

L'assimilation de voisement en français : elle vaut pour les non-mots autant que les mots

Pierre Hallé,^{1,2} Kaja Androjna³ et Juan Seguí²

(1) LPP (CNRS-Paris 3), 19 rue des Bernardins, 75005 Paris

(2) LMC, 71 Avenue Édouard Vaillant, 92774 Boulogne Billancourt

(3) DEC (ENS), 45 rue d'Ulm, 75005 Paris

pierre.halle@univ-paris3.fr, kaja.androjna@ens.fr,

juan.segui@parisdescartes.fr

RESUME

Des études récentes sur l'assimilation de voisement en français ont tenté de répondre à plusieurs questions essentielles. Tout d'abord, s'agit-il d'un processus graduel ou catégoriel ? Des analyses récentes basées sur les changements distributionnels de taux de voisement (v-ratio) lors de l'assimilation suggèrent que l'assimilation est optionnelle mais est complète lorsqu'elle a lieu. Elle serait donc catégorielle. Dans cette étude, nous raffinons ces observations par l'analyse d'indices secondaires de voisement (e.g., durée d'occlusion). De plus, nous abordons la question de savoir si la réalisation de l'assimilation est motivée par des règles phonologiques généralisables plutôt que par une connaissance réductible à la compilation statistique des énoncés produits ou entendus. Nous utilisons pour cela une situation extrême : celle de non-mots en situation ou non d'assimilation. En parole lue, ils sont assimilés autant que les mots, ce qui est en faveur des règles plutôt que des statistiques.

ABSTRACT

Voice assimilation in French: It applies to nonwords just like to words

Recent studies on voice assimilation in French have addressed several important issues. Firstly, is this assimilation process gradient or categorical? Recent analyses of how distributions of voicing ratio change with assimilating to non-assimilating contexts suggest that assimilation is optional but is complete when it occurs. It would therefore be categorical. In this study, we refine these observations with the analysis of some secondary cues to voicing (e.g., closure duration). We also address the issue of whether the occurrence of assimilation is motivated by phonological rules that can be generalized to any item or, rather, is determined by internalized statistics on heard or produced utterances. We use the extreme situation of nonwords in assimilating vs. non-assimilating context. In read speech, voice assimilation affects nonwords as much as words, supporting rule-based rather than statistical accounts.

MOTS-CLES : assimilation de voisement, français, parole lue, non-mots.

KEYWORDS : voice assimilation, French, read speech, nonwords.

1 Introduction

Baucoup d'aspects de l'assimilation de voisement en français sont assez consensuels : elle est régressive et restreinte aux contacts entre obstruantes. Elle est plus systématique (ou plus complète) à l'intérieur des mots qu'entre les mots. Enfin, l'idée que les

obstruantes assimilées conservent leur caractère lenis ou fortis originel a fait place à la notion d'assimilation partielle au niveau phonétique, suggérée par l'observation de taux de voisement intermédiaires (Gow et Im, 2004 ; Snoeren, Hallé et Segui, 2006). Mais cette notion même a été remise en question récemment par des analyses distributionnelles du taux de voisement, ou *v-ratio* (Hallé et Adda-Decker, 2007, 2011), suggérant, du moins pour le français, un processus optionnel mais catégoriel : lorsque l'assimilation a lieu, il s'agit d'un échange catégoriel en termes de *v-ratio*. Cependant, il semble que d'autres indices liés au voisement laissent des traces du voisement sous-jacent, même en cas d'assimilation complète en termes de *v-ratio* (Snoeren et al. 2008). Dans cette étude, nous analyserons systématiquement ces autres indices dits "secondaires" ainsi que les taux de voisement (*v-ratio*) et d'harmonicité (HNR).

Cette étude pose aussi une question plus générale sur les variations des formes parlées. Deux points de vue s'opposent. Selon un point de vue générativiste, il existe un niveau de représentation abstraite (par exemple en termes de traits distinctifs) sur lequel opèrent des processus phonologiques imperméables aux fréquences d'occurrence, qui produisent des formes de surface. Ces processus —par exemple l'assimilation— sont supposés généralisables à toute forme nouvelle, en particulier à des non-mots lus. Selon un point de vue constructiviste, les variantes sont apprises dans leur contexte, formant avec ce contexte des constructions d'autant plus stables et non analysables qu'elles sont fréquentes (Bybee, 2001). Dans cette optique, les non-mots, dont la fréquence est nulle, ne peuvent être appris dans des constructions où ils seraient modifiés par rapport à leur forme "libre". Pour tenter de contribuer à ce débat, nous comparerons mots et non-mots en parole lue dans des contextes motivant ou non une assimilation de voisement.

