

Variations temporelles des phonèmes en parole conversationnelle : propriétés phonétiques et facteurs lexicaux

Temporal variations of phonemes in conversational speech:
phonetic properties and lexical factors

Christine Meunier¹
Brigitte Bigi²

Abstract: This study aims to identify certain phonetic and lexical determinants of major durational variations in spontaneous speech. Three duration samples (SHORT, STANDARD and LONG) were extracted from a corpus of French conversational speech. With high concentrations of liquids and glides in SHORT, vowels show strong temporal distortion towards lengthening. Conversely, fricatives and plosives represent stable temporal supports. Some function words tend to promote reduction while the lengthening of phonemes is due, in large part, to their presence in some frequent items in conversation (discourse markers, interjections). The duration of specific phonemes (/l/, /r/, /w/, /m/) is strongly conditioned by their lexical or communicative context. Phonemic durations are therefore determined, among other things, by the nature of phonemes and by the properties of lexicon and its use in conversation.

Key words: phonetic duration, corpus, conversation, lexicon, automatic alignment.

1. Introduction

Les variations temporelles des unités linguistiques (phonèmes, mots, syllabes, etc.) sont la conséquence de processus à la fois précis et complexes. Précis car lorsque cette organisation temporelle est perturbée (comme c'est parfois le cas dans les pathologies de la parole), la compréhension du locuteur s'en voit dégradée. Complexes car les variations temporelles des petites unités sont la conséquence d'un grand nombre de facteurs qui peuvent interagir entre eux (en se cumulant ou en s'inhibant). Les durées des segments phonétiques

¹ Aix Marseille Univ, CNRS, LPL, Aix-en-Provence; christine.meunier@univ-amu.fr.

² Aix Marseille Univ, CNRS, LPL, Aix-en-Provence; brigitte.bigi@univ-amu.fr.

sont fortement dépendantes de leur place dans la syllabe, la durée de la syllabe étant elle-même déterminée par la structure accentuelle (Greenberg *et al.* 2003). Ce timing millimétré est essentiel à la communication. On observe toutefois, et particulièrement en parole spontanée, des variations très importantes des patterns temporels dont l'origine dépasse probablement le cadre de la syllabe ou de l'organisation accentuelle. Comment ces variations « extrêmes » sont-elles interprétables ? A quels types de facteurs sont-elles dues ?

La production d'une parole spontanée, non préparée, engendre des variations temporelles bien supérieures à celles que l'on peut observer en parole contrôlée ou lue. Depuis plusieurs décennies, les recherches phonétiques sur la parole spontanée se sont considérablement accrues. Le développement des outils automatiques d'enrichissement de corpus (aussi bien pour la transcription, l'alignement, que pour les analyses acoustiques) a permis le développement des analyses sur de très grandes quantités de données, ce qui était impossible avant. Ces avancées ont permis à la fois d'augmenter la taille des observations, mais aussi de changer la nature de celles-ci. Sont alors apparus des phénomènes peu connus tels que des formes phonétiques non prototypiques, un nombre considérable d'omissions ou de réductions phonétiques (Johnson 2004) ou encore des allongements également peu prototypiques. Ces variations sont dues à un contexte plus riche et plus varié mais aussi à une structuration syntaxique et prosodique des énoncés spontanés très différente de celle de l'écrit ou de l'oral préparé. La parole spontanée recouvre un très large éventail de types de parole différents, allant de la parole produite par un professeur devant une classe à une conversation familière entre amis. Il a ainsi été montré que la proportion de durées extra-courtes des unités segmentales est bien plus importante en parole conversationnelle en comparaison avec une parole journalistique plus préparée et ce, aussi bien en anglais qu'en français (Adda-Decker & Snoeren 2011). La « préparation » en production de la parole tend ainsi à normaliser les durées. On peut faire l'hypothèse que la conversation, très riche en information contextuelle, permet une sous-spécification phonétique plus importante mais non dommageable à la compréhension. Notre objectif, dans cette étude, vise à mieux comprendre le fonctionnement phonétique des variations temporelles spécifiques à la parole spontanée. Pour ce faire, nous analysons les variations extra-courtes et extra-longues des segments phonétiques en parole conversationnelle au regard de leur **nature** et du **contexte lexical**.

La **nature** des phonèmes, autrement dit leur caractéristiques phonético-articulatoires, est rarement évoquée dans les travaux sur la réduction ou l'allongement phonétique, alors qu'il est probable qu'elle influe sur la propension des phonèmes à varier temporellement. Adda-Decker & Snoeren (2011) évaluent cette propension dans deux types de corpus : contrôlé et spontané. Les résultats montrent une

augmentation de la proportion de phonèmes courts au détriment des durées moyennes en parole spontanée. C'est le cas pour les voyelles, les liquides et les glissantes, mais moins pour les fricatives. On note également une plus forte proportion de voyelles très longues en spontané mais pas pour les autres phonèmes, à l'exception de la glissante /w/. Ces résultats suggèrent que certains phonèmes, indépendamment de facteurs, absorbent davantage que d'autres les variations temporelles. Les caractéristiques articulatoires des phonèmes pourraient, en partie, déterminer leur empan temporel. Or ces caractéristiques articulatoires sont universelles et l'on peut faire l'hypothèse que les variations temporelles des segments phonétiques pourraient être similaires dans toutes les langues. Toutefois, les durées segmentales sont soumises à de nombreux facteurs dont leur distribution dans les mots du lexique.

L'étude de l'influence du **lexique** sur les variations temporelles des segments phonétiques s'est particulièrement développée dans les analyses portant sur des corpus de parole spontanée. Tous les mots ne sont pas produits de la même façon dans un discours oral non préparé. Il a été montré que les caractéristiques lexicales telles que la fréquence (Jurafsky *et al.* 2001, Pluymakers *et al.* 2005) ou la place dans le mot (position interne; Meunier & Espesser 2011) favorisent la réduction phonétique. Certaines formes lexicales très fréquentes sont par ailleurs très souvent réduites (Johnson 2004, Torreira & Ernestus 2011, Bybee *et al.* 2016). Plus particulièrement, les propriétés morphosyntaxiques semblent avoir une influence sur les durées phonétiques. Le taux de segments vocaliques supérieurs à 100ms est bien moins important dans des mots outils (fonctionnels) que dans des mots de contenu (Adda-Decker *et al.* 2008). On observe même une diminution moyenne de 25ms dans la durée de voyelles de mots fonction monosyllabiques comparées à celles des voyelles de mots de contenu, également monosyllabiques (Meunier & Espesser 2011). La fonction attribuée aux mots dans un discours oral spontané joue ainsi un rôle important dans la réduction temporelle des segments phonétiques. Mais le discours spontané fait également apparaître des allongements de durée très importants et qui n'apparaissent pas dans des productions orales contrôlées. Ils peuvent être dus à des disfluences (hésitations), aux positions en final d'énoncé ou encore à des marqueurs discursifs. Nous faisons donc l'hypothèse ici que les propriétés du lexique (morphosyntaxiques ou discursives) ont un impact sur la durée des segments phonétiques. Plus encore, nous cherchons à observer si les variations temporelles de certains phonèmes pourraient être la conséquence de leur présence massive dans certains types de mots.

Notre objectif dans ce papier est d'explorer **les durées extrêmes** de l'empan temporel dans la parole conversationnelle afin de comprendre quels phonèmes sont situés à chacune des extrémités

et quels sont les facteurs qui favorisent ces durées extrêmes. Si la nature des phonèmes peut indiquer leur propension universelle à varier temporellement, la composition phonotactique et morphologique des mots est, en revanche, spécifique à chaque langue. C'est cette compensation que nous cherchons à mieux comprendre ici. Nous proposons d'aborder la question des variations temporelles via une approche ascendante, quantitative et descriptive. Pour ce faire, nous exploitons les données de l'alignement automatique sur un corpus de **parole conversationnelle** avec un focus spécifique sur les **segments extra-courts** (moins de 30ms) et **extra-longs** (plus de 200ms). Nous posons deux questions : 1/ la réduction et l'allongement phonétique affectent-ils indifféremment tous les phonèmes ? 2/ les segments réduits ou allongés se concentrent-ils plus volontiers sur certains mots ou catégories de mots ? Dans un premier temps une description du corpus utilisé sera présentée (annotation, distribution et durées phonétiques globales). Les résultats portant sur les cohortes de durées extrêmes seront ensuite détaillés. Enfin, les durées phonétiques seront analysées à la lumière des propriétés morphosyntaxiques.

2. Le corpus CID

Nous avons choisi d'observer les variations temporelles via une extraction automatique des segments phonétiques selon leur durée dans un corpus de parole conversationnelle. Les données ont été extraites du *Corpus of Interactional Data - CID* (Bertrand *et al.* 2008). Il s'agit d'un enregistrement audio-vidéo de dialogues spontanés (8h au total) entre des locuteurs français natifs partageant une certaine familiarité (8 conversations d'une heure chacune entre deux locuteurs, soit 16 locuteurs, 10 femmes et 6 hommes). Le CID est doté d'annotations à de multiples niveaux (Blache *et al.* 2010), ce qui permet d'analyser les interactions entre différents secteurs linguistiques. Nos descriptions se basent donc sur un style de parole spontanée et relâchée (conversations familières). L'extraction des données a été effectuée sur une version du corpus phonétisée avec un système à base de règles à partir d'une Transcription Orthographique Enrichie. L'alignement automatique en mots et phonèmes a été réalisé avec *SPPAS* (Bigi 2015), et *Julius* comme décodeur (Lee *et al.* 2001). Une version anonymisée de ces annotations du CID est publiquement disponible³.

