Actes de JEP-TALN-RECITAL 2024
Toulouse, France, 8 au 12 juillet 2024

Réductions temporelles en français parlé : Où peut-on trouver les zones de réduction ?

Yaru Wu[1,2,3] Kim Gerdes[2] Martine Adda-Decker[2,3]
(1) UR4255 CRISCO, Université de Caen Normandie, France
(2) LISN (CNRS), Université Paris-Saclay, 91405 Orsay cedex, France
(3) Laboratoire de Phonétique et Phonologie (UMR7018, CNRS-Sorbonne Nouvelle), France
yaru.wu@unicaen.fr, gerdes@lisn.fr, madda@limsi.fr

RÉSUMÉ

Cet article examine la réduction dans la parole continue en français, ainsi que les différents facteurs qui contribuent au phénomène, tels que le style de parole, le débit de parole, la catégorie de mots, la position du phone dans le mot et la position du mot dans les groupes syntaxiques. L'étude utilise trois corpus de parole continue en français, couvrant la parole journalistique formelle, la parole journalistique informelle et la parole familière. La méthode utilisée comprend l'alignement forcé et l'étiquetage automatique des zones de réduction. Les résultats suggèrent que la réduction de la parole est présente dans tous les styles de parole, mais moins fréquente dans la parole formelle, et que la réduction est plus susceptible d'être observée lorsque le débit de parole est plus élevé. La position médiane des mots ou des groupes syntaxiques a tendance à favoriser la réduction.

ABSTRACT

Temporal Reductions in Spoken French : where can we find the reduction zones ?

This article investigates speech reduction in continuous speech in French, as well as different factors that contribute to the phenomenon, such as speech style, speech rate, word category, phone position in the word, and location of words within syntactic groups. The study uses three corpora of continuous speech in French, covering formal, less formal, and casual speech. The method used involves forced alignment and the labeling of reduction zones. The results suggest that speech reduction is present in all speech styles but less common in formal speech and that reduction is more likely to be observed in speech utterances with a high speech rate. The middle position of either words or syntactic groups tends to favor reduction.

MOTS-CLÉS : variation, réduction, style de parole, découpage syntaxique, parole continue.

KEYWORDS: variation, reduction, speech style, syntactic chunking, continuous speech.

1 Introduction

La variation de la parole peut être observée de manière continue, en particulier dans la parole familière (Duez, 1997; Ernestus, 2000; Meunier & Espesser, 2011). Dans la communication quotidienne, les mots et les phrases ne sont pas toujours articulés comme on s'y attend dans leur forme canonique complète (c'est-à-dire la forme de référence). Dans la parole naturelle, les mots sont souvent hypo-articulés, ce qui donne des productions avec des segments affaiblis ou avec moins de segments que prévu. En effet, si nous comparons la réalisation phonétique des mots à leurs prononciations canoniques, nous pouvons trouver des parties de mots supprimées ou fusionnées dans le flux continu

Actes des 35èmes Journées d'Études sur la Parole, pages 153–162.

de la parole. En outre, les résidus articulatoires et les détails phonétiques peuvent souvent être retracés dans des séquences de mots réduites (Kohler, 1999; Hawkins, 2003), même lorsque des segments sont supprimés (Niebuhr & Kohler, 2011). Par exemple, Johnson (2004) a montré que la réduction massive était courante dans l'anglais américain conversationnel en parole continue. Schuppler *et al.* (2012) ont constaté que seulement 11,7% des occurrences totales de /t/ en fin de mot sont prononcées de manière canonique en néerlandais conversationnel. Un exemple typique de réduction temporelle de la parole, s'étendant sur plusieurs segments contigus dans la parole familière française, concerne la séquence de mots « je ne sais pas » (/ʒənəsɛpa/, 4 syllabes en français dans la forme canonique), souvent prononcée comme une séquence monosyllabique similaire à [ʃpa]. Afin de mieux comprendre la réduction de la parole en parole continue, nous étudions différents facteurs de variation contribuant à la réduction de la parole, à savoir le style de parole, le débit de parole, la catégorie de mots, la position du phone dans le mot et la position des mots dans les groupes syntaxiques.