2 Étude de production (parole lue)

2.1 Méthode

2.1.1 Locuteurs

Huit locuteurs (4 hommes et 4 femmes), étudiants à l'Université Paris 3 (âge moyen 27 ans), ont participé aux enregistrements. Tous étaient originaires de la région parisienne ou du nord de la France et vivaient à Paris au moins depuis 3 ans. Nous avons pris soin d'éviter des locuteurs originaires du sud de la France pour éviter l'insertion de schwas ou ceux originaires du nord-est de la France pour éviter le dévoisement des consonnes finales de mot. Aucun des locuteurs n'avait jamais souffert de troubles du langage.

2.1.2 Matériel

Nous avons utilisé 48 séquences test nom-adjectif (e.g., *vide partiel*) avec un contact occlusive-occlusive entre les deux mots (dans l'exemple ci-dessus, /d/#/p/). Pour la moitié de ces séquences, le contact C1#C2 était non-assimilant (e.g., *mythe païen, guide bavard*) et pour l'autre moitié, le contact était assimilant (e.g., *mythe barbare, guide patient*). La comparaison de ces deux conditions, contrôle et assimilation, est en effet nécessaire pour quantifier les effets possibles de l'assimilation.

Les contacts C1#C2 respectaient plusieurs contraintes. (1) Les lieux d'articulation de C1 et C2 étaient différents pour éviter les géminées ; nous nous sommes d'autre part limités

aux dentales et labiales ; les contacts utilisés étaient donc /b#d, b#t, d#b, d#p, p#t, p#d, t#p, t#b/ avec 6 séquences nom-adjectif pour chaque contact. (2) La voyelle précédent C1 était soit /a/ soit /i/, ceci pour tester l'influence possible du contexte vocalique sur le degré de voisement (Ohala, 1983) ; la voyelle suivant C2 était dans la mesure du possible /a/ (80%) et parfois /o/ (20% : e.g., *crabe dodu*). (3) Pour toutes les séquences nom-adjectif, le nom était monosyllabique et l'adjectif dissyllabique, pour maintenir constante la structure prosodique de la séquence qui peut influencer le voisement (Slis, 1986). Étant donné le nombre des contraintes segmentales et prosodiques à satisfaire, nous n'avons pas contrôlé la fréquence lexicale des mots cibles. Aux séquences de mots étaient appariées 48 séquences de deux non-mots se rapprochant autant que possible des séquences nom-adjectif du point de vue de leur structure phonétique (e.g., *zite pajotte, chide bafique, zite bagonne, chide palcotte*). En plus des 96 séquences test, nous avons utilisé 48 séquences distracteur, 24 séquences nom-adjectif et 24 séquences de deux non-mots, avec une variété de contacts C1#C2 autres que occlusive-occlusive (e.g., *rire naïf, rêve cruel ; zile facarde, siffe lovette*).

Les séquences nom-adjectif ou de non-mots étaient produites dans le cadre de la phrase porteuse *On parle jamais de ____*. L'omission de la particule de négation *ne* était délibérée pour inciter les locuteurs à produire les phrases avec spontanéité. De plus, cette phrase permettait d'introduire les séquences de non-mots. La structure syntaxique des énoncés à lire était donc maintenue constante, ce qui permettait d'éviter des variations indésirables liées à la structure syntaxique (Kuzla, Cho et Ernestus, 2007).

2.1.3 Procédure d'enregistrement

Les locuteurs ont été enregistrés individuellement dans une chambre sourde. Ils ont été enregistrés directement sur ordinateur (16 bits, 44.1 kHz) avec le logiciel Sound Studio, via un microphone-casque (MicroMic C520L) relié à une carte son externe (EDIROL).

Les locuteurs avaient pour consigne de lire les phrases de la manière la plus fluide possible et d'éviter toute pause à l'intérieur d'une même phrase, notamment entre les derniers mots de la phrase. Pour leur donner une idée de la vitesse d'élocution attendue (5-6 syllabes/seconde), nous leur avons présenté cinq phrases à un débit de 6 syllabes/seconde à l'aide d'un métronome (<http://www.metronomeonline.com/>).

