Y a-t-il des géminées en français ? Une étude acoustique des mots en <iCC-> dans de grands corpus de parole

Adèle Jatteau[1,2*], *Jinyu* Li[1,3], et *Lori* Lamel[4]

[1]STL, UMR 8163, CNRS & Université de Lille, France
[2]Institut Universitaire de France
[3]LPP, UMR 7018, CNRS & Université Sorbonne Nouvelle, France
[4]LISN, UMR 9015, CNRS & Université Paris-Saclay, France

Abstract. Bien que la possibilité de produire une consonne longue dans des mots comme *irréel* ou *immense* soit mentionnée dans les grammaires et les dictionnaires du français, il n'existe pas à notre connaissance d'étude phonétique sur leur réalisation. Le présent travail explore la durée des consonnes dans les mots en <iCC-> à travers le rôle de l'orthographe (<iC-> vs. <iCC->), du style de parole et de la morphologie (mots complexes comme *irréel* vs. mots simples comme *immense*) dans de grands corpus de parole naturelle en français. Les résultats suggèrent que <mm> et <ll> sont associés à des durées plus longues <rr> et <nn>, et que les mots qui présentent des consonnes longues sont les mots préfixés et certains mots lexicalement marqués pour la consonne longue.

1 Introduction

Le français n'a pas de contraste de gémination : il ne peut pas distinguer deux mots uniquement par la durée d'une consonne, contrairement à d'autres langues comme l'italien (*papa* 'pape' vs. *pappa* 'bouillie'). Pourtant, les grammaires françaises rapportent la possibilité d'avoir des consonnes longues ou doubles dans plusieurs contextes, recensés dans le Tableau 1.

Le phénomène est suffisamment saillant pour être enregistré dans les dictionnaires : le Grand Robert [1] par exemple, qui inclut les catégories 2-3b-5-6-7 du Tableau 1, recense à peu près 1000 mots susceptibles d'être prononcés avec une géminée sur 20 000 mots contenant une consonne double (ou deux consonnes identiques séparées par un schwa).[i]

Ces catégories recouvrent des réalités diverses, que l'on peut regrouper en deux grandes classes : celles qui sont créées par la rencontre de deux consonnes sous-jacentes (catégories 1-2-3-4), et celles qu'on peut considérer comme des 'prononciations de lecture' (catégories 6-7-8). Alors que les premières sont considérées comme un fait de langue normal, les secondes sont marquées sociolinguistiquement, et globalement rejetées comme pédantes.

* Autrice correspondante : adele.jatteau@univ-lille.fr

Entre ces deux groupes, les consonnes longues qui peuvent apparaître à la frontière de préfixe (catégorie 5) ne se laissent pas facilement classer. Le présent travail s'intéresse à cette catégorie, qui n'a fait à notre connaissance l'objet d'aucune étude phonétique jusqu'à présent. L'objectif est de décrire l'usage d'un sous-ensemble de ces mots préfixés, les mots en <iCC-> (type *irréel, immense*), dans de grands corpus de parole naturelle. Avant de présenter les spécificités de cette catégorie dans la section 1.3, nous présentons un peu plus en détail les deux grandes classes dans les sections 1.1 et 1.2.

Tableau 1. Types de consonnes doubles en français.

1. Frontière de mot		*mer Rouge, frappe pas*
2. Mots composés		*sèche-cheveux, protège-genou*
3. Chute d'un schwa	a. entre deux mots b. dans le mot	*pas c'soir, pas d'drap* *là-d'dans, nett'té*
4. Frontière de suffixe	a. *-r-* du futur et conditionnel b. *-i-* [j] de l'imparfait	*acquerrai, courra, mourrais* *croyions*
5. Frontière de préfixe		*transsibérien, irréel, immigration, commisération, alléger, collaborer, syllabe*
6. Mots savants avec <CC>		*malléable, attique, appétence, reddition, additif*
7. Mots courants avec <CC>		*sommet, grammaire, collège, année, attendre, effet*
8. Sous l'emphase		*accablé*

1.1 Consonnes longues recouvrant deux consonnes sous-jacentes

Les seules catégories de consonnes doubles françaises qui ont fait l'objet d'études phonétiques sont, à notre connaissance, les consonnes à la frontière de mot (catégorie 1 dans le Tableau 1) et celles à la frontière de suffixe (catégorie 4a dans le Tableau 1)[ii]. Rialland (1994, [2]) propose également une analyse phonologique des catégories 1 et 3a. La catégorie 2, celle des mots composés, n'est mentionnée dans aucune des sources que nous avons consultées[iii]. Dans cet ensemble de catégories, la consonne a une origine phonologique : il y a deux consonnes sous-jacentes dans *mer Rouge* ou *nett(e)té*.

Les cas de C#C sont les plus étudiés. Dans les années 1990, un débat oppose Marchal & De Negro (1991, [3]) et Smith (1996, [4]) pour savoir si les consonnes à la frontière de mot en français fusionnent en une seule consonne longue (Marchal & De Negro 1991) ou bien sont constituées de deux consonnes enchaînées (Smith 1996). Marchal & De Negro (1991) proposent une étude palatographique et acoustique des occlusives sourdes et voisées (/t, d, k, g/) en contact à la frontière de mot, à partir de 36 phrases naturelles prononcées par 10 locuteurs en vitesse normale et rapide. Leurs résultats présentent une étonnante homogénéité : les consonnes identiques à la frontière C#C sont deux fois plus longues que les consonnes simples V#C (80ms vs. 160ms pour les voisées, plus de 90ms vs. 180ms pour les sourdes) ; ils observent un seul mouvement lingual dans C#C, et pas de différence dans

la durée de l'explosion ni dans la force de l'articulation entre C#C et V#C (contrairement à ce qui est observé pour les géminées contrastives lexicales d'autres langues, par exemple en italien, Burroni et al. 2024, [5]). L'étude conclut que C#C est prononcé comme une seule consonne longue. A l'opposé, Smith (1996) observe, à l'aide de l'articulatographe de l'Institut de la Communication Parlée de Grenoble et de mesures acoustiques, et à partir des données d'un seul locuteur, plusieurs indices suggérant que les C#C sont articulées comme deux consonnes distinctes : dans les séquences *le pape passait, le cap passait*, elle observe un relâchement de la première occlusion dans 2 lectures sur 6 (vs. 9/14 dans les séquences p#k, du type *le pape cassait*), un léger recul de la mandibule inférieure au milieu de la séquence dans 12 répétitions sur 15, et aucune différence dans la durée, vitesse ou amplitude d'élévation de la mandibule inférieure entre les séquences p#p et p#k, suggérant que le /p/ est articulé de la même manière dans les deux cas. Cette conclusion rejoint d'ailleurs l'intuition de Martinet (1971 : 188, [6]), pour qui le premier phonème dans ces séquences « doit à tout prix garder son indépendance phonologique ». Smith (1996) souligne toutefois que ces indices de deux gestes consonantiques s'affaiblissent avec la vitesse d'élocution ; on peut se demander si cela contribue à ses résultats divergents d'avec l'étude de Marchal & Del Negro, puisque les occlusives simples de son étude (dans *dis pas*) atteignent une durée de 116ms (contre moins de 100ms pour les /t, k/ de Marchal & Del Negro) et les doubles entre 194 et 229ms (contre 180ms chez Marchal & Del Negro), suggérant un débit légèrement plus lent dans son expérience.

