

Lénition et fortition des occlusives en coda finale dans deux langues romanes : le français et le roumain

Mathilde Hutin¹, Adèle Jatteau², Ioana Vasilescu¹, Lori Lamel¹, Martine Adda-Decker^{1,3}

(1) Université Paris-Saclay, CNRS, LIMSI, Bât. 507, rue du Belvédère, 91405 Orsay, France

(2) Université de Lille, CNRS, UMR 8163, STL, Lille, France

(3) Université Paris 3 Sorbonne Nouvelle, CNRS, UMR 7018, LPP, 19 rue des Bernardins,
75005 Paris, France

{mathilde.hutin, ioana.vasilescu, lori.lamel, martine.adda}@limsi.fr,
adele.jatteau@univ-lille.fr

RÉSUMÉ

L'exploration automatisée de grands corpus permet d'analyser plus finement la relation entre motifs de variation phonétique synchronique et changements diachroniques : les erreurs dans les transcriptions automatiques sont riches d'enseignements sur la variation contextuelle en parole continue et sur les possibles mutations systémiques sur le point d'apparaître. Dès lors, il est intéressant de se pencher sur des phénomènes phonologiques largement attestés dans les langues en diachronie comme en synchronie pour établir leur émergence ou non dans des langues qui n'y sont pas encore sujettes. La présente étude propose donc d'utiliser l'alignement forcé avec variantes de prononciation pour observer les alternances de voisement en coda finale de mot dans deux langues romanes : le français et le roumain. Il sera mis en évidence, notamment, que voisement et dévoisement non-canoniques des codas françaises comme roumaines ne sont pas le fruit du hasard mais bien des instances de dévoisement final et d'assimilation régressive de trait laryngal, qu'il s'agisse de voisement ou de non-voisement.

ABSTRACT

Lenition and fortition of word-final stops in two Romance languages: French and Romanian.

Automatic transcription and speech-text alignment of large corpora enables a fine investigation of the relationship between synchronic phonetic variation and diachronic change: errors in automatic transcripts can provide valuable insights about indicative possible systemic change that is in the process of appearing. It ensues that such automatic processing can be useful to investigate cross-linguistic attested phonological phenomena to establish whether or not they are occurring in languages that are not yet subject to them. The present study proposes to use forced alignment to explore voicing alternation in word-final position in two Romance languages: French and Romanian. It will be shown that non-canonical voicing and devoicing of French and Romanian coda stops are not random but are instances of final devoicing and regressive assimilation of voicing and voicelessness.

MOTS-CLÉS : Grands corpus, alignement automatique, alignement forcé, lénition, fortition, voisement, dévoisement, français, roumain

KEYWORDS: Large corpora, automatic alignment, forced alignment, lenition, fortition, voicing, devoicing, French, Romanian

1 Introduction

Les études sur la variation phonétique et sur la relation entre motifs de variation synchronique et changements diachroniques connaissent récemment un nouvel essor grâce à l'exploration automatisée de corpus toujours plus grands. D'autres études ont montré que les erreurs dans les transcriptions automatiques sont riches d'enseignements sur la variation contextuelle en parole continue, et plus généralement sur les possibles mutations systémiques sur le point d'apparaître (Adda-Decker, 2006 ; Ohala, 1997). Dans ce cadre, il est intéressant de se pencher sur des phénomènes phonologiques largement attestés dans les langues du monde en diachronie comme en synchronie pour établir leur émergence ou non dans des langues qui n'y sont pas encore sujettes.

Les processus de lénition et de fortition figurent parmi ces phénomènes. Un segment est considéré comme subissant une lénition si sa transformation suit un chemin aboutissant à sa syncope pure et simple ; un segment est considéré comme subissant une fortition s'il suit le chemin inverse (bien que cette définition soit débattue ; Honeybone, 2008). On sait grâce à l'observation du changement diachronique dans les langues romanes (Brandão de Carvalho, 2008) qu'un segment non-voisé devra d'abord devenir voisé, par exemple, avant de disparaître, et qu'ainsi le premier est « plus fort » que le second. Il s'ensuit qu'un segment non-voisé qui devient voisé est un segment lénifié, tandis qu'un segment voisé qui devient non-voisé est un segment fortifié. En synchronie, de tels phénomènes de lénition et fortition consonantiques peuvent avoir lieu à divers degrés dans les langues romanes occidentales (Ryant & Liberman, 2016 ; Vasilescu & al., 2018 pour l'espagnol ; Hualde & Prieto, 2014 pour l'espagnol et le catalan ; Hualde & Nadeu, 2011 pour l'italien ; Jatteau et al. 2019a,b,c pour le français). Pour les langues romanes orientales en revanche, la situation est plus délicate : Chitoran et al. (2015), observant la lénition en roumain chez 8 locuteurs natifs, ne fait état que de quelques rares cas d'affaiblissement consonantique ; Niculescu et al. (soumis), étudiant lénition et fortition dans des grands corpus du roumain, ne montre que peu de variation, si ce n'est en coda, mais n'entre pas davantage dans le détail. La position finale de mot, désormais coda, est en effet, à travers les langues du monde, connue pour être une position sensible aux phénomènes d'assimilation ou de neutralisation (Keating et al. 1983 ; Blevins, 2006 ; Myers, 2012), et c'est pourquoi elle est au centre de cette étude.