1.1. Annotation phonétique du CID

La phonétisation du corpus fournit une annotation phonétique du corpus en SAMPA⁴ (Wells 1997). Elle ne permet pas la distinction de

³ <https://hdl.handle.net/11403/sldr000720>.

⁴ Code SAMPA pour le français : <https://www.phon.ucl.ac.uk/home/sampa/french.htm>.

l'ensemble des voyelles moyennes. Nous avons ainsi une seule voyelles (moyenne) pour les variantes mi-ouvertes et mi-fermées. De même, les deux variantes de la voyelle ouverte, /a/, ne sont codées que par un seul phonème. Enfin, les deux variantes de la voyelle nasale antérieure sont également regroupées sous un même symbole phonétique. Dans ce papier, le codage des voyelles sera donc le suivant :

- e pour /e/ et /ɛ/
- o pour /o/ et /ɔ/
- ø pour /ø/, /œ/ et /ə/
- a pour /a/ et /ɑ/
- ɛ̃ pour /ɛ̃/ et /œ̃/

L'alignement automatique du corpus a permis l'annotation de 302745 segments. L'annotation des rires et des bruits n'est pas prise en compte dans l'inventaire. Nous avons exclu de ce premier jeu de données les annotations phonétiques qui ne représentaient pas de données pertinentes pour notre étude. Il s'agit des catégories *unknown* (34 occurrences) et *punctuation* (7 occurrences). En revanche, nous avons conservé les annotations des interjections et les mots tronqués pour lesquels la réalisation est constituée de phonèmes identifiables, bien définis et que nous considérons comme prenant part à la production orale des phonèmes dans une conversation. Nous avons choisi de traiter à part les pauses remplies (*filled pauses*, 5040 occurrences) exclusivement produites sous la forme « euh » de façon à ne pas les associer à la voyelle /ø/. Elles seront donc évoquées plus loin dans cet article mais ne sont pas conservées dans le jeu de données final. Le jeu de données conservé pour l'analyse représente donc 297664 phonèmes.

Nous avons bien conscience que la durée des phonèmes dans un corpus aligné automatiquement est liée aux caractéristiques, voire aux biais de l'aligneur. Notamment, les segments extra-courts de 30ms peuvent être dus à des erreurs d'alignement ainsi qu'à des segments non réalisés mais présents dans l'annotation car transcrits par les annotateurs. Toutefois, les omissions ne représentent pas un problème ici : un phonème transcrit mais non réalisé est, de toute façon, considéré comme une réalisation extra-courte sous sa forme extrême. Ce n'est donc pas uniquement la durée des phonèmes elle-même qui nous intéresse ici, mais l'indication qu'un phonème est produit en version extra-courte.

Par ailleurs, la trame minimale de l'aligneur est de 30ms. Toutefois, l'alignement des rires (1542 occurrences) et des pauses remplies (5040 occurrences) ayant été corrigé manuellement, l'annotation de certains phonèmes adjacents a parfois été modifiée. Cela implique que quelques durées puissent être inférieures à 30ms. Nous avons donc traité indifféremment tous les phonèmes de 30ms

ou inférieurs à 30ms (voir plus loin 3.1. *Méthode de comparaison des cohortes*) dans la mesure où tous les phonèmes n'ont pas été corrigés.

Il est possible que l'aligneur sous-estime ou, à l'inverse, surestime la durée de certains phonèmes. Pour cette raison, les frontières produites par l'aligneur ont été comparées avec une segmentation manuelle faite par deux experts. La différence entre les deux annotations ne montre pas d'écart majeur pouvant avoir un impact sur la présente étude (Bigi & Meunier 2018).

1.2. Distribution des phonèmes dans le corpus

La répartition des phonèmes fréquents ou rares dans les grands corpus de parole, qu'ils soient oraux ou écrits, est, en général, assez stable. Une comparaison avec les fréquences relevées dans *Lexique* (New *et al.* 2001) montre que, quels que soient les corpus, on retrouve toujours parmi les phonèmes les plus fréquents les voyelles /e/, /a/, /i/, /ø/ et les consonnes /ʀ/, /s/, /t/, /l/ (figure 1). Dans nos données, la hiérarchie habituelle est globalement respectée (tableau 1). Les phonèmes /ɛ/, /a/, /ʀ/ restent en tête des phonèmes les plus fréquents. On notera la présence massive de /e/ dans notre corpus qui, rappelons-le, regroupe les deux voyelles moyennes antérieures non-arrondies.

Pho	e	a	ʀ	s	t	l	i	k	ø	m	p	d	ã	w	
Occ.	38589	28231	17512	17257	16503	15457	14035	13659	12183	11978	11549	10989	8924	8554	
	o	y	n	õ	v	ẽ	u	f	ʒ	j	b	z	ʃ	g	ç
	7944	6943	6749	6716	6575	6029	5358	5230	4804	4271	3926	3067	2206	1374	1052

Tableau 1 : Nombre d'occurrences pour chacun des phonèmes dans le corpus (en ordre décroissant)

Si la fréquence des phonèmes dans le corpus CID est relativement proche de celle observée dans *Lexique*, on notera malgré tout quelques disparités. Notamment la surreprésentation de /w/ ainsi qu'une légère sous-représentation de /ø/, /d/ et /l/ dans le CID (figure 1). La rareté des glissantes /ç/, /j/, des occlusives sonores /g/ et /b/ et des fricatives /ʃ/, /f/, /z/, /ʒ/ est similaire dans les deux jeux de données.

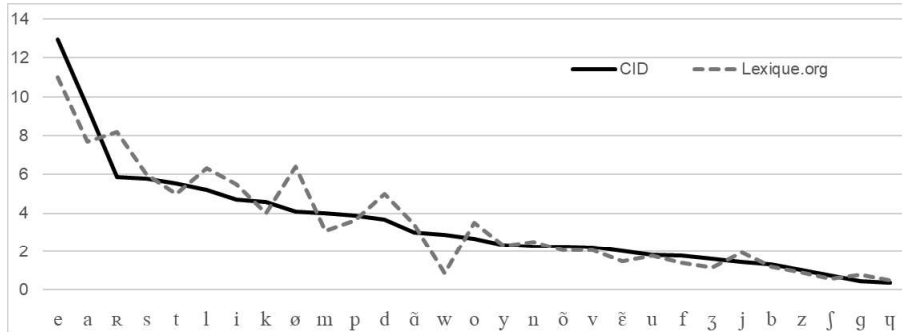


Figure 1 : Fréquence (en %) des phonèmes dans le corpus CID et dans la base Lexique

On constate ainsi que la production orale des phonèmes n'est clairement pas uniforme, ce qui pourrait suggérer des routines articulatoires plus ancrées (car répétées plus souvent) pour les phonèmes les plus fréquents. La forte disparité de la distribution des phonèmes dans les grands corpus (qu'ils soient écrits ou oraux) est, bien entendu, essentiellement due à leur fréquence d'apparition dans le lexique.

1.3. Durées des phonèmes dans le corpus

Si de nombreux travaux ont exploré les variations des durées phonétiques, peu nous livrent des valeurs effectives des durées des phonèmes en français. Bartkova & Sorin (1987) ont déterminé des durées intrinsèques des phonèmes du français suggérant ainsi qu'ils étaient plus courts ou plus longs selon leur mode d'articulation, les fricatives étant les phonèmes les plus longs et les liquides les plus courts. Ces données, issues de parole lue (syllabes isolées), nous fournissent un point de départ pour évaluer les variations temporelles de chaque phonème.

La moyenne des durées des phonèmes du CID est de 79ms, ce qui représente moins de la moitié de celles extraites de Bartkova & Sorin (1987) (178ms). 75% des phonèmes durent moins de 90ms (figure 2, à gauche) alors que dans des données de parole lue, peu de phonèmes ont une durée moyenne inférieure à 100ms. Les segments inférieurs ou égaux à 30ms représentent une proportion très importante et l'on observe un décrochage du nombre d'occurrences ayant une durée légèrement plus longue, ce qui est conforme aux observations précédentes en parole conversationnelle (figure 2, à droite, Adda-Decker & Snoeren 2011). Cette surreprésentation des phonèmes inférieurs ou égaux à 30ms tient à deux raisons : 1/ les phonèmes

extra-courts sont très nombreux en parole spontanée et 2/ il n'est pas possible de sous-catégoriser les phonèmes de 0 à 30ms (trame minimale de l'aligneur). Par ailleurs, il semble que les phonèmes très longs soient plus nombreux dans le CID que dans le corpus de conversation téléphonique de Adda-Decker & Snoeren (2011). Cela peut être dû au fait que nous avons conservé, dans nos cohortes, des occurrences non lexicales comme les interjections et troncations (voir plus loin 4.1. *Les mots du corpus*).