La moitié des sujets ont lu d'abord les phrases avec séquences nom-adjectif mots et ensuite celles avec séquences de non-mots ; l'autre moitié ont fait l'inverse. L'ordre semi-aléatoire était le même pour tous les locuteurs.

2.1.4 Segmentation des séquences nom-adjectif et non-mots

Pour effectuer les analyses acoustiques décrites en 2.1.5, il était nécessaire de procéder à un étiquetage phonétique des séquences cibles de mots ou de non-mots. Nous nous sommes limités à étiqueter les événements suivants : début de la voyelle V précédant C1 dans les contacts C1#C2 ; fin de V ou début de l'occlusion de C1 ; fin de l'occlusion de C1 ou début du relâchement ; fin du relâchement de C1. La segmentation du signal de parole a été faite manuellement par inspection visuelle de spectrogramme (logiciel Praat : Boersma, 2001) et, dans les cas difficiles, à l'aide de courbes de dérivée spectrale, d'énergie, et/ou de suivi de formants. L'apparition/disparition du deuxième

formant a été retenue comme critère principal pour localiser le début et la fin de V. Le début de V était parfois difficile à localiser dans le contexte d'un /r/ précédent (e.g., *rite banni*), mais jamais pour les autres consonnes. La fin du formant 2 de V permettait de localiser sans difficulté particulière à la fois la fin de V et le début de l'occlusion de C1. Par contre, la fin de l'occlusion c'est à dire le début du relâchement de l'occlusive C1 était parfois difficile à repérer. Dans 10 % des cas, le relâchement n'était pas visible sur le spectrogramme, mais il était parfois possible, lorsque l'occlusion était voisée, de détecter des indices de changement de lieu soit dans le signal de parole (discontinuité d'amplitude ou de phase) soit dans la courbe de dérivée spectrale (pic local). Dans les cas où même ces indices n'étaient d'aucune aide (environ 6% des cas), nous avons pris comme frontière entre C1 et C2 le milieu de l'occlusion entre V et la voyelle suivante. Nous avons placé la fin du relâchement de C1 à la fin du bruit de relâchement lorsqu'il était visible. Nous avons sinon considéré que la durée de relâchement était nulle et placé une marque de fin, par convention, 1 ms après celle du début de relâchement.

2.1.5 Analyses acoustiques

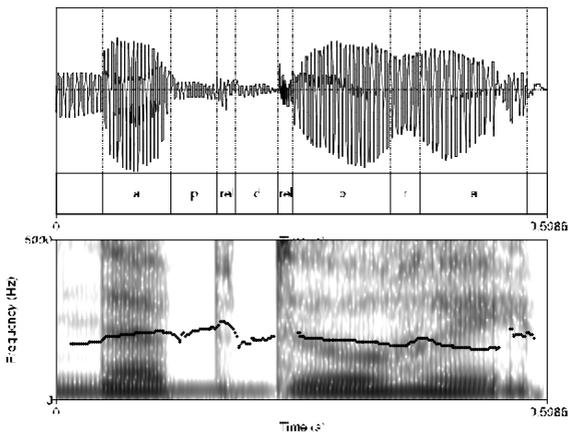


FIGURE 1 – Exemple de mesures pour la séquence *nappe dorée* : v -ratio = 1 ; HNR = 7.78 dB ; durées : 82, 55, 25 ms (V, occlusion, relâchement) ; intensités : 71.7 et 72.7 dB (occlusion et relâchement) ; F0 en début et fin de V : 196 et 204 Hz.

Les mesures acoustiques ont porté sur la *consonne C1*, potentiellement assimilée ou non selon le voisement de C2, et sur la *voyelle précédente V*. Nous distinguons entre indices primaires et secondaires de voisement, et pour les indices secondaires, entre indices locaux et non-locaux (Goldrick et Blumstein, 2006). Nous avons retenu comme indice primaire le taux de voisement, ou v -ratio : la proportion de signal voisé dans l'occlusion qui reflète directement la présence/absence de vibration laryngale (Barry et Teifour, 1999 ; Hallé et Decker, 2007 ; Snoeren et al., 2006) basée sur la présence/absence de périodicité dans le signal telle qu'indiqué par un algorithme de détection de F0

(méthode de cross-corrélation de Praat, avec réglages par défaut sauf la plage F0 ([60, 400] Hz) et le pas d'analyse (2 ms)). Nous avons également utilisé une mesure HNR ('harmonic to noise ratio') sur l'occlusion de C1, permettant de quantifier la périodicité du signal de façon non binaire (réglages dans Praat : pas d'analyse 2 ms, fenêtres d'analyse de 50 ms, 4,5 périodes par fenêtre d'analyse, seuil de silence 0.003).