La réduction est largement considérée comme un phénomène qui se produit dans la parole familière (voir par exemple Johnson, 2004; Pluymaekers *et al.*, 2005; Schuppler *et al.*, 2012). Cependant, ce phénomène peut également être observé dans des contextes plus formels (Adda-Decker & Lamel, 2017). Nous nous attendons à observer moins de réductions dans une parole plus formelle. En outre, on sait peu de choses sur la façon dont la réduction est distribuée, pour différents styles de parole, en fonction de différents facteurs linguistiques. Le débit de parole s'est avéré être un facteur important pour la réalisation/suppression des occlusives alvéolaires à l'intérieur des mots en anglais spontané dans Raymond *et al.* (2006). Dans cet article, nous souhaitons étudier l'influence du débit de parole sur la réduction de la parole en français continu, couvrant la parole journalistique formelle, la parole journalistique informelle et la parole familière. Outre les deux facteurs de variation précédents, nous avons également ajouté à notre étude des analyses sur l'influence de la catégorie de mots, de la position du phone dans le mot et de la position des mots au sein des groupes syntaxiques, facteurs peu étudiés dans la littérature. Pour ce faire, nous utilisons le découpage syntaxique pour regrouper des suites de mots et étudier les phénomènes de réduction. De plus, Gendrot *et al.* (2016) ont montré que le découpage syntaxique pouvait aider à détecter automatiquement la hiérarchie prosodique dans un corpus de parole journalistique. Nos analyses sur les groupes syntaxiques correspondent donc dans une large mesure aux analyses sur les groupes prosodiques.

Dans ce qui suit, le corpus, l'annotation de la réduction et le traitement des données sont décrits dans la section 2. Les résultats sont présentés dans la section 3, avant les conclusions de la section 4.

2 Méthode

2.1 Corpus et alignement

Trois grands corpus de parole continue en français ont été utilisés pour notre étude, c'est-à-dire le corpus de parole journalistique formelle ESTER (Galliano *et al.*, 2006), le corpus de parole journalistique informelle ETAPE (Gravier *et al.*, 2012), et le corpus de parole familière NCCFr (Torreira *et al.*, 2010). Le corpus ESTER couvre environ 100 heures d'émissions radiophoniques en français. Le corpus ETAPE contient environ 42,5 heures de débats et de conversations à la radio et à la télévision. Le *Nijmegen Corpus of Casual French* (NCCFr) contient 35 heures de conversations informelles entre amis. Le système de transcription de la parole LISN (anciennement LIMSI, Gauvain *et al.*, 2002) a été utilisé pour segmenter automatiquement les données en mots et en phones, en mode d'alignement forcé. Avec cette méthode, la durée minimale d'un segment est de 30 ms (Adda-Decker & Lamel,

2000). L'association optimale des segments de parole avec une transcription phonémique (obtenue via un lexique de prononciation) a été choisie en fonction de la transcription du segment au niveau du mot et du modèle acoustique. Les transcriptions orthographiques ont été fournies au système de transcription en mode d'alignement forcé, ce qui nous a permis d'obtenir les frontières des phones et des mots. Le système a sélectionné automatiquement la prononciation la plus adaptée pour chaque mot à partir du dictionnaire de prononciation. Les pauses, les hésitations ou les respirations étaient également détectées automatiquement. Par la suite, nous nous référons aux pauses, hésitations ou respirations comme étant des pauses en général.

2.2 Localisation des zones de réduction

Au-delà de la réduction segmentale (par exemple, la réduction des voyelles) qui étudie la réalisation acoustique des segments affaiblis, nous définissons la réduction segmentale comme une réduction temporelle (ou contraction) de phones contigus, dont les durées restent inférieures à 40 ms chacun. Nous nous concentrons ici sur les séquences de segments qui sont réduites (i.e. temporellement courtes). La réduction temporelle est localisée grâce aux caractéristiques du système d'alignement forcé. Tout d'abord, l'alignement forcé fait correspondre les segments de parole les plus adaptés sur la base des modèles acoustiques et des variations de prononciation fournies dans le dictionnaire de prononciation. Deuxièmement, le système de transcription de la parole du LISN génère des phones d'une durée minimale de 30 ms, ce qui correspond à 3 cadres (Adda-Decker & Lamel, 2000). Lorsque des mots ou des séquences de mots sont produits avec une réduction (par exemple, des phones fusionnés ou supprimés), le système force toujours l'alignement de tous les phones présents dans la prononciation sélectionnée. Les variantes de prononciation comprennent généralement la voyelle schwa facultative et les consonnes de liaison. Pour des séquences de mots comme « par exemple » (/paʁɛgzɑ̃pl/) prononcées comme des séquences de phones similaires à [paʁɑ̃p], l'alignement mettra en évidence la réduction qui s'est produite dans la parole en forçant la présence de tous les segments ([paʁɛgzɑ̃pl]), mais en leur attribuant des durées très courtes. Par conséquent, le résultat de l'alignement forcé sera une séquence de segments d'une durée minimale de 30 (ou 40ms) afin de placer tous les phones de son modèle acoustique correspondant à la prononciation complète [paʁɛgzɑ̃pl].

