/R/ LENITION IN QUEBEC FRENCH : EVIDENCE FROM THE DISTRIBUTION OF 9 ALLOPHONES IN LARGE CORPORA

Lancien Mélanie[12], Hutin Mathilde[35], Stuart-Smith Jane[4], Adda-Decker Martine[25], Vasilescu Ioana[5]

[1] FNSRS, Switzerland, [2] LPP, France, [3] ILC-FNRS, Belgium, [4] GULP, UK, [5] LISN-CNRS, France

melanie.lancien@unil.ch, mathilde.hutin@uclouvain.be, jane.stuart-smith@glasgow.ac.uk,
martine.adda-decker@sorbonne-nouvelle.fr, ioana.vasilescu@lisn.upsaclay.fr

ABSTRACT

Lenition is a process whereby a segment shifts to a "weaker" variant (i.e., closer to deletion in the history of languages). Lenition is also a positional phenomenon, typically affecting intervocalic or coda consonants before post-coda or post-pausal onsets. While the lenition of stops is well studied in Romance languages, investigations about other segments are rare. We propose to fill this gap by focusing on /R/ in Québec French (QF), a variety documented to exhibit up to 9 allophones. We examine 50K+ read words from the PFC-Québec Corpus [1] that we manually annotated (for voicing, manner and place of articulation – based on perception and spectrograms). The analysis of the distribution of /R/ in different syllabic positions shows that lenited (approximantized, vocalized and non-realized) variants indeed appear in leniting positions (coda and intervocalic), thus showing that /R/ realizations in QF are not in free variation but indeed an instance of lenition.

Keywords: Lenition, French, rhotics, big data phonetics, laboratory phonology

1. INTRODUCTION

Lenition is a well-known process in historical phonetics whereby a segment shifts to a "weaker" variant. A variant is considered as "weaker" than its former realization if it can be found on the path from the former realization to complete deletion in the history of languages. Lenition and its opposite, fortition, are positional phenomena, i.e., some syllabic positions are, regardless of phonetic context, weakening positions, favoring lenition, while other positions are strengthening positions, favoring fortition, or at least resistance to lenition [2]. Typically, the intervocalic context (henceforth, V_V as /R/ in Fr. "bureau" /byRo/, *office*) and the coda position, either word-internal (V_C as in Fr. "verdit" /vɛRdi/, *turns green*) or word-final (V_# as in Fr. "dur" /dyR/, *hard*) are leniting contexts, as in (1), while the onset position, either word-initial (#_V as in Fr. "rouge" /Ruʒ/, *red*) or word-internal (C_V as in the family name "Malraux" /malRo/ *Malraux*[1]), are fortifying ones as in (2).

(1) Lat. /t/ → Fr. /∅/:
V_V = Lat. pa**t**er → pa**d**rem → pe**ð**re → Fr. père

(2) Lat. /t/ → Fr. /t/:
C_V = Lat. por**t**a → Fr. por**t**e

Studies on lenition in Romance languages however rarely focus on rhotics (to the exception of some studies on Catalan [3] or Brazilian Portuguese [4, 5, 6, 7]). The reason behind the scarcity of these studies probably lies in the fact that /R/[2] in Standard French shows little variation besides the well-known voicing assimilation (e.g., "quatre" [katχ], *four* but "cadre" [kadʁ], *frame*), arguably an instance of lenition itself [8]. However, this relative uniformity is not shared by all varieties of French, and Canadian French shows an impressive amount of allophones in the realization of the rhotic ([9, 10] for instance).

In Canada, most of the native speakers of French are found in the province of Québec (inhabitants of Québec representing 23.2% of Canada's population). The last survey published by the "Institut de la statistique du Québec" about Francophonie (2016), stated that 79% of the population of Québec were identifiable as native speakers of French, i.e., approximately 6.4M out of 8.1M inhabitants. In Québec, the population is mainly distributed along the shores of the Saint-Laurent river, which leads to the terminology Eastern/Western Québec French in sociolinguistic studies (see [10] for instance). Another group of speakers is to be mentioned here: Acadian speakers, who are gathered in the North-East part of Québec and considered belonging to a different regiolect (see [11] among others). However, according to a demographic study conducted by Laval University [12], in the last decade, 97% of the population was concentrated in 20% of the territory, out of which 80% in the cities of Sept-Îles and Montreal.

Previous works [13, 14, 15, 16] show that there are both apical and dorsal realizations of

/R/ co-occurring as allophones in Québec French (henceforth QF). The use of these allophones varies across speakers depending on their region (apical variants being a marker of Western QF and uvular fricatives of Eastern QF) but also their age (younger speakers displaying close to no apical allophones), thus making /R/ a sociolinguistic marker [16] evolving over time (tending towards the disappearance of the apical variant). The articulation mode also varies (fricatives, trills, approximants...) resulting in up to 9 allophones of /R/ in Québec French [17].