Enfin, l'étude de Meisenburg (2006, [7]) a l'originalité d'inclure le contraste entre *il courait* et *il courrait* (catégorie 4a du Tableau 1) en plus des consonnes doubles à la frontière de mot dans *frappa* vs. *frappe pas* et *il a dit* vs. *il l'a dit*. Elle inclut également une expérience de perception pour savoir si les différences de durée produites sont exploitables ou non par les locuteurs. Ses résultats sur le contexte C#C confirment et nuancent ceux de Marchal & Del Negro (1991) et Smith (1996) : les séquences C#C sont effectivement plus longues que V#C, mais cela dépend des locuteurs (7 locuteurs sur 11 ont un /p/ plus long dans *frappe pas*, et 8/12 ont un /l/ plus long dans *il l'a dit*). De plus, tous les locuteurs qui produisaient une différence de durée se sont montrés sensibles à cette différence en perception.

Le cas de *courait/courrait* a un statut très différent. D'abord, nous ne sommes plus à la frontière de mot, mais à l'intérieur de mot, où les règles phonologiques peuvent être différentes. Dans ce contexte, *courais/courrais* et *mourais/mourrais* sont à notre connaissance les seules 'paires minimales' possibles reposant sur la longueur de la consonne. De plus, la rhotique a une histoire différente des autres consonnes en français. Alors que la majorité des consonnes géminées du latin sont perdues en proto-français autour du 7e s. (Zink 1986 : 154, [8], Scheer 2020 : 388-389, §294, [9]), /rr/ reste distinct de /r/ au sein des radicaux au moins jusqu'au 12e s., où des graphies non étymologiques commencent à apparaître dans le Nord. Sa simplification n'est généralisée qu'au 17e s., ce qui correspond également au passage d'un *r* apical à un *r* dorsal (Zink 1986 : 156, Scheer 2020 : 389, n. 4). Puisque Thurot ([1881-3] 2012 : 377, [10]) explique que le *r* long de *courrai, pourrai* est déjà signalé comme contrastif au 16e s., on peut se demander si ces /rr/ ne sont pas restés contrastifs tout au long de l'histoire du français.[iv] Fouché (1959 : 319, [12]) considère d'ailleurs que tous les <rr> graphiques doivent être prononcés « [ʀ·] (= [ʀ] légèrement plus long et plus fort) ». Quoi qu'il en soit, les résultats de Meisenburg (2006) suggèrent que ce contraste est possible en français contemporain, mais qu'il est plus fragile que celui des consonnes longues à la frontière de mot : dans son expérience, seuls 3 locuteurs sur 12 produisent une différence de durée entre le /r/ de *courait* et celui de *courrait* - et les mêmes 3 locuteurs sont capables de la percevoir.

Pour conclure, le français semble bien capable de produire une consonne longue (ou une suite de consonnes identiques) lorsque la séquence correspond à deux consonnes sous-jacentes. L'étude de Rialland (1994) montre par ailleurs que la formation de ces consonnes

obéit à des contraintes phonologiques : on ne forme pas de géminée C#C avec les consonnes qu'elle analyse comme extrasyllabiques, comme le /s/ initial de *sport* (**pas c'sport* vs. *pas c'soir*). La situation est beaucoup moins claire pour les catégories 6-7-8 du Tableau 1, où la consonne longue est optionnelle et liée à un marquage sociolinguistique particulier.

1.2 Les consonnes longues issues d'une prononciation de lecture

A l'autre bout du Tableau 1, les catégories 6-7 n'ont, à notre connaissance, jamais fait l'objet d'une étude phonétique dans la littérature récente.[v] La description la plus précise que nous ayons trouvée pour l'instant est celle du questionnaire de Martinet (1971) sur les mots *sonnez, sommet, addition* (les résultats sont détaillés ci-dessous). Ces consonnes longues sont en revanche fréquemment citées dans les grammaires, où elles sont souvent associées à un jugement prescriptif. Si, comme nous l'avons mentionné plus haut, la plupart des géminées héritées du latin se sont simplifiées avant l'ancien français, ces consonnes longues sont analysées comme des prononciations de lecture ou « effet Buben » (Chevrot et Malderez 1999, [12] – ce qui n'empêche pas qu'elles aient pu intégrer le lexique pour certains locuteurs). Pour plusieurs auteurs, elles sont présentées comme d'apparition récente en français, et en voie de développement au 20e s. (Bruneau 1927 : 52, [14], Carton 1974 : 216, [15]). Dauzat (1940 : 106, [16]) est même très précis sur la question, lorsqu'il affirme que « l'usage de doubler les consonnes, inconnu il y a un siècle, s'est beaucoup développé depuis quelques décades : il a commencé dans le Nord, et chez les éducateurs (en partie sous l'influence de la dictée) ». C'est effectivement chez les officiers originaires du Nord que Martinet (1971) trouve le plus de gémination en 1940-41.

Ces consonnes longues sont optionnelles en français, et leur production est souvent perçue, à l'image des liaisons interdites, comme des hypercorrections de mauvais goût. Dauzat (1940 : 106) représente bien ce jugement général, en voyant dans ces géminées un « abus » qui « se manifeste chez les personnes à demi-instruites, désireuses de montrer qu'elles connaissent l'orthographe ».

Les grammaires distinguent cependant les mots étrangers (notamment Fouché 1959) ou savants, comme *malléable, grammatical, collégiale, appétence, reddition, attique, décennal*, d'une part, et les mots « fréquemment employés » comme *sommet, sommaire, grammaire, nommer, appas, collège, année, attendre, accabler, aggraver, effet* d'autre part (Straka 1952 : 17-18, [17]). Dans le premier cas, qui correspond à la « prononciation plus ou moins solennelle des discours, des sermons, des conférences » (Straka 1952 : 17-18), si la géminée est considérée comme « affectée » (Grevisse 2008 : 44, §36, [18]) ou « pédante » (Bruneau 1927 : 50), elle reste plus ou moins acceptable, même si la prononciation simple est toujours indiquée comme correcte. Dans le cas des mots courants, en revanche, la prononciation longue de la consonne est fortement récriée, et taxée de snobisme et de pédantisme (même chez Carton 1974, qui reste pourtant plutôt neutre). Le Grevisse, quant à lui, s'étonne poliment qu'on puisse « entendre si souvent en France » un /l/ long dans *Hollande*.