Le français et le roumain sont toutes deux des langues romanes, la première appartenant à la branche occidentale, la seconde à la branche orientale. Ces deux langues sont notables, dans leur famille, pour permettre à la totalité de leurs occlusives, d'apparaître en coda, y compris en toute fin de mot, comme le montrent les bandes noires dans les tableaux 1.a et 1.b.

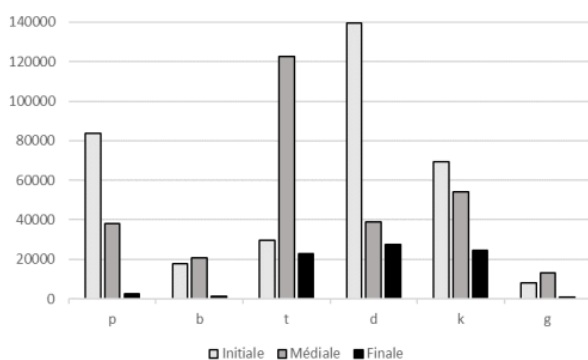


FIGURE 1.a : Répartition des occlusives du français selon leur position dans le mot (Adda-Decker, 2019 ; Vasilescu et al., soumis)

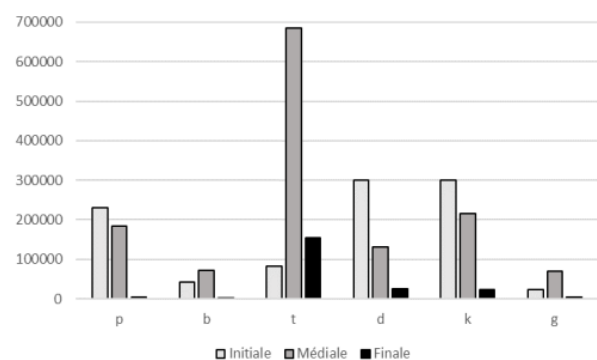


FIGURE 1.b : Répartition des occlusives du roumain selon leur position dans le mot (Adda-Decker, 2019 ; Vasilescu et al., soumis)

De plus, français (1) comme roumain (2) ont une opposition phonologique de voisement, c'est-à-dire que les oppositions p~b, t~d, k~g... donnent lieu à des paires minimales.

(1) Français : /pu/, *pou* ~ /bu/, *boue* ; /tu/, *tout* ~ /du/, *doux* ; /ku/, *cou* ~ /gu/, *goût*

(2) Roumain : /palə/, *lame* ~ /balə/, *bave* ; /talə/, *rassemblement* ~ /dalə/, *dalle* ;
/kalə/, *cale* ~ /galə/, *gala*

Dans la mesure où cette opposition est canoniquement maintenue en coda, ces deux langues sont idéales pour observer les alternances de voisement dans cette position.

La présente étude se propose donc d'observer les alternances de voisement en coda dans deux langues romanes : le français et le roumain. Dans la Section 2, nous présenterons les données et la méthode utilisées. La Section 3 sera consacrée aux phénomènes de fortition, la Section 4 aux phénomènes de lénition. La Section 5 conclut et propose un regard critique sur les résultats.

2 Données et méthode

Pour examiner la variation de voisement en coda, nous proposons ici une étude des occlusives non-voisées et voisées dans des grands corpus du français et du roumain. Des phénomènes aussi spécifiques et multifactoriels que la variation accidentelle de voisement peuvent être observés de façon plus fiable grâce à de très grands corpus (Coleman et al. 2016). Outre l'observation à grande échelle qui permet des résultats statistiquement significatifs, l'avantage des grands corpus repose sur le fait que la parole y est plus naturelle que dans des enregistrements opérés lors d'expériences ou d'enquêtes car récoltée à partir de conversations réelles.

Pour ce qui est du français, nous avons utilisé trois grands corpus transcrits manuellement. Le corpus ESTER (Galliano et al. 2005) comprend originellement 80 heures de discours semi-préparé récoltées en 2003, 2005 et 2009, mais nous l'avons filtré pour enlever les données en français non-métropolitain (RFI et RTM) et ne retenir que les quelque 40 heures en français standard. Le corpus bipartite ETAPE 1 et 2 (Gravier et al. 2012) contient 13,5 heures de radio et 29 heures de télévision en français, notamment des débats et des conversations, récoltées dans les années 2010. Enfin, NCCFr (Torreira et al. 2010) contient 31 heures d'interaction en face-à-face, spontanée, entre amis, enregistrées entre 2007 et 2008.