In the present paper, we propose to fill the gap regarding the phonetics and phonology of /R/ in French, and in particular in Québec French. We aim to analyse 50k+ items pronounced by native QF speakers to identify their phonetic realizations and possible patterns of variation. In particular, we focus on positional factors to establish whether this variation in the realization of rhotics in QF can be considered an instance of lenition.

In the following, we present our data and methodology in Section 2, our results in Section 3 and our conclusions and discussion in Section 4.

2. METHOD

2.1. Corpus

For this study, we use a subset of the PFC-Québec corpus [1, 18], in which speakers had to read 2 word lists (max. 304 items)[3] specifically designed to bring out commonly discussed variation patterns in French phonology.

The corpus covers 32 locations, though only 29 were selected for this paper (we ruled out the data from incomplete list readings or from illiterate speakers). Among them, 13 are Western Laurentian, 12 Eastern Laurentian, and 4 Acadian, for a total of 396 speakers (206 females, 190 males) born between 1921 and 1999.

2.2. Coding

For this study, we manually coded the 58,083 exemplars of /R/ occurring in the recorded lexical words of the corpus. The purpose of this coding was to differentiate the 9 allophones of /R/ displayed by QF speakers. A trained phonetician listened to the recordings and looked at the spectrograms and oscillograms generated by Praat [19] in order to segment and code the phone. The choice of allophone was made according to both the perception of the annotator and a set of features:

- apical tap [ɾ] or trill [r][4]: Perceived as apical + 1 to 3 flappings in the spectrogram,
- uvular trill [ʀ]: Perceived as uvular + 2 to 3 flappings in the spectrogram,
- voiceless fricative [χ]: Friction noise and no F0 or voicing bar in the spectrogram,
- voiced fricative [ʁ]: Friction noise in the spectrogram and F0 detected + voicing bar,
- retroflex [ɻ]: Perceived as a retroflex + formants in the spectrogram,
- approximant variant [ɹ]: Formants in the spectrogram, glide-like,
- vocalized variant [ɚ]: Very stable formants in the spectrogram, schwa-like,
- deleted: No visible trace of a phone in the spectrogram - very often in clusters.

For the purpose of our analysis, the phonological contexts in which each phone appears is also automatically coded. One of the 6 contexts listed below (by order of strength) was attributed to each token depending on its position in the word:

- word-initial onset: #_V (e.g., rouge /Ruʒ/, *red*),
- part of an onset TR-cluster: C_V (e.g., bras /bRɒ/, *arm*),
- intervocalic onset: V_V (e.g., bureau /byRo/, *office*),
- word-internal coda: V_C (e.g., verdit /vɛRdi/, *turns green*),
- word-final coda: V_# (e.g., dur /dyR/, *hard*),
- part of a coda TR-cluster: C_# (e.g., quatre /katR/, *four*).

2.3. Data

For the present study, we excluded retroflex variants (only 21 tokens in a reduced number of word-forms) as well as word-final clusters (almost exclusively /tR/ clusters thus biasing the results), which results in a subset of 55,672 tokens.

The remaining variants are grouped into 4 categories according to strength of /R/[20]: Apical (apical trills [r] and taps [ɾ]), Uvular (uvular trills [ʀ]), Fricatives ([ʁ] and [χ]), and Lenified ([ɹ], [ɚ] and non-realized). Counts are given in Table 1.

3. RESULTS

In this section, we present our results regarding the distribution of our 4 /R/ categories (Apical, Uvular, Fricatives and Lenited) as a function of their syllabic position in the word. However, it appears that apical and uvular variants are not generalized among all speakers: Most speakers (n=153) use only the voiced and voiceless uvular fricatives and no

	Apical	Uvular	Fricatives	Lenited	Total
	9.9%	6%	49.1%	35%	100%
#_V	16.6%	33.1%	16.5%	10.8%	14.13%
	786	475	3917	2793	7971
	10.6%	3.4%	50.7%	35.3%	100%
C_V	44.8%	47.8%	42.8%	27.4%	36%
	2116	686	10159	7065	20026
	10.7%	1.7%	28.2%	59.4%	100%
V_V	25.5%	13.1%	13.3%	25.9%	20.2%
	1207	188	3169	6681	11245
	6.8%	0.5%	46.9%	45.7%	100%
V_C	5.9%	1.5%	8.1%	7.3%	7.4%
	281	21	1926	1876	4104
	2.7%	0.5 %	37.1%	59.7%	100%
V_#	7.1%	4.5 %	19.2 %	28.5%	22.1%
	338	65	4569	7354	12326
	8.5%	2.6%	42.6%	46.3%	100%
Total	100%	100 %	100 %	100%	100%
	4728	1435	23740	25769	55672