L'enquête de Martinet (1971) permet de donner une idée de la fréquence de quelques-unes de ces géminées en 1940-41 chez des membres masculins de la bourgeoisie française. Les sujets enquêtés déclarent prononcer un /n/ long dans *sonnez* pour 12% des non-méridionaux, et 13% des méridionaux. Les chiffres sont beaucoup plus importants pour *sommet* (45% des non-méridionaux, 40% des méridionaux) et pour *addition* (35% des non-méridionaux, 17% des méridionaux). On peut également signaler que Bruneau (1927 : 50) juge que [ll] est « assez commun », alors que les autres géminées sont exceptionnelles ; c'est aussi la liquide qui a la plus longue liste chez Fouché (1959). La liquide est également la mieux représentée dans les géminées optionnelles signalées par le Grand Robert, avec 590 mots recensés vs. 208 pour /m/, 176 pour /ʁ/ et seulement 39 pour /n/.

Plusieurs de ces auteurs signalent toutefois que ce type de prononciation est en développement ; dans les termes de Carton (1974 : 216), « [l]e développement des consonnes longues en français contemporain n'est-il pas l'indice d'un acheminement vers un stade ou elles auraient un statut phonologique ? ». On peut donc supposer qu'à la période qui nous intéresse (pour notre corpus, celle des années 2000-2010), les consonnes longues sont plus fréquentes, et que leur valeur sociolinguistique a évolué.

Enfin, la catégorie 8 concerne les consonnes longues qui sont prononcées sous l'emphase ou « dans le style affectif » (Carton 1974 : 216). Comme le note Carton avec humour, « l'immensité avec « deux » m, c'est plus vaste ! ». Cette catégorie appartient au même groupe que les 6 et 7 parce que la consonne longue vient d'une prononciation de lecture. Toutefois, son marquage stylistique est très différent, puisqu'il est potentiellement indépendant du registre formel ou non de l'énoncé. Nous y reviendrons dans les sections 4 et 5.

1.3 Les mots préfixés

La catégorie 5 du Tableau 1, celle des mots préfixés, constitue une zone intermédiaire et mixte entre les deux classes présentées en 1.1 et 1.2. Contrairement aux géminées purement graphiques, les mots préfixés sont dans les grammaires françaises soit associés aux mots savants, comme *collaboration, syllabe, syllogisme*, soit classés dans une catégorie spécifique, dans laquelle la géminée est « étymologique » et directement liée au « sens de la composition » (Bruneau 1927 : 51), ou bien sert à « insister sur la valeur sémantique d'un préfixe » (Carton 1974 : 216) – autrement dit, la consonne longue recouvrirait deux consonnes sous-jacentes. Walker (1975, [19]), qui tente de caractériser phonologiquement la strate savante du lexique français, considère la gémination comme un trait de cette strate, mais doit écarter les mots préfixés qui ne rentrent pas tous dans cette catégorie. A ces différents titres, les consonnes longues à la frontière de préfixe sont généralement considérées comme acceptables dans les grammaires.

Le problème est que le « sens de la composition » est très variable selon les mots et les préfixes. D'un côté, des préfixes comme *sur-* dans *surréaliste* ou *trans-* dans *transsibérien* gardent toujours la même consonne finale, et jouissent d'une certaine productivité. De l'autre, il faut avoir étudié le grec pour savoir que *syllabe* est étymologiquement un mot préfixé, et l'on peut douter de la possibilité pour un francophone contemporain d'y reconstruire une frontière morphologique. Or, de ce caractère composé ou non dépend largement l'origine de la consonne longue. Si l'on analyse le mot comme contenant un préfixe à consonne finale, comme *sur-*, alors on peut supposer que le /ʁ/ de *surréaliste* correspond à deux consonnes sous-jacentes /syʁ+ʁealist/, comme les catégories 1-2-3-4. Si, au contraire, le mot est analysé comme simple, alors l'éventuelle consonne longue de *syllabe* doit être à l'origine un « effet Buben », comme les catégories 6-7-8.

Les mots en <iCC-> se situent à leur tour au centre de cette zone mixte. Ils incluent des mots étymologiquement préfixés par le /in/ négatif, comme *irréel, immature*, ou par le /in/ locatif, comme pour *immigré, illustre*. L'enquête de Martinet (1971) montre que la consonne longue dans les mots à préfixe négatif *illogique* et *irrémédiable* est très bien acceptée par les locuteurs interrogés : entre 77 et 94% déclarent pouvoir la produire. Tranel (1976, [20]) propose d'analyser le préfixe en synchronie comme /in/. Devant sonante, la nasale finale subirait une assimilation régressive, donnant lieu à une géminée qui peut ensuite faire l'objet d'une règle de dégémination optionnelle. Cela revient à placer dans la grammaire synchronique le développement diachronique du préfixe. Cette analyse permet à Tranel de proposer une représentation unifiée du préfixe, qu'il soit réalisé [iC-] devant sonante, [in-] devant voyelle (ex. *inutile*) ou [ɛ̃] devant obstruante (*impossible*). Son objectif est alors de rendre compte de l'hétérogénéité des mots préfixés en /in-/ : ceux dont la sémantique est peu

transparente, comme *innombrable* (qui ne signifie pas exactement « non-nombrable »), sont préfixés dès le lexique ; ceux dont la sémantique est transparente, mais qui ne sont pas productifs, comme *irrégulier*, sont séparés par une frontière de morphème ; enfin, la catégorie des mots en *in*-Verbe-*able*, qui est pleinement productive mais qui ne connaît pas l'assimilation devant sonante (*inratable, inlavable*), est scindée par une frontière de mot.

On peut se demander si ce raisonnement s'étend aux mots préfixés de l'ancien /in/ locatif, dont le caractère compositionnel est généralement opaque en français contemporain. A l'exception peut-être de quelques paires comme *immerger/émerger, immigrer/émigrer*, on ne peut pas supposer une frontière morphologique dans *illusion, immunité*. Plus généralement, de nombreux mots en <iCC-> ne se laissent plus analyser en synchronie, y compris des mots qui contenaient étymologiquement le préfixe négatif (*immense, innocent*). Dans ces mots, l'éventuelle consonne longue semble plutôt d'origine orthographique. Dans ce cas, puisque les géminées d'origine graphique sont bien attestées en français, est-il légitime de supposer deux consonnes sous-jacentes même dans les mots de type *irréel* ? L'alternative est de renoncer à l'unité du préfixe négatif et de reconstruire dans ces mots un préfixe /i-/ simple, comme le fait par exemple Apothéloz (2003, [21]).

