



HAL
open science

Alternances de voisement et processus de lénition et de fortition : une étude automatisée de grands corpus en cinq langues romanes

Ioana Vasilescu, Yaru Wu, Adèle Jatteau, Martine Adda-Decker, Lori Lamel

► **To cite this version:**

Ioana Vasilescu, Yaru Wu, Adèle Jatteau, Martine Adda-Decker, Lori Lamel. Alternances de voisement et processus de lénition et de fortition : une étude automatisée de grands corpus en cinq langues romanes. *Revue TAL : traitement automatique des langues*, 2020. hal-03040177

HAL Id: hal-03040177

<https://hal.science/hal-03040177v1>

Submitted on 4 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Alternances de voisement et processus de lénition et de fortition : une étude automatisée de grands corpus en cinq langues romanes

Ioana Vasilescu* — Yaru Wu* et ** — Adèle Jatteau*** — Martine Adda-Decker** — Lori Lamel*

* CNRS, LIMSI, Université Paris-Saclay

** Laboratoire de Phnétique et de Phonologie, Université Paris 3 Sorbonne Nouvelle

*** Université de Lille

RÉSUMÉ. Cette étude porte sur l'analyse des alternances de voisement en relation avec les processus de lénition et de fortition dans cinq langues romanes : le français, l'espagnol, l'italien, le portugais et le roumain. Nous mobilisons près de mille heures de données orales et nous faisons appel à des systèmes de transcription automatique spécifiques à chaque langue. La méthode consiste en l'alignement de variantes de prononciation non canoniques qui permet d'examiner la propension des occlusives sourdes canoniques /ptk/ à être réalisées en tant que segments sonores (lénition) et inversement, des occlusives sonores /bdg/ à être réalisées comme sourdes (fortition). Les résultats montrent que la position médiane est un site privilégié de la lénition, tandis que l'initiale de mot est propice à la fortition.

ABSTRACT. This study focuses on the analysis of voicing alternations as predictors of lenition and fortition phenomena, in five Romance languages : French, Spanish, Italian, Portuguese and Romanian. We use about 1k hours of oral data and language dependent speech recognition systems to explore voicing alternations. The method consists of the alignment of non canonical pronunciation variants that allows to examine the propensity of canonical voiceless stops /ptk/ to be realized as voiced segments (lenition), and of voiced stops /bdg/ to be realized as voiceless (fortition). Results show that word-internal stops are likely to undergo lenition, whereas word initial ones are rather prone to fortition.

MOTS-CLÉS : lénition, fortition, occlusives, voisement, langues romanes.

KEYWORDS: lenition, fortition, stops, voicing, Romance Languages.

1. Introduction

La parole continue représente un site privilégié des phénomènes de variation phonétique et permet d’observer de nombreux processus de réduction (Adda-Decker et Lamel, 2017). Ces processus se traduisent par une prononciation réduite de mots, où certains segments sont substitués par des variantes non canoniques ou complètement supprimés. Le débit et le style de parole, la fréquence lexicale, le contexte ou encore les propriétés morphologiques d’un mot sont autant de facteurs jouant un rôle dans l’occurrence et le type de motifs de réduction segmentale rencontrés (Ernestus, 2011 ; Meunier et Espesser, 2011 ; Adda-Decker et Snoeren, 2011). L’étude de ces phénomènes est révélatrice de l’articulation entre les niveaux phonétiques et les lois phonologiques régissant le fonctionnement d’une langue. Elle l’est aussi de l’articulation entre la synchronie et la diachronie, car la variation observée à un instant donné (variation synchronique) contient les précurseurs phonétiques des changements durables (variation diachronique) (Ohala, 1989). L’étude proposée ici s’inscrit dans cette mouvance : nous abordons les alternances de voisement des consonnes occlusives sourdes (p, t, k) et sonores (b, d, g) dans cinq langues romanes (français, espagnol, italien, portugais, roumain). Nous interrogeons ainsi la réalisation non canonique d’une consonne occlusive sourde prononcée en tant que sonore (par exemple, français *cape* prononcé *ca[b]e*) ou d’une consonne sonore prononcée en tant que sourde (par exemple, français *robe* prononcé *ro[p]e*) dans la parole continue. Ces phénomènes sont comparés aux processus historiques de lénition, qui fait référence à l’affaiblissement des consonnes et qui se traduit par la réalisation sonore d’une consonne sourde, et de fortition, associée, quant à elle, au renforcement consonantique et traduit par la réalisation sourde d’une consonne sonore. Les deux processus sont discutés en tant que variation synchronique observée dans des enregistrements oraux et ils sont comparés aux processus historiques qui ont régi l’émergence des langues romanes modernes à partir du latin. Nous prenons en compte la position des consonnes dans le mot, à savoir à l’initiale, à l’intérieur (médiane) et en position finale.

Dans le sillon de l’approche initiée par Ohala (1989), nous estimons la propension des cinq langues romanes à déployer les deux motifs de variation sonore qui font écho à des tendances historiques. En effet, les deux phénomènes retenus et en particulier la lénition, ont bénéficié d’une longue tradition linguistique, à la fois descriptive et expérimentale. Notre étude s’inscrit dans une approche expérimentale et se démarque des études précédentes : (i) par la taille des corpus, car nous faisons appel à des centaines d’heures d’enregistrements par langue, acquises dans le cadre de projets dédiés à la transcription automatique des langues, et par (ii) l’approche méthodologique, car nous traitons de manière comparable toutes les langues avec des techniques issues de la reconnaissance vocale. Les corpus utilisés ont été développés pour l’entraînement des systèmes de reconnaissance vocale et correspondent, pour la plupart, à des sources radiophoniques et télévisuelles. En explorant ce type de données nous faisons l’hypothèse d’une analyse des variétés standard des langues romanes, car les données de type médias s’adressent au plus grand nombre.

Les grands corpus et l'étude automatisée de la variation orale, reposant surtout sur des systèmes de reconnaissance vocale, apportent une dimension nouvelle à l'étude expérimentale des langues romanes. Cette méthode permet une exploration uniforme des données, car les systèmes de reconnaissance vocale, de par leurs architectures comparables, opèrent des choix concernant les qualités acoustiques des consonnes analysées sur la base d'un ensemble de traits homogènes. Ils permettent aussi d'ancrer une représentation statistique des phénomènes étudiés sur une grande quantité de données et, de ce fait, de proposer une estimation réaliste de leur portée dans la langue. En effet, même si les données de type médias s'appuient sur des variétés standard des langues, la diversité des locuteurs et des sources permet de faire des hypothèses sur l'impact des aspects diatopiques (géographiques) et sociolinguistiques (les caractéristiques des locuteurs) sur la réalisation des motifs sonores analysés. Dans une phase ultérieure, ces acquis nous permettront de formuler des hypothèses diachroniques ancrées dans la variation synchronique au sujet de la propension d'un motif de variation à devenir une mutation sonore durable. Ils s'intègrent aussi dans une démarche vertueuse, laquelle, *in fine*, peut bénéficier aux technologies de la parole. À titre d'exemple, les motifs de variation identifiés peuvent être intégrés dans l'architecture des systèmes notamment sous la forme de variantes de prononciation s'il s'avère qu'ils sont très fréquents dans la langue et qu'ils affectent la performance des systèmes (Vasilescu *et al.*, 2018).

Le travail est organisé de la manière suivante : les paragraphes 1.1 et 1.2 sont dédiés à la description des processus de changement de voisement dans les langues du monde et tout particulièrement dans les langues romanes. Nous dédions la section 2 aux données mobilisées pour cette étude et à la méthodologie automatique adoptée. La section 3 est dédiée à l'étude des consonnes occlusives dans les langues romanes et constitue un préambule aux expérimentations concernant les alternances de voisement. Les alternances de voisement en lien avec les phénomènes de lénition et de fortition dans les langues romanes sont ainsi discutées dans la section 4, laquelle est suivie par une dernière section 5, consacrée aux conclusions et aux perspectives.

1.1. *Lénition et fortition dans les langues*

La lénition et la fortition sont des processus fréquents dans les langues du monde, à la fois d'un point de vue diachronique, en tant que changements sonores, et synchronique, en tant que motifs de variation phonétique propre à la parole continue.

La lénition est un processus de réduction segmentale, se traduisant, entre autres, par un affaiblissement consonantique sous l'influence d'un contexte vocalique environnant. Parmi les définitions de la lénition ayant reçu un consensus dans la littérature linguistique de tous bords, celle de Hyman (1975) décrit le processus de la manière suivante : on dit qu'un segment X est plus faible qu'un segment Y si Y passe par X dans son évolution vers une suppression totale. Si l'on prend comme exemple les consonnes occlusives qui font l'objet de ce travail, Y est l'occlusive sourde phonologique et X est sa réalisation sonore suite à la lénition (par exemple lat. *patrem* > es.

padre « père »). Hyman (1975) s'appuie lui-même sur la première définition « scientifique » du processus, proposée par Rudolf Thurneysen au début du vingtième siècle. Ce dernier, dans son *Handbuch des Altirischen* (« Manuel de vieil irlandais ») édité en 1909, définit le phénomène comme une mutation des consonnes qui concerne la réduction de l'énergie utilisée dans leur articulation (cité par Martinet (1952)).