Pour ce qui est des indices secondaires, nous avons mesuré la durée et l'énergie de l'occlusion et du relâchement de C1 (indices locaux) ainsi que la durée et le contour F0 de la voyelle V précédent C1 (indices non-locaux). La Figure 1 illustre le calcul des différents indices pour la séquence *nappe dorée*.

2.2 Résultats

Nous avons effectué des analyses de variance par sujets sur chacun des indices décrits plus haut. Les facteurs (tous intra-sujet) étaient : *lexicalité* de la séquence cible (mots vs. non-mots), voisement de C1 et voisement de C2 (*vc1* et *vc2*, respectivement).

v-ratio. La Table 1 résume les données de v-ratio pour l'occlusion de C1, pour les séquences de mots et de non-mots, selon le contact de voisement.

	NV-NV	NV-V	variation	V-V	V-NV	variation
mots	0.19	0.65	+ 0.46	0.99	0.50	- 0.49
non-mots	0.19	0.51	+ 0.32	0.92	0.48	- 0.44

TABLE 1 – v-ratios (occlusion de C1) dans les conditions contrôle (en grisé) et assimilation. V (voisé) et NV (non-voisé) indiquent le voisement des consonnes en contact ; "variation" (en bleuté) : changement de v-ratio entre contrôle et assimilation.

Entre les conditions contrôle et assimilation, le v-ratio change en moyenne de 0.44, il augmente ou diminue selon que C2 est voisé ou non. Ce changement mesure l'effet du facteur *vc2*, qui est très significatif globalement, $F(1,7) = 169.90$, $p < .00001$, et aussi en détail, tant pour les non-mots que les mots, que C1 soit voisé ou non, $ps < .0005$. L'assimilation n'est pas plus forte dans un sens que dans l'autre, comme le montre l'interaction non-significative entre *vc2* et *vc1*, $F(1,7) = 2.88$, $p = .13$, avec cependant une tendance marginale vers davantage de dévoisement que de voisement pour les séquences de non-mots, $F(1,7) = 3.68$, $p = .097$. Enfin, l'assimilation tend à être plus forte pour les mots que les non-mots, une tendance marginalement significative seulement pour C1 non-voisée, $F(1,7) = 4.53$, $p = 0.07$.

Les valeurs moyennes de v-ratio pour NV-V ou V-NV (assimilation) sont intermédiaires entre celles pour NV-NV et V-V (contrôle). Cependant, avant de conclure au caractère graduel de l'assimilation, il faut examiner les distributions de v-ratio. Les distributions pour NV-NV et V-V montrent que les v-ratios sont concentrés dans le dernier intervalle ([.875, 1]) pour les C1 sonores et dans les trois premiers ([0, .375]) pour les C1 sourdes, définissant les catégories voisée et non-voisée en contexte non-assimilant. Les distributions pour NV-V suggèrent un passage d'une catégorie à l'autre pour les C1 sourdes, tant pour les mots que les non-mots. L'assimilation semble donc catégorielle bien qu'optionnelle dans le sens du voisement. Tel n'est pas le cas dans le sens du dévoisement : la catégorie voisée (intervalle [.875, 1]) ne se redistribue pas dans la non-voisée (intervalle [0, .375]) mais plutôt dans la région intermédiaire [.375, 0.725].

L'assimilation dans ce sens semble donc incomplète ou graduelle, ce qui est contraire aux résultats de Hallé et Adda-Decker (2007, 2011).