Deux hypothèses émergent. Dans la première, il y a bien deux consonnes sous-jacentes dans les mots comme *irréel*, *illégal*, dans lesquels une frontière morphologique active peut être raisonnablement posée. La consonne longue de ces mots relèverait donc de la classe examinée dans la section 1.1. Dans les mots morphologiquement simples (en tout cas non-préfixés), comme *illusion*, la consonne longue aurait par contre une origine graphique, et sa réalisation pourrait être encouragée par un mécanisme d'analogie sur un gabarit /iCC-/ fréquent dans la langue. On note, d'ailleurs, que la grande majorité des consonnes longues mentionnées dans la littérature se trouve entre les deux premières syllabes du mot, comme si un gabarit (C)VCC- gouvernait l'apparition des géminées potentielles. Si cette hypothèse est correcte, on devrait trouver plus de consonnes longues dans les mots dont la composition est « sensible » que dans les mots simples. Dans la deuxième hypothèse, toutes ces géminées sont purement d'origine graphique, à classer avec celles de la section 1.2. Dans ce cas, on ne devrait pas trouver d'asymétrie entre le taux de gémination de mots comme *illégal* et *illusion*.

1.4 Questions de recherche

Le présent travail se concentre sur les mots commençant graphiquement par <iCC->. Nous proposons de répondre aux questions suivantes :

Q1. Y a-t-il effectivement des consonnes longues dans ces mots en français contemporain, et à quelle fréquence ?

Q2. Sont-elles liées à un style de parole particulier (discours « savant ») ?

Q3. Observe-t-on des taux de gémination différents selon le type de frontière morphologique ?

Les variantes sociolinguistiques sont très difficiles à éliciter en laboratoire, surtout lorsqu'elles sont conscientisées par les locuteurs. Pour comprendre ce qui se passe dans le langage courant, notre étude se fonde sur de grands corpus de parole « naturelle », au sens où ils n'ont pas été enregistrés pour les besoins de la recherche en linguistique. Nous avons mené deux études sur ces corpus. Dans la première, nous observons l'effet de l'orthographe et du style de parole en comparant les mots en <iC-> et en <iCC-> dans trois corpus assimilés à des styles de parole différents. Dans la deuxième, nous nous concentrons sur les <iCC->, en faisant l'hypothèse que la gémination optionnelle est un phénomène lexical (certains mots peuvent avoir une géminée, d'autres non) et catégorique (une consonne est longue ou simple, comme en italien, avec peu de valeurs intermédiaires). Nous menons une étude lexicale destinée à relever quels types de mots préfixés présentent la gémination.

La section 2 présente les corpus et la méthodologie adoptée. L'étude 1 et l'étude 2 font l'objet respectivement des sections 3 et 4. La section 5 discute ces résultats et conclut.

2 Méthodologie

Trois corpus ont été utilisés pour cette étude : ESTER, ETAPE et NCCFr. Le corpus ESTER (Galliano et al. 2005, [22]) contient 90h de journaux d'information radiophoniques enregistrés entre 1998 et 2003. Le corpus ETAPE (Gravier et al. 2012, [23]) contient 13,5h d'émissions de radio et 29h d'émission de TV, pour un total d'environ 38h de parole, enregistrées en 2011-2012. Les émissions ont été sélectionnées pour inclure essentiellement du discours non planifié. Enfin, le corpus NCCFr (Nijmegen Corpus of Casual French, Torreira et al. 2010, [24]) comprend 35h d'échanges informels entre amis, essentiellement des étudiants d'une université parisienne, enregistré en 2008.

Ces trois corpus ont été alignés automatiquement avec le système de traitement automatique du laboratoire (Gauvain et al. 2002, [25]). Au sein de cet alignement, deux sous-corpus équilibrés de mots ont été constitués, de manière à tester l'effet de l'orthographe sur la prononciation. Le premier corpus contient les <iCC-> : tous les mots commençant graphiquement en *inn-* (115 occurrences), *irr-* (113 occurrences)*, ill-* (256 occurrences) ont été sélectionnés. Les mots en *imm-* étant beaucoup plus nombreux (721 occurrences), seul un sous-ensemble de 524 mots a été retenu (avec sélection aléatoire). Ce sous-corpus contient des composés en *in-* négatif, comme *immérité, innombrable, irrattrapable*, des composés en *in-* locatif, comme *immigré, innover*, dont la frontière morphologique en français contemporain peut être très claire (comme dans *immortel, irrégulier*) ou complètement obscure (*innocent, immonde*).

Le second sous-corpus contient les <iC-> : des mots commençant graphiquement par *im-, in-* et*, ir-* (la base de données ne contenait pas de mots en *il-*). Une sélection semi-aléatoire a été opérée parmi les <iC-> pour former un sous-corpus comparable à celui des <iCC-> : le but était d'avoir le même nombre de mots pour chaque consonne et chaque corpus (ESTER, ETAPE et NCCFr), et la même distribution du débit phonémique. Le débit phonémique a été estimé localement à partir d'une fenêtre glissante de trois mots (précédent, courant et suivant). Pour chaque occurrence, le débit a été calculé en divisant le nombre total de phonèmes de la fenêtre, moins le phonème cible, par la durée totale de la fenêtre, amputée de la durée de ce même phonème. Pour chaque corpus, le débit phonémique des <iCC-> a été segmenté en quartiles afin de définir des intervalles, du plus lent au plus rapide. Ensuite, un nombre équivalent d'<iC-> a été échantillonné dans les mêmes intervalles. Lorsque certains intervalles ne contenaient pas suffisamment de données, le déficit a été compensé par des intervalles plus fournis. Enfin, lorsque le nombre total d'<iC-> obtenus était inférieur à celui des <iCC->, des <iC-> restants ont été ajoutés aléatoirement afin d'atteindre la taille cible, c'est-à-dire la même taille que celle des <iCC->. L'ensemble des <iC-> contient essentiellement des mots morphologiquement simples (non-préfixés), comme *image, irakien.* Pour la consonne /n/, la plupart des mots (tous sauf les dérivés de *initier*) ont été à un moment plus ou moins ancien de leur histoire des composés : *inaugurer, inopiné, inédit* ; d'autres présentent un préfixe négatif reconstructible en français contemporain, comme *inhabituel, inaction, inutile*. L'ensemble de la sélection (<iC-> + <iCC->) contient 1760 mots.

Pour finir, la segmentation de l'ensemble de ces deux sous-corpus a été vérifiée manuellement par le premier auteur sous Praat (Boersma & Weenink 2017, [26]). Cette vérification a conduit à écarter 36 items, trop difficiles à segmenter en général à cause d'une superposition de parole. La sélection finale contient 1724 mots, dont la répartition par consonne est détaillée dans le Tableau 2a. La disparité entre les types d'<iCC-> appelle un commentaire : manifestement, dans notre corpus totalisant 160h de parole, les mots en *imm-* sont beaucoup plus fréquents (en occurrences) que les mots en *ill-*, eux-mêmes plus fréquents

que les mots en *inn-* et *irr-*. Pour comparer avec la fréquence de type, les mots <iCC-> ont été lemmatisés pour le nombre et le genre : '*illégal*' par exemple inclut les occurrences de *illégal, illégale*, *illégaux, illégales* ; *illustrer*, seul cas de verbe, inclut les formes conjuguées et les participes du verbe. La fréquence en type (nombre de lemmes par consonne) pour les <iCC-> est donnée dans le Tableau 2b.