La fortition est souvent décrite comme un processus contraire à la lénition qui a bénéficié de moins d'intérêt que le premier (Bybee et Easterday, 2019). Il est également moins consensuel, les cas de fortition sont plus débattus (Honeybone, 2008). La fortition fait référence à un renforcement consonantique se manifestant par une augmentation de la constriction orale : une consonne sonore tend à devenir sourde. Cependant, la fortition n'est pas uniquement le contraire de la lénition, même si les deux phénomènes partagent des traits communs (Lavoie, 2001). La fortition est un phénomène plus restrictif, car si la lénition peut concerner toutes les consonnes occlusives, la fortition concerne souvent le renforcement articuloire des glides qui peuvent devenir des consonnes fricatives, affriquées ou encore des occlusives (Bybee et Easterday, 2019). La lénition et la fortition partagent le conditionnement par de nombreux facteurs contextuels tels que le contexte segmental, la position dans le mot et la longueur de ce dernier, les patrons prosodiques spécifiques aux mots concernés, la fréquence lexicale, etc. (Gurevich, 2004 ; Bybee et Easterday, 2019 ; de Carvalho *et al.*, 2008 ; Honeybone, 2008).

La lénition en tant que processus plus fréquent que la fortition a fait l'objet d'une riche littérature linguistique. Ainsi, des approches formelles aux analyses fonctionnelles, en passant par des travaux en phonétique diachronique et synchronique, de nombreux auteurs se sont penchés sur les différents aspects linguistiques impliqués dans la chaîne de processus réunis sous le nom de lénition avec un intérêt tout particulier pour les langues romanes (Honeybone, 2008 ; Lavoie, 2001 ; Gurevich, 2004 ; de Carvalho *et al.*, 2008 ; Bybee et Easterday, 2019 ; Hualde *et al.*, 2011 ; Carrasco *et al.*, 2012 ; Hualde et Nadeu, 2011 ; Recasens, 2016 ; Hualde et Prieto, 2014 ; Eager, 2017 ; Katz, 2016). La comparaison de phénomènes de lénition dans des dizaines de langues menée par des typologistes comme Lavoie (2001), Gurevich (2004) et plus récemment Bybee et Easterday (2019) a montré que la lénition peut être mise en lien avec une diversité de phénomènes de réduction. Ces phénomènes se succèdent parfois l'un à l'autre, un processus de lénition établi pouvant être à l'origine d'une nouvelle mutation sonore, ce qui encourage Gurevich (2004) à proposer une vision « domino » de la lénition. La lénition serait une chaîne d'affaiblissements consonantiques successifs qui naissent dans la parole continue. Elle peut commencer avec une simplification de consonnes géminées, comme dans le cas du mot latin */koppa/* où les consonnes géminées sont réduites à une seule, le mot devenant *copa /kopa/* en italien et *coupe /kup/* en français. Une telle simplification peut entraîner des confusions lexicales et ouvrir la voie à un autre phénomène de lénition qui consiste en la sonorisation des consonnes sourdes afin d'éviter la confusion avec d'anciennes géminées (la sonorisation concerne le mot latin *vita* lorsqu'il est devenu *vida* en espagnol). La phase finale, illustrée par l'évolution du même mot en français, est tout naturellement la suppression (*vie /vi/*).

1.2. La lénition dans les langues romanes : mutations historiques et motifs de variation synchronique

La lénition des consonnes occlusives intervocaliques (par exemple, lat. *focus* > fr. feu, la consonne occlusive /k/ située entre les voyelles /o/ et /u/ disparaissant en français) représente un des phénomènes fondateurs dans l'histoire des langues romanes modernes (von Wartburg, 1950 ; Vaissière, 1996 ; de Carvalho *et al.*, 2008). En effet, l'une des classifications des langues romanes les plus connues proposée par Walther von Wartburg remonte aux années 1950 et sépare les membres de la famille latine selon une frontière géographique virtuelle, reliant les villes italiennes La Spezia et Rimini (von Wartburg, 1950). Au nord de cette ligne les langues romanes ont subi une lénition des occlusives sourdes intervocaliques (dans notre corpus, français, espagnol et portugais). Au sud, le contraste sourd/sonore dans cette position a été préservé (dans notre corpus, italien et roumain). Le processus est plus large et il a comme point de départ la simplification des géminées du latin (par exemple, lat. *vacca* « vache » > it. *vaca*, la consonne double /kk/ devenant simple /k/), ce qui entraîne dans une deuxième phase une confusion lexicale résultant de la perte de contraste (comme dans les exemples de la section précédente). La confusion s'instaure entre des mots possédant la série d'occlusives sourdes proprement dite et celle issue de la gémination. La spirantisation des dernières intervient comme un processus naturel de maintien de contraste phonologique (Martinet, 1955). Par la suite, en français les spirantes sont supprimées dans certaines positions à l'intérieur du mot, et ce, dès le Moyen Âge. Dans d'autres langues romanes occidentales la spirantisation devient phonologique comme en espagnol et en portugais (par exemple, es. *vida* « vie » se prononce /viða/ et non pas /vida/).

Lat. classique	Français	Espagnol	Italien	Portugais	Roumain
caput « tête »	chef	cabo	capo	cabo	cap
focus « feu »	feu	fuego	fuoco	fogo	foc
joco,-are « jouer »	jouer	jugar	giocare	joga	juca
vita « vie »	vie	vida	vita	vida	viața
copertus « couvert »	couvert	cubierto	coperto	coberto	acoperit

Tableau 1. Lénition et maintien des occlusives non voisées /p t k/ dans l'évolution du latin vers les langues romanes modernes

Le tableau 1 résume les mutations sonores ayant affecté les occlusives intervocaliques dans l'évolution du latin vers les langues romanes modernes. Si les exemples cités en latin correspondent à sa phase classique, les processus de lénition ont évolué pendant plusieurs siècles avant d'aboutir aux formes connues des langues romanes modernes. À titre d'exemple, en français, la spirantisation des occlusives sonores est datée du III^e siècle (lat. *nudum* > gallo-rom. * *nuðu* > ancien fr. *nu*), tandis que les occlusives sourdes seraient devenues sonores et ensuite spirantes entre le IV^e et le VI^e siècle (lat. *ripa* > gallo-rom. * *riba*, *riβa* > ancien fr. *rive*) (Richter, 1934). Le tableau 1

oppose l'italien et le roumain, d'un côté, au français, espagnol et portugais, de l'autre, les premières n'ayant pas subi de processus de lénition.

La lénition dans les langues romanes modernes continue à recevoir beaucoup d'attention aujourd'hui, en tant que processus synchronique permettant de faire un lien entre la phonétique et la phonologie. En effet, l'analyse des phénomènes de lénition permet de dresser un lien entre les lois phonologiques profondes ayant façonné une langue et la variation synchronique, donc de mettre en parallèle diachronie et synchronie, car, pour citer Hualde et Chitoran (2016), « ce qui est un processus actif à l'interface phonétique-phonologie dans une langue, peut refléter un changement sonore finalisé dans une autre langue ». Dans cette littérature, l'espagnol a bénéficié d'une attention particulière car le processus y reste très actif. On parle ici d'un « deuxième tour » de la lénition, le premier étant celui qui s'est produit lors de l'évolution du latin vers les différentes langues romanes (Oftedal, 1985). Selon Hualde *et al.* (2011) et Carrasco *et al.* (2012), la spirantisation des occlusives intervocaliques représente un phénomène d'allophonie spécifique à l'espagnol qui caractérise à la fois la position intralexicale (VCV) (par exemple, l'espagnol *cada* prononcé /caða/) et interlexicale (V#CV) (par exemple, la prononciation /la ðuða/ du syntagme *la duda*). Cette réalisation est systématique et concerne tous les styles de parole, de sorte qu'une réalisation en tant que consonne occlusive d'un segment consonantique en position intervocalique serait décodée par les auditeurs natifs comme emphatique ou anormale. L'espagnol n'est pas la seule langue romane occidentale affectée par la lénition synchronique. Une réalisation proche du mode d'articulation fricatif des occlusives intervocaliques est recensée comme phénomène de variation dans d'autres idiomes latins. Ainsi Recasens (2016) et Hualde et Prieto (2014) trouvent des indices de lénition des occlusives après fricatives et liquides en catalan. Hualde et Nadeu (2011) mènent une étude expérimentale portant sur la lénition en italien de Rome qui montre que ce type de réduction caractérise la variété analysée. Enfin, les alternances de voisement sont aussi examinées en français et mettent en valeur des contextes de lénition (Hallé et Adda-Decker, 2007a ; Hallé et Adda-Decker, 2007b ; Jatteau *et al.*, 2019a ; Jatteau *et al.*, 2019b). Pour ce qui est du roumain, la question est brièvement abordée par Hualde *et al.* (2015) en lien avec la notion de non-réalisation phonétique (*phonetic undershoot*), les auteurs comparant des contextes propices à la lénition à partir de données orales en espagnol et en roumain. Sur des exemples du corpus de thèse du troisième auteur, ils observent qu'il n'y a pas de voisement intervocalique en roumain, que la non-réalisation phonétique apparaît de manière marginale, mais qu'on observe plutôt une tendance à la suppression massive de groupes consonantiques et même de syllabes de type CV (Hualde *et al.*, 2015). Les consonnes sont souvent non identifiables sur les spectrogrammes, ce qui suggère qu'elles ne sont pas réalisées. Enfin, pour ce qui est des mesures physiques de la lénition, notons à l'instar de Eager (2017) qui dresse un bilan des paramètres acoustiques discutés dans la littérature, que la lénition est un processus contextuel. Par conséquent, la lénition est estimée à travers l'interaction avec les segments environnants. Les mesures acoustiques concernent la durée, l'intensité, le degré de voisement des occlusives concernées, mais aussi le rap-

port à l'accentuation, les propriétés acoustiques du contexte vocalique, la fréquence des mots, le débit et le style de parole.

