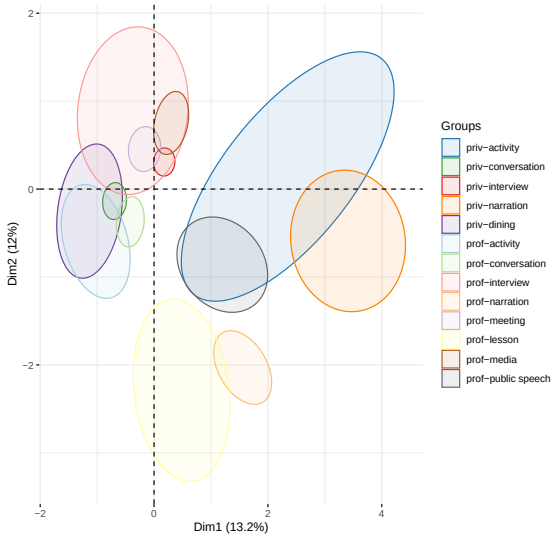

FIGURE 10 – Deux premières CP et ellipses de confiance pour chaque situation / style de parole.

4 Conclusion

Dans cet article, nous avons présenté une analyse de la variation prosodique des styles de parole dans un corpus à grande échelle (300 heures, plus de 2500 locuteurs) du français parlé. Alors que les tendances générales des études antérieures sur les corpus plus petits sont largement confirmées, l'analyse de ces «big data» indique qu'une approche beaucoup plus nuancée de la variation du style de parole est nécessaire : certaines situations de communication ont des contraintes très spécifiques (par exemple, le discours professionnel dans les médias, ou le discours politique en public, mais aussi narrations spontanées d'expériences personnelles) qui influencent plusieurs paramètres prosodiques

à la fois et sont ainsi plus faciles à distinguer et à classer. Cependant, la variation individuelle des conversations spontanées, voire des entretiens sociolinguistiques, est plus élevée. Ce résultat de notre étude sur grand corpus renforce les appels à encore plus de diversité dans la recherche sur la parole : un effort doit être fait non seulement pour étudier les phénomènes phonétiques et prosodiques à travers les différents styles de parole, mais aussi pour intégrer la variation individuelle dans les analyses. De plus, dans des études futures, nous envisageons une analyse plus détaillée de l'interface prosodie-syntaxe (entre autres, en incluant plusieurs niveaux d'unités syntaxiques et prosodiques) ; une exploration de méthodes alternatives pour décrire les styles de parole (un autre regroupement au niveau des métadonnées) ; et étendre les mesures extraites (en particulier au niveau segmental). Enfin, notre travail a enrichi et amélioré l'annotation d'un grand corpus du français parlé : le corpus est disponible sous la licence Creative Commons, et notre travail est rendu disponible sous les mêmes conditions.

Références

AVANZI M., BÉGUELIN M.-J. & DIÉMOZ F. (2016). De l'archive de parole au corpus de référence. Le corpus oral de français de Suisse romande (OFROM). *Corpus*, **15**, 309–342. Actes du colloque Corpus de Français Parlés et Français Parlés des Corpus.

BALDAUF-QUILLIATRE H., COLON DE CARVAJAL I., ETIENNE C., JOUIN-CHARDON E., TESTON-BONNARD S. & TRAVERSO V. (2016). CLAPI, une base de données multimodale pour la parole en interaction : apports et dilemmes. *Corpus*, **15**, 165–194. Actes du colloque Corpus de Français Parlés et Français Parlés des Corpus.

BELIAO J., KAHANE S. & LACHERET A. (2013). Modéliser l'interface intonosyntaxique. In *Prosody-Discourse Interface Conference 2013, Proceedings*. DOI : 10.13140/2.1.1701.1205.

BENZITOUN C., DEBAISIEUX J.-M. & DEULOFEU H.-J. (2016). Le projet ORFÉO : un corpus d'étude pour le français contemporain. *Corpus*, **15**, 91–114. Actes du colloque Corpus de Français Parlés et Français Parlés des Corpus.

BRANCA-ROSOFF S., FLEURY S., LEFEUVRE F. & PIRES M. (2012). *Discours sur la ville. Présentation du Corpus de Français Parlé Parisien des années 2000 (CFPP2000).*

CARRUTHERS J. (2008). Annotating an oral corpus using the Text Encoding Initiative : Methodology, problems, solutions. *Journal of French Language Studies*, **18**(1), 103–119.