Cette caractéristique spécifique du système d'alignement (c'est-à-dire la génération de segments courts consécutifs (30 ou 40 ms) pour les mots ou les séquences de mots réduits) peut être exploitée en tant qu'indicateur de réduction. Nous considérons qu'une séquence de plusieurs segments courts consécutifs (30 ou 40 ms) générée par l'alignement forcé est une zone de réduction, et plus il y a de segments dans la zone donnée, plus la réduction est considérée comme importante. Afin de mieux comprendre la réduction et le degré de réduction, nous avons décidé de diviser la réduction en quatre catégories. Les segments courts au sein d'une séquence de 3 segments courts consécutifs ou plus, de 30 ou 40 ms, sont étiquetés « 3 ». De même, les segments courts au sein d'une séquence de 2 segments courts consécutifs de 30 ou 40 ms sont étiquetés « 2 » ; les segments courts de 30 ms (c'est-à-dire la durée minimale autorisée par le système) qui ne sont pas précédés ou suivis d'un autre segment court de 30 ou 40 ms sont étiquetés « 1 ». Tous les autres phones sont étiquetés « 0 ». Le degré de réduction est donc composé de quatre catégories, à savoir « 0 », « 1 », « 2 » et « 3 ». Plus le chiffre est élevé, plus la zone de réduction est importante. Cette méthode ascendante permet d'identifier les mots et les séquences de mots qui sont réduits et de localiser les zones de réduction. Par exemple, le mot « ministre » en français (/ministʁ/) peut être réalisé sous forme de séquences phonétiques similaires à [miz] dans des mots fortement réduits. Avec l'alignement forcé, le système est « forcé » d'aligner les séquences entières ([ministʁ]) et nous obtenons 3 segments courts consécutifs ou plus de 30 ou 40 ms (c'est-à-dire le degré « 3 »). Ensuite, nous avons regroupé les trois niveaux

de réduction présentés dans cette section (c'est-à-dire « 1 », « 2 », « 3 ») dans un groupe nommé « Réduit » et « 0 » est appelé « Normal » (c'est-à-dire sans réduction). Le taux de réduction est donc la proportion de segments réduits (c'est-à-dire 1, 2, 3) par rapport à l'ensemble des points de données (c'est-à-dire 0, 1, 2, 3).

2.3 Préparation des données

Le débit de parole a été mesuré pour chaque énoncé entre les pauses (unités interpausales) et exclut donc les pauses. Le débit de parole est donc de 0 pour les pauses. Le calcul du débit basé sur les unités interpausales nous permet de vérifier si la réduction se produit davantage dans les énoncés interpausaux ayant un débit de parole plus élevé. Les mots ont également été étiquetés automatiquement en fonction de leurs catégories grammaticales de manière traditionnelle, sur la base d'un dictionnaire français des formes fléchies (Lefff, Sagot, 2010). Les mots ambigus obtiennent des étiquettes multiples, utilisées comme telles dans les règles de combinaison. Prenons par exemple le segment ambigu « les portions », qui peut être interprété comme un déterminant/nom ('les portions') ou comme un pronom/verbe ('(nous) les portions'). Le segment est annoté comme $det - pron\ noun - verb$ et regroupé sans réellement résoudre l'ambiguïté. Les cas d'ambiguïtés dont la résolution aboutirait en fait à un découpage différent sont suffisamment rares pour être ignorés. Le marquage de la partie du discours basé sur Lefff a permis de regrouper les tokens en trois catégories plus larges : 'gram' pour les mots grammaticaux, 'lex' pour les mots lexicaux et 'ponct' pour la ponctuation. [1]