Tableau 2. Corpus sélectionné : nombre de mots en <iC-> et <iCC-> présents dans le corpus (1a) et nombre de lemmes pour les <iCC-> (1b).

	a. Fréquence d'occurrence		**b.** Fréquence de type
Consonne	**<iC->**	**<iCC->**	**<iCC->**
im(m)-	517	520	34
in(n)-	108	113	8
ir(r)-	110	110	28
il(l)-	-	246	21

Pour tenir compte des différences de débit de parole, la durée normalisée de ces consonnes a été calculée en divisant leur durée par le débit phonémique local. Il faut souligner que la durée des segments autres que la sonante dans la séquence n'a pas été contrôlée manuellement : il s'agit des valeurs de l'alignement automatique.

3 Etude 1 : effet de l'orthographe et du style de parole

La première étude porte sur les consonnes du Tableau 1 représentées dans les deux sous-corpus, <iC-> et <iCC->, afin de tester l'effet de l'orthographe et du style de parole sur la longueur des consonnes. Elle inclut donc *im(m)-*, *in(n)-* et *ir(r)-*, pour un total de 1478 mots. L'hypothèse est que dans un style plus formel, les consonnes doubles à l'écrit seront plus souvent réalisées plus longues que leurs équivalents simples. Les trois corpus inclus dans l'étude sont utilisés comme proxys pour le style de parole : ESTER correspond à des journaux radio, délivrés par des locuteurs professionnels probablement à partir d'un script ; ETAPE contient de la parole non planifiée, mais produite par des locuteurs professionnels de radio et TV ; et NCCFr a été constitué de manière à cibler le style de parole le plus informel, par des conversations entre amis.

La durée normalisée des consonnes /m, n/ et /ʁ/ a été modélisée à l'aide d'un modèle linéaire mixte (package lmer [27] dans R) selon la formule suivante : durée normalisée ~ consonne * double_lettre * corpus + (1|locuteur) + (1|lemme). Le premier effet aléatoire permet de tenir compte de la variabilité inter-locuteur pour modéliser les différences globales de rythme entre individus. Le deuxième effet aléatoire fait l'hypothèse que la variation pourrait être d'origine lexicale, et que certains lemmes pourraient être associés à des consonnes plus longues que d'autres. L'effet de la consonne (/m, n/ ou /ʁ/), de la graphie et du corpus ont été placés en interaction dans l'hypothèse que l'effet de l'orthographe sur la longueur des consonnes est lié au style de parole distinct des trois corpus. Les prédictions marginales du modèle, calculées avec le package *emmeans* [28] ont ensuite été utilisées pour générer les moyennes et les intervalles de confiance à 95%. Les durées normalisées des /m, n, ʁ/ après /i/ sont représentées dans la Figure 1, et les résultats du modèle apparaissent dans le Tableau 3.

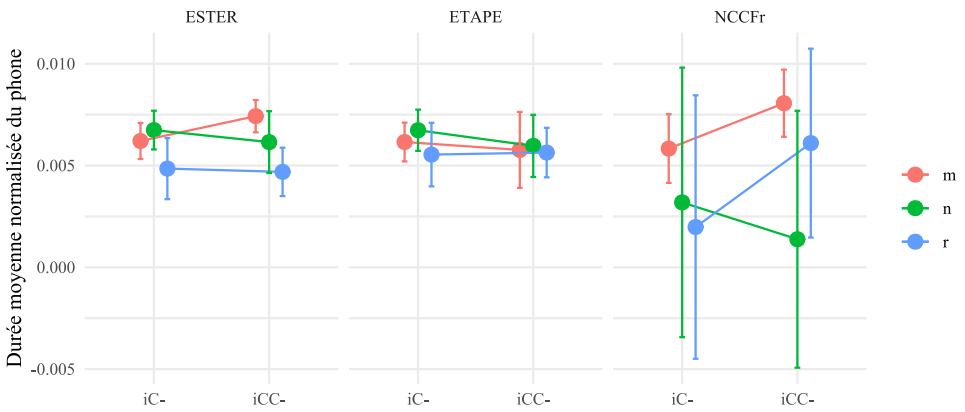

Fig. 1. Durées moyennes prédites par le modèle pour chaque consonne selon la présence d'une lettre double (<iC-> vs. <iCC->) et selon le corpus. Les barres verticales représentent les intervalles de confiance à 95%.

Tableau 3. Comparaison des durées normalisées des <iC-> et des <iCC-> pour chaque consonne et chaque corpus. La colonne "Effet" indique la différence moyenne normalisée, et "IC 95% bas/haut" correspondent aux intervalles de confiance à 95 %.

Corpus	Con-sonne	Effet (normalisé)	IC 95% bas	IC 95% haut	p-value
ESTER	m	0.00122	0.00003	0.00241	0.044
	n	-0.00059	-0.00238	0.00120	0.515
	r	-0.00016	-0.00207	0.00175	0.868
ETAPE	m	-0.00039	-0.00248	0.00170	0.716
	n	-0.00077	-0.00260	0.00106	0.408
	r	0.00010	-0.00188	0.00208	0.922
NCCFr	m	0.00223	0.00013	0.00433	0.038
	n	-0.00181	-0.01093	0.00732	0.697
	r	0.00412	-0.00382	0.01207	0.309

Comme le montre la Figure 1, les consonnes doubles à l'écrit ne correspondent à des consonnes légèrement plus longues à l'oral que dans une minorité de cas : /m/ dans ESTER, /ʁ/ dans ETAPE et /m, ʁ/ dans NCCFr. Dans les autres cas, la consonne double à l'écrit correspond de manière inattendue à une consonne légèrement plus brève à l'oral. Toutefois, le modèle statistique suggère que la plupart de ces effets sont très faibles et ne reflètent pas une différence consistante : seul /m/ dans ESTER et NCCFr est significativement produit plus long lorsqu'il correspond à une double consonne à l'écrit. Il faut toutefois nuancer ce résultat en tenant compte de la taille d'échantillon : le corpus contient beaucoup plus de *im(m)-* que de *in(n)-* et de *ir(r)-*, ce qui joue sur la précision des estimations et sur la significativité des effets.

La Figure 1 montre également une très grande variabilité des durées normalisées dans NCCFr. C'est lié au fait que le corpus est beaucoup plus petit : notre étude contient 1042 mots en provenance d'ESTER, 376 en provenance d'ETAPE, et seulement 60 de NCCFr. L'absence de significativité de l'effet sur /n/ et /ʁ/ dans NCCFr vient du fait qu'il y a très peu d'occurrences de *in(n)-* et *ir(r)-* dans ce corpus (5 en tout). De plus, nous verrons dans

l'étude 2 que l'effet sur /m/ dans ce corpus est largement liée à plusieurs occurrences très longues du /m/ du mot *immonde*. Les résultats pour ce corpus doivent donc être pris avec précaution.