Table 1: Distribution of variants for QF /R/ according to predicted context of lenition (top to bottom). The rates in blue are the rates of each variant in the given position, the rates in green are the rates of each variant across positions. The numbers in black are the counts of each variant in each position.

apical nor uvular variants, some speakers (n=47) use almost exclusively the apical variant[5], a few fricatives and never the uvular, while other speakers (n=49) use the apical variant only marginally and prefer the uvular. We will therefore present the results for the "Fricative speakers" (21,523 tokens) in subsection 3.1, then for the "Apical speakers" (6,111 tokens) in subsection 3.2 and for the "Uvular speakers" (6,852 tokens) in subsection 3.3.

Our statistical analyses consist in multinomial logistic mixed models (mclogit package [21] under R 4.0). We started with a full model with speaker and word variables as random effects as in (3), and used model comparison through log likelihood comparison to find the best-fit model. In this paper we present only the results for linguistic context.

(3) Formula : $Coding \sim Context * Birthyear * Sex$

3.1. Fricative Speakers

Regarding speakers using exclusively fricative and lenited variants, as can be seen in Fig.1, lenited /R/ appears more in weakening positions, i.e., intervocalic onset and both word-internal and word-final codas, than in strengthening positions, i.e., word-initial onsets ($\Delta=28.2\%$, $\Delta=7.4\%$ and $\Delta=17.0\%$ respectively).

The multinomial mixed effect model shows that there is indeed a main effect for Context ($\chi^2(4$, n=21,523)= 24.576, p<0.001).

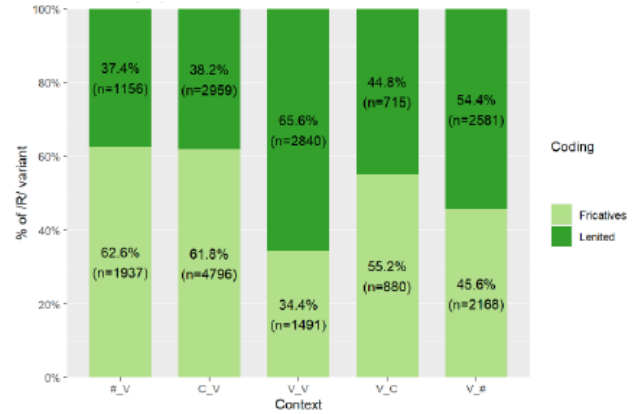Posthoc tests then show that fricatives are more

Figure 1: Rates of each variant of /R/ in QF according to predicted context for lenition (from left to right) for the subset of fricative /R/ users (n=153 speakers, n=21523 /R/).

likely to be used in every context except V_V and V_# where they are realized as often as lenited variants.

It thus seems that /R/ variation patterns qualify as lenition, although it has not yet fully expanded from word-final to word-internal coda position, which is not surprising since the latter usually undergoes change later than the former [2, 22].

3.2. Apical Speakers

Regarding speakers with at least 5% apical /R/ in their productions, Fig.2 shows that they do not display any uvular trills, and few uvular fricatives. Moreover, it is clear that word-internal and word-final codas, i.e., weakening positions, display more lenited variants than word-initial onsets, i.e., strengthening positions ($\Delta=21.7\%$ and $\Delta=58.0\%$ respectively).

The multinomial mixed effect model shows that there is indeed a main effect for Context ($\chi^2(4$, n=6,111)=17.3072, p<0.01).

Posthoc tests then show that all the tested pairs differ in each context (p<0.001). Thus, V_# is the context that triggers the most lenited variants (p<0.001), followed by V_C. Very few lenified variants are used in the other contexts, including the intervocalic position, where apicals are largely preferred. It therefore seems that Apical speakers also display a variation pattern that qualifies as lenition.

3.3. Uvular Speakers

Regarding speakers with at least 5% uvular trilled /R/ in their productions, as can be seen in Fig.3, they display very few apical variants and many uvular
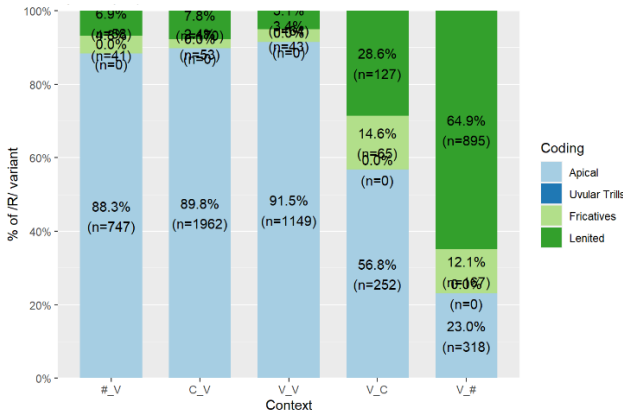
Figure 2: Rates of each variant of /R/ in QF according to predicted context for lenition (from left to right) for the subset of apical /R/ users (n=47 speakers, n=6,111 /R/).