La position du phone dans le mot a également été étudiée dans cette étude. Nous distinguons les phones situés au début des mots (« Deb ») des phones situés au milieu (« Mil ») ou à la fin (« Fin ») du mot. Les mots ne contenant qu'un seul phone ont été étiquetés « 1Ph ». Selon Gendrot *et al.* (2016), il est utile de détecter automatiquement la hiérarchie prosodique à travers le découpage syntaxique en utilisant un corpus de parole journalistique. Dans cet article, nous utilisons un système similaire basé sur des règles pour le découpage syntaxique qui regroupe les mots de manière à maximiser les groupes de mots avec un nombre donné de syllabes.

Le nombre de syllabes pour chaque groupe syntaxique a été calculé sur la base de la transcription et d'un système basé sur des règles, comptant les schwas finaux des mots comme une demi-syllabe. Nous avons exécuté de manière itérative l'algorithme de découpage NLTK [2] basé sur des règles avec pour objectif d'obtenir un nombre maximum de syllabes fixé à 7 syllabes, conformément à la littérature (Gendrot *et al.*, 2016). Lorsque l'algorithme atteint ou dépasse l'objectif donné, c.-à-d. le nombre maximal de sept syllabes, le chunk n'est plus qualifié pour être combiné avec les mots voisins. Nous obtenons ainsi une annotation BILU pour chaque mot (et donc aussi pour chaque phone). [3]

3 Résultats

Dans cette section, nous présentons d'abord la distribution de la réduction temporelle pour les trois corpus. Ensuite, nous présentons l'impact du débit de parole, de la catégorie de mots, de la position du phone dans le mot, de la position du mot dans le groupe syntaxique et du style de parole. En ce

1. Les catégories grammaticales de Lefff sont cl, det, pre, pro, pri, coo, aux, et csu ; les catégories lexicales sont nc, np, adj, v, adv ; $ponct$, $parent$, $epsilon$, et b pour la ponctuation.
2. NLTK est une bibliothèque Python spécialement conçue pour le traitement naturel du langage.
3. « B » signifie « Begin », représentant les premiers mots du groupe syntaxique ; « I » signifie « In », indiquant que le mot est situé au milieu du groupe syntaxique ; « L »signifie « Last », représentant le dernier mot du groupe syntaxique ; « U » représente les groupes syntaxiques qui ne contiennent qu'un seul mot.

qui concerne les analyses statistiques, nous avons utilisé le modèle linéaire généralisé (Nelder & Wedderburn, 1972) dans R (R Development Core Team, 2019). Les facteurs mentionnés ci-dessus ont été inclus comme effets fixes dans le modèle. En outre, nous avons inclus le nombre de syllabes dans le mot et la fréquence de mots comme variables de contrôle.

3.1 Distribution de la réduction

Le taux global de réduction temporelle est présenté dans cette section. Le tableau 1 montre le taux de durée pour les phones de 30 ou 40 ms (c'est-à-dire des segments extrêmement courts, donc réduits) et pour les phones de 50 ms+ (c'est-à-dire 50 ms ou plus). Le taux de réduction temporelle le plus élevé est observé pour le corpus de parole familière (∼30%), suivi par le corpus journalistique informel ETAPE (22%) et le corpus journalistique formel ESTER (18%). En d'autres termes, moins le style de parole est formel, plus la réduction est présente dans la parole.

	Durée : 30∼40ms	Durée : 50+ms
ESTER	18%	82%
ETAPE	22%	78%
NCCFr	30%	70%

TABLE 1 – Le taux de durée pour les phones (1) de 30 ou 40 ms et (2) de 50 ms ou plus.

Comme indiqué dans la section 2.2, les phones ont également obtenu une étiquette de réduction selon les conventions. Les trois corpus mis en commun, notre base de données comprend 7,6 millions de phones, dont 6,2 millions ne présentent aucune réduction. Plus précisément, 900 000 phones ont une réduction de degré « 1 », 360 000 de degré « 2 », et seulement 157 000 de degré « 3 », ce qui montre une distribution déséquilibrée des degrés de réduction.