Pour ce qui est du roumain, nous avons utilisé le corpus bipartite Quaero qui illustre la variante standard (dialecte daco-roumain du sud de la Roumanie). Il consiste en 3,5 heures de parole journalistique et 3,5 heures de débats et d'interviews du début des années 2010. La première partie a été récoltée sur des chaînes de télévision et des émissions radiophoniques roumaines provenant essentiellement des antennes radio RFI Journal et RRA (Radio România Actualități) et de l'agence de presse Euranet, et est représentative d'un discours plutôt préparé, voire lu. La deuxième partie consiste en débats et interviews récoltés sur la chaîne de télévision nationale roumaine Antena 3 et est plutôt représentative d'un discours spontané, non-préparé. Les émissions contenant trop de chevauchement de parole ou de bruit de fond ont été enlevées des données.

Comme ces données ne sauraient être segmentées à la main en raison de leur quantité, elles ont été traitées suivant la méthode décrite dans Gauvain et al. (2002, 2005) et Hallé & Adda-Decker (2007, 2011). Deux systèmes de reconnaissance automatique de la parole, un pour le français et un pour le roumain, ont été utilisés pour forcer un alignement pour lequel des variantes de prononciation ont été introduites pour les codas voisées et non-voisées (cf. Jatteau et al. 2019a,c pour le français ; Vasilescu et al. 2014 pour le roumain). Les systèmes ont été autorisés à sélectionner la coda dite

canonique (celle qu'on trouve dans la forme sous-jacente) ou sa variante altérée si la réalisation acoustique correspondait mieux à cette dernière. Par exemple /p/ peut être étiqueté [p] ou [b], et /b/ [b] ou [p]. Ainsi, selon ce que le système estimait correspondre le mieux à l'audio, un mot français comme *soupe* pouvait être aligné avec les transcriptions [sup], [supə], [sub] ou [subə] et un mot roumain comme *grup* avec les transcriptions [grup] ou [grub] ; de même, un mot français comme *tube* pouvait être aligné avec les transcriptions [tyb], [tybə], [typ] ou [typə] et un mot roumain comme *dialog* avec les transcriptions [dialog] ou [dialok].

Dans la mesure où il a été montré que schwa avait tendance à bloquer les effets d'adjacence (Hutin et al., 2020, soumis), nous avons décidé de ne pas conserver les items réalisés avec un schwa final en français dans nos analyses. Cela permet aussi une comparaison plus fiable avec le roumain, où il n'y a pas de schwa final optionnel. De plus, dans les 7 heures de parole en roumain, 37 items étaient considérés comme étant suivis du mot *-ul*, qui est en fait un morphème ici mal segmenté pour des raisons graphiques après les sigles (ex. *SMURD-ul*, « le SAMU » segmenté [smurd + ul] et non [smurdul] à cause du tiret). Ces 37 items ont été enlevés de la base de données finale.

Ainsi, les corpus du français utilisés dans la présente analyse comprennent 58911 occlusives en coda et les corpus du roumain 4529. Leur répartition est donnée dans le Tableau 1.

	p	t	k	b	d	g	Total
Français	3178	23113	21701	2024	7827	1068	58911
Roumain	110	3288	486	60	487	98	4529

TABLEAU 1. Répartition des occlusives en coda en français et en roumain.

On constate entre autres que /t/ est sur-représenté dans les deux langues. On remarque aussi que les données du roumain sont beaucoup moins nombreuses que celles du français, ce qui signifie que les analyses pour cette langue risquent de ne pas toujours être significatives.

3 Phénomènes de fortition

Le dévoisement en position finale de mot peut être considéré comme une instance de fortition. Elle peut typiquement avoir deux manifestations : l'assimilation régressive de non-voisement due à un segment non-voisé dans l'environnement immédiat (dans le cas des codas, à droite), ou le dévoisement en fin de domaine prosodique, attendu notamment en fin d'énoncé devant pause (Myers 2012, Jatteau et al. 2019b). Dans nos données, 22,29% des occlusives voisées /b, d, g/ sont réalisées dévoisées en français et 51,78% en roumain. De telles proportions de dévoisement méritent une étude plus poussée.

3.1 Dévoisement final vs assimilation de non-voisement

Ce que nous entendons par dévoisement final ici consiste en la réalisation non-voisée d'une obstruante canoniquement voisée lorsque celle-ci apparaît devant pause ; l'assimilation de non-voisement consiste en la réalisation non-voisée d'une obstruante canoniquement voisée lorsqu'elle est dans l'environnement immédiat d'une autre obstruante non-voisée. Il a été démontré que le français a une tendance variable au dévoisement des obstruantes devant pause (Jatteau et al. 2019a,b,c) comme à l'assimilation de non-voisement (Snoeren et al., 2006) : étant donnée la proportion de dévoisement des codas roumaines, il est intéressant de voir si ces deux phénomènes pourraient avoir lieu aussi en roumain.