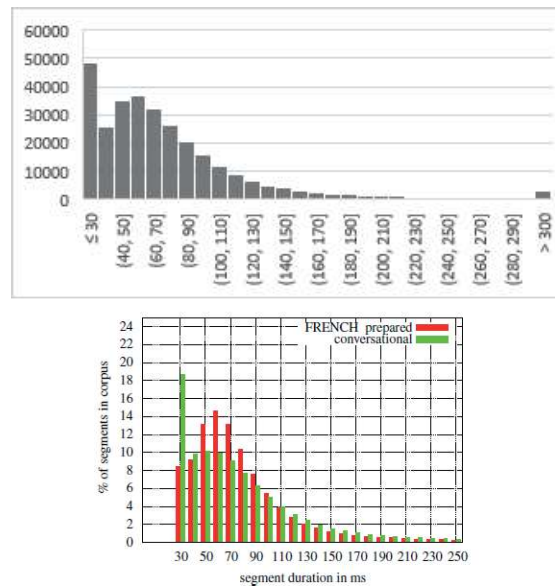


Figure 2 : Distribution des durées des segments phonétiques dans le corpus CID⁵ (à droite) et en parole journalistique préparée et en conversation téléphonique⁶ (à gauche) (d'après Adda & Snoeren 2011)

Les durées des phonèmes sont globalement dépendantes de leur classe articulatoire. On l'observe clairement sur les données de logatomes lus (figure 3, en blanc). Les fricatives sourdes sont les plus longues alors que les liquides et les glissantes sont les plus courtes. On observe une nette différence entre les consonnes sonores (courtes) et leurs homologues sourdes (très longues). Les voyelles présentent des valeurs moyennes, exceptées les trois nasales, bien plus longues. Malgré de très fortes différences concernant les durées moyennes (les logatomes de Bartkova & Sorin sont 2,25 fois plus longs que ceux du CID), les patterns de durées sont globalement conservés pour les phonèmes du CID (figure 3, en gris).

⁵ Axe y : nombre d'occurrences ; axe x : durées segmentales.

⁶ Axe y : proportion des occurrences.

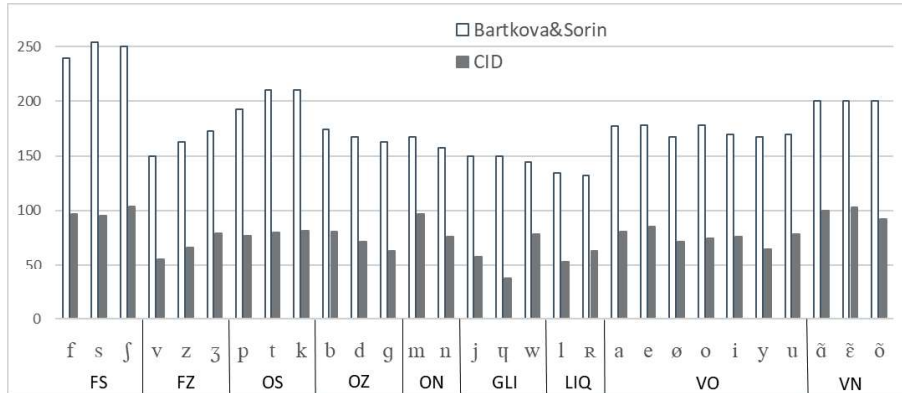


Figure 3 : Durées moyennes (en ms) des phonèmes du français mesurées 1/ sur des logatomes lus de type CVC (en blanc; d'après de Bartkova & Sorin 1987) et 2/ sur le corpus CID (en gris)⁷

Le CID fait apparaître des durées élevées pour les fricatives et occlusives sourdes et pour les voyelles nasales. Les glissantes, liquides et consonnes nasales sont globalement plus courtes, mais avec des disparités importantes. On notera que /ɥ/ est le phonème le plus court (moyenne = 37ms), suivi par /l/ (moyenne = 53ms), puis par la fricative /v/ (moyenne = 55ms). On observe de fortes différences de durées à l'intérieur du groupe des trois glissantes. Les phonèmes les plus longs sont les voyelles nasales /ã/, /ê/ et /ô/, ainsi que les trois fricatives sourdes /f/, /s/ et surtout /ʃ/. La durée moyenne de /m/ (97ms) est extrêmement surprenante pour une consonne nasale, qui est habituellement assez courte. Les phonèmes dont les valeurs sont peu attendues montrent en fait des médianes bien supérieures aux moyennes. C'est le cas de /w/ (moyenne : 78ms; médiane : 50ms), /ø/ (moyenne : 71ms; médiane : 50ms) et /m/ (moyenne : 97ms; médiane : 70ms). Cela suggère que ces phonèmes sont généralement plutôt courts mais que plusieurs réalisations très allongées tirent la moyenne vers des valeurs plus hautes. Ces aspects seront abordés dans les sections suivantes.

2. Cohortes SHORT et LONG

2.1. Méthode de comparaison des cohortes

L'annotation des 297664 phonèmes nous a permis de créer trois cohortes regroupant les phonèmes selon leur durée :

⁷ Les phonèmes sont regroupés en macro-classes.

Cohorte **SHORT** (segments extra-courts) \leq 30ms (48677 phonèmes, 16,35% du corpus)
 Cohorte **STANDARD** (segments standards) entre 30ms et 200ms (238290 phonèmes, 80,06% du corpus)
 Cohorte **LONG** (segments extra-long) \geq 200ms (10697 phonèmes, 3,59% du corpus)

Adda-Decker & Snoeren (2011) ont distingué quatre cohortes de durée dans leur étude (*short* : \leq 40ms, *medium* : 50-110, *long* : 120-240, *very long* : \geq 250). Leur distinction fournit un panorama plus détaillé de la distribution des phonèmes sur la totalité du corpus. Notre objet d'étude portant sur les durées extrêmes, nous focalisons notre attention sur les cohortes SHORT et LONG, dont nous avons élargi l'empan (\geq 200ms) de façon à intégrer un plus grand nombre de données. Dans nos trois catégories, les phonèmes sont distribués de façon très inégale. Effectivement, il n'est pas question ici de constituer des catégories homogènes mais plutôt d'observer la distribution des phonèmes dans des cohortes de durées extrêmes. La cohorte SHORT représente une proportion considérable du corpus pour des raisons évoquées plus haut (trame minimale de l'aligneur, voir 2.3. *Durées des phonèmes dans le corpus*). Pour la cohorte LONG, nous avons choisi de ne considérer que les phonèmes dont la durée était approximativement supérieure aux durées des phonèmes les plus longs observées en parole lue (Bartkova & Sorin 1987). Cette limite de durée (200ms) n'est pourtant pas une durée extrême en parole lue, mais elle devient très rare en parole spontanée (3,59% des occurrences de phonèmes). Si la cohorte SHORT est caractérisée par un empan de durée très restreint (0 à 30ms), les durées des cohortes LONG vont de 200ms à plus de 3 secondes. SHORT est donc caractérisée par un nombre important de données comprenant des durées très similaires. LONG est caractérisée par un nombre restreint de données portant sur un empan de valeurs temporelles très large.

Pour chaque cohorte, la fréquence (en pourcentage) de chaque phonème est extraite. Ensuite, la différence (D) entre chaque cohorte extrême (SHORT et LONG) et la cohorte STANDARD est calculée ainsi :

$D_SHORT = \%SHORT - \%STANDARD$
$D_LONG = \%LONG - \%STANDARD$

D_SHORT et D_LONG fournissent ainsi respectivement la sur- et la sous-représentation de chaque phonème dans une cohorte extrême (SHORT ou LONG). Par exemple, si la fréquence d'un phonème x est de 9% dans STANDARD, de 16% dans SHORT et de 4% dans LONG, cela signifie que x est surreprésenté en SHORT ($D_SHORT =$

16-9 = **7**) mais sous-représenté en LONG ($D_LONG = 4-9 = -5$). On voit alors que les valeurs **positives** expriment la **surreprésentation** alors que les valeurs **négatives** expriment une **sous-représentation**.

Le score obtenu n'est pas directement lié à la fréquence des phonèmes, puisque chaque phonème est comparé à lui-même. Toutefois, la fréquence joue partiellement un rôle dans ces calculs : 80% des réalisations de la glissante /ɥ/ sont dans SHORT, mais ces 80% ne représentent que 1,72% des phonèmes de la cohorte SHORT, et 0,09% de STANDARD ; ainsi, la valeur de D_SHORT pour /ɥ/ est de 1,6, ce qui est relativement peu élevé. En fait, /ɥ/ représente 0,35% des phonèmes du corpus et sa forte proportion dans SHORT est minimisée par sa faible présence dans le corpus. Les valeurs de D expriment donc la possibilité qu'a un phonème d'être réduit ou allongé, pondérée par sa fréquence dans le corpus.

2.2. Distribution des phonèmes dans les cohortes extrêmes

Rappelons que les effectifs des trois cohortes sont très hétérogènes (SHORT : 16,35% ; STANDARD : 80,05% ; LONG : 3,59%). Tous les phonèmes sont présents dans les trois cohortes, même s'ils ne le sont pas tous dans les mêmes proportions. Dans les deux parties suivantes consacrées aux cohortes SHORT et LONG nous présenteront les différences D_SHORT et D_LONG aussi bien pour les macro-classes de phonèmes que pour les phonèmes eux-mêmes. Le regroupement des phonèmes en macro-classes (représentant globalement le mode d'articulation) permet de rendre visibles les tendances à l'allongement ou à la réduction selon des grandes classes articulatoires. Les macro-classes utilisées sont les suivantes :

- **OS** : Occlusives sourdes
- **OZ** : Occlusives sonores
- **ON** : Occlusives nasales
- **FS** : Fricatives sourdes
- **FZ** : Fricatives sonores
- **LIQ** : Liquides
- **GLI** : Glissantes
- **VO** : Voyelles orales
- **VN** : Voyelles nasales

2.2.1. Phonèmes surreprésentés dans SHORT (D_SHORT)

Les liquides, glissantes et voyelles orales sont surreprésentées dans SHORT (D_SHORT positif, figure 4). On notera que ces trois classes de phonèmes sont également identifiées par Adda-Decker & Snoeren (2011) comme celles dont la proportion est pratiquement aussi importante en short qu'en medium en conversation. En revanche, les

occlusives et fricatives sourdes (essentiellement /s/ et /t/) ainsi que les nasales (voyelles et consonnes) sont plutôt sous-représentées dans cette cohorte (figure 4). Pour les autres macro-classes les différences entre les deux cohortes sont moins tranchées. Ces observations tendraient à montrer que les liquides et glissantes sont les classes de phonèmes les plus réduits ici.