Degré de réduction	**ESTER**		**ETAPE**		**NCCFr**	
	Occ.	%	Occ.	%	Occ.	%
0	3058667	82,1	2029469	77,67	893096	69,76
1	429280	11,52	304808	11,67	170410	13,31
2	142062	3,81	123430	4,72	93752	7,32
3	95337	**2,56**	155273	**5,94**	123066	**9,61**

TABLE 2 – Distribution de la réduction pour le corpus de parole journalistique formelle ESTER, le corpus de parole journalistique informelle ETAPE et le corpus de parole familière NCCFr.

Le tableau 2 présente la distribution de la réduction pour différents styles de parole, à savoir la parole journalistique formelle ESTER, la parole journalistique informelle ETAPE et la parole familière NCCFr. La parole familière (NCCFr) présente un taux de réduction de degré « 3 » (en gras) plus élevé que la parole journalistique formelle (ESTER) et la parole journalistique informelle (ETAPE) : ESTER 2,56 % vs ETAPE 5,94 % vs NCCFr 9,61 %. Les mêmes tendances sont observées pour les degrés « 1 » et « 2 » en ce qui concerne les styles de parole.

Il est intéressant de noter que les taux de réduction des segments étiquetés « 1 » restent relativement constants à travers les styles, alors qu'une augmentation importante peut être observée pour les segments étiquetés « 2 » et surtout pour ceux étiquetés « 3 » lorsque l'on passe de la parole formelle (ESTER) à la parole familière (NCCFr). Cela suggère que la réduction tend à impliquer des séquences plus longues plutôt que de multiples segments d'un seul phone pour une parole moins formelle.

La figure 1 illustre la distribution des durées pour les trois styles de parole (de gauche à droite : (1) la parole journalistique formelle ESTER, (2) la parole journalistique informelle ETAPE, (3) la parole familière NCCFr). La durée des phones en millisecondes (ms) est indiquée sur l'axe des

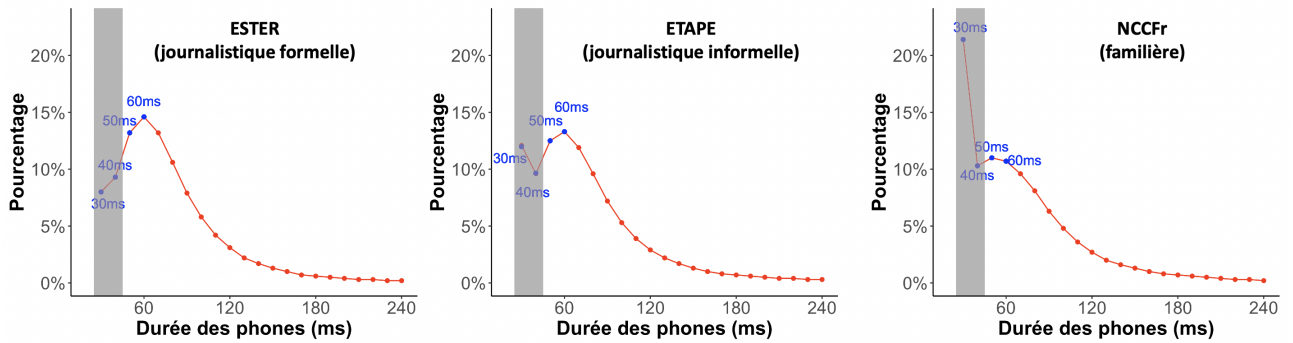

FIGURE 1 – Distribution de la durée des segments dans les trois corpus (de gauche à droite : parole journalistique formelle ESTER, parole journalistique informelle ETAPE, parole familière NCCFr).

abscisses. Des schémas différents de distribution des durées des phones sont observés pour les trois styles de parole. Pour la parole journalistique formelle ESTER, la distribution des durées de phones correspond à une courbe globale en forme de cloche et le pic de la courbe est situé à 60 ms (voir l'abscisse). Cela suggère que la durée la plus fréquemment observée dans la parole journalistique formelle est de 60 ms. Pour la parole journalistique informelle ETAPE, la proportion de la durée minimale autorisée par le système (30 ms) a augmenté de 5%. En ce qui concerne la parole familière NCCFr, le pic de la courbe correspond à la durée minimale de 30 ms, ce qui suggère que la durée la plus fréquemment observée ici est de 30 ms (>20 %). Ces observations sont cohérentes avec les résultats trouvés dans Adda-Decker & Lamel (2017) sur la durée des phones dans la parole préparée et spontanée (en français et en anglais).