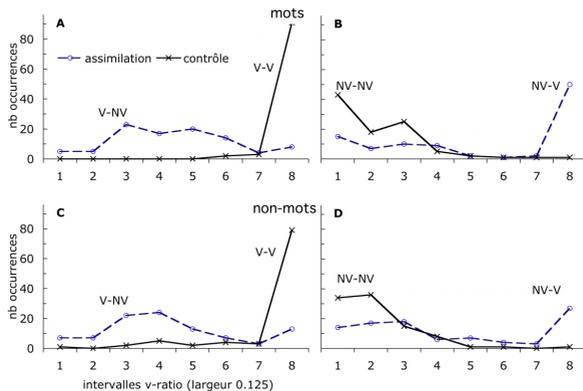


FIGURE 2 – Distributions du v-ratio pour les C1 sonores (A et C) vs. sourdes (B et D) en fin de mot (A et B) ou de non-mot (C et D) dans les conditions contrôle et assimilation.

HNR. La Table 2 résume les données de HNR selon le modèle de la Table 1.

	NV-NV	NV-V	variation	V-V	V-NV	variation
mots	- 95 dB	- 26 dB	+ 69 dB	11 dB	- 58 dB	- 69 dB
non-mots	- 92 dB	- 49 dB	+ 43 dB	2 dB	- 58 dB	- 60 dB

TABLE 2 – HNR moyen (occlusion de C1) dans les conditions contrôle (en grisé) et assimilation. V et NV indiquent le voisement des consonnes en contact ; “variation” (en bleuté) : changement de HNR moyen entre contrôle et assimilation.

Les HNR moyens sur C1 et leurs distributions sont en bon accord avec les v-ratios moyens et leurs distributions. Entre les conditions contrôle et assimilation, le HNR change en moyenne de 60 dB, augmentant si C2 est voisé et diminuant sinon. Comme pour le v-ratio, l'effet du facteur $vc2$ est très significatif, reflétant une forte assimilation tant pour les non-mots que les mots, que C1 soit voisé ou non, $ps < .005$. L'assimilation n'est pas plus forte dans un sens que dans l'autre (interactions $vc2 \times vc1$ non-significatives), avec une tendance non significative vers davantage de dévoisement que de voisement pour les non-mots, $F(1,7) = 2.64$, $p = .15$. L'assimilation tend à être plus forte pour les mots que les non-mots. Cette tendance est significative uniquement pour C1 non-voisée, $F(1,7) = 8.75$, $p < .05$.

durées. Les patterns de durée pour la voyelle V précédent C1 et l'occlusion de C1 sont typiques du voisement de C1 (voyelle plus longue, $p < .005$, occlusion plus courte, $p < .05$, pour C1 voisée). Ces patterns ne changent pas globalement en condition d'assimilation. Autrement dit, les durées de V et de l'occlusion de C1 résistent à l'assimilation. Ce n'est pas le cas de la durée de relâchement de C1. Le relâchement est globalement plus long pour C1 non-voisée que voisée, $p < .05$. Le contexte assimilant

neutralise cette tendance, sans l'inverser, comme le montre la significativité du facteur $vc2$, $p < .0005$. Tout ceci vaut aussi bien pour les non-mots que les mots. Notons que les durées sont plus longues pour les non-mots que les mots (V : $75 > 67$ ms, $p < .005$; occlusion : $70 > 64$ ms, $p < .05$; relâchement : $19 > 14$ ms, $p < .05$), reflétant un débit plus lent pour les non-mots. La Table 3 montre les durées de V et de C1 (occlusion et relâchement).

		NV-NV	NV-V	variation	V-V	V-NV	variation
voyelle V	mots	64	63	- 1	71	69	- 2
	non-mots	74	74	0	78	75	- 3
occlusion de C1	mots	69	67	- 2	60	60	0
	non-mots	70	73	+ 3	66	69	+ 3
relâchement de C1	mots	20	16	- 4	10	13	+ 3
	non-mots	25	20	- 5	12	20	+ 8

TABLE 3 – durée (ms) de la voyelle V, de l'occlusion de C1 et du relâchement de C1.

intensités. Globalement, l'intensité de l'occlusion est plus forte pour C1 voisée que non-voisée, $p < .0005$. Ceci vaut aussi pour le relâchement, $p < .01$. Le contexte assimilant neutralise cette tendance sans toutefois l'inverser, comme le montre la significativité du facteur $vc2$, $ps < .005$ (sauf pour le relâchement de C1 dans les non-mots). La Table 4 montre les intensités moyennes pour V et l'occlusion de C1.