Il est intéressant de noter qu'un précédent modèle sans effet aléatoire pour le lemme montrait les mêmes résultats (seul /m/ dans ESTER et NCCFr est statistiquement plus long lorsqu'il correspond à une double consonne graphique), mais avec des effets plus nets. Cela suggère que certains lemmes sont régulièrement prononcés avec une consonne longue, et d'autres non. La deuxième étude explore cette piste de recherche.

4 Etude 2 : effet du lexique et de la morphologie

La deuxième étude se concentre sur les mots orthographiés avec une consonne double, les <iCC->. L'hypothèse à l'origine de cette étude est que la possibilité de géminer une consonne est une propriété lexicale : certains mots l'autorisent, d'autres non. Dans l'intuition de la première autrice de cet article, locutrice native du français, il est possible d'avoir une consonne longue dans *immense* ou *irréaliste*, mais pas dans *innocent* ou *immunité*. La gémination serait donc une propriété lexicale, variable selon les locuteurs, qui pourraient formuler des jugements de grammaticalité sur sa réalisation. C'est du reste le parti pris implicite des dictionnaires, lorsqu'ils étiquettent certaines <CC> comme pouvant être prononcées longues et d'autres non. Dans cette perspective, la gémination en français est envisagée comme un phénomène catégorique : bien que la durée ne soit pas contrastive dans le système, la consonne est prononcée longue ou brève, avec peu de valeurs intermédiaires, comme dans les langues où la durée est contrastive.

Pour explorer ces hypothèses, tous les mots en <iCC-> ont été écoutés par le premier auteur et classés, de manière qualitative et subjective, dans la catégorie 'simple' ou 'géminé'. Cela inclut, en plus des chiffres donnés dans le Tableau 2 pour *imm-, inn-, irr*, 246 mots en *ill-*, du type *illettré, illusion, illicite*. Dans quelques cas, l'appartenance de la consonne à la catégorie géminée/simple a été jugée indécidable, et les exemples ont été écartés. Cela concerne 21 *imm-,* 2 *inn-*, 7 *ill-* et 6 *irr-*. Le nombre d'occurrences pour chaque consonne inclus dans cette seconde étude apparaît dans le Tableau 4.

Les durées moyennes normalisées des CC- classées comme 'simples' et comme 'géminées' sont représentées dans la Figure 2, et comparées avec celle des C-. Cela ne concerne donc que *im(m)-, in(n)-* et *ir(r)-*. Un test de Kruskal-Wallis (choisi pour gérer les différences de taille entre les groupes) montre que les trois catégories sont statistiquement distinctes. Les comparaisons post-hoc par tests de Wilcoxon indiquent que la différence entre les deux types de 'simples' (<iCC> classés comme « simples » et <iC->) est statistiquement significative mais de faible ampleur ($r = 0.16$), tandis que les 'géminées' se distinguent nettement des deux autres catégories, avec des tailles d'effet fortes (différence entre <iC-> et <iCC-> étiquetés comme 'géminés' : $r = 0.41$; différence entre <iCC-> étiquetés comme 'simples' et <iCC-> étiquetés comme 'géminés' : $r = 0.55$).

Si l'on se concentre maintenant sur les <iCC->, le Tableau 4 présente les taux de gémination de *imm-, inn-, ill-* et *irr-*. Ces chiffres confirment les résultats de la première étude : les mots en *imm-* présentent de forts taux de gémination comparés à *inn-*, *irr-*. L'inclusion de *ill-* montre toutefois que l'intuition de Bruneau (1927) était correcte : c'est la liquide qui présente le plus fort taux de gémination. Le taux de gémination semble corrélé à la fréquence d'occurrence (cf. Tableau 2a) : plus une consonne « géminable » est utilisée, et plus elle est géminée.

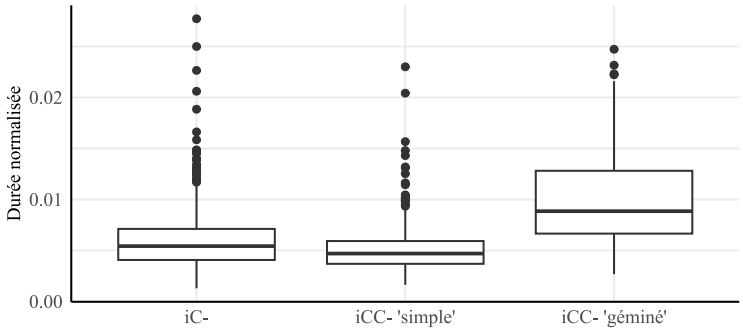

Fig. 2. Durées normalisées des consonnes <iC-> et leurs équivalents <iCC-> étiquetés comme ‘simples’ et comme ‘géminés’.

Tableau 4. Nombre d’occurrences et taux de gémination (= occurrences étiquetées comme ‘géminées’) par consonne.

	Nb. occurrences	% gémination
imm-	499	31%
inn-	111	3%
ill-	239	40%
irr-	104	12%

Le Tableau 5 permet de préciser cette estimation en examinant le taux de gémination des lemmes apparaissant plus de 10 fois dans le corpus.

Tableau 5. Nombre d’occurrences et taux de gémination (= occurrences étiquetées comme ‘géminées’) des lemmes <iCC-> apparaissant au moins 10 fois.

Lemme	Nb. total	% gém.
immense	38	55
immobilier	16	50
immonde	10	50
immédiat	73	48
immédiatement	69	48
immeuble	55	38
imminent	14	21
immigré	43	14
immunité	37	8
immatriculation	17	6
immigration	84	1

Lemme	Nb. total	% gém.
innombrables	10	30
innocence	30	0
innovation	29	0
innocent	21	0
illusion	29	62
illégal	49	45
illicite	12	42
illustre	25	40
illustrer	26	27
illustration	26	19
illimité	21	10
irresponsabilité	12	17
irrégulier	11	9
irresponsable	23	4

La gémination optionnelle est-elle une propriété lexicale ? Le Tableau 5 suggère que oui : la propension à géminer varie de 0% à 62% selon les lemmes. Si l'on avait affaire à un effet purement orthographique, on s'attendrait à une variation plus ou moins équivalente pour les différents lemmes. Par ailleurs, il est remarquable que les mots dérivés d'une même base ne connaissent pas tous le même taux de gémination : si *immédiat* et *immédiatement* peuvent géminer tous les deux à peu près une fois sur deux, *immigration* montre un taux presque nul à côté de *immigré*. C'est également ce qui ressort de l'écoute des fichiers sons : un même locuteur peut géminer *i*[ll]*usion* et ne pas géminer *i*[ll]*usoire* dans la même phrase.