2. Données et méthodologie

Cette étude s'appuie sur les présupposés théoriques de Ohala (1989) et de Greenberg (1969) selon lesquels les motifs de variation synchronique suivent des tendances historiques et contiennent les prémices des changements sonores durables. Nos travaux y apportent la dimension « très grands corpus » qui permet d'aborder deux dimensions théoriques associées, à savoir la relation entre l'étendue d'un motif de variation dans l'usage linguistique et sa « chance » de devenir un changement durable, et le lien entre la fréquence observée et des facteurs extralinguistiques comme la variation géographique et sociolinguistique. La première dimension concerne la fréquence d'occurrences d'un motif de variation en lien avec son statut dans la langue : un motif de variation est-il suffisamment fréquent pour qu'on puisse le considérer comme établi ? La seconde dimension s'intéresse à la manière dont la langue adopte le phénomène étudié : concerne-t-il un style de parole en particulier ou est-il largement répandu ?

La présente proposition s'inscrit parmi les approches automatisées qui font appel à des systèmes de reconnaissance vocale en tant qu'outils d'étude linguistique de la variation phonétique. Nous faisons appel à la méthode connue comme *alignement forcé de variantes de prononciation* qui permet de « simuler » un changement sonore à grande échelle. Ainsi, la méthode a fourni des résultats prometteurs dans l'exploration des alternances de voisement en français et en espagnol (Hallé et Adda-Decker, 2007a ; Hallé et Adda-Decker, 2007b ; Vasilescu *et al.*, 2018 ; Jatteau *et al.*, 2019a ; Jatteau *et al.*, 2019b). Hallé et Adda-Decker (2007a) et Hallé et Adda-Decker (2007b) se sont penchés sur les alternances de voisement en français en lien avec les processus d'assimilation consonantique intervenant à la frontière des mots. Les auteurs ont fait appel à un corpus journalistique segmenté automatiquement. Premièrement, ils ont calculé le taux de voisement en tant que pourcentage de trames voisées des occlusives et des fricatives présentes en fin de mot selon le contexte segmental droit (l'attaque du mot suivant). Ensuite, ils ont procédé à l'alignement forcé de variantes de prononciation par un système de transcription automatique du français paramétré pour encoder la prononciation non canonique alternative sourde ou sonore (ainsi, dans le cas de la lénition, dans un exemple comme *sept janvier* l'augmentation du taux de voisement de l'occlusive sourde /t/ suggère une réalisation sonore [d], par conséquent le système de transcription automatique aligne la séquence comme [sɛd#zãvjɛ]). Les résultats montrent que la frontière lexicale s'avère un site privilégié des processus d'assimilation de voisement : les consonnes analysées changent de caractéristiques acoustiques en fonction de l'attaque du mot suivant. Les études ultérieures poursuivent dans cette direction et s'intéressent au dévoisement final en tant que processus phonologique et à ses précurseurs phonétiques (Jatteau *et al.*, 2019a ; Jatteau *et al.*, 2019b). Ryant et Liberman (2016) et Vasilescu *et al.*

(2018) mènent des études comparables au sujet de la lénition en espagnol péninsulaire et d'Amérique latine.

Pour ce qui est des données, l'étude représente une exploitation écologique de grands corpus acquis dans le cadre de projets portant sur la reconnaissance vocale en différentes langues. Il s'agit pour la plupart de corpus ayant servi à l'entraînement des systèmes, bénéficiant d'un alignement automatique au niveau phonémique et permettant une identification automatique des contextes segmentaux pertinents pour l'étude. Par exploitation écologique, nous entendons ainsi une valorisation linguistique de ces données peu ou plus utilisées à visée technologique, dans l'esprit des initiatives telles que le corpus ESTER (Galliano *et al.*, 2006), qui a ouvert la voie à la validation d'hypothèses linguistiques à partir d'une quantité de données jusqu'alors inaccessibles aux phonéticiens et aux phonologues.

2.1. Données

Nous faisons appel à 981 heures de données verbales couvrant des variétés *a priori* standard des cinq langues romanes suivantes : français, espagnol, italien, portugais et roumain. Les caractéristiques des variétés disponibles correspondent au style de parole de type médias, en général destiné à une large audience et présentant des traits linguistiques partagés par le plus grand nombre. Le tableau 2 illustre des aspects quantitatifs des données utilisées. Les données sont majoritairement des journaux radiophoniques ou télévisuels couvrant à la fois de la parole (semi)préparée, des interviews et des débats. Cette sélection portée par les caractéristiques des données habituellement utilisées dans la reconnaissance vocale permet de formuler l'hypothèse de variétés standard illustrant les cinq langues romanes. Pour cette étude préliminaire, les données sont également utilisées indistinctement par rapport aux variétés géographiques, par exemple les données en espagnol et en portugais couvrent à la fois des variétés européennes et d'Amérique latine.

Langue	nbr. heures	mots (M)	mots uniques (k)	var/mot	var v/nv
Fr	176	2,4	55,7	2	6,8
Es	223	2,6	61,9	1	4,4
It	168	1,8	57,0	1	5,3
Po	114	1,0	40,0	1	3,7
Ro	300	3,6	48,0	1	3,7

Tableau 2. *Caractéristiques des données utilisées : langues, durée/corpus, nombre de mots (en millions M), nombre de mots uniques (en milliers k), variantes par mot dans le dictionnaire d'origine, variantes par mot permettant des alternances de voisement pour toutes les occlusives (moyennes)*

Les différents corpus utilisés dans l'étude viennent de plusieurs sources : ils ont été acquis auprès de « Linguistic Data Consortium » et d'ELRA, ou ils ont été développés dans le cadre des projets internationaux suivants : IST ALERT ¹, pour une partie des données en français et en portugais (da Silva *et al.*, 2011 ; da Silva *et al.*, 2013) ; IRST pour une partie des données en italien (Després *et al.*, 2013 ; Marcello *et al.*, 2000) ; OSEO Quaero (Lamel, 2012 ; Lamel *et al.*, 2011) pour toutes les langues ².

2.2. Méthodologie

La méthode consiste à prévoir des variantes de prononciation non canoniques dans les dictionnaires de prononciation et à autoriser les systèmes de reconnaissance vocale à les sélectionner lors du processus d'alignement forcé. Les variantes sont intégrées de manière automatique suivant la règle de l'alternance de voisement, à savoir que tout mot présentant des occlusives sourdes peut être aligné avec des occlusives sonores et réciproquement (un mot présentant des occlusives sonores peut être aligné avec des réalisations sourdes). Développés dans le cadre des projets mentionnés dans la partie 2.1, les systèmes sont entraînés sur le type de corpus mobilisés pour cette étude (Lamel et Gauvain, 2002). Ils sont comparables en termes d'architecture et ils couvrent toutes les langues retenues, à savoir le français (Gauvain *et al.*, 2005), l'espagnol (Vasilescu *et al.*, 2018), l'italien (Després *et al.*, 2013), le portugais (da Silva *et al.*, 2011 ; da Silva *et al.*, 2013) et le roumain (Vasilescu *et al.*, 2014). Les segmentations au niveau des mots et des phones ont été générées à l'aide de modèles acoustiques similaires pour toutes les langues, chacune étant estimée à l'aide de données d'entraînement annotées spécifiques à la langue et de dictionnaires de prononciation. Les modèles acoustiques sont indépendants du locuteur, ce sont des modèles monophoniques indépendants du contexte et de la position des mots. Chaque modèle de phone est un modèle de Markov caché (HMM) à densité continue de gauche à droite à trois états avec des mélanges de gaussiennes comprenant jusqu'à 32 gaussiennes par état, sauf pour le silence qui est modélisé par un seul état avec 256 gaussiennes. Le paramétrage acoustique est basé sur une représentation cepstrale (Hermansky, 1990) du signal vocal. L'indépendance vis-à-vis du locuteur est obtenue en estimant les paramètres du modèle sur de grands corpus de parole contenant des données provenant d'une grande population de locuteurs (plusieurs centaines de locuteurs). Concernant l'analyse de la parole, le signal acoustique est transformé en suite de vecteurs avec un vecteur toutes les 10 ms. Chaque vecteur est composé de 13 coefficients cepstraux augmentés de leurs dérivées de premier et de second ordre résultant dans des vecteurs de dimension 39. Les coefficients cepstraux encodent de manière compacte l'information spectrale locale, alors que les dérivées d'ordre 1 et 2 informent sur la dynamique d'évolution temporelle de ces caractéristiques spectrales dans une fenêtre temporelle d'environ 100 ms. Cette analyse a été utilisée pour toutes les langues et est utilisée au LIMSI depuis 1996 (Gauvain *et al.*, 2002). Les modèles acoustiques sont formés à

1. <https://cordis.europa.eu/project/id/IST-1999-10354/en>

2. www.quaero.org

partir de 58 à 248 heures de données vocales, provenant de diverses émissions d'information, de débats, de « talk-shows » et de podcasts. Le nombre de phonèmes varie de 27 pour l'espagnol à 35 pour le portugais. Des unités supplémentaires sont utilisées pour représenter le silence, la respiration et les pauses remplies. À l'exception du français, pour lequel nous disposons d'un important dictionnaire de prononciation interne, les prononciations des mots sont générées à l'aide de règles de conversion des lettres en sons, complétées par quelques dictionnaires d'exceptions spécifiques à la langue. Les données utilisées pour entraîner les modèles acoustiques sont résumées dans le tableau 3.