CHRISTODOULIDES G. (2014). Praaline : integrating tools for speech corpus research. In *LREC 2014 – 9th International Conference on Language Resources and Evaluation, May 26–31, Reykjavik, Iceland, Proceedings*, p. 31–34.

CHRISTODOULIDES G. (2018). Acoustic correlates of prosodic boundaries in french : A review of corpus data. *Revista de Estudos da Linguagem, Belo Horizonte*, **26**(4), 1531–1549. aop13597.2018.

CHRISTODOULIDES G. & AVANZI M. (2014). An evaluation of machine learning methods for prominence detection in french. In *Interspeech 2014 – 15th Annual Conference of the International Speech Communication Association, September 14–18, Singapore, Proceedings*, p. 116–119.

CHRISTODOULIDES G. & BARRECA G. (2017). Expériences sur l'analyse morphosyntaxique des corpus oraux avec l'annotateur multi-niveaux DisMo. *Corela : Cognition, Représentation, Langage*, **HS-21**. journals.openedition.org/corela/4867.

CRESTI E., BACELAR DO NASCIMENTO F., MORENO SANDOVAL A., VERONIS J., MARTIN P. & KALID C. (2005). *The C-ORAL-ROM Corpus. A Multilingual Resource of Spontaneous Speech for Romance Languages*. John Benjamins Publishing Company.

ÉQUIPE DELIC (2004). *Autour du Corpus de référence du français parlé*. Publications de l'université de Provence. Recherches sur le français parlé No 18, 265 pp.

ESKENAZI M. (1993). Trends in speaking styles research. In *Proceedings of Eurospeech*, p. 501–509.

GOLDMAN J.-P., PRŠIR T., CHRISTODOULIDES G. & AUCHLIN A. (2014). Speaking style prosodic variation : an 8-hour 9-style corpus study. In *7th International Conference on Speech Prosody, May 20–23, Dublin, Ireland, Proceedings*, p. 105–109.

HATHOUT N., SAJOUS F. & CALDERONE B. (2014). GLÀFF, a Large Versatile French Lexicon. In *LREC 2014 – 9th International Conference on Language Resources and Evaluation, May 26–31, Reykjavik, Iceland, Proceedings*.

KOCH P. & ÖSTERREICHER W. (2012). Language of immediacy – language of distance : Orality and literarcy from perspective of language theory and linguistic history. In C. LANGE, B. WEBER & G. WOLF, Éds., *Communicative spaces : Variation, contact, and change*, p. 441–473. Frankfurt : Peter Lang.

LLISTERRI J. (1992). Speaking styles in speech research. In *ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language, Dublin, Ireland, 15–17 July 1992*, p. 15–17.

LÉON P. (1993). *Précis de phonostylistique, Parole et expressivité*. Paris : Nathan Université.

MARTIN L., DEGAND L. & SIMON A. (2014). Forme et fonction de la périphérie gauche dans un corpus oral multigenres annoté. *Corpus*, **13**, 243—-265.

MERTENS P. (2004). The Prosogram : Semi-automatic transcription of prosody based on a tonal perception model. In *Proc. of Speech Prosody 2004, March 23–26, Nara, Japan*, p. 549–552.

POVEY D., GHOSHAL A., BOULIANNE G., BURGET L., GLEMBEK O., GOEL N., HANNEMANN M., MOTLICEK P., QIAN Y., SCHWARZ P., SILOVSKY J., STEMMER G. & VESELY K. (2011). The Kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding* : IEEE Signal Processing Society. IEEE Catalog No. : CFP11SRW-USB.

R CORE TEAM (2017). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

VEIGA A., CELORICO D., PROENÇA J., CANDEIAS S. & PERDIGÃO F. (2012). Prosodic and phonetic features for speaking styles classification and detection. In D. TORRE TOLEDANO, Éd., *Advances in Speech and Language Technologies for Iberian Languages. Communications in Computer and Information Science*, volume 328, p. 15–17. Berlin, Heidelberg : Springer.

WAGNER P., TROUVAIN J. & ZIMMERER F. (2015). In defense of stylistic diversity in speech research. *Journal of Phonetics*, **48**, 1–12. DOI : 10.1016/j.wocn.2014.11.001.