Pour démontrer qu'il y a bien une tendance au dévoisement final, il faut montrer qu'il y a significativement plus de codas dévoisées devant pause que dans tout autre contexte ; pour démontrer qu'il y a bien de l'assimilation de trait laryngal, il faut montrer qu'il y a significativement plus de codas dévoisées devant obstruante non-voisée. Pour ce faire, les données ont été réparties en cinq catégories, selon que le mot contenant la coda d'intérêt était suivi d'une pause ou d'un autre mot commençant par une voyelle, une sonante, une obstruante voisée ou une obstruante non-voisée.

Les Figures 2.a et 2.b présentent les proportions de codas non-canoniquement dévoisées selon leur contexte de droite en français et en roumain respectivement.

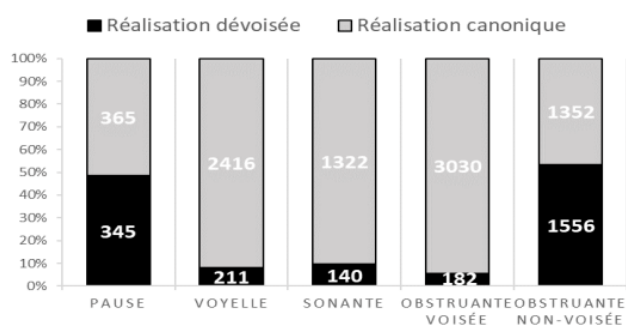


FIGURE 2.a : Proportion de /b, d, g/ en fin de mot réalisés dévoisés selon le contexte de droite en français.

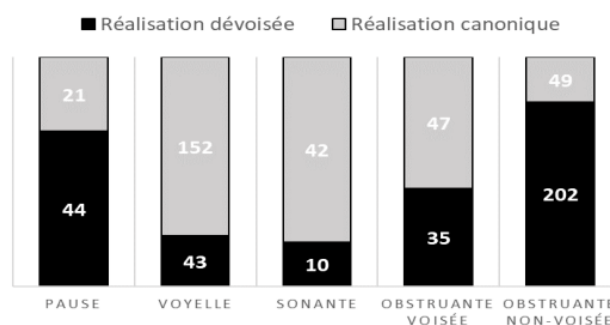


FIGURE 2.b : Proportion de /b, d, g/ en fin de mot réalisés dévoisés selon le contexte de droite en roumain.

Il est intéressant de constater que le français et le roumain montrent les mêmes tendances. Dans les deux langues, les occlusives voisées /b, d, g/ ont tendance à être largement dévoisées devant pause (48,59% en français, 67,69% en roumain) et devant obstruante non-voisée (53,51% en français et 80,48% en roumain). Inversement, dans les deux langues, le dévoisement est en moins grande proportion devant voyelle (8,03% en français, 22,05% en roumain), sonante (9,58% en français, 19,23% en roumain) et obstruante voisée (5,67% en français et 42,68% en roumain), avec néanmoins des différences de proportions qui pourront faire l'objet d'investigations ultérieures.

Les résultats semblent moins nets pour le roumain, où on trouve des proportions importantes de dévoisement dans des contextes où ce dernier n'est pas du tout attendu – ce qui est peut-être dû à la présence avant la coda d'une sonante, voisée par défaut, qui propagerait son trait laryngal à droite (ex. *scurt*, « court » ; *presimt*, « pressentiment » ; *moment*, « moment ») ou encore à l'accent lexical, qui n'a pas été pris en compte ici. Cependant, les différences de voisement selon le contexte de droite sont significatives en roumain ($\chi^2=183.19$, $p<.0001$) comme en français ($\chi^2=2876.7$, $p<.0001$), ce qui nous permet de conclure que ces deux langues sont sujettes à dévoisement final proprement dit et à assimilation de non-voisement.

3.2 L'effet du lieu d'articulation

La littérature existante pointe du doigt les facteurs phonétiques en jeu dans le dévoisement final. Entre autres, comme il est plus difficile de maintenir le différentiel de pression nécessaire à la vibration des plis vocaux dans un tractus vocal plus court (Ohala, 1997), on suppose que les vélaires auront davantage tendance à être dévoisées.

Les Figures 3.a et 3.b montrent les proportions d'occlusives canoniquement voisées qui se trouvent dévoisées devant pause (cas de dévoisement final) selon leur lieu d'articulation.