Données	Français	Espagnol	Italien	Portugais	Roumain
Heures	248	139	71	77	58
Phones	34 + 3	29 + 1	27 + 3	35 + 4	29 + 1

Tableau 3. Données d'entraînement par langue : heures et nombre de phones incluant des unités non verbales (+ n)

Les tendances à la lénition et à la fortition sont étudiées en tant qu'alternances de voisement voisé → non-voisé (fortition) et non-voisé → voisé (lénition). La figure 1 montre les choix que le système peut faire afin d'aligner des variantes voisées et non-voisées compte tenu de leurs particularités acoustiques. Les modèles acoustiques sont spécifiques à la langue et ils ont été préalablement entraînés sur des données réelles. Par exemple, dans le cas d'un syntagme comme *sept janvier* en français le système peut opérer un choix entre les transcriptions [sɛt], [sɛd], [zɛt] ou [zɛd] pour le mot *sept*. À chaque occurrence du mot, le système compare les consonnes concernées par l'alternance de voisement aux modèles acoustiques qui leur sont associés. Ainsi, ces consonnes peuvent être substituées par leurs réalisations non canoniques si le modèle correspond plus à de telles réalisations (dans le cas de l'occlusive /t/ du mot *sept*, le système peut choisir un [d] si la réalisation est voisée). Par la suite, l'analyse des taux et des contextes d'incidence des processus d'alternances permet de mettre en évidence à la fois les conditions d'occurrences et la portée dans la langue.

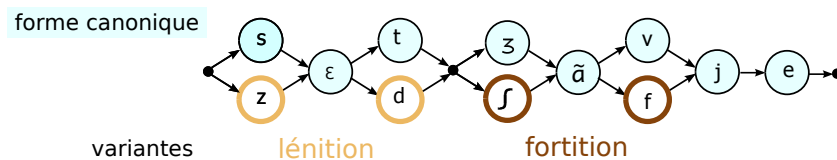


Figure 1. Alternances de voisement possibles et alignement forcé de variantes de prononciation du syntagme « sept janvier »

3. Fréquence des occlusives sourdes et sonores dans les langues romanes

La modélisation phonologique des changements sonores et son interaction avec le niveau phonétique de la variation synchronique font encore débat dans les différentes théories dédiées à la phonologie diachronique. Le rôle de la fréquence lexicale dans la propagation d'une mutation sonore est cependant accepté par de nombreux auteurs, d'autant plus que la mesure des effets de fréquence dans les grands corpus permet d'étayer de manière fiable cette hypothèse (Bybee, 2002 ; Labov *et al.*, 2013).

La taille des corpus mobilisés pour cette étude permet d'offrir un aperçu convaincant de la relation entre la fréquence des segments analysés et les motifs de variation. Ainsi nous dédions cette section à l'étude de la fréquence des occlusives dans les langues romanes en fonction de facteurs linguistiques tels que la position du phonème dans le mot et la fréquence des mots dans les corpus. Les taux sont le résultat des occurrences estimées *via* l'alignement forcé des données audio. L'objectif de ces comparaisons est d'identifier les conditions potentielles d'émergence des changements sonores liées à l'alternance de voisement. Plus précisément, les mesures effectuées ont comme objectif de répondre aux questions suivantes :

- Quelles sont les occlusives les plus fréquentes dans les langues romanes analysées ?
- Quelle est la relation entre la fréquence observée, la position dans le mot et la fréquence du mot dans le corpus ?
- Quelle est la relation entre la fréquence intra et interlangue romane ? Y a-t-il des motifs généraux spécifiques au domaine latin ou, au contraire, les effets de fréquence sont-ils spécifiques à chaque langue ?

Pour répondre à ces questions nous nous sommes penchées sur la fréquence des consonnes occlusives dans les langues romanes en général et sur la représentation des consonnes analysées dans les mots grammaticaux et lexicaux les plus fréquents dans les corpus. La figure 2 résume la fréquence dans les données des différentes classes d'occlusives sourdes et sonores dans les langues romanes, tous contextes lexicaux confondus. Les taux d'occlusives sont comparables à travers les langues. Le lieu d'articulation coronal (/t/, /d/) est le mieux représenté dans toutes les langues analysées (plus de 50 % du total d'occlusives), suivi par les occlusives labiales (/p/, /b/) et vélaires (/k/, /g/). Si ce classement tient compte des traits laryngaux, les corpus mettent en avant la fréquence comparable des occlusives coronales sonores et sourdes en français, espagnol et portugais, suivies par les occlusives sonores en première et les sourdes en deuxième place en italien et en roumain. Dans les deux langues romanes orientales, la coronale sourde /t/ est de loin le segment le plus fréquent aussi bien au sein des occlusives de la langue que par comparaison avec les autres langues romanes. Dans ce classement, les sourdes vélaires (/k/) et labiales (/p/) occupent une troisième et une quatrième place respectivement, et elles sont suivies par les sonores labiales (/b/) et vélaires (/g/). Les fréquences observées permettent de former deux groupes qui opposent les langues romanes occidentales (français, espagnol et portugais) au groupe oriental (italien, roumain).

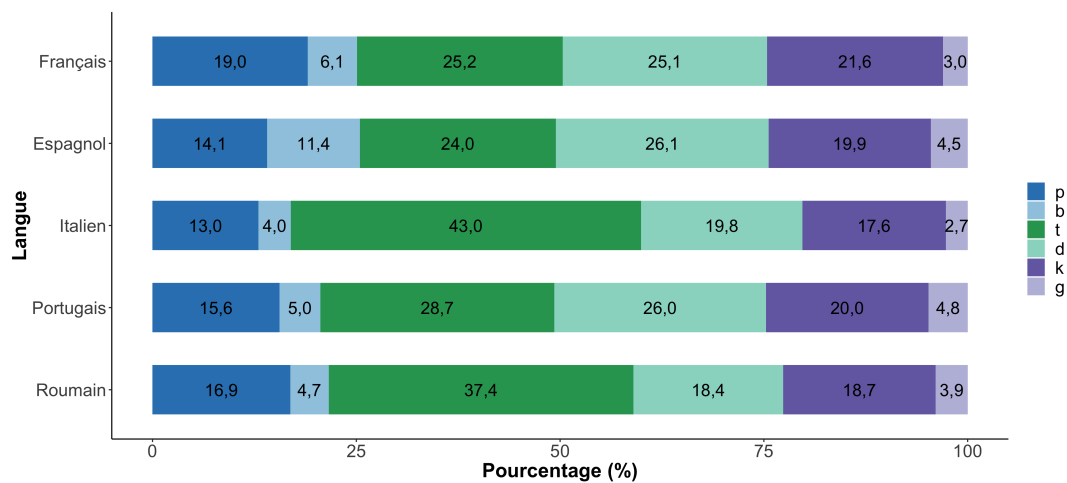


Figure 2. *Fréquences des consonnes occlusives dans nos corpus de langues romanes*

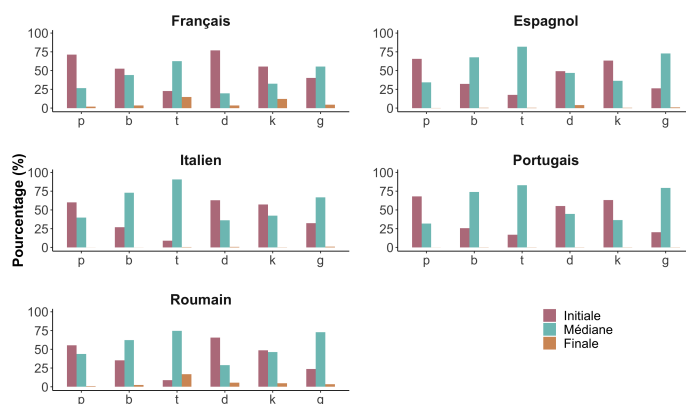


Figure 3. *Fréquences des consonnes occlusives dans les langues romanes selon la position dans le mot (initiale, médiane, finale)*

La prise en compte de la position dans le mot permet de mettre en évidence des tendances spécifiques aux langues prises individuellement. La figure 3 résume ces tendances pour toutes les langues analysées. Nous observons que les positions initiale

et médiane concentrent le plus de consonnes occlusives, et ce, dans toutes les langues analysées. Au contraire, les occlusives sont moins fréquentes en position finale et cette tendance est le résultat des changements sonores spécifiques à l'évolution du latin vers les langues romanes modernes, lors de laquelle les consonnes finales ont disparu favorisant des fins de mot vocaliques (de Carvalho *et al.*, 2008). Les quelques fréquences élevées de codas consonantiques en espagnol, en français et en roumain peuvent être mises en lien avec des effets de fréquence lexicale spécifiques à ces langues. Ainsi, en espagnol, il s'agit des noms finissant en /d/, conséquence de la lénition historique de l'occlusive sourde comme dans lat. *civitate(m)* > esp. *ciudad* (par exemple, *ciudad* « ville », *actitud* « attitude », *calidad* « qualité », *autoridad* « autorité », *salud* « santé », *voluntad* « volonté », etc.). En français, les codas consonantiques sont le résultat de l'amuïssement de la voyelle finale (par exemple, lat. *costa* > fr. *côte*), ils sont présents dans des noms communs, des noms propres, des participes passés, mais également dans de fréquents emprunts anglophones ayant une coda consonantique (*Web*, *week-end*, *download*, *blog*, etc.). Enfin, en roumain, les participes passés alimentent la classe de mots ayant comme coda un /t/ (par exemple, *dat* « donné », *plecat* « parti », *rupt* « cassé », etc.). La prépondérance des occlusives en position initiale et médiane dans les langues romanes peut être ainsi mise en lien avec des changements historiques spécifiques à cette famille de langues. Nous pouvons également noter des effets de symétrie dans les effets de fréquence par rapport à la position dans le mot : ainsi, si les consonnes sourdes sont plus fréquentes à l'initiale de mot, leurs contreparties sonores prédominent en position médiane. Par la suite nous avons considéré les mots les plus fréquents contenant les consonnes d'intérêt. Cette estimation permet d'établir une relation entre la fréquence lexicale et la fréquence phonétique. Elle s'appuie sur l'hypothèse que les phénomènes de variation étudiés sont fortement liés au vocabulaire de la langue. En effet, il a déjà été montré que le statut lexical et la fréquence des mots dans un corpus contribuent aux processus de réduction segmentale qui caractérisent la parole continue (Adda-Decker et Lamel, 2017). Les mots grammaticaux, en général plus courts et plus fréquents, ainsi que certains mots lexicaux fréquents, sont plus susceptibles d'être affectés par les phénomènes d'affaiblissement articulatoire, tels que la lénition, et par la réduction segmentale.