Cet effet lexical est-il lié à la structure morphologique des mots ? Nous nous sommes interrogées, à la fin de la section 1.3, sur le statut spécial des mots à préfixe /in/ négatif, ou plus largement sur les mots dans lesquels une frontière morphologique peut être facilement reconstruite par les locuteurs du français contemporain. Force est de constater que peu de lemmes de ce type remontent dans notre filet : trois *ill-* (*illégal, illicite, illimité*), trois *irr-* (*irresponsabilité, irresponsable, irrégulier*) et un *inn-* (si l'on considère que *nombre* est reconnaissable dans *innombrable*, bien que le mot ne signifie pas « non-nombrable »). Dans cette liste, les trois *rr-* se détachent par des taux de gémination plus bas. On peut se demander toutefois s'il ne s'agit pas d'un cas particulier. La rhotique, en français comme dans d'autres langues, est connue pour la variabilité de ses réalisations (Gendrot et al. 2015, [29]). On pourrait imaginer que la « longueur » de /ʁ/ puisse être réinterprétée par un autre paramètre acoustique, comme une baisse du voisement ou plus de friction, et que la différence entre *ir**r**anien* et *ir**r**ationnel* ne passerait pas forcément par la durée (cf. aussi la citation de Fouché en §1.1, pour qui <rr> est plus long mais aussi « plus fort »). Cette hypothèse, suggérée par l'écoute des fichiers et l'examen des spectrogrammes, demanderait une étude spécifique. Si elle est correcte, on pourrait mettre de côté les *irr-* et conclure que les mots à préfixe négatif saillant présentent de forts taux de gémination, comme en ont l'intuition les grammaires du français.[vi] D'un autre côté, la plupart des lemmes du Tableau 5 présentent une frontière morphologique faible ou inexistante, et plusieurs de ces lemmes présentent également de forts taux de gémination, comme *immense* ou *illusion*. Ces résultats suggèrent que les mots dont la frontière morphologique est saillante sont privilégiés pour la gémination, et que les mots formés sur un gabarit similaire en <iCC-> peuvent être marqués lexicalement pour la gémination – auquel cas ils géminent tout autant que leurs « modèles » présumés.

Enfin, si l'on observe le taux de gémination dans chaque corpus (c'est-à-dire le taux de <iCC-> étiquetés comme 'géminés'), on trouve 31% de gémination dans ESTER, 19% dans ETAPE et 19% dans NCCFr. Dans ce dernier corpus, il n'y a en réalité que 7 'géminées', dont 5 viennent du mot *immonde* prononcé avec emphase. Dans ce corpus, il semble bien que la gémination existe essentiellement comme effet de style pour marquer une émotion (catégorie 8 dans le Tableau 1). Cette remarque met en évidence une limite importante de notre étude : nous n'avons pas pris en compte l'effet de l'emphase. Or, ce facteur entre en concurrence avec celui du style de parole que nous avons associé à chaque corpus : on peut parler avec emphase aussi bien dans un style formel que dans un style informel. A l'écoute des fichiers, notre intuition est que ce facteur joue un rôle non négligeable dans le discours journalistique, qui a tendance à marteler les mots lexicaux. Nous avons toutefois noté de nombreux passages dans lesquels cette « accentuation » ne s'accompagnait pas de gémination. Nos résultats sur le style de parole doivent donc être pris avec précaution, en attendant une étude plus complète.

5 Conclusion

Ce travail s'est intéressé à la description des consonnes longues en français contemporain, en se concentrant sur le cas particulier des mots en <iCC->. Un premier apport a consisté à clarifier la typologie des géminées possibles en français, et à passer en revue une littérature

plutôt maigre. Cet état des lieux nous a permis de mettre en évidence les spécificités de la sous-catégorie des mots en <iCC->, dont les membres oscillent entre des mots morphologiquement complexes ou morphologiquement simples (non-préfixés) en synchronie, et entre un jugement plus ou moins tolérant de la part des grammairiens. Une première étude a évalué l'effet de l'orthographe en comparant la durée de la consonne dans les mots en <iC-> et en <iCC->. Les résultats montrent que seul /m/ (sur un sous-ensemble /m, n, ʁ/) montre un effet significatif de l'orthographe sur la durée. L'étude 2 s'est concentrée sur une étude qualitative des mots en <iCC->, annotés par la première autrice de l'article comme 'simples' ou 'géminés'. Cette deuxième étude suggère que /l/ et /m/ sont plus souvent longs que /n/ et /ʁ/. L'étude lexicale montre également que les mots dans lesquels un préfixe est facilement reconstructible en synchronie montrent généralement de plus grandes durées, à l'exception des mots en *irr-*. Pour la rhotique, nous nous sommes demandé si la « longueur » de la consonne ne pouvait toutefois pas être réalisée par d'autres paramètres que la durée. Les mots dans lesquels la frontière de préfixe est faible ou inexistante, quant à eux, semblent être marqués lexicalement pour la gémination (la consonne CC peut être prononcée longue ou non). Enfin, l'effet du style de parole n'est pas clair. D'une part, nous ne trouvons un effet de l'orthographe sur /m/ que dans les corpus ESTER et NCCFr, qui correspondent respectivement aux styles de parole le plus formel et le plus informel. D'autre part, ce résultat est remis en cause par l'existence d'un autre facteur que nous n'avons pas pris en compte, l'emphase. S'il a été possible de le mettre en évidence dans NCCFr, du fait du faible nombre de consonnes longues observé, nous ne savons pas à quel point il motive les consonnes longues des deux autres corpus. La difficulté, sur ce point, est de trouver comment annoter l'emphase sur des critères objectifs.

Enfin, il faut signaler que plusieurs facteurs susceptibles d'interagir avec la durée des consonnes n'ont pas été pris en compte ici. La voyelle suivante, par exemple, n'a pas pu être équilibrée dans les différents corpus ; les mots en *inn-*, par exemple, sont presque tous des dérivés de *innocent* et *innover*, donc suivis d'un /o/, tandis que les mots en *in-* montrent des voyelles suivantes plus variées. Le nombre de syllabes dans le mot n'a pas non plus été pris en compte, alors qu'il peut affecter la durée des segments : plus le mot est long, plus les segments sont courts (Yang 1998, [30]). Il faut toutefois noter que la durée de la voyelle précédente et sa relation temporelle avec la durée de la consonne, le renforcement articulatoire de la consonne et différents degrés de préfixation sont examinés dans une étude parallèle des mêmes autrices, Li, Jatteau & Lamel (à paraître, [31]). Ces paramètres sont loin d'épuiser la question, et cette première étude sur la durée des consonnes à la frontière de préfixe dessine de nombreuses pistes pour des recherches ultérieures.