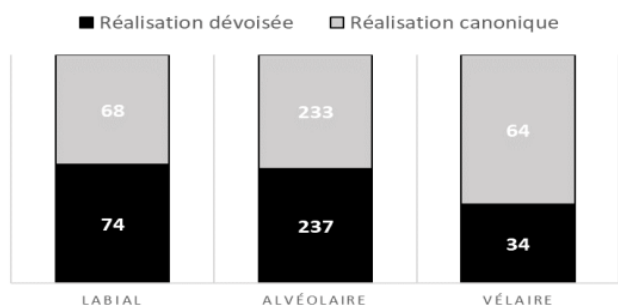


FIGURE 3.a : Proportion de réalisation dévoisée de /b, d, g/ devant pause en français

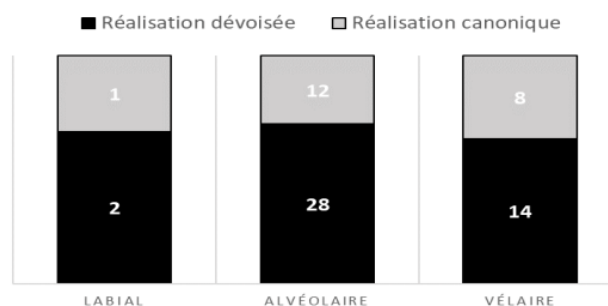


FIGURE 3.b : Proportion de réalisation dévoisée de /b, d, g/ devant pause en roumain

Le dévoisement final en roumain semble toucher davantage les alvéolaires (70,00%) que les labiales (66,67%) et les vélaires (63,64%). La tendance attendue des vélaires à favoriser le dévoisement est ici compromise mais la différence n'est pas significative ($\chi^2=0.26$, $p=0.88$). Pour le français, labiales et alvéolaires sont également plus souvent dévoisées (52,11% et 50,43% respectivement) que les vélaires (34,69%). L'hypothèse que les vélaires devraient dévoiser plus spontanément que les autres occlusives est mise à mal cette fois-ci de façon significative ($\chi^2=8.91$, $p=.01$).

Les Figures 4.a et 4.b montrent les proportions d'occlusives canoniquement voisées qui se trouvent dévoisées devant obstruante non-voisée (cas d'assimilation de non-voisement) selon leur lieu d'articulation.

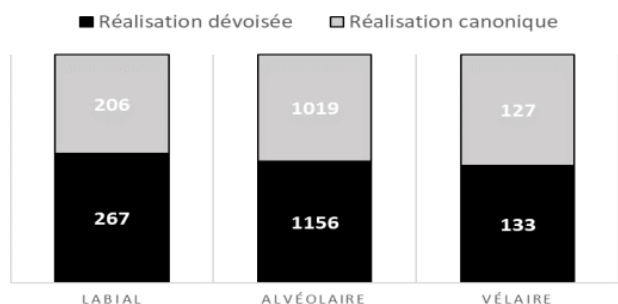


FIGURE 4.a : Proportion de réalisation dévoisée de /b, d, g/ devant obstruante non-voisée en français

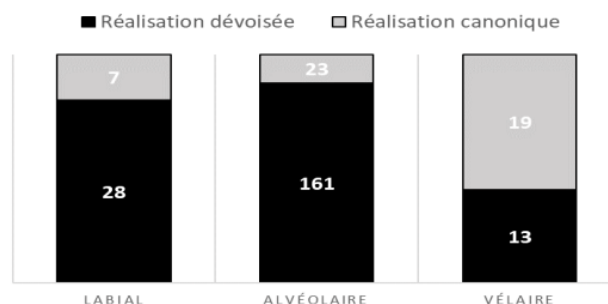


FIGURE 4.b : Proportion de réalisation dévoisée de /b, d, g/ devant obstruante non-voisée en roumain

Dans le cas de l'assimilation régressive de non-voisement, c'est-à-dire quand les codas /b, d, g/ sont réalisées dévoisées devant une obstruante non-voisée, on constate qu'en roumain, le dévoisement touche en priorité, comme pour le dévoisement final, les alvéolaires (87,50%) puis les labiales (80,00%) et en dernier lieu les vélaires (40,63%). Cette fois cependant, les résultats sont statistiquement significatifs ($\chi^2=38.13$, $p<.0001$). En français, là aussi, les tendances sont les mêmes que pour le dévoisement final : les labiales sont les plus dévoisées (56,45%), puis les alvéolaires (53,15%) et enfin les vélaires (51,15%), mais ces résultats ne sont pas significatifs ($\chi^2=2.33$, $p=.31$).

L'assimilation régressive de non-voisement semble donc corroborer les tendances observées pour le dévoisement final dans chacune des langues étudiées. Cependant, les effets du point d'articulation sont peu significatifs pour le dévoisement final en roumain et pour l'assimilation de non-voisement en français. Dans le premier cas, on peut supposer que le peu d'occurrences d'occlusives labiales en coda en roumain soit en cause. Dans le second, on peut supposer qu'il n'y a simplement pas de

différence selon le lieu d'articulation. En tous cas, l'hypothèse des vélaires favorisant le dévoisement est mise à mal dans les deux langues.

4 Phénomènes de lénition

Le voisement non-canonique d'une obstruante sous-jacement non-voisée peut être considéré comme une instance de lénition. Si le voisement final spontané, c'est-à-dire devant pause, est largement controversé, l'assimilation régressive de voisement, due à un segment voisé dans son environnement immédiat (dans le cas des codas, à droite) est largement attesté. Dans nos données, 23,87% des occlusives non-voisées /p, t, k/ sont réalisées voisées en français et 16,45% en roumain.