Le tableau 4 montre les mots lexicaux et grammaticaux ³ les plus fréquents ayant au moins deux syllabes et contenant au moins une occlusive sourde ou sonore. Les mots monosyllabiques ne figurent pas dans le tableau car ils sont, par définition, les candidats les plus pertinents pour les phénomènes de réduction (Adda-Decker et Snoeren, 2011). Le tableau indique le nombre d'occurrences disponibles dans les corpus et le rang de fréquence. Le tableau met en avant des similarités à travers les langues : les occlusives sourdes et sonores sont bien représentées à travers un vocabulaire en grande partie commun à toutes langues romanes. À l'exception du français où les mots monosyllabiques prédominent dans les premiers rangs de fréquence, toutes les

3. Nous avons ainsi inclus dans cette classe des mots grammaticaux définis en tant que tels mais également les formes conjuguées des verbes *avoir* et *être* qui peuvent avoir le statut d'auxiliaires.

Français					
Grammatical	#occ.	Rang	Lexical	#occ.	Rang
après	3 014	111	aujourd'hui	3 495	99
depuis	1 958	147	pays	2 396	125
déjà	1 835	152	beaucoup	2 184	134
avec	1 618	165	premier	2 839	144
comment	1 421	178	président	1 814	154
Espagnol					
Grammatical	#occ.	Rang	Lexical	#occ.	Rang
para	22 963	27	gobierno	4 038	79
como	15 855	31	gracias	4 017	80
también	10 549	34	decir	4 011	81
porque	9 628	40	presidente	3 985	82
todo	7 588	51	puede	3 823	90
Italien					
Grammatical	#occ.	Rang	Lexical	#occ.	Rang
della	13 320	21	stato	3 242	53
anche	9 103	30	Italia	2 524	73
questo	6 938	34	presidente	2 182	84
delle	5 678	38	parte	2 156	87
come	5 595	39	così	2 047	92
Portugais					
Grammatical	#occ.	Rang	Lexical	#occ.	Rang
como	4 953	34	peçoas	2 204	63
muito	4 706	36	Portugal	2 020	69
também	3 344	37	dizer	1 670	79
porque	3 562	45	país	1 420	85
portanto	3 200	49	tempo	1 136	113
Roumain					
Grammatical	#occ.	Rang	Lexical	#occ.	Rang
care	4 6567	10	spune	5 922	52
pentru	36 024	12	astăzi	4 108	77
dacă	10 070	32	putea	4 071	80
foarte	8 509	33	timp	4 037	82
după	8 318	42	declarat	3 419	101

Tableau 4. Mots grammaticaux et lexicaux fréquents dans les langues romanes (bi ou polysyllabiques, contenant au moins une occlusive)

langues possèdent des mots bi ou polysyllabiques comportant au moins une occlusive dans les 50 premiers mots les plus fréquents. Pour ce qui est des mots lexicaux ils correspondent soit à des mots appartenant au vocabulaire fondamental des langues (par exemple, les verbes *dire* ou *pouvoir*) soit à des items spécifiques aux corpus journalistiques (par exemple, *président*, *gouvernement*). Enfin, nous pouvons noter que

toutes les occlusives sont représentées aussi bien à l'initiale qu'en position médiane dans les mots. De ce fait, il est possible que tous ces mots alimentent à la fois les processus d'affaiblissement consonantique, qu'il s'agisse de lénition proprement dite ou de suppression (Adda-Decker et Lamel, 2017), et des processus de renforcement, donc de fortition, pour ce qui est de la position à l'initiale de mot. Les analyses de fréquence effectuées offrent des éléments de réponses aux questions formulées plus haut et permettent de formuler des hypothèses au sujet de la lénition et de la fortition. La fréquence des occlusives comparable à travers nos cinq langues romanes est soutenue, entre autres, par des items grammaticaux appartenant au vocabulaire fondamental hérité du latin et par des mots propres au contexte journalistique. Ces observations concernent toutes les langues romanes. Nous observons toutefois des effets de fréquence qui correspondent à la classification en langues romanes occidentales et orientales. Des effets de symétrie dans la distribution consonantique au sein des langues sont également à noter : par exemple, nous observons que si les occlusives sourdes prédominent à l'initiale de mot, les contreparties sonores sont plus fréquentes en position médiane. Les positions initiale et médiane dans le mot concentrent le plus de consonnes occlusives. Ces positions permettent de faire des hypothèses quant aux alternances de voisement. Historiquement, la position médiane est un site privilégié de la lénition, il est pertinent d'interroger cette position en lien avec la lénition. L'initiale de mot est décrite comme une position forte, historiquement préservée dans les langues romanes de la chaîne de phénomènes de lénition, nous pouvons faire l'hypothèse que les occlusives sourdes garderont leur forme canonique dans cette position.

4. Analyse statistique des alternances de voisement dans les langues romanes

Cette section est dédiée à l'analyse des alternances de voisement dans les cinq langues romanes en lien avec les processus de lénition et de fortition. Les corrélats acoustiques des processus de lénition et de fortition ont été examinés dans de nombreuses langues. Si la littérature linguistique ne converge pas vers un consensus quant aux motifs généraux et à l'interaction des deux processus, les chercheurs de tous bords s'accordent pour dire que les traits laryngaux doivent être pris en compte (Bybee et Easterday, 2019 ; Katz, 2016). Les sites d'émergence des deux processus présentent également des régularités à travers les langues : ainsi, la position médiane est un contexte privilégié de la lénition (Lavoie, 2001 ; Gurevich, 2004), tandis que le renforcement consonantique ou la fortition est plutôt rencontrée à l'initiale de mot ou de domaine prosodique (Katz, 2016). Afin de tester les hypothèses concernant les conditions et les sites d'émergence de ces patrons de variation, nous estimons le processus d'alternances dans son intégralité, selon le contexte environnant et selon la position dans le mot. L'analyse, tous contextes confondus, offrira un aperçu global des tendances de changement de voisement selon les langues. La prise en compte des effets d'adjacence, à travers l'analyse des alternances de voisement par rapport aux segments présents à droite et à gauche des consonnes d'intérêt permettra d'étayer ces processus. La prise en compte de la position permet d'estimer si, dans la lignée des études

antérieures, la position initiale de mot est un site propice à la fortition et si la position médiane peut favoriser la lénition.

Nous estimons la pertinence statistique de ces tendances avec des modèles linéaires généralisés (MLG) implémentés dans R (R Core Team, 2013). Ils nous permettront de répondre aux questions suivantes : y a-t-il des différences dans le traitement du voisement selon le type de consonne (sourde, sonore) et la position dans le mot ? Est-ce que le contexte a une pertinence statistique ? Est-ce que les différences dépendent de la langue ? Les facteurs pris en compte sont (1) le voisement des occlusives (sonores et sourdes, référence : occlusive sourde), (2) le contexte gauche (consonne sourde, consonne sonore, voyelle et pause, référence : consonne sourde), (3) le contexte droit (consonne sourde, consonne sonore, voyelle, et pause, référence : consonne sourde), (4) la position dans le mot (initiale, médiane et finale, référence : initiale du mot) et (5) la langue (français, espagnol, italien, portugais et roumain, référence : français). Des tests *post hoc* basés sur le modèle ont été effectués pour comparer les différents niveaux des effets fixes.

La figure 4 montre le taux global de consonnes occlusives ayant changé de mode de voisement dans chacune des langues romanes. Ainsi, Occ.N correspond aux pourcentages d'occlusives sourdes pour lesquelles le système a proposé une variante sonore (par exemple, le mot français *képi* a été aligné comme *ké[b]i*). Occ.V regroupe naturellement les occlusives sonores ayant été alignées avec une variante sourde (par exemple, le mot français *ride* a été aligné comme *ri[t]e*). L'espagnol et l'italien montrent des taux d'alternances au sens du changement sourde → sonore légèrement plus importants, mais en moyenne les variantes non canoniques sourdes sont plus souvent proposées à la place des formes sous-jacentes sonores. La figure met en avant que le dévoisement est plus fréquent que le voisement. Les résultats du MLG confirment que les consonnes occlusives sonores sont plus sujettes à l'alternance de voisement que leurs contreparties sourdes [$\beta = 0,112024$, ET = 0,002826, $p < 0,001$]⁴. Le résultat conforte des observations antérieures concernant les phénomènes de réduction spécifiques à la parole continue (Adda-Decker et Lamel, 2017). En portugais cette tendance s'avère très forte et invite à une analyse postérieure plus appuyée des spécificités consonantiques de cette langue. L'alternance de voisement varie également en fonction des cinq langues romanes. Les résultats suggèrent moins d'alternances de voisement pour l'espagnol [$\beta = -0,063427$, ET = 0,003791, $p < 0,001$], l'italien [$\beta = -0,280088$, ET = 0,006636, $p < 0,001$] et le roumain [$\beta = -0,268180$, ET = 0,003856, $p < 0,001$], par rapport à celles observées en français. Au contraire, l'alternance de voisement est plus importante en portugais [$\beta = 0,365511$, ET = 0,005315, $p < 0,001$], par rapport à celle observée en français.