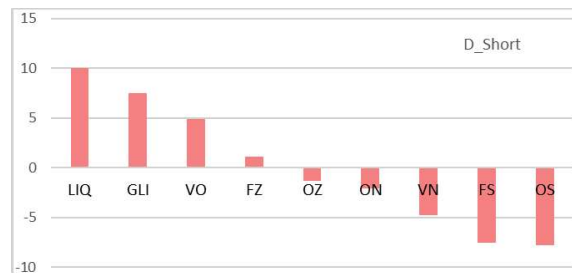


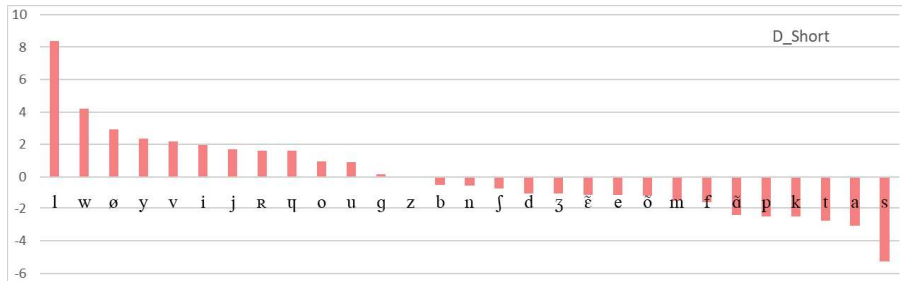
Figure 4 : Valeurs de D_SHORT pour chaque macro-classe⁸

Une observation plus fine de D_SHORT par phonème (figure 5) montre que la liquide /l/ présente la plus forte valeur de D_SHORT (8,4), nettement supérieure à celle des autres phonèmes. Effectivement, si /l/ représente 5,19% des phonèmes de la cohorte STANDARD, il passe à 12,28% des phonèmes de SHORT. Il est le phonème le plus fréquent de la cohorte SHORT. Notons également que 39 % de ses réalisations se trouvent dans cette cohorte. Pour Adda-Decker *et al.* (2008), /l/ est également le phonème présentant le plus fort pourcentage de durée minimale.

Le cas du /r/ est assez surprenant et intéressant, car cette liquide est connue pour sa forte variabilité et sa propension à réduire en parole continue en français. Certes elle est surreprésentée dans SHORT mais de façon modérée (D_SHORT= 1,6), notamment si on la compare à l'autre liquide /l/. 21% de ses réalisations sont dans SHORT et 78% dans STANDARD. Il semble donc que la grande majorité des /r/ produits soient peu réduits, comparé à d'autres phonèmes.

La glissante /w/ montre une valeur D_SHORT de 4,2 et 36% de ses réalisations se trouvent dans la cohorte SHORT. Elle est le deuxième phonème dont la surreprésentation est la plus importante dans SHORT. On rappellera que /w/ est assez fréquent dans le CID alors qu'il est plus rare dans *Lexique* (figure 3).

⁸ Les classes en valeurs positives sont surreprésentées dans la cohorte SHORT, alors que les valeurs négatives expriment une sous-représentation.

Figure 5 : Valeurs de D_SHORT pour chaque phonème⁹

Parmi les voyelles orales, ce sont essentiellement les trois voyelles hautes (/i/, /y/, /u/) qui sont surreprésentées dans SHORT. La présence de ces voyelles dans SHORT n'est pas surprenante dans la mesure où elles sont très fréquemment assourdis et/ou réduites dans les productions de parole spontanée. On notera par ailleurs que leur pattern articulatoire est similaire à celui des trois glissantes, elles-mêmes surreprésentées dans SHORT. La voyelle /ø/ (regroupant /ø/, /œ/ et /ə/), est également surreprésentée dans SHORT (D_SHORT=2,9). Ces résultats ne confirment pas clairement les données de Adda-Decker *et al.* (2008) et Adda-Decker & Snoeren (2011). Toutefois, la phonétisation des voyelles n'étant pas similaire dans les deux corpus, il serait hasardeux de comparer les deux jeux de données.

Finalement, seules les deux voyelles antérieures et non arrondies (/a/ et /e/) ne sont pas surreprésentées dans SHORT. Avec 11,62% des phonèmes de SHORT, la voyelle /e/ est le deuxième phonème le plus fréquent de cette cohorte. Toutefois, elle ne fait pas partie des phonèmes considérés comme surreprésentés dans SHORT, car il s'agit du phonème le plus fréquent du corpus et sa proportion dans STANDARD est également très importante (12,80% des phonèmes de cette cohorte). D_SHORT fournit donc une valeur peu significative (-1,1) suggérant que la proportion de /e/ dans les deux cohortes varie très peu. Mais étant donnée la présence massive de /e/ dans le corpus, on constate que de très nombreux exemplaires de cette voyelle sont réduits.

En revanche, la fricative /s/ (D_SHORT= -5,2) et la voyelle /a/ (D_SHORT= -3,1) sont plutôt sous-représentées dans la cohorte SHORT. De même, les trois occlusives sourdes ont une valeur D_SHORT négative, ce qui suggère une sous-représentation dans SHORT. Les durées moyennes des occlusives sourdes dans le corpus sont tout à fait dans la moyenne (81 à 83ms) mais très peu de leurs réalisations sont extra-courtes.

⁹ Les phonèmes en valeurs positives sont surreprésentés dans la cohorte SHORT, alors que les valeurs négatives expriment une sous-représentation.

2.2.2. Phonèmes surreprésentés dans LONG (D_LONG)

La Cohorte LONG et le calcul de D_LONG nous permettent de rendre visible les phonèmes dont l'allongement est fréquent. Les voyelles (orales et nasales) montrent une valeur de D_LONG positive et particulièrement élevée pour VO (D_LONG = 15, figure 6). C'est également le cas pour les occlusives nasales. On aurait pu s'attendre à une présence massive des trois voyelles nasales dans la mesure où leur durée moyenne est très élevée, mais la très forte prépondérance de /e/ dans LONG (figure 7) masque en partie leur surreprésentation. En effet, /e/ présente à la fois le plus fort taux de surreprésentation (D_LONG= 9,9) mais il est aussi le phonème le plus représenté dans LONG (23% des occurrences de LONG).

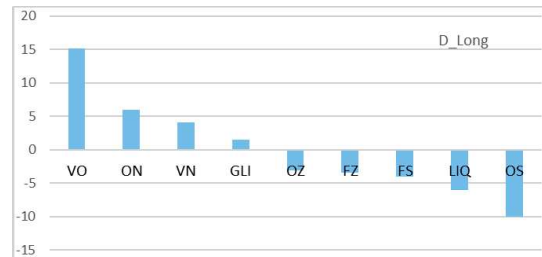
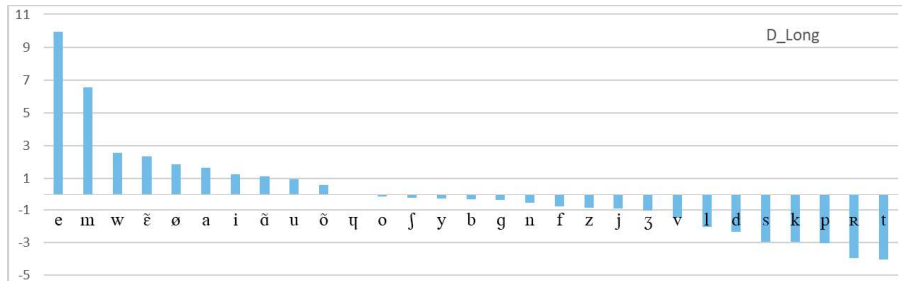


Figure 6 : Valeurs de D_LONG pour chaque macro-classe¹⁰

Alors que l'ensemble des valeurs positives de D_LONG sont représentées par des voyelles, on note l'étonnante surreprésentation de /m/ (D_LONG = 6,6) et de /w/ (D_LONG = 2,53) dont les durées intrinsèques sont plutôt courtes. /w/ était déjà surreprésenté dans D_SHORT, ce qui suggère une forte variabilité temporelle pour cette glissante. Dans Adda-Decker & Snoeren (2011) /w/ est le seul phonème à être clairement surreprésenté dans *long* et *very long* en parole spontanée spécifiquement.

C'est également le cas pour /ø/, surreprésenté dans SHORT et LONG. Nous devons toutefois rappeler que les pauses remplies « euh », phonétiquement réalisées /ø/, ne sont pas prises en compte pour l'analyse de cette voyelle. Or les pauses remplies représentent 5040 occurrences (1,7% des segments du corpus) et se retrouvent majoritairement dans la cohorte LONG (2929 occurrences, soit 58% des réalisations de « euh ») avec une durée moyenne de 295ms, donc supérieure à la durée minimale de la cohorte LONG. On peut donc considérer que la configuration articulaire de cette voyelle apparaît très souvent dans une réalisation allongée.

¹⁰ Les classes en valeurs positives sont surreprésentées dans la cohorte LONG, alors que les valeurs négatives expriment une sous-représentation.

Figure 7 : Valeurs de D_LONG pour chaque phonème¹¹

A l’opposé, on constate que les occlusives sourdes sont sous-représentées dans LONG ($D_LONG = -10,07$, figure 6). Or nous avons également constaté la sous-représentation de cette macro-classe dans D_SHORT (-7,8). Il apparaît ainsi que les OS pourraient représenter la classe la plus stable d’un point de vue temporel. De même, les fricatives sourdes, intrinsèquement longues (voir 4.1. *Durées des phonèmes dans le corpus*) sont plutôt sous-représentées dans LONG. La fricative /s/, très fréquente dans le corpus, est ainsi sous-représentée dans les deux cohortes SHORT et LONG, signe là encore d’une certaine stabilité temporelle.