4.1 L'assimilation régressive de voisement

L'assimilation de voisement consiste en la réalisation voisée d'une obstruante canoniquement non-voisée lorsqu'elle est dans l'environnement immédiat d'une obstruante voisée, ou éventuellement d'une sonante, voisée par défaut. Il a été démontré que le français pouvait être sujet à assimilation de voisement (Snoeren et al. 2006, Hallé & Adda-Decker 2007, 2011) : étant donnée la proportion de voisement des codas roumaines et leur propension, par ailleurs, à l'assimilation de non-voisement, il est intéressant de voir si cette langue est également sujette à assimilation de voisement.

Pour démontrer qu'il y a bien de l'assimilation de trait laryngal, en l'occurrence de voisement, il faut montrer qu'il y a significativement plus de codas non-canoniquement voisés devant obstruante voisée ou encore sonante. Pour ce faire, les données ont été réparties en cinq catégories, selon que le mot contenant la coda d'intérêt était suivi d'une pause ou d'un autre mot commençant par une voyelle, une sonante, une obstruante voisée ou une obstruante non-voisée.

Les Figures 5.a et 5.b présentent les proportions de codas non-canoniquement voisés selon leur contexte de droite en français et en roumain respectivement.

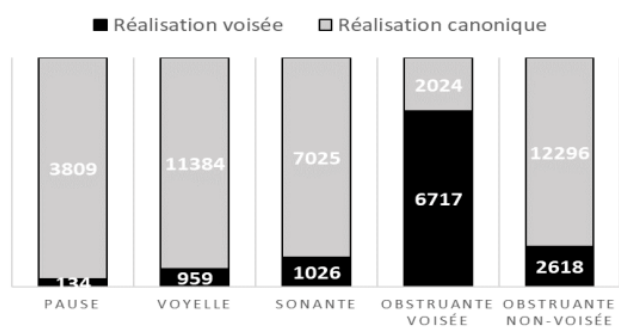


FIGURE 5.a : Proportion de /p, t, k/ en fin de mot réalisés voisés selon le contexte de droite en français.

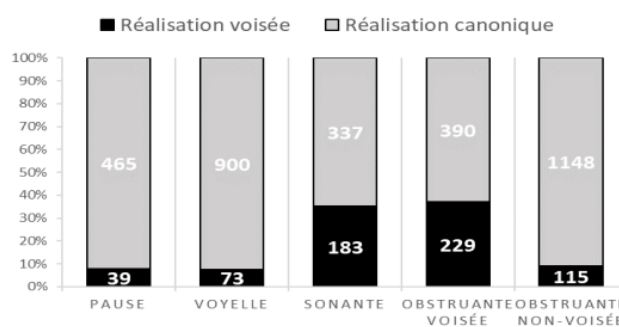


FIGURE 5.b : Proportion de /p, t, k/ en fin de mot réalisés voisés selon le contexte de droite en roumain.

Dans ces figures, on voit que le voisement de /p, t, k/ en coda de mot a lieu essentiellement devant obstruante voisée, en français (76,84%) comme en roumain (37,00%). L'effet est moins marqué en roumain, mais dans cette langue on trouve aussi du voisement devant sonante (35,19%), qui est un autre contexte favorisant l'assimilation de voisement. De plus, la différence n'est pas seulement significative pour le français ($\chi^2=17046$, $p<.0001$) mais aussi pour le roumain ($\chi^2=456.57$, $p<.0001$),

ce qui nous permet de conclure que le voisement non-canonique de /p, t, k/ en fin de mot est bien, dans ces deux langues, une instance d'assimilation régressive de voisement.

4.2 L'effet du lieu d'articulation

En français, les études sur l'assimilation de voisement proposent que les labiales seraient les plus propices au voisement, puis les alvéolaires et enfin seulement les vélares (Hallé & Adda-Decker 2007). Les Figures 6.a et 6.b indiquent les proportions de voisement de /p, t, k/ dans nos corpus.

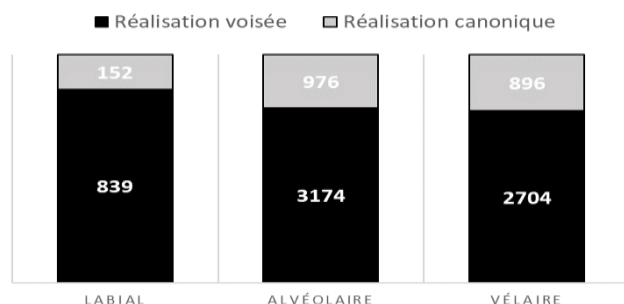


FIGURE 6.a : Proportion de réalisations voisées de /p, t, k/ devant obstruante voisée en français