4. β , appelé aussi coefficient, est un paramètre libre du modèle. ET correspond à l'erreur type (*Standard Error* en anglais).

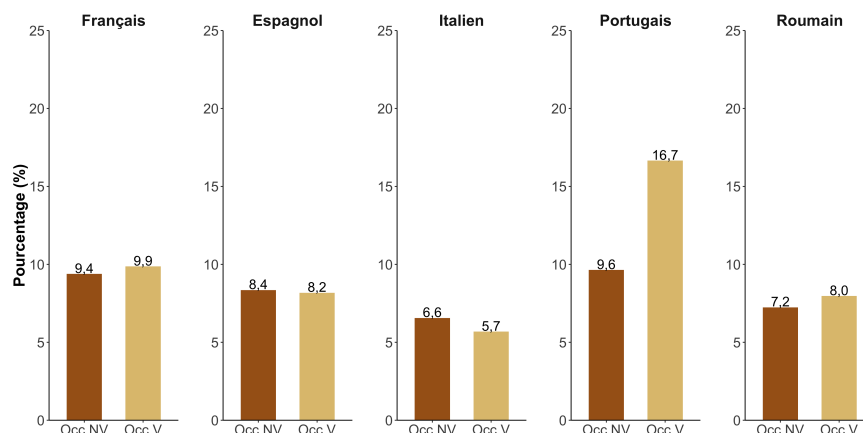


Figure 4. *Pourcentage d’occlusives ayant changé de mode de voisement : Occ.NV = occlusives sourdes ayant été alignées avec une variante sonore par les systèmes de transcription automatique de la parole, Occ.V = occlusives sonores ayant été alignées avec une variante sourde*

Afin d’examiner l’impact des phénomènes d’adjacence, à savoir l’influence du contexte segmental environnant, dans l’émergence des processus de lénition et de fortition, les deux figures suivantes déclinent les taux d’alternances de voisement par langue et par direction de l’alternance en fonction du contexte gauche et droit des consonnes analysées. Ainsi, la figure 5 montre les pourcentages d’occlusives ayant été alignées avec une variante non canonique sonore (Occ.NV) ou sourde (Occ.V) par les systèmes de transcription automatique de la parole en fonction du contexte segmental gauche (pause, consonne sourde, consonne sonore, voyelle). Le graphique met en avant la prévalence des phénomènes d’assimilation contextuelle dans le changement de classe de voisement. Ainsi, un contexte gauche voisé (une consonne sonore ou une voyelle) contribue à l’augmentation du pourcentage d’occlusives sourdes qui sont alignées avec des variantes sonores (par exemple, en français *robe [g]ourte* serait proposé à la place de *robe courte*). Réciproquement, un contexte gauche non voisé (consonne sourde) influence le choix de variantes non canoniques sourdes à la place des occlusives sonores (par exemple, *sac [p]leu* serait proposé à la place de *sac bleu*). L’analyse statistique dédiée à l’effet du contexte gauche suggère moins d’alternances de voisement induites par les consonnes sonores [$\beta = -0,138643$, ET = 0,004788, $p < 0,001$] et par les voyelles [$\beta = -0,179700$, ET = 0,004266, $p < 0,001$], comparée à l’effet observé pour les consonnes sourdes, confirmant l’hypothèse du dévoisement prévalent. Cependant, les pauses entraînent encore plus d’alternances de voisement que les consonnes sourdes [$\beta = 0,164395$, ET = 0,005851, $p < 0,001$].

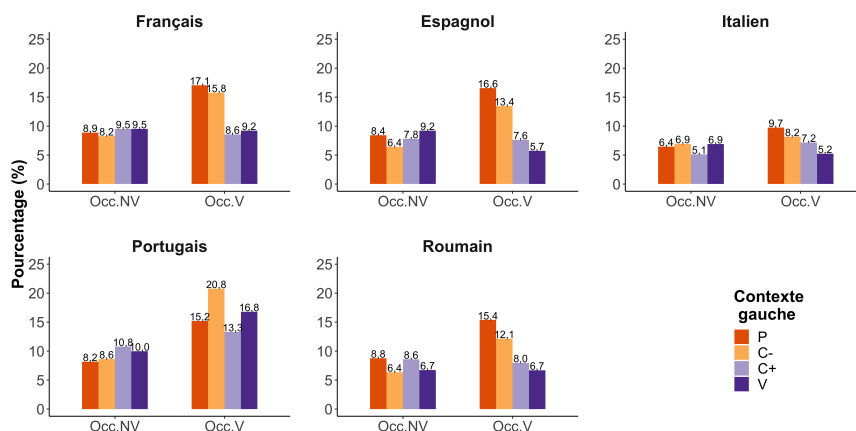


Figure 5. Pourcentage d’occlusives ayant changé de mode de voisement en fonction du contexte segmental gauche (P = Pause, C- = consonne sourde, C+ = consonne sonore, V = voyelle)

La figure 6 montre les pourcentages d’occlusives ayant été alignées avec une variante non canonique sonore (Occ.NV) ou sourde (Occ.V) par les systèmes de transcription automatique de la parole en fonction du contexte segmental droit (pause, consonne sourde, consonne sonore, voyelle) et montre les mêmes tendances que la figure précédente (figure 5). Ainsi, les phénomènes d’assimilation sont en grande partie responsables des processus d’affaiblissement ou de changement de classe de voisement. Pour ce qui est de l’influence du contexte droit, notons qu’une occlusive voisée influence encore plus qu’une voyelle le segment non voisé précédent. En ce qui concerne le contexte droit, les résultats du modèle statistique suggèrent que les consonnes sourdes sont plus susceptibles d’entraîner des alternances de voisement que les voyelles [$\beta = -0.165738$, ET = 0,003653, $p < 0,001$] et les pauses [$\beta = -0,096536$, ET = 0,004649, $p < 0,001$]. En revanche, les consonnes sonores entraînent plus d’alternances de voisement que les consonnes sourdes [$\beta = 0,019931$, ET = 0,003329, $p < 0,001$]. Les résultats issus de l’analyse du rôle du contexte mettent en avant l’importance des effets d’adjacence dus aux consonnes dans les alternances de voisement. L’analyse statistique montre que le contexte gauche est plus propice au dévoisement, tandis que le contexte droit favorise le voisement. Ainsi, le contexte gauche pourrait être mis en lien avec les phénomènes de fortition, tandis que le contexte droit pourrait être associé à la lénition. Ces résultats ouvrent la voie à de nouvelles hypothèses de recherche étayées par de grands corpus, et des études ultérieures pourraient prendre plus spécifiquement en compte les asymétries observées. Notons que l’effet de la pause sur les phénomènes d’adjacence fait l’objet de travaux sur le dévoisement en français (Jatteau *et al.*, 2019a ; Jatteau *et al.*, 2019b ; Hutin *et al.*, 2020a) et plus récem-

ment sur les effets d'adjacence en fin de mot en français comparé au roumain (Hutin *et al.*, 2020b ; Hutin *et al.*, 2020c).

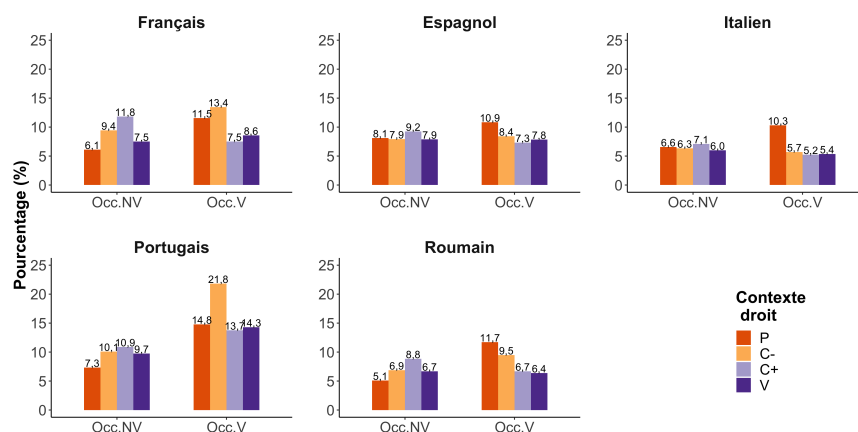


Figure 6. Pourcentage de consonnes occlusives ayant changé de mode de voisement en fonction du contexte segmental droit (P = Pause, C- = consonne sourde, C+ = consonne sonore, V = voyelle)

Nous avons examiné de plus près les contextes lexicaux identifiés comme propices à la fortition et à la lénition, à savoir les positions initiale et médiane⁵. Pour ce qui est de la fortition, la figure 7 montre les pourcentages d'occlusives en position initiale de mot ayant été alignées avec une variante non canonique sonore (Occ.NV) ou sourde (Occ.V) par les systèmes de transcription automatique de la parole. Le dévoisement (une variante non canonique sourde alignée à la place d'une occlusive sous-jacente sonore) prédomine pour la quasi-totalité des langues romanes, à l'exception de l'italien. Ce résultat conforte l'hypothèse que l'initiale de mot est un site propice au renforcement consonantique, en relation avec les phénomènes de fortition. Pour ce qui est de l'italien, le pourcentage d'occlusives sourdes qui deviennent sonores étant plus important, nous pouvons faire l'hypothèse que d'autres processus caractéristiques de la frontière lexicale peuvent potentiellement intervenir, tels que le redoublement phono-syntaxique. Le redoublement phono-syntaxique est un phénomène de gémination consonantique à l'initiale de mot qui se produit dans le contexte d'un mot qui finit en voyelle accentuée ou non, suivi par un mot dont l'initiale est une obstruante (par exemple, *it. città piccola* *città* [pp] *picola* « petite ville », *it. a casa* a [kk] *asa* « à la maison »). Ces processus, ainsi que leur interaction avec les motifs de variation généraux, méritent une analyse plus poussée.