3. Facteurs lexicaux

3.1. Les mots du corpus

Afin de mettre en correspondance les phonèmes avec les mots produits, il est nécessaire que l’annotation phonétique soit synchronisée avec une annotation 1/ des mots (tokens) et 2/ des catégories morphosyntaxique des mots. La transcription du corpus a permis de créer ces niveaux grâce à *MarsaTag* (Rauzy *et al.* 2014). *MarsaTag* est un outil permettant de créer un niveau d’annotation morphosyntaxique adapté au français soit à partir de la transcription soit à partir des mots alignés.

Les 8 heures du CID contiennent 6611 formes lexicales différentes, tandis que la totalité des occurrences est de 114397. La moitié des occurrences du corpus totalise seulement 39 formes différentes. Les mots du corpus CID se répartissent dans différentes catégories morphosyntaxiques (désormais CMS). Certains mots sont caractérisés par un poids sémantique plus fort que d’autres. C’est le cas des verbes, noms, adjectifs et adverbes que l’on qualifie habituellement de mots de contenu (désormais MC). A l’inverse, les

¹¹ Les phonèmes en valeurs positives sont surreprésentés dans la cohorte LONG, alors que les valeurs négatives expriment une sous-représentation.

prépositions, conjonctions, pronoms ou déterminants ont une fonction plus syntaxique permettant le lien entre les différents éléments d'une phrase (Bertrand & Espesser 2017) ; ce sont les mots de fonction (désormais MF). Certes, cette répartition est grossière, tant il est possible de trouver en conversation des adverbes (*pas, enfin, donc*) ou des verbes dont la charge sémantique est plutôt faible et dont la fonction dans le discours se rapporte plus à des MF. Toutefois, cette distinction reflète assez bien la distribution des mots dans la parole. Les MC représentent 46,8% de la totalité des mots du corpus, tandis que les MF en représentent 48,7%. Ces derniers sont peu nombreux, souvent courts (monosyllabiques) et produits de très nombreuses fois. A l'inverse, les MC sont très nombreux et leur nombre est, par nature, illimité. Ils sont souvent plus longs et beaucoup d'entre eux sont produits en nombre très limité.

Catégorie	Occurrences	%
Pronom	25199	22,0
Verbe	20436	17,9
Adverbe	15123	13,2
Nom	13419	11,7
Déterminant	9427	8,2
Conjonction	9390	8,2
Préposition	8693	7,6
<i>Interjection</i>	5068	4,4
Adjectif	4585	4,0
Auxiliaire	3057	2,7

Tableau 2 : Répartition (par fréquence décroissante) des mots dans le corpus en fonction de leur catégorie morphosyntaxique (d'après Blache *et al.* 2017)

La distribution des formes lexicales présentes dans le CID est donc semblable à celle que l'on trouve dans l'ensemble des corpus oraux spontanés et des grandes bases de données, telles que *Lexique* (New *et al.* 2001)¹². On y retrouve les caractéristiques de la loi de Zipf avec des mots dont la fréquence est inversement proportionnelle à leur rang dans le corpus. Les formes lexicales n'apparaissant qu'une seule fois dans le corpus sont au nombre de 3259 tandis que la forme la plus fréquente apparaît 3130 fois, ce qui est comparable à d'autres corpus de parole spontanée tel que NccFr (Torreira *et al.* 2010). La distribution des types de mots dans le corpus CID montre que leur fréquence d'apparition n'est pas homogène et dépend fortement de leur CMS (tableau 2). On notera le nombre très important de pronoms, représentant près d'un quart des occurrences lexicales du corpus.

Un focus sur les dix mots les plus fréquents du corpus (tableau

¹² *Lexique*, site réalisé par Boris New & Christophe Pallier (<http://www.lexique.org>).

3) montre qu'ils sont courts (monosyllabiques) et que la plupart sont des MF. Il est possible de penser que la réalisation des phonèmes présents dans les mots les plus fréquents pourrait être affectée par certaines propriétés de ces mots comme la répétition, la nature morphosyntaxique ou la longueur.

Forme	<i>est</i>	<i>c'</i>	<i>ouais</i>	<i>et</i>	<i>de</i>	<i>tu</i>	<i>pas</i>	<i>je</i>	<i>ça</i>	<i>le</i>
Occurrences	3130	3018	2916	2679	2033	2027	1895	1893	1817	1655
%	3,05	2,95	2,85	2,61	1,98	1,98	1,85	1,85	1,77	1,62

Tableau 3 : Occurrences des 10 formes lexicales les plus fréquentes avec leur fréquence d'apparition dans le corpus (d'après Meunier 2012)

Par ailleurs, la spécificité d'un corpus de parole conversationnelle tient également en la présence d'un certain nombre d'occurrences qui ne sont pas considérées comme des entrées lexicales. Leur présence n'est pas anodine, car ils représentent 10% des occurrences (tokens) de ce type de corpus (Bertrand & Espesser 2017). Il s'agit des interjections (*ah*, *bouh*, etc.) dont les *Feedbacks* (*mh*) qui accompagnent en général le narrateur et l'invitent à poursuivre. Nous avons souhaité les conserver dans notre analyse, car nous cherchons à inventorier l'ensemble des variations temporelles fondées sur des réalisations à base de phonèmes.

Nous disposons ainsi pour nos analyses d'une annotation sur trois niveaux comprenant : 1/ l'alignement phonétique, 2/ l'alignement des mots et 3/ l'annotation en catégories morphosyntaxiques (CMS). Notre objectif, dans cette étape, est de mettre en relation les variations temporelles des phonèmes avec les propriétés des mots dans lesquels ils sont produits.

3.2. Distributions phonétiques dans SHORT et LONG selon les mots porteurs

L'objectif de cette partie est de rendre compte des distributions temporelles présentées précédemment en fonction de la spécificité morphosyntaxique des mots porteurs ainsi que de leur fréquence. Nous présenterons ici le cas de quelques phonèmes dont les variations de durées peuvent être mises en perspectives avec des propriétés lexicales.

Le cas des liquides est particulièrement intéressant à observer à la lumière des mots dans lesquels elles sont produites. /l/ est majoritairement présent dans des MF, tels que les déterminants, les pronoms ou certains adverbes très fréquents (figure 8). L'ensemble de ces mots ont la particularité d'être très peu nombreux mais répétés de très nombreuses fois. *La*, *le*, *les* représentent à eux seuls 82% (4205 occurrences) de l'ensemble des déterminants dans lesquels /l/ est présent. *Elle(s)*, *il* et *l'*, *la*, *le*, *les*, *lui* représentent 89% (2433

occurrences) des pronoms dans lesquels /l/ est présent. Et *là, alors, plus*, représentent à eux trois 57% (1254 occurrences) des adverbes où /l/ est présent. Autrement dit, la production du phonème /l/ est concentrée dans un nombre très restreint de MF répétés de très nombreuses fois. Le contexte lexical de /r/ est à l'opposé. Cette liquide apparaît presque exclusivement dans des mots de contenu tels que verbes, noms (figure 9). Même si certains verbes sont fréquents (*faire, dire, voir*), ils représentent seulement 11% à eux trois des verbes dans lesquels /r/ est présent. Pour les noms, seul *truc* est répété 383 fois dans le corpus, les autres noms étant peu fréquents. Autrement dit, la consonne /r/ est plutôt produite dans un nombre de mots très important mais qui sont répétés peu de fois dans le corpus.

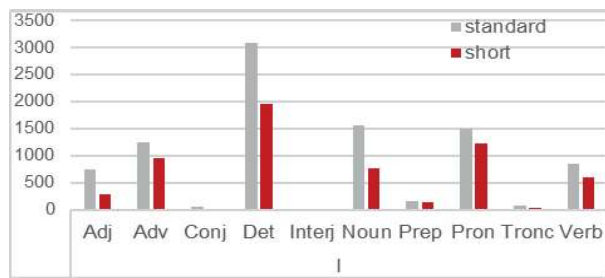


Figure 8 : Répartition des occurrences de /l/ dans STANDARD et SHORT en fonction des CMS

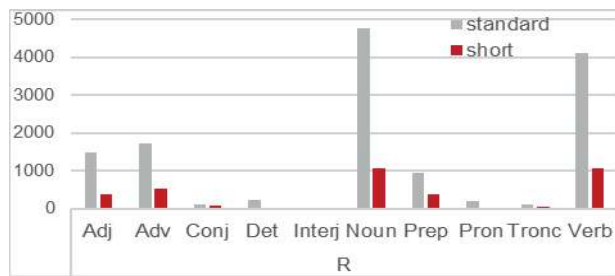


Figure 9 : Répartition des occurrences de /r/ dans STANDARD et SHORT en fonction des CMS

Les deux liquides montrent donc des contextes lexicaux très différents. La distribution des /l/ extra-courts (cohorte SHORT) concerne essentiellement les déterminants et les pronoms et dans des proportions très importantes. Les /r/ de la cohorte SHORT concernent essentiellement les noms et verbes mais la proportion de /r/ extra-courts est bien moindre que pour /l/. Rappelons que /l/ est la consonne la plus surreprésentée dans SHORT. Il semble donc que la forte présence de /l/ dans des MF entraîne un nombre important de

phonèmes extra-court. A l'inverse, la présence quasi-exclusive de /ʁ/ dans des mots de contenu pourrait limiter ses formes réduites.