Nos analyses sur la distribution de la durée nous permettent de mieux comprendre la spontanéité des trois types de parole et donc le phénomène de réduction dans la production de la parole. Dans ce qui suit, nous nous concentrerons sur ces zones grises et nous étudierons les facteurs de variation qui conditionnent la réduction.

La figure 2 présente les résultats sur le débit de parole. Les figures 3 à 5 illustrent le taux de réduction en fonction des catégories de mots, de la position du segment dans le mot et de la position du mot dans le groupe syntaxique, pour les trois styles de parole. Pour chaque figure, les résultats sur la parole journalistique formelle (ESTER) sont présentés sur le panneau de gauche, suivis de la parole journalistique informelle (ETAPE) au milieu et de la parole familière (NCCFr) sur le panneau de droite.

3.2 Débit de parole

La figure 2 présente le taux de réduction en fonction du débit de parole pour les trois styles de parole, à savoir la parole journalistique formelle ESTER, la parole journalistique informelle ETAPE et la parole familière NCCFr. Les résultats montrent que la réduction de la parole (« Réduit ») est associée à un débit de parole plus élevé que celui observé dans « Normal ». Cette tendance est observée pour les trois corpus. Les résultats du GLM confirment qu'il est plus probable d'observer la réduction de parole avec un débit de parole plus élevé [log odds ratio = 0,0776, $|Z|$ = 383,27, p<0,001].

3.3 Catégorie de mots

Le taux de réduction en fonction des catégories de mots (c.-à-d. les mots grammaticaux par rapport aux mots lexicaux) est présenté dans la figure 3. Les mots grammaticaux (« gram ») sont plus réduits que

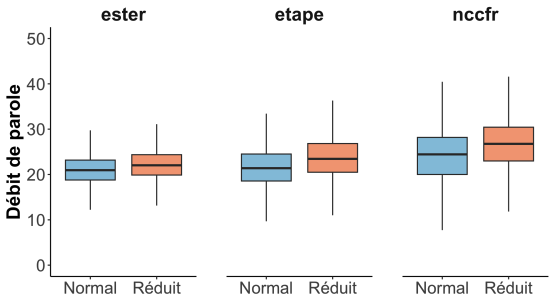

FIGURE 2 – Réduction en fonction du débit de parole pour les trois styles de parole (de gauche à droite : parole journalistique formelle ESTER, parole journalistique informelle ETAPE, parole familière NCCFr).

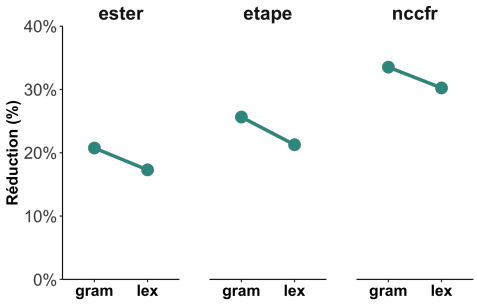

FIGURE 3 – Taux de réduction en fonction des catégories de mots (mots grammaticaux (gram) vs. mots lexicaux (lex)) pour les trois corpus (de gauche à droite : parole journalistique formelle ESTER, parole journalistique informelle ETAPE, parole familière NCCFr).

les mots lexicaux (« lex »). Ce résultat n'est pas surprenant, étant donné que les mots grammaticaux sont plus fréquemment utilisés dans la parole continue et donc plus enclins à la réduction. Les résultats du GLM confirment que la probabilité d'observer une réduction diminue significativement pour « lex » [log odds ratio = -0,1911, $|Z|$= 63,94, p<0,001], par rapport à celle observée pour « gram ».

3.4 Position du phone dans le mot

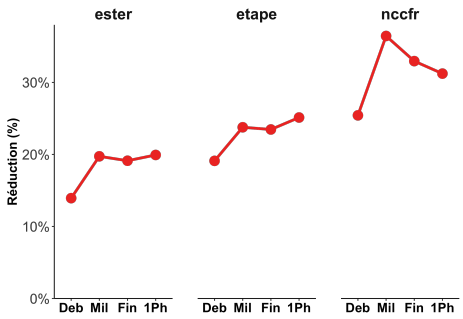

FIGURE 4 – Taux de réduction en fonction de la position du segment dans le mot (Début (« Deb ») vs. Milieu (« Mil ») vs. Fin (« Fin ») du mot vs. mot avec un phone (« 1ph »)) pour les trois corpus (de gauche à droite : parole journalistique formelle ESTER, parole journalistique informelle ETAPE, parole familière NCCFr).