5. Comme mentionné dans le paragraphe précédent, la position finale faisant déjà l'objet d'études approfondies en français et en roumain, n'a pas été considérée dans cette étude.

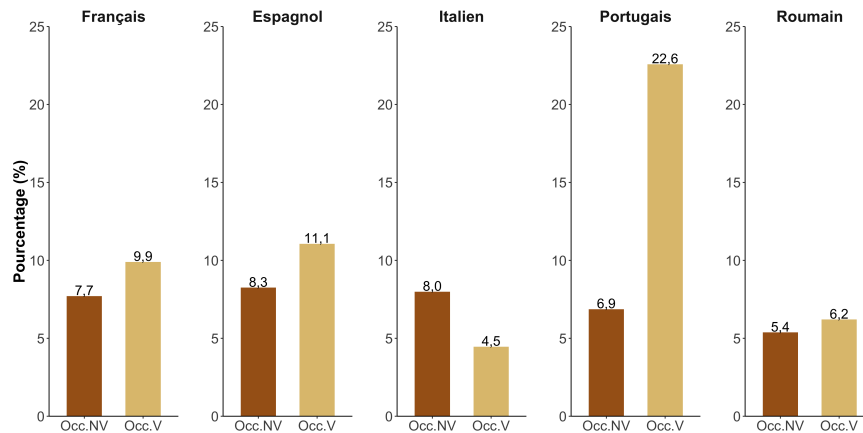


Figure 7. *Pourcentage d’occlusives ayant changé de mode de voisement en position initiale de mot*

La figure 8 montre les pourcentages d’occlusives ayant été alignées avec une variante non canonique sonore (Occ.NV) ou sourde (Occ.V) par les systèmes de transcription automatique de la parole en position médiane. En moyenne, cette position est propice au changement de voisement dans le sens de la lénition, car à l’exception du portugais toutes les langues romanes montrent un taux plus important d’alternances dans la direction segment non voisé → voisé que l’inverse. Pour ce qui est de la significativité statistique de ces résultats, le modèle MLG suggère que la position médiane du mot déclenche moins d’alternances de voisement par rapport à la position initiale du mot [$\beta = -0,063226$, $ET = 0,002863$, $p < 0,001$].

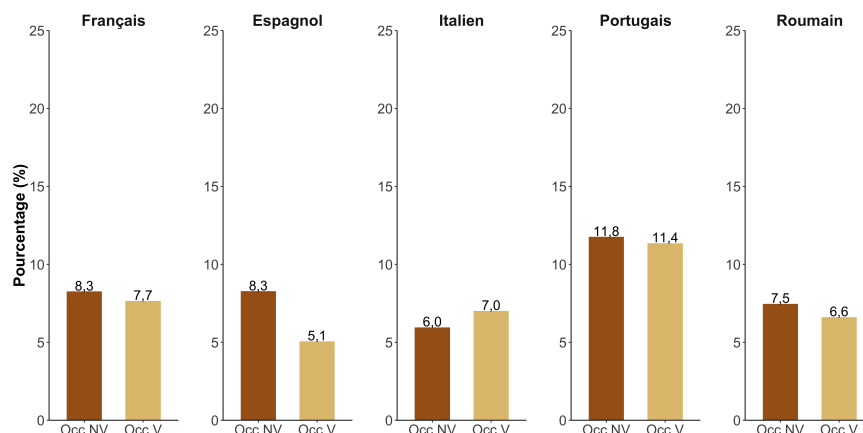


Figure 8. *Pourcentage d'occlusives ayant changé de mode de voisement : position médiane*

Enfin, des tests *post hoc* basés sur le modèle MLG montrent que les occlusives sonores et sourdes sont significativement différentes les unes des autres, en ce qui concerne l'alternance de voisement ($p < 0,001$). Les quatre types de contextes gauche et droit examinés sont significativement différents les uns des autres (pour toutes les comparaisons par paires : $p < 0,001$). Les positions dans les mots sont également significativement différentes les unes des autres (pour toutes les comparaisons par paires : $p < 0,001$). En ce qui concerne l'effet de la langue, les cinq langues sont significativement différentes les unes des autres ($p < 0,001$), à l'exception de la paire roumain-italien. Par ailleurs, l'analyse statistique confirme que les alternances de voisement et la fréquence des occlusives étaient fortement et positivement corrélées ($r(88) = 0,93$, $p < 0,001$). Plus les occlusives sont fréquentes, plus nous avons tendance à observer les alternances de voisement. Ces résultats, bien que préliminaires, confortent l'hypothèse principale de ce travail : il est possible d'observer dans de très grands corpus des alternances de voisement qui peuvent être mises en relation avec les pré-curseurs phonétiques des processus de lénition et de fortition dans les langues. Les tendances observées sont cohérentes avec la littérature dédiée à la fois aux réductions segmentales propres à la parole continue, et avec celle concernant les sites et les particularités acoustiques de la lénition et de la fortition en tant que motifs de variation synchronique. Ainsi, nous avons noté que le dévoisement est prépondérant et qu'il s'agit d'un trait général rencontré dans toutes les langues. Les alternances de voisement sont également fortement dues à des phénomènes d'assimilation qui semblent se produire aussi bien en position intralexicale qu'à la frontière de mot. La position intralexicale est un site privilégié de la lénition, l'initiale de mot est propice au renforcement consonantique. Enfin, des tendances spécifiques aux différentes langues

romanes émergent, qui nécessitent des explorations plus poussées et qui convoquent les aspects diachroniques.

5. Discussion

Nous avons proposé une étude préliminaire des alternances de voisement en lien avec les processus de lénition et de fortition dans cinq langues romanes : le français, l'espagnol, l'italien, le portugais et le roumain. Nous avons pris en compte la série d'occlusives sourdes /ptk/ et sonores /bdg/, communes aux cinq langues. Nous avons fait appel à de très grands corpus (près de mille heures de données orales transcrites manuellement) que nous avons examinés avec une méthode automatique qui se base sur l'utilisation de systèmes de transcription automatique de la parole, spécifiques à chaque langue, en tant qu'instruments linguistiques. La méthode consiste en l'alignement de variantes de prononciation non canoniques à la place des segments phonologiquement attendus et permet de « simuler » un changement sonore. Dans le cas de ce travail, cette méthode permet d'examiner la propension des consonnes occlusives sourdes canoniques à être réalisées en tant que segments voisés (lénition) et inversement, celles des segments sonores à être réalisées comme sourds (fortition). Les alternances de voisement ont également été estimées en lien avec la fréquence des segments dans les langues romanes. Nous avons inscrit notre démarche dans une perspective à la fois synchronique, de la variation phonétique propre à l'oral, et diachronique, dans la mesure où la lénition est un phénomène fondateur des langues romanes dans leur évolution historique, à partir du latin. Enfin, la dimension « très grands corpus » ouvre la voie à des questions théoriques associées, à savoir la relation entre la fréquence d'un motif de variation et sa « chance » de devenir un changement durable, et le lien entre la fréquence observée et des facteurs extralinguistiques comme la variation sociolinguistique.

Les résultats obtenus montrent qu'il est possible d'observer dans de très grands corpus des alternances de voisement qui peuvent être mises en relation avec les pré-curseurs phonétiques des processus de lénition et de fortition dans les langues. Les alternances observées concernent différentes positions lexicales et soulignent la pertinence des processus d'assimilation contextuelle dans l'émergence des motifs de variation. Ce travail convoque, bien évidemment, de nombreuses questions théoriques et pratiques. Il ouvre la porte à de nombreuses démarches expérimentales afin d'éclaircir les aspects spécifiques à chacune des langues et de valider les tendances observées. Si les grands corpus permettent d'observer la portée d'un motif de variation dans les langues en offrant des tendances statistiques, il reste néanmoins qu'une vérification plus fine des observations à partir de sous-corpus se prêtant à une vérification manuelle reste indispensable. Nous prévoyons cette démarche conjointe, ainsi que des analyses acoustiques complémentaires, portant par exemple sur le voisement, ou perceptives portant sur la perception de ces tendances par des auditeurs natifs et non natifs des langues analysées. Enfin, ce travail ouvre la voie à des démarches pouvant bénéficier aux technologies de la parole, comme la reconnaissance vocale à travers

l'amélioration des dictionnaires de prononciation, et l'identification automatique des langues, à travers la prise en compte des spécificités acoustiques plus fines de langues issues d'une même famille.