		NV-NV	NV-V	variation	V-V	V-NV	variation
occlusion de C1	mots	61	67	+ 6	74	65	- 9
	non-mots	60	65	+ 5	72	65	- 7
relâchement de C1	mots	63	68	+ 5	73	60	- 13
	non-mots	66	66	0	72	63	- 9

TABLE 4 – intensité moyenne (dB) de l'occlusion de C1 et du relâchement de C1.

3 Discussion

Nos résultats suggèrent d'abord que l'assimilation de voisement inter-mot en français, pour des contacts entre occlusives, est catégorielle dans le sens du voisement mais graduelle dans celui du dévoisement pour ce qui est du v-ratio ou du HNR. (Nos données montrent d'autre part l'équivalence de ces deux mesures.) En termes de v-ratio ou HNR *moyens*, l'assimilation est d'ampleur équivalente dans les deux sens. Pour les indices secondaires de voisement, nous trouvons que les durées d'occlusion de la consonne C1 et de la voyelle précédente ne sont pas affectées par l'assimilation. Les mêmes "traces" du voisement sous-jacent ont été trouvées dans d'autres études (Goldrick et Blumstein, 2006 ; Snoeren et al., 2008). Par contre, les autres indices secondaires "locaux", durée de relâchement, intensité moyenne de l'occlusion et du relâchement sont affectés par l'assimilation dans le sens du voisement opposé.

Nos résultats suggèrent ensuite que l'assimilation a des caractéristiques quantitatives et qualitatives quasi-identiques pour les mots et les non-mots. Nous trouvons une légère différence de force d'assimilation (v-ratio ou HNR) à l'avantage des mots. Mais cette

différence est sans doute due au débit légèrement plus lent trouvé pour les non-mots (cf. Barry et Teifour, 1999). Ernestus et Baayen (2006) rapportent des résultats similaires pour la neutralisation de voisement en fin de mot (hollandais) : elle est identique pour des mots et des non-mots lus. Ces données et les nôtres renforcent donc une vision selon laquelle la production des énoncés de parole passe par l'application de règles phonologiques à des représentations phonémiques abstraites plutôt que par le rappel direct de formes de surface stockées en mémoire qui serait sensible à des statistiques de cooccurrence (Bybee, 2001).

Références

- BARRY, M. et TEIFOUR, R. (1999). Temporal patterns in Syrian Arabic voicing assimilation. In *Proceedings of the 14th ICPHS*, 2429-2432. San Francisco.
- BOERSMA, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5 (9/10), 341-345.
- BYBEE, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.
- ERNESTUS, M. et BAAYEN, H. (2006). The functionality of incomplete neutralisation in Dutch: The case of past-tense formation. In L. Goldstein, D. Whalen, C. Best (eds.), *Papers in Laboratory phonology VIII* (pp. 27-49). Berlin: Mouton de Gruyter.
- GOLDRICK, M., & BLUMSTEIN, S. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, 21, 649-683.
- GOW, D. et IM, A. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language*, 51, 279-296.
- HALLÉ, P. et ADDA-DECKER, M. (2007). Voicing assimilation in journalistic speech. *Proceedings of 16th ICPHS*, 493-496. Saarbrücken.
- HALLÉ, P. et ADDA-DECKER, M. (2011). Voice assimilation in French obstruents: A gradient or a categorical process? In J. Goldsmith, E. Hume & L. Wetzels (eds.), *Tones and features: A festschrift for Nick Clements* (pp. 149-175). Berlin : De Gruyter.
- KUZLA, C., CHO, T. et ERNESTUS, M. (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics*, 35, 301-320.
- OHALA, J. (1983). The origin of sound patterns in vocal tract constraints. In P. MacNeilage (ed.), *The production of speech* (pp. 189-216). New York : Springer Verlag.
- SLIS, I. (1986). Assimilation of voice in Dutch as function of stress, word boundaries, and sex of speaker and listener. *Journal of Phonetics*, 14, 311-326.
- SNOEREN, N., SEGUI, J. et HALLÉ, P. (2006). A voice for the voiceless : Production and perception of assimilated speech in French. *Journal of Phonetics*, 34, 241-268.
- SNOEREN, N., SEGUI, J. et HALLÉ, P. (2008). On the role of regular phonological variation in lexical access: Evidence from voice assimilation in French. *Cognition*, 108, 512-521.