Références

1. Editions Le Robert (s.d.). *Le Grand Robert de la langue française* [Dictionnaire électronique]. https://www.lerobert.com/dictionnaires/francais/langue/dictionnaire-le-grand-robert-de-la-langue-francaise-edition-abonnes-3133099010289.html.
2. Rialland, A. (1994). The phonology and phonetics of extrasyllabicity in French. In P. A. Keating (éd.), *Phonological Structure and Phonetic Form* (136-159). Cambridge University Press.
3. Marchal, A., & Del Negro, A. (1991). Gémination phonétique en frontière de mots. *Proceedings of 12th Congress of Phonetic Sciences*, 438-441.
4. Smith, C. (1996). La production des consonnes doubles en français. *Travaux de l'Institut de phonétique de Strasbourg*, *26*, 145-156.

5. Burroni, F., Maspong, S., Benker, N., Hoole, P. A., & Kirby, J. (2024). Spatiotemporal features of bilabial geminate and singleton consonants in Italian. *ISSP 2024 - 13th International Seminar on Speech Production*, 181-184.
6. Martinet, A. (1971). *La prononciation du français contemporain*. Droz.
7. Meisenburg, T. (2006). Fake geminates in French: A production and perception study. In Hoffmann, Rüdiger & Mixdorff, Hansjörg (éds.), *Speech Prosody 2006*, 599–602.
8. Zink, G. (1986). *Phonétique historique du français*. Presses Universitaires de France.
9. Scheer, T. (2019). Consonnes en coda (_.C). In C. Marchello-Nizia, B. Combettes, S. Prévost, & T. Scheer (Éds.), *Grande grammaire historique du français* (387-398). Mouton de Gruyter.
10. Thurot, C. (1881, réimp. 2012). De la prononciation française depuis le commencement du seizième siècle, d'après le témoignage des grammairiens (Vol. 2). Slatkine.
11. Bradley, T. G. (2001). The phonetics and phonology of rhotic duration contrast and neutralization. Pennsylvania State University.
12. Fouché, P. (1959). *Traité de prononciation française* (2e éd.). Klincksieck.
13. Chevrot, J.-P., & Malderez, I. (1999). L'effet Buben : De la linguistique diachronique à l'approche cognitive (et retour). *Langue française*, *124*(1), 104-125.
14. Bruneau, C. (1927). *Manuel de phonétique*. Berger-Levrault.
15. Carton, F. (1974). *Introduction à la phonétique du français*. Bordas.
16. Dauzat, A. (1940). La prononciation des speakers à la radio. *Le français moderne*, 8, 105-108.
17. Straka, G. (1952). La prononciation parisienne. Ses divers aspects et ses traits généraux (Introduction à l'étude de la prononciation du français moderne). *Bulletin de la Faculté des Lettres de Strasbourg*, t. 30, vol. 5-6, 212-253.
18. Grevisse, M., & Goosse, A. (2008). *Le bon usage. Grammaire française* (14e éd.). De Boeck & Duculot.
19. Walker, D. C. (1975). Lexical stratification in French phonology. *Lingua*, *37*(2-3), 177-196.
20. Tranel, B. (1976). A Generative Treatment of the Prefix in- of Modern French. *Language*, *52*(2), 345-369.
21. Apothéloz, D. (2003). Le rôle de l'iconicité constructionnelle dans le fonctionnement du préfixe négatif in-. *Cahiers de Linguistique Analogique*, *1*, 35-63.
22. Galliano, S., Geoffrois, E., Mostefa, D., Choukri, K., Bonastre, J.-F., & Gravier, G. (2005). The ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast News. *Proceedings of ISCA Interspeech, Lisbon*, 1149-1152.
23. 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. *LREC - 8th international conference on Language Resources and Evaluation*, 3995-3999.
24. Torreira, F., Adda-Decker, M., & Ernestus, M. (2010). The Nijmegen corpus of casual French. *Speech Communication*, *10*(3), 201-212.
25. Gauvain, J.-L., Lamel, L., & Adda, G. (2002). The LIMSI broadcast news transcription system. *Speech Communication*, *37*(1-2), 89-108.
26. Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [Logiciel]. http://www.praat.org

27. Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1).
28. Lenth, R. V., & Piaskowski, J. (2017). *emmeans : Estimated Marginal Means, aka Least-Squares Means* (2.0.1) https://CRAN.R-project.org/package=emmeans
29. Gendrot, C., Kühnert, B., & Demolin, D. (2015). Aerodynamic, articulatory and acoustic realization of French /ʁ/. *Proceedings of the 18th ICPhS*, Glasgow.
30. Yang, L.-C. (1998). *Contextual Effects on Syllable Duration*. 37-42.
31. Li, J., Jatteau, A. & Lamel, L. (à paraître). Effets de la complexité morpho-orthographique sur la structuration temporelle de la parole : le cas des pseudo-géminées en français. *JEP* 2026, Montpellier.

[i] Nous remercions vivement Julien Lemaire (U. de Lille) pour l'extraction des données du Robert.
[ii] Mais cf. la note v pour des travaux en cours, ainsi que la référence à paraître citée en conclusion.
[iii] Nous n'avons malheureusement pas été en mesure de nous procurer un autre des rares travaux sur cette question : Abry, C., Orliaguet, J.-P., & Sock, R. (1990). Patterns of speech phasing. Their robustness in the production of a timed linguistic task: Single vs. Double (abutted) consonants in French. *European Bulletin of Cognitive Psychology*, 10(3), 269-288 (la revue est également appelée CPC, pour *Cahiers de Psychologie Cognitive* ou *Current Psychology of Cognition*).
[iv] Pas nécessairement avec le même statut : si l'opposition entre vibrante simple (ou battue) et vibrante longue était la seule opposition de longueur de la langue, on peut se demander si elle fonctionnait comme une opposition positionnelle (une consonne sous-jacente vs. deux consonnes) ou mélodique (deux consonnes contenant des traits différents), comme dans le débat sur la rhotique espagnole (ex. Bradley 2001, [11]). D'après Martinet (1971 : 196) et Thurot ([1881-3] 2012), c'est d'abord le /rr/ qui a basculé vers l'articulation uvulaire, comme en portugais brésilien contemporain : pendant un temps au moins, le contraste /r/-/rr/ a été réalisé par une différence mélodique.
[v] Une étude est toutefois en cours sur les géminées dans les discours d'E. Macron, menée par Anke Grutschus (Université de Bonn), Paolo Mairano (Université de Tours), Adèle Jatteau (Université de Lille) et Jinyu Li (Université de Lille & Sorbonne Nouvelle).
[vi] Si cette hypothèse est correcte, et que des /ʁ/ 'fortis' sont réalisés par d'autres paramètres acoustiques que la longueur, alors il est possible que fréquence de type, fréquence d'occurrence et taux de gémination soient corrélés (cf. Tableaux 2 et 4).