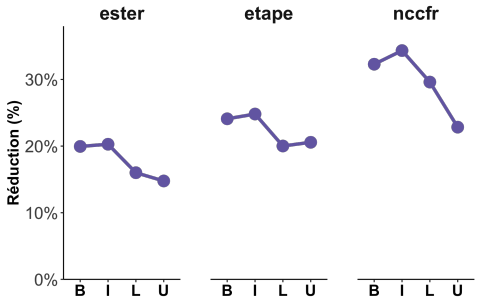

FIGURE 5 – Taux de réduction en fonction de la position du mot dans le groupe syntaxique : Début (B : *Begin*) vs. Milieu (I : *In*) vs. Fin (L : *Last*) du groupe syntaxique vs. groupe syntaxique contenant un seul mot (U : *Unique*), pour les trois corpus (de gauche à droite : parole journalistique formelle ESTER, parole journalistique informelle ETAPE, parole familière NCCFr).

La figure 4 montre le taux de réduction des phones en fonction de la position du phone dans le mot (« Deb », « Mil », « Fin », « 1Ph »). Les résultats montrent que, de manière générale, la position qui favorise le plus la réduction est « Mil ». Les résultats du GLM montrent que, tout en considérant d'autres facteurs de variation, la probabilité d'observer une réduction diminue significativement pour « Deb » [log odds ratio = -0,5401, $|Z|$ = 203,96, p<0, 001], pour « Fin » [log odds ratio = -0,1913,

$|Z|$ = 76,85, p<0,001] et pour « 1Ph » [log odds ratio = -0,4052, $|Z|$ = 97,31, p<0,001], par rapport à celle observée pour « Mil ».

3.5 Position du mot dans le groupe syntaxique

La figure 5 illustre le taux de réduction en fonction de la position du mot dans le groupe syntaxique. Il est intéressant de noter que, comme le montre la figure 4, la position qui favorise le plus la réduction est « I » (milieu). Plus précisément, les phones à l'intérieur des mots qui se situent au milieu d'un groupe syntaxique sont plus susceptibles de subir la réduction. Les résultats du GLM confirment que la probabilité d'observer la réduction diminue significativement pour « B » [log odds ratio = -0,0654, $|Z|$= 23,11, p<0,001], pour « L » [log odds ratio = -0,2532, $|Z|$= 108,25, p<0,001] et pour « U » [log odds ratio = -0,1747, $|Z|$= 49,35, p<0,001], par rapport à ce qui est observé pour « I ». Il convient de mentionner que le groupe « U » est composé de mots multisyllabiques tels que « aujourd'hui », « également », « effectivement », ainsi que de mots monosyllabiques tels que « et » et « tout ».

3.6 Style de parole

L'influence du style de parole sur la réduction peut être observée à partir des figures 2 à 5. Les résultats du GLM confirment que la probabilité d'observer une réduction augmente de manière significative pour la parole journalistique informelle [log odds ratio = 0,1882, $|Z|$= 89,59, p<0,001] et pour la parole familière [log odds ratio = 0,4298, $|Z|$= 159,98, p<0,001], par rapport à ce qui a été observé pour la parole journalistique formelle.