La fricative voisée /v/ est également intéressante, car elle se comporte, à bien des égards, de façon très différente des autres fricatives. Exceptée /s/, les fricatives, qu'elles soient sourdes ou sonores, sont plutôt rares. /v/ ne fait pas ici exception, même si elle est la deuxième fricative la plus fréquente (figure 3). Toutefois, la fréquence de /v/ est essentiellement due à sa présence dans la catégorie Verbe (figure 10). Et plus précisément, 35% des réalisations de /v/ dans la catégorie Verbe se trouvent dans deux verbes très fréquents, *avoir* et *voir*. La réalisation de /v/ dans ces deux verbes est également surreprésentées dans SHORT (D_SHORT pour /v/ = 2,2). En effet, la réalisation de /v/ dans les multiples déclinaisons de *avoir* est très souvent réduite. De même, l'utilisation extrêmement fréquente de la séquence *tu vois* en parole conversationnelle entraîne une réalisation souvent approximante, réduite (voire omise), de cette fricative. Notons que /v/ est également présent dans les verbes fréquents tels que *vouloir* et *aller* (*vas, va, vais, etc.*).



Figure 10 : Répartition des occurrences de /v/ dans STANDARD et SHORT en fonction des CMS

La consonne nasale /m/ apparaît très peu en SHORT, mais nous avons remarqué qu'elle était surreprésentée en LONG. Rappelons que la consonne /m/ présente une moyenne (97ms) bien supérieure à sa médiane (70ms), ce qui laisse supposer des réalisations globalement courtes et un nombre réduit d'occurrences très longues. Une analyse plus fine nous permet de constater que ces occurrences longues apparaissent essentiellement dans la catégorie *Interjection* (figure 11). Il s'agit en fait de l'interjection *mh* extrêmement fréquente à l'oral en conversation sous forme de *Feedbacks* (Prévot *et al.* 2016) et qui représente 76% des réalisations LONG de cette consonne. On notera que /m/ est présent dans la conjonction *mais*, mot très fréquent puisqu'il représente 12% des réalisations de cette consonne. Toutefois, la fréquence de ce mot n'a pas d'influence sur sa durée, car il n'apparaît pas préférentiellement dans SHORT ou dans LONG.

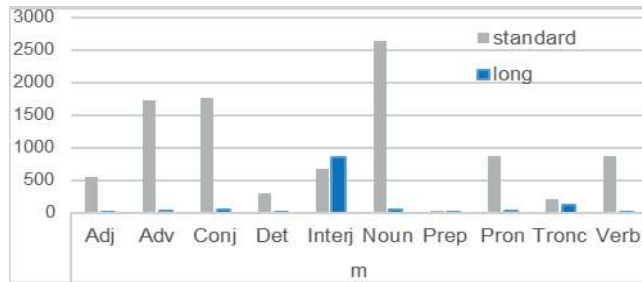


Figure 11 : Répartition des occurrences de /m/ dans STANDARD et LONG en fonction des CMS

Accordons maintenant notre attention au comportement des deux phonèmes /ɥ/ et /e/, respectivement le plus rare et le plus fréquent du corpus. 1052 occurrences de /ɥ/ sont présentes dans le corpus, ce qui représente 0,35% des phonèmes. 85% de ses réalisations se trouvent dans la cohorte SHORT, ce qui en fait le phonème le plus réduit, sa durée moyenne étant de 37ms. Cette glissante est principalement présente dans la catégorie *Pronom* (figure 12) et plus précisément dans le pronom *lui*, qui représente 27% de ses réalisations. Elle est également présente dans le verbe *être* avec la forme *suis* (12% de ses réalisations), fréquente en conversation. Enfin, on la retrouve dans l'adverbe *puis* (11% de ses réalisations), très fréquent en conversation sous sa forme *et puis*, très souvent réalisé /epi/. /ɥ/ est donc un phonème rare mais présent dans quelques MF fréquents. Ainsi, la quasi totalité de ses rares réalisations se font sous des formes extra-courtes (figure 12).

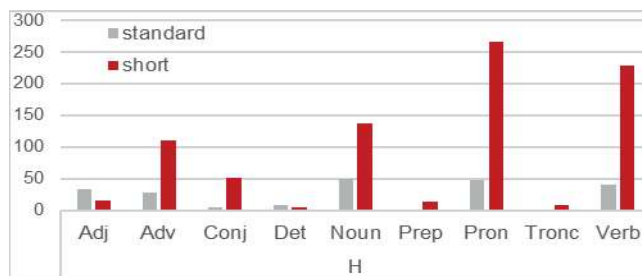


Figure 12 : Répartition des occurrences de /ɥ/ dans STANDARD (en gris) et SHORT (en rouge) en fonction des CMS

A l'inverse, avec ses 38589 occurrences, /e/ est de loin le phonème le plus fréquent du corpus. Il représente 13% des phonèmes du corpus et présente un très grand nombre d'occurrences dans les trois cohortes (12% de SHORT, 13% de STANDARD et 23% de LONG). Sa fréquence est due, majoritairement, à sa forte présence dans les verbes (figure 13) et notamment dans les verbes fréquents (*être*, *faire*, etc.) mais

également dans les formes verbales des temps de l'imparfait (*avais*), du conditionnel (*ferais*) et du futur (*serai*). Mais /e/ est également présent dans des mots très utilisés en conversation tels que les conjonctions *mais* et *et*, l'interjection *ouais* ou encore le déterminant *les*.

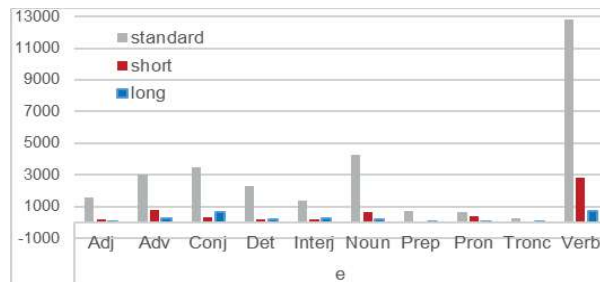


Figure 13 : Répartition des occurrences de /e/ dans STANDARD, SHORT et LONG en fonction des CMS

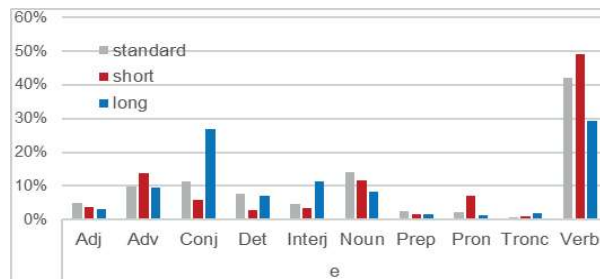


Figure 14 : Répartition (en %) des réalisations de /e/ dans chacune des cohortes en fonction des CMS

Nous avons vu que /e/ n'était pas sureprésenté dans SHORT ($D_SHORT = -1,1$) mais qu'il l'était dans LONG ($D_LONG = 9,9$). Cette surreprésentation est peu visible sur la figure 13, car la cohorte LONG représente seulement 3,6% du corpus. En revanche, sur la figure 14 (représentant la proportion de réalisation de la voyelle à l'intérieur de chaque cohorte), on observe que /e/ est très souvent allongé dans les catégories Verbe et Conjonction. La conjonction *et* représente 2913 occurrences, soit 7,5% des réalisations de /e/. Alors que ce mot est très fréquent, les réalisations de /e/ y sont peu réduites (seulement 5,8%) alors qu'elles sont très souvent allongées (16,7%). *Et* et *mais* sont très souvent utilisés comme marqueurs discursifs et peuvent, dans ces cas, être très allongés. Ainsi, /e/ est allongé (cohorte LONG) dans 11% des réalisations de *mais*¹³.

¹³ Notons ici que, dans la conjonction *mais*, seul le /e/ est allongé, alors que nous avons vu plus haut que, dans ce même mot, le /m/ ne présente pas d'allongement spécifique.

La voyelle /ø/ (regroupant les voyelles antérieures mi-ouvertes et mi-fermées) est relativement fréquente puisqu'elle représente 4% des phonèmes du corpus. Seuls les schwas transcrits dans la TOE apparaissent dans le corpus. Il est donc probable que cette voyelle prendrait une part beaucoup plus importante dans le corpus si l'ensemble des schwas y était inventorié.

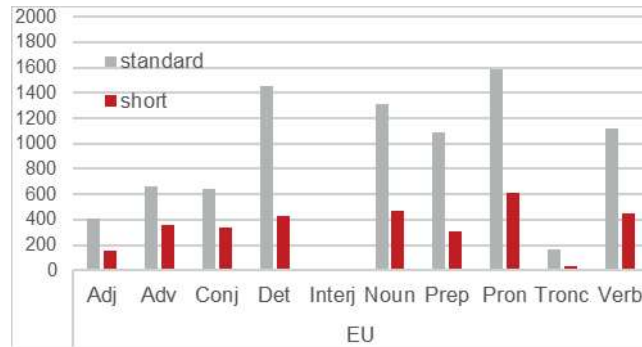


Figure 15 : Répartition des occurrences de /ø/ dans STANDARD et SHORT en fonction des CMS

La voyelle /ø/ est répartie dans de nombreuses CMS et l'on observe sur la figure 15 que, pour chacune de ces catégories, une forte proportion de réalisations réduites y est associée. Nous avons effectivement remarqué plus haut que /ø/ est surreprésenté en SHORT ($D_{SHORT} = 2,9$). Dans l'ensemble des CMS, /ø/ est présent dans des mots monosyllabiques fréquents tels que : *que* (conjonction : 952 occurrences dont 33% dans SHORT) ; *le* (déterminant : 1324 occurrences dont 25% dans SHORT), *de* (préposition : 1378 occurrences dont 20% dans SHORT), *je* (pronom : 892 occurrences dont 35% dans SHORT). A eux seuls, ces quatre mots représentent 37% des réalisations de /ø/ avec, chaque fois, une très forte proportion de réalisations réduites.