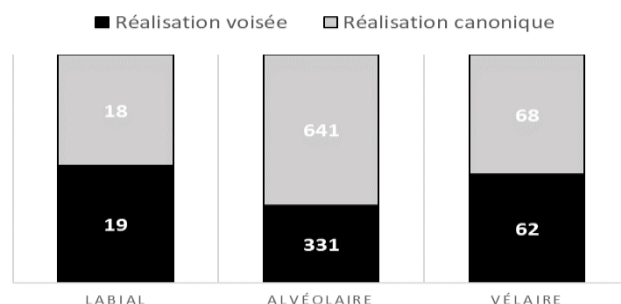


FIGURE 6.b : Proportion de réalisations voisées de /p, t, k/ devant obstruante voisée et sonante en roumain

Les figures 6a et 6b montrent que les labiales sont les plus enclines à subir une assimilation de voisement en roumain (51,35%) comme en français (84,66%). En revanche, les vélares sont, en roumain, assez sujettes à l'assimilation de voisement (47,69%) tandis qu'en français, elles y sont les moins sujettes (75,11%). Nos données confirment donc l'échelle proposée par Hallé & Adda-Decker (2007).

Enfin, il est intéressant de voir qu'en roumain, la coda la moins couramment voisée est l'alvéolaire (34,05%), alors que sa contrepartie voisée était la plus couramment dévoisée. Il s'agit là peut-être d'un effet de la grande fréquence de /t/ en coda en roumain. Dans tous les cas, les différences sont significatives pour le roumain ($\chi^2=13.05$, $p=.002$) comme pour le français ($\chi^2=40.42$, $p<.0001$).

5 Conclusion et discussion

Dans cette étude, il a été mis en évidence que le voisement et le dévoisement non-canoniques des codas du français comme du roumain ne sont pas le fruit du hasard mais bien des instances de dévoisement final et d'assimilation régressive de trait laryngal, qu'il s'agisse de voisement ou de non-voisement, surtout en français. Le point d'articulation, pour sa part, a un effet significatif pour l'assimilation de voisement dans les deux langues, l'assimilation de non-voisement en roumain et le dévoisement final en français. De plus, plusieurs points intéressants sont à noter. Tout d'abord, en français comme en roumain, le dévoisement final semble défavoriser la vélaire, alors que l'état de l'art nous invitait à postuler le contraire. Par ailleurs, les deux phénomènes de dévoisement, le dévoisement final et l'assimilation de non-voisement, ne se comportent pas de la même façon selon la langue, mais montrent les mêmes tendances à l'intérieur d'une même langue – tendances qui s'opposent systématiquement à celles observées pour l'assimilation de voisement.

Malheureusement, certaines des observations proposées reposent sur des différences statistiquement non-significatives. La cause en est le plus souvent un effectif de données roumaines trop peu élevé. Il serait intéressant de confirmer les résultats avec de plus grands corpus du roumain.

Maintenant que la présence de dévoisement final et d'assimilation de trait laryngal a été montrée en roumain et en français, et l'effet du lieu d'articulation de la coda exploré, de plus amples recherches s'imposent sur le comportement des fricatives, ou encore sur divers paramètres sociolinguistiques comme le style de parole ou le sexe du locuteur.

Remerciements

Cette recherche a été en partie financée par le labex DigiCosme (projet ANR-11-LABEX-0045DIGICOSME) opéré par l'ANR dans le cadre du programme « Investissement d'Avenir » Idex Paris Saclay (ANR-11-IDEX-0003-02).

Références

- ADDA-DECKER, M. (2006). De la reconnaissance automatique de la parole à l'analyse linguistique des corpus oraux. Papier présenté aux Journées d'Étude sur la Parole (JEP2006), Dinard, France, 12–16 juin
- ADDA-DECKER, M. (2019). Variation in Romance languages: insights from large corpora. Invited speech at *49th Linguistic Symposium on Romance Languages – LSRL 49*.
- BLEVINS, J. (2006). Theoretical synopsis of Evolutionary Phonology, *Theoretical Linguistics*, vol.32, no. 2, pp.117–166
- BRANDÃO DE CARVALHO, Joaquim. (2008). Western Romance. In BRANDÃO DE CARVALHO, J., SCHEER, T. & SÉGÉRAL, P. (eds) *Lenition and Fortition*. Berlin: Mouton de Gruyter.
- CHITORAN, I., HUALDE, J. & NICULESCU, O. (2015). Gestural undershoot and gestural intrusion – from perceptual errors to historical sound change. *Proceedings of 2nd ERRARE Worskhop* (Sinaia, Romania)
- COLEMAN, J., RENWICK, M. E.L. & TEMPLE, R.A.M. (2016). Probabilistic underspecification in nasal place assimilation. *Phonology*, 33(3), 425–458
- 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. *9th European Conference on Speech Communication and Technology*
- GAUVAIN, J.-L., LAMEL, L. & ADDA, G. (2002). The LIMSI broadcast news transcription system. *Speech communication* 37 (1-2), 89–108
- GAUVAIN, J.-L., ADDA, G., ADDA-DECKER, M., ALLAUZEN, A., GENDNER, V., LAMEL, L. & SCHWENK, H. (2005). Where Are We in Transcribing French Broadcast News? *Proceedings of ISCA Eurospeech'05*, Lisbon, Sep 2005.
- 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*
- HALLÉ, P. & ADDA-DECKER, M. (2007). Voicing assimilation in journalistic speech. *16th International Congress of Phonetic Sciences*, 2007, 493–496.
- HALLÉ, P. & ADDA-DECKER, M. (2011). Voice assimilation in French obstruents: A gradient or a categorical process? *Tones and features: A festschrift for Nick Clements*, De Gruyter, 149–175