4 Conclusions

Cette étude porte sur la réduction de la parole et sur les facteurs susceptibles de favoriser la réduction en parole continue en français. Trois grands corpus de styles de parole différents (du formel au familier) ont été analysés. Les données de parole ont été automatiquement segmentées en mots et en phones, en mode d'alignement forcé. L'alignement automatique permet de localiser de manière innovante et efficace les zones de réduction de la parole. Les résultats montrent que le débit de parole, la catégorie de mots, la position du phone dans le mot, la position du mot dans le groupe syntaxique et le style de parole ont tous un impact significatif sur la réduction de la parole. Plus précisément, il est plus probable d'observer la réduction dans la parole produite avec un débit de parole élevé. En outre, il est plus probable d'observer la réduction dans les mots grammaticaux que dans les mots lexicaux. La position interne des mots ou des groupes syntaxiques favorise également la réduction. Enfin, le style de parole joue un rôle important en termes de réduction de la parole. Les contextes moins formels ont tendance à déclencher plus de réductions. En effet, la réduction peut prendre la forme d'une fusion, d'une suppression ou d'une recombinaison de segments ou de syllabes. Ces mécanismes peuvent potentiellement être liés à divers processus phonologiques, ce qui pourrait être intéressant sur le plan phonologique. Dans l'ensemble, cet article nous permet d'acquérir des connaissances sur la réduction et sur sa localisation en parole continue. En outre, notre recherche peut fournir des indications précieuses sur la manière dont les segments de la parole peuvent être réduits tout en restant compréhensibles pour les auditeurs.

Références

ADDA-DECKER M. & LAMEL L. (2000). The use of lexica in automatic speech recognition. In *Lexicon Development for Speech and Language Processing*, p. 235–266. Springer.

ADDA-DECKER M. & LAMEL L. (2017). Discovering speech reductions across speaking styles and languages. In *Rethinking reduction : Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*. Walter de Gruyter GmbH & Co KG.

DUEZ D. (1997). Acoustic markers of political power. *Journal of Psycholinguistic Research*, **26**(6), 641–654.

ERNESTUS M. T. C. (2000). *Voice assimilation and segment reduction in casual Dutch : A corpus-based study of the phonology-phonetics interface*. Thèse de doctorat, LOT, Utrecht.

GALLIANO S., GEOFFROIS E., GRAVIER G., BONASTRE J.-F., MOSTEFA D. & CHOUKRI K. (2006). Corpus description of the ester evaluation campaign for the rich transcription of french broadcast news. In *Proceedings of LREC*, volume 6, p. 315–320.

GAUVAIN J.-L., LAMEL L. & ADDA G. (2002). The limsi broadcast news transcription system. *Speech communication*, **37**(1), 89–108.

GENDROT C., GERDES K. & ADDA-DECKER M. (2016). Détection automatique d'une hiérarchie prosodique dans un corpus de parole journalistique. *Langue Francaise*, **191**, 123–147.

GRAVIER G., ADDA G., PAULSON N., CARRÉ M., GIRAUDEL A. & GALIBERT O. (2012). The etape corpus for the evaluation of speech-based tv content processing in the french language. In *LREC-Eighth international conference on Language Resources and Evaluation*.

HAWKINS S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of phonetics*, **31**(3-4), 373–405.

JOHNSON K. (2004). Massive reduction in conversational american english. In *Spontaneous speech : Data and analysis. Proceedings of the 1st session of the 10th international symposium*, p. 29–54 : Tokyo, Japan : The National International Institute for Japanese Language.

KOHLER K. J. (1999). Articulatory prosodies in german reduced speech. p. 89–92.

MEUNIER C. & ESPESSER R. (2011). Vowel reduction in conversational speech in french : The role of lexical factors. *Journal of Phonetics*, **39**(3), 271–278.

NELDER J. A. & WEDDERBURN R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society : Series A (General)*, **135**(3), 370–384.

NIEBUHR O. & KOHLER K. J. (2011). Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics*, **39**(3), 319–329.

PLUYMAEKERS M., ERNESTUS M. & BAAYEN R. H. (2005). Lexical frequency and acoustic reduction in spoken dutch. *The Journal of the Acoustical Society of America*, **118**(4), 2561–2569.

R DEVELOPMENT CORE TEAM (2019). *R : A La nguage and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RAYMOND W. D., DAUTRICOURT R. & HUME E. (2006). Word-internal/t, d/deletion in spontaneous speech : Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language variation and change*, **18**(1), 55–97.

SAGOT B. (2010). The lefff, a freely available and large-coverage morphological and syntactic lexicon for french. In *7th international conference on Language Resources and Evaluation (LREC 2010)*.

SCHUPPLER B., VAN DOMMELEN W. A., KOREMAN J. & ERNESTUS M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final/t : Studies at the phonemic and sub-phonemic level. *Journal of Phonetics*, **40**(4), 595–607.

TORREIRA F., ADDA-DECKER M. & ERNESTUS M. (2010). The nijmegen corpus of casual french. *Speech Communication*, **52**(3), 201–212.