/ø/ est légèrement surreprésenté en LONG ($D_{LONG} = 1,9$) et cette surreprésentation concernent essentiellement sa présence dans les prépositions (18% des réalisations de /ø/ dans LONG ; *de*, par exemple) et dans les déterminants (14% des réalisations de /ø/ dans LONG ; *le*, par exemple). Il est possible d'attribuer ces allongements à des phénomènes de disfluece souvent localisés avant une pause et générant des allongements importants (Blache *et al.* 2010).

Enfin, nous avons également observé la distribution des CMS dans D_{SHORT} et D_{LONG} . On remarquera la surreprésentation dans SHORT des phonèmes présents dans les pronoms, verbes et déterminants (figure 16, à gauche). A l'inverse, les phonèmes produits dans des noms sont nettement sous-représentés dans cette cohorte, ce qui suggère que les noms sont moins affectés par la réduction. Notons

que la catégorie Verbes contient les auxiliaires (*est, étais, a*) ainsi que les formes très fréquentes dans une conversation (*vois, crois, etc.*).

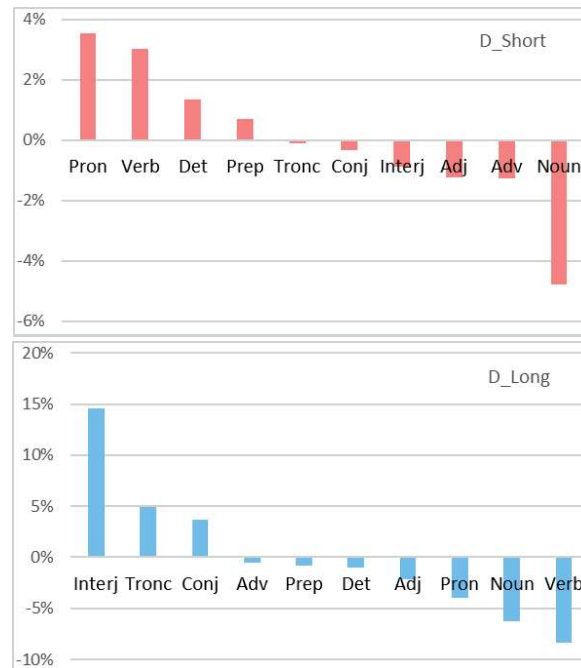


Figure 16 : Valeurs de D_SHORT (à gauche) et D_LONG (à droite) pour l'ensemble des phonèmes à l'intérieur de chaque CMS

Les phonèmes présents dans les interjections et les troncations sont clairement surreprésentés dans LONG. Pour les troncations, tous les phonèmes peuvent être concernés, mais les deux interjections *mh* et *ouais* représentent 75% des réalisations de cette catégorie. Les phonèmes présents ici sont donc clairement impactés. Les conjonctions, avec *et* et *mais*, sont également très fréquentes en LONG, mais leur allongement est probablement dû, dans ce cas, à leur fonction discursive et/ou à des disfluences, très souvent responsables d'allongements (Blache *et al.* 2010). On remarquera enfin que les noms sont sous-représentés dans LONG, comme dans SHORT, ce qui leur confère une certaine stabilité temporelle.

4. Discussion

Nous avons entrepris d'analyser les durées de l'ensemble des phonèmes du français dans un corpus de parole conversationnelle. Notre objectif visait à mieux déterminer les parts respectives des propriétés phonétiques et/ou lexicales dans des productions extra-

courtes ou extra-longues dans ce type de parole. Le jeu de données a été divisé en trois cohortes temporelles représentant respectivement des phonèmes extra-courts (SHORT \leq 30ms), des phonèmes de durée standard (30 < STANDARD < 200ms) et des phonèmes extra-longues (LONG \geq 200ms). Tous les phonèmes sont présents dans les trois cohortes, mais leur proportion dans chacune des cohortes est variable. La distribution des phonèmes nous a permis de mettre en évidence les phonèmes qui sont surreprésentés dans la cohorte SHORT (donc souvent réduits) et ceux qui sont surreprésentés dans la cohorte LONG (donc souvent allongés). L'ensemble de nos résultats tendent à montrer une image complexe de ces facteurs, conjointement responsables de ces variations temporelles.

Considérons dans un premier temps que la fréquence des phonèmes n'est pas homogène dans un corpus oral de parole spontanée. Cette hétérogénéité pourrait impacter les routines articulatoires mises en œuvre dans la réalisation des segments phonétiques : une réalisation très fréquente d'un phonème pourrait entraîner une hypo-articulation importante. Mais d'après nos résultats, la fréquence des phonèmes, elle-même, ne semble pas avoir d'influence quant à leur réalisation. La glissante /ɥ/, très peu fréquente, est réalisée dans 85% des cas dans la cohorte SHORT alors que la voyelle /e/, la plus fréquente du corpus n'est pas surreprésentée en SHORT.

Il semble en revanche que les propriétés phonético-articulatoires aient un impact sur les variations temporelles extrêmes des phonèmes, notamment pour ce qui concerne leur mode d'articulation. Nous avons noté que les liquides et glissantes sont les macro-classes présentant les valeurs de D_SHORT les plus importantes, tandis que les voyelles orales et les nasales (voyelles et consonnes) sont surreprésentées dans la cohorte LONG. On notera par ailleurs que les occlusives (sonores et sourdes) et les fricatives (sonores et sourdes) sont plutôt sous-représentées dans les deux cohortes SHORT et LONG. Ces résultats recouvrent partiellement ceux obtenus dans des études précédentes (Adda-Decker *et al.* 2008, Adda-Decker & Snoeren 2011), dans lesquels la réduction des glissantes et des liquides était observée. Notre méthodologie nous permet de caractériser les phonèmes en fonction de leur propension à absorber, ou non, les variations temporelles. Le tableau 4 exprime cette propension en fonction des catégories articulatoires¹⁴. Les phonèmes plutôt ouverts sont sensibles à la variation (réduction ou allongement), tandis que les phonèmes dont la fermeture buccale est critique (fricatives) ou totale (occlusives) présentent essentiellement des durées standard et sont peu représentés dans les cohortes de durées extrêmes (SHORT ou LONG). Les durées extra-courtes semblent plutôt concerner les liquides et les glissantes qui se trouvent dans une position intermédiaire sur

¹⁴ Selon les résultats présentés dans 4.3.1. et 4.3.2.

l'échelle d'ouverture buccale et sont, par ailleurs, considérées comme des phonèmes variables et instables (Chafcouloff 1983).

OS	OZ	ON	FS	FZ	LIQ	GLI	VO-f	VO-o	VN
p t k	b d g	m n	f s ʃ	v z ʒ	l r	j w ɥ	i y u	e ø o a	ã õ ẽ

Tableau 4 : Représentation des macro-classes et phonèmes en fonction de leur surreprésentation dans SHORT ou LONG¹⁵

Toutefois, il semble que l'allongement affecte essentiellement les voyelles. Même si l'on peut supposer que les voyelles sont sans doute les meilleurs supports phonétiques pour une production tenue¹⁶, il est probable que leur position de noyau syllabique joue un rôle important dans l'allongement temporel en portant l'ensemble des marques prosodiques (accentuation, allongement final, etc.). Il semble ainsi que les voyelles (pour l'allongement) et les liquides et glissantes (pour la réduction) soient les meilleurs supports phonétiques concernant la latitude dont peut disposer un locuteur dans l'optimisation de sa production. A l'opposé, les fricatives et occlusives pourraient constituer des îlots de stabilité temporelle.

Il nous faut également évoquer les durées intrinsèques des phonèmes en relation avec leur propension à varier. On aurait pu s'attendre à ce que les phonèmes courts se trouvent surreprésentés dans SHORT et les longs surreprésentés dans LONG (et inversement)¹⁷. C'est globalement le cas pour les liquides et les glissantes (en SHORT). C'est également le cas pour certaines voyelles (en LONG). Mais, comme évoqué ci-dessus, ce n'est le cas ni pour les occlusives, ni pour les fricatives. Et on observe, par ailleurs, des comportements temporels inattendus, comme la forte présence de /w/ et /m/ dans LONG ou encore la surreprésentation de certains phonèmes (/ø/, /w/) dans les deux cohortes extrêmes. Nos résultats montrent que ces comportements inattendus sont dus à des supports lexicaux spécifiques à la langue orale.

Les nombreux travaux concernant la nature morphosyntaxique des mots montrent leur impact sur la réduction phonétique (Johnson 2004, Jurafsky *et al.* 2001, Meunier & Espesser 2011, Wu & Adda-Decker 2020). Cet effet réducteur des MF est également identifiable dans notre corpus. Les phonèmes présents dans les pronoms (*je, tu, me, le, lui, etc.*), verbes (*est, étais, etc.*) et déterminants (*le, la, les, etc.*) sont surreprésentés dans SHORT. La forte surreprésentation de /l/ dans SHORT est ainsi due à sa nature phonétique et à sa présence dans des MF courts et

¹⁵ En rouge, les phonèmes surreprésentés dans SHORT ; en bleu, les phonèmes surreprésentés dans LONG ; en vert, les phonèmes surreprésentés dans les deux cohortes.

¹⁶ La voyelle /a/ est souvent choisie pour être produite sur une longue durée dans le cadre d'évaluation de la qualité vocale.