- HONEYBONE, P. (2008). Lenition, weakening and consonantal strength: tracing concepts through the history of phonology. In BRANDÃO DE CARVALHO, J., SCHEER, T. & SÉGÉRAL, P. (eds) *Lenition and Fortition*. Berlin: Mouton de Gruyter.
- HUALDE, J. & NADEU, M. (2011). Lenition and phonemic overlap in Rome Italian. *Phonetica*, 68, 215–242
- HUALDE, J. & PRIETO, P. (2014). Lenition of intervocalic alveolar fricatives in Catalan and Spanish. *Phonetica*, 71, 109–127
- HUTIN, M., JATTEAU, A., VASILESCU, I., LAMEL, L. & ADDA-DECKER, M. (Soumis). Overview of word-final schwa in Standard French and its shielding effect from fortition and lenition. *Linguistic Vanguard*.
- HUTIN, M., JATTEAU, A., VASILESCU, I., LAMEL, L. & ADDA-DECKER, M. (2020). Le schwa final en français standard est-il un lubrifiant phonétique ? *Actes du Congrès Mondial de Linguistique française, CMLF 2020*.
- JATTEAU, A., VASILESCU, I., LAMEL, L. & ADDA-DECKER, M. (2019a). "Gra[f] ! Le dévoisement final dans les grands corpus de français". Paper presented at the SRPP seminar, Université Sorbonne Nouvelle, Feb. 15th 2019
- JATTEAU, A., VASILESCU, I., LAMEL, L., ADDA-DECKER, M. & AUDIBERT, N. (2019b). "Gra[f]e!" Word-final devoicing of obstruents in Standard French: An acoustic study based on large corpora. *Proceedings of Interspeech*, 1726–1730. Graz, Austria.
- JATTEAU, A., VASILESCU, I., LAMEL, L. & ADDA-DECKER, M. (2019c). Final devoicing of fricatives in French : Studying variation in large corpora with automatic alignment. *International Congress of Phonetic Sciences*, Melbourne, Australie.
- KEATING, P., LINKER, P. & HUFFMANN, M.-K. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* 11(3). 277–290.
- MYERS, S. (2012). Final devoicing: Production and perception studies. In T. BOROWSKY, S. KAWAHARA, T. SHINYA, M. SUGAHARA (Eds.), *Prosody matters: Essays in honor of Elisabeth Selkirk*, Equinox, pp.148–180.
- NICULESCU, O., VASILESCU, I., CHITORAN, I., ADDA-DECKER, M. & LAMEL, L. (Soumis). Romanian obstruents still strong after all these years: Obstruent voicing and devoicing in a large corpus study of Romanian. *Labphon 17*, 2020.
- OHALA, J. J. (1997). Aerodynamics of phonology. *Proceedings of the Seoul International Conference on Linguistics*. Seoul: Linguistic Society of Korea, pp. 92–97
- RYANT, N. & LIBERMAN, M. (2016). Large-scale analysis of Spanish /s/-lenition using audiobooks. *Proceedings of the 22nd International Congress on Acoustics* (Buenos Aires, Argentina)
- SNOEREN, N., HALLÉ, P. & SEGUI, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics*, vol.34, pp. 241–268
- TORREIRA, F., ADDA-DECKER, M. & ERNESTUS, M. (2010). The Nijmegen corpus of casual French. *Speech Communication*, vol. 52, no. 3, pp.201–212.
- VASILESCU, I., HERNANDEZ, N., VIERU, B. & LAMEL, L. (2018). Exploring Temporal Reduction in Dialectal Spanish: A Large-scale Study of Lenition of Voiced Stops and Coda-s. *Proceedings of Interspeech* (Hyderabad, India)
- VASILESCU, I., VIERU, B. & LAMEL, L. (2014). Exploring Pronunciation Variants for Romanian Speech-to-text Transcription. *Proceedings of SLTU* (St. Petersburg, Russia)
- VASILESCU, I., WU, Y., JATTEAU, A., ADDA-DECKER & LAMEL, L. (Soumis). Alternances de voisement et processus de lénition et de fortition : une étude automatisée de grands corpus en cinq langues romanes. *Revue TAL*.