6. Bibliographie

- Adda-Decker M., Lamel L., « Discovering speech reductions across speaking styles and languages », in F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, M. Zellers (eds), *Rethinking reduction : Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*, De Mouton Gruyter, 2017.
- Adda-Decker M., Snoeren N., « Quantifying temporal speech reduction in French using forced speech alignment », *Journal of Phonetics*, vol. 39, n° 3, p. 261-270, 2011.
- Bybee J., « Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change », *Language Variation and Change*, Cambridge University Press, p. 261-290, 2002.
- Bybee J., Easterday S., « Consonant strengthening : A crosslinguistic survey and articulatory proposal », *Linguistic Typology*, vol. 2, n° 23, p. 263-302, 2019.
- Carrasco P., Hualde J. I., Simonet M., « Dialectal differences in Spanish voiced obstruent allophony », *Phonetica*, vol. 69, p. 149-179, 2012.
- da Silva T. F., Gauvain J.-L., Lamel L., « Lattice-based Unsupervised Acoustic Model Training », *Proceedings of the 36th International Conference on Acoustics, Speech and Signal Processing-ICASSP*, p. 4656-4659, 2011.
- da Silva T. F., Gauvain J.-L., Lamel L., « Interpolation of acoustic models for speech recognition », *Proceedings of Interspeech*, p. 3347-3351, 2013.
- de Carvalho J. B., Scheer T., Ségéral P. (eds), *Lenition and Fortition*, Francke, Berlin, 2008.
- Després J., Lamel L., Gauvain J.-L., Vieru-Dimulescu B., Woehrling C., Le V. B., Oparin I., « The Vocapia Research ASR systems for Evalita 2011 », in C. Bosco, A. Mazzei (eds), *Lecture Notes in Computer Science*, vol. 7689, Springer, Berlin Heidelberg, p. 286-294, 2013.
- Eager C., *Contrast preservation and constraints on individual phonetic variation*, Doctoral dissertation, Urbana-Champaign, IL, University of Illinois at Urbana-Champaign, 2017.
- Ernestus M., « An introduction to reduced pronunciation variants », *Journal of Phonetics*, n° 39, p. 253-260, 2011.
- Galliano S., Geoffrois E., Gravier G., f. Bonastre J., Mostefa D., Choukri K., « Corpus description of the ester evaluation campaign for the rich transcription of french broadcast news », *In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC 2006)*, p. 315-320, 2006.
- Gauvain J.-L., Adda G., Adda-Decker M., Allauzen A., Gendner V., Lamel L., Schwenk H., « Where Are We in Transcribing French Broadcast News ? », *Proceedings of Interspeech*, p. 1665-1668, 2005.
- Gauvain J.-L., Lamel L., Adda G., « The LIMSI Broadcast News transcription system », *Speech Communication*, vol. 37, n° 1-2, p. 89-108, 2002.

- Greenberg J., « Some methods of dynamic comparison in linguistics », in J. Puhvel (ed.), *Substance and structure of language*, University of California Press, Berkeley and Los Angeles, p. 147-203, 1969.
- Gurevich N. (ed.), *Lenition and Contrast. The functional consequences of certain phonetically motivated sound changes*, Routledge, New York and Londres, 2004.
- Hallé P.-A., Adda-Decker M., « Bayesian framework for voicing alternation and assimilation studies on large corpora in French », *Proceedings of the 15th international congress of phonetic sciences*, 2007a.
- Hallé P.-A., Adda-Decker M., « Voicing assimilation in journalistic speech », *Proceedings of the 15th international congress of phonetic sciences*, p. 493-496, 2007b.
- Hermansky H., « Perceptual Linear Predictive (PLP) Analysis of Speech », *J. Acoust. Soc. Am.*, vol. 57, n° 4, p. 1738-52, 1990.
- Honeybone P., « Lenition, weakening and consonantal strength : tracing concepts through the history of phonology », in J. B. de Carvalho, T. Scheer, P. Segeal (eds), *Lenition and Fortition*, Mouton de Gruyter, Berlin, p. 9-92, 2008.
- Hualde J., Chitoran I., « Surface sound and underlying structure : The phonetics-phonology interface in Romance languages », in S. Fisher, C. Gabriel (eds), *Manual of grammatical interfaces in Romance*, vol. 10, De Mouton Gruyter, p. 23-44, 2016.
- Hualde J. I., Chitoran I., Niculescu O., « Gestural undershoot and gestural intrusion – from perceptual errors to historical sound change », *2nd ERRARE Worskhop : Errors by humans and machines in multimedia, multimodal, multilingual data processing*, p. 123-136, 2015.
- Hualde J. I., Simonet M., Nadeu M., « Consonant lenition and phonological recategorization », *Laboratory Phonology*, vol. 2, n° 2, p. 301-329, 2011.
- Hualde J., Nadeu M., « Lenition and Phonemic Overlap in Rome Italian », *Phonetica*, vol. 68, p. 215-242, 2011.
- Hualde J., Prieto P., « Lenition of Intervocalic Alveolar Fricatives in Catalan and Spanish », *Phonetica*, vol. 71, p. 109-127, 2014.
- Hutin M., Jatteau A., Lamel L., Adda-Decker M., « Le schwa final en français standard est-il un lubrifiant phonétique ? », *Actes du Congrès Mondial de Linguistique Française - CMLF 2020*, 2020a.
- Hutin M., Jatteau A., Vasilescu I., Lamel L., Adda-Decker M., « Lénition et fortition des occlusives en coda dans deux langues romanes : le roumain et le français », *Actes des 33èmes Journées d'Etude sur la Parole - JEP2020*, 2020b.
- Hutin M., Niculescu O., Vasilescu I., Lamel L., Adda-Decker M., « Lenition and fortition of stop codas in Romanian », *Proceedings of SLTU-CCURL*, 2020c.
- Hyman L. (ed.), *Phonology : Theory and Analysis*, Holt, Rinehart and Winston, New York, 1975.
- Jatteau A., Vasilescu I., Lamel L., Adda-Decker M., « Final devoicing of fricatives in French : Studying variation in large-scale corpora with automatic alignment », in S. Calhoun, P. Escudero, M. Tabain, P. Warren (eds), *19th International Congress of Phonetic Sciences*, Australasian Speech Science and Technology Association Inc., Melbourne, Australia, p. 295-299, 2019a.

- Jatteau A., Vasilescu I., Lamel L., Adda-Decker M., Audibert N., « “Graff] e !” Word-Final Devoicing of Obstruents in Standard French : An Acoustic Study Based on Large Corpora », *Proceedings of Interspeech*, p. 1726-1730, 2019b.
- Katz J., « Lenition, perception and neutralisation », *Phonology*, vol. 33, p. 43-85, 2016.
- Labov W., Rosenfelder I., Fruehwald J., « One Hundred Years of Sound Change in Philadelphia : Linear Incrementation, Reversal, and Reanalysis », *Language*, vol. 89, p. 30-65, 2013.
- Lamel L., « Multilingual speech processing activities in Quaero : application to multimedia search in unstructured data », *International Conference on Human Language Technologies - The Baltic Perspective*, IOS Press, Tartu, Estonia, 2012.
- Lamel L., Courcinous S., Despres J., Gauvain J.-L., Josse Y., Kilgour K., Kraft F., Bac L. V., Ney H., Nußbaum-Thom M., Oparin I., Schlippe T., Schlüter R., Schultz T., Silva T. F. D., Stücker S., Sundermeyer M., Vieru B., Vu N. T., Waibel A., Woehrlich C., « Speech Recognition for Machine Translation in Quaero », *The International Workshop on Spoken Language Translation - IWSLT 2011, San Francisco, USA*, 2011.
- Lamel L., Gauvain J.-L., « Automatic Processing of Broadcast Audio in Multiple Languages », *Proceedings of EUSIPCO*, 2002.
- Lavoie L., *Consonant Strength. Phonological Patterns and Phonetic Manifestations*, Psychology Press, Routledge, 2001.
- Marcello F., Giordani D., Coletti P., « Development and Evaluation of an Italian Broadcast News Corpus », *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC 2000)*, p. 921-924, 2000.
- Martinet A., « Celtic Lenition and Western Romance Consonants », *Language*, vol. 28, n° 2, p. 192-2017, 1952.
- Martinet A., *Économie des changements phonétiques*, Francke, Bern, 1955.
- Meunier C., Espesser R., « Vowel reduction in conversational speech in French : the role of lexical factors », *Journal of Phonetics*, vol. 39, n° 3, p. 271-278, July, 2011.
- Oftedal M., *Lenition in Celtic and in Insular Spanish : The second voicing of stops in Gran Canaria*, Universitetsforlaget, Oslo, 1985.
- Ohala J., « Sound Change is Drawn from a Pool of Synchronic Variation », in L. E. Breivik, E. H. Jahr (eds), *Language Change : Contributions to the study of its causes*, The Hague, Mouton de Gruyter, p. 173-198, 1989.
- R Core Team, *R : A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. 2013.
- Recasens D., « The Effect of Contextual Consonants on Voiced Stop Lenition : Evidence from Catalan », *Language and Speech*, vol. 1, n° 59, p. 139-161, 2016.
- Richter E., « Chronologische Phonetik des Französischen bis zum Ende des 8. Jahrhunderts », in Halle (ed.), *Zeitschrift für Romanische Philologie, Beiheft 82*, Max Niemeyer Verlag, 1934.
- Ryant N., Liberman M., « Large-scale analysis of Spanish /s/-lenition using audiobooks », *Proceedings of the 22d International Congress on Acoustics*, 2016.
- Vaissière J., « From Latin to Modern French : on diachronic changes and synchronic Variations », in U. Kiel (ed.), *AIPUK, Arbetisberitche*, Institut für Phonetik und digitale Sprachverarbeitung, p. 61-74, 1996.

Vasilescu I., Hernandez N., Vieru B., Lamel L., « Exploring Temporal Reduction in Dialectal Spanish : A Large-scale Study of Lenition of Voiced Stops and Coda-s », *Proc. Interspeech 2018*, p. 2728-2732, 2018.

Vasilescu I., Vieru B., Lamel L., « Exploring Pronunciation Variants for Romanian Speech-to-Text Transcription », *Proceedings of SLTU-2014*, p. 161-168, 2014.

von Wartburg W., *Die Ausgliederung der romanischen Sprachräume*, Francke, Bern, 1950.