¹⁷ On pouvait d'ailleurs penser que la distribution des phonèmes dans les cohortes SHORT et LONG ne refléterait que leur durée intrinsèque.

fréquents. /r/ semble ainsi moins réduit en raison de sa rareté dans des MF. Il est de rigueur d'attribuer aux mots fréquents un rôle réducteur (Pluymakers *et al.* 2005), ce qui est vrai pour bon nombre d'entre eux. Mais la langue orale génère des items spécifiques et très fréquents tels que les interjections *mh* et *ouais* ou des disfluences (*de*, *le*, etc. allongés suivis de pauses). La fréquence dans ce cas n'est pas synonyme de réduction mais plutôt d'allongement extrême, ce qui fait de /m/ et de /w/ des phonèmes potentiellement extra-longes alors que leur nature phonétique ne le prédirait pas. Enfin, en plus de leur propriété morphosyntaxique, les mots peuvent se voir attribuer une fonction spécifique dans le discours oral. C'est le cas des marqueurs discursifs tels que *et* et *mais* dont la fréquence n'est pas, là non plus, synonyme de réduction. Le phonème /e/ se voit ainsi très souvent allongé en conversation. Nous voyons également que la simple distinction MF et MC ne suffit pas pour rendre explicites les variations temporelles. Les noms (MC), à l'inverse des autres CMS, semblent générer très peu de variation et peuvent ainsi représenter des supports de stabilité temporelle pour les phonèmes, ce qui n'est pas le cas des verbes (également MC). Les phonèmes présents dans les pronoms (MF) sont surreprésentés dans SHORT, alors que ceux présents dans les conjonctions (également MF) sont surreprésentés dans LONG.

Enfin, la réduction phonétique concerne également des séquences qui dépassent le cadre du mot (Meunier 2013, Wu & Adda-Decker 2020) et qui sont, la plupart du temps, impossible à identifier perceptivement. Ces séquences sont caractérisées par des zones réduites dont l'origine est probablement plus à attribuer à l'agencement du discours oral qu'au lexique. Dans ce cas tous les phonèmes, et toutes les catégories de mots, sont potentiellement affectés et c'est probablement pour cette raison que tous les phonèmes sont présents dans la cohorte SHORT.

Les variations temporelles des phonèmes font donc intervenir des facteurs multiples. Le mode de production des macro-classes de phonèmes étant commun à toutes les langues, on peut s'attendre à ce que leur propension à varier ou à rester stables soit observable dans des corpus de différentes langues. En revanche, la distribution des phonèmes dans le lexique étant spécifique à chaque langue (ou famille de langue), il est probable que la sur- ou sous-représentation des phonèmes dans des cohortes de durées extrêmes soit modulée en fonction de cette distribution.

5. Conclusion

Ce travail avait pour objectif d'analyser deux facteurs qui génèrent de la variation temporelle au niveau phonétique dans un corpus de parole conversationnelle et nos résultats tendent vers une interprétation combinée de ces facteurs. Les liquides et les glissantes,

intrinsèquement courtes et instables, se retrouvent massivement dans les réalisations extra-courtes. Les voyelles, intrinsèquement longues, se retrouve de façon importante dans les phonèmes allongés. Les occlusives et fricatives représentent des supports temporels stables. La présence de phonèmes dans certains mots fonction (pronoms, déterminants) favorise leur réduction alors que l’allongement des phonèmes est dû, en grande partie, à leur présence dans certains items fréquents et spécifiques à la langue orale, tels que les marqueurs discursifs, les interjections ou les troncations. Ces supports favorisent l’allongement peu prédictible de certains phonèmes tels que /w/ ou /m/. Il est difficile d’établir un rôle spécifique à la fréquence des mots qui peut, selon les cas, occasionner réduction ou allongement. Les durées phonétiques sont donc sensibles à des facteurs multiples. La structure discursive de la langue orale génère des contraintes bien plus diversifiées que celle de la parole contrôlée. Si la nature articuloire de chaque phonème indique une propension à varier potentiellement universelle, la distribution des phonèmes dans les mots impose une modulation temporelle qui est spécifique à chaque langue.

Références bibliographiques

- Adda-Decker, M., Gendrot, C., Nguyen, N. (2008), « Contributions du traitement automatique de la parole à l’étude des voyelles orales du français », *Traitement Automatique des Langues*, 49/3, p. 13-46.
- Adda-Decker, M., Snoeren, N. D. (2011), “Quantifying temporal speech reduction in French using forced speech alignment” *Journal of Phonetics*, 39, p. 261-270.
- Bartkova, K., Sorin, C. (1987), “A model of segmental duration for speech synthesis in French”, *Speech Communication*, 6/3, p. 245-260.
- Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., Rauzy, S. (2008), « Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle », *Traitement Automatique des Langues*, 49/3, p. 105-134.
- Bertrand, R., Espesser, R. (2017), “Co-narration in French conversation storytelling : A quantitative insight”, *Journal of Pragmatics*, 111, p. 33-53.
- Bigi, B. (2015), “SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech. *The Phonetician*”, 111-112, p. 54-69.
- Bigi, B., Meunier, C. (2018), “Automatic segmentation of spontaneous speech”, *Revista de Estudos da Linguagem, International Thematic Issue: Speech Segmentation*, 26/4, p. 1489-1530.
- Blache, P., Bertrand, R., Bigi, B., Bruno, E., Cela, E., Espesser, R., Ferré, G., Guardiola, M., Hirst, D., Magro, E.-P., Martin, J. C., Meunier, C., Others (2010), “Multimodal annotation of conversational data”, in Xue, N., Poesio, M. (eds), *Proceedings of the Fourth Linguistic Annotation Workshop*, p. 186-191.
- Blache, P., Bertrand, R., Ferré, G., Pallaud, B., Prévot, L., Rauzy, S. (2017), “The Corpus of Interactional Data : A Large Multimodal Annotated Resource”, in Ide, N., Pustejovsky, J. (eds), *Handbook of Linguistic Annotation*, Springer, p. 1323-1356.

- Bybee, J., File-Muriel, R. J., Napoleão De Souza, R. (2016), "Special reduction : A usage-based approach", *Language and Cognition*, 8/3, p. 421-446.
- Chafcouloff, M. (1983), « A propos des indices de distinction /1-r/ en français », *Speech Communication*, 2/2-3, p. 137-139.
- Greenberg, S., Carvey, H., Hitchcock, L., Chang, S. (2003), "Temporal properties of spontaneous speech – A syllable-centric perspective", *Journal of Phonetics*, 31/3-4, p. 465-485.
- Johnson, K. (2004), "Massive reduction in conversational American English", in Yoneyama, K., Maekawa, K. (eds), *Spontaneous speech: Data and analysis*, The National International Institute for Japanese Language, Tokyo, p. 29-54.
- Jurafsky, D., Bell, A., Gregory, M., Raymond, W. D. (2001), "Probabilistic relations between words : Evidence from reduction in lexical production", in Bybee, L., Hopper, P., *Frequency and the emergence of linguistic structure*, John Benjamins, p. 229-254.
- Lee, A., Kawahara, T., Shikano, K. (2001), "Julius – An open source real-time large vocabulary recognition engine", in Dalsgaard, P., Lindberg, B., Benner, H. (eds), *Proceedings European Conference on Speech Communication and Technology (EUROSPEECH)*, Aalborg University, p. 1691-1694.
- Meunier, C. (2012), « Contexte et nature des réalisations phonétiques en parole conversationnelle », Besacier, L., Lecouteux, B., Sérasset, G., *Actes de la conférence conjointe JEP-TALN-RECITAL*, volume 1: JEP, Université de Grenoble, p. 1-8.
- Meunier, C., Espesser, R. (2011), "Vowel reduction in conversational speech in French : The role of lexical factors", *Journal of Phonetics*, 39/3, p. 271-278.
- New, B., Pallier, C., Ferrand, L., Matos, R. (2001), « Une base de données lexicales du français contemporain sur internet : LEXIQUE », *L'Année Psychologique*, 101, p. 447-462.
- Pluymakers, M., Ernestus, M., Baayen, H. (2005), "Lexical frequency and acoustic reduction in spoken Dutch", *Journal of the Acoustical Society of America*, 118, p. 2561-2569.
- Prevot, L., Gorisch, J., Bertrand, R. (2016), "A CUP of COFEE: A Large Collection of Feedback Utterances Provided with Communicative Function Annotations", in Calzolari, N. et al. (eds), *Proceedings of the 10th Language Resources and Evaluation Conference*, European Language Resources Association.
- Rauzy, S., Montcheuil, G., Blache, P. (2014), "MarsaTag, a tagger for French written texts and speech transcriptions", *Second Asian Pacific Corpus Linguistics Conference*, Hong Kong, p.220-220.
- Torreira, F., Adda-Decker, M., Ernestus, M. (2010), "The Nijmegen Corpus of Casual French", *Speech Communication*, 52/3, p. 201-212.
- Torreira, F., Ernestus, M. (2011), "Vowel elision in casual French : The case of vowel /e/ in the word *c'était*", *Journal of Phonetics*, 39/1, p. 50-58.
- Wells, J. C. (1997), "SAMPA computer readable phonetic alphabet", in Gibbon, D., Moore, R., Winski, R. (eds), *Handbook of Standards and Resources for Spoken Language Systems*, Mouton de Gruyter, Berlin & New York.
- Wu, Y., Adda-Decker, M. (2020), « Réduction temporelle en français spontané : Où se cache-t-elle ? Une étude des segments, des mots et séquences de mots fréquemment réduits », in Benzitoun, C. et al., *6e conférence conjointe JEP-TELN-RECITAL. Volume 1 : Journées d'Études sur la Parole*, ATALA et AFCP Publisher, p.627-635.