fricatives. The rates of lenited variants are higher in all positions than those in Fig.2, suggesting that these speakers have a stronger tendency to lenition than Apical speakers. However, there are still more lenited variants in weakening positions, i.e., word-internal coda, word-final coda and even intervocalic positions, than in strengthening positions, i.e., word-initial onsets (Δ=22.9%, Δ=57.2% and Δ=28.9% respectively). One additional interesting observation is that rates of uvular fricatives seem rather constant across all positions while the variant losing ground *vis-à-vis* lenited variants is the uvular trill, which can therefore be identified as a properly "strong" variant rather than as a "default" variant.
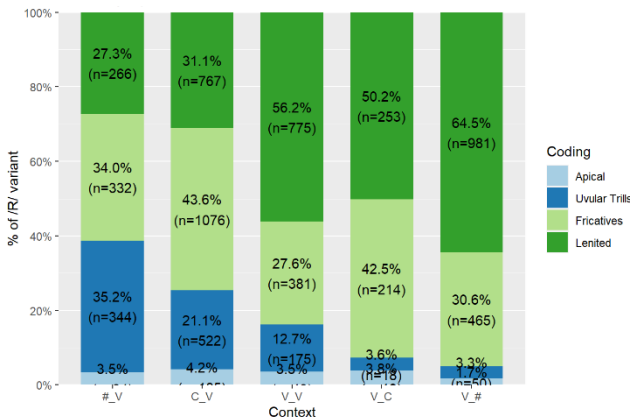

Figure 3: Rates of each variant of /R/ in QF according to predicted context for lenition (from left to right) for the subset of uvular trill /R/ users (n=49 speakers, n=6,852 /R/).

The multinomial mixed effect model shows that there is indeed a main effect for Context ($\chi^2(4$, n=6,852)=39.105, p<0.001).

Posthoc tests then show that all rates are significantly different, except for Uvular Fricatives *vs* Uvular Trills in word-initial onset #_V positions, Uvular Fricatives *vs* Lenited variants in intervocalic V_V and word-internal coda V_C positions, and finally Apical *vs* Uvular Trills in word-internal V_C and word-final V_# coda positions. In sum, uvular fricatives are equally distributed with uvular trills in strong onset position but with lenited variants in weak intervocalic and internal coda positions, while apical and uvular trills are equally distributed in both word-internal and word-final coda position.

For these speakers too, then, the distribution of /R/ realizations qualifies as lenition, where the lenited variants are in quasi complementary distribution with a strong uvular variant.

4. CONCLUSION AND DISCUSSION

In the present paper, we analyze the distribution of 55,672 realizations of /R/ in Québec French with regards to their syllabic positions. We find that "weaker", i.e., approximant-like, vowel-like or even non-realized variants of /R/ indeed occur in syllabic positions identified by historical linguists as weakening positions, thus advocating in favor of an ongoing lenition of /R/ in this variety of French. Our work corroborates the idea of intervocalic and coda positions being triggering positions for lenition overall in Romance languages [2, 22]. It also sheds light on the fact that QF speakers can be grouped according to their preferred "strong" allophone of /R/ (i.e., apical, uvular, or fricative) and that these groups all behave similarly regarding lenition.

In future work, even though single transcribers have been shown to be internally consistent [23], additional corroboration from random sample coding will be undertaken. Future studies will also allow us to explore the geo- and socio-linguistic factors at play in the lenition of /R/ and even compare these patterns with those found in other varieties such as Metropolitan French. We also intend to provide observations regarding each /R/ variant separately, and to include in our models frequency and various other phonetic characteristics (duration, formants, speech-rate, etc.), which have been shown to correlate with lenition [24, 25].

5. ACKNOWLEDGEMENTS

consortium.

6. REFERENCES

[1] M.-H. Côté, "Le projet PFC et la géophonologie du français laurentien," J. Durand, G. Kristoffersen, and B. Laks, Eds. Nanterre : Presses Universitaires de Paris Ouest, 2014, pp. 173–198.

[2] P. Ségéral and T. Scheer, "Positional factors in lenition and fortition," in *Lenition & fortition*, J. B. de Carvalho, T. Scheer, and P. Ségéral, Eds. Mouton de Gruyter, 2008, pp. 131–172.

[3] D. Recasens and A. Espinosa, "Phonetic typology and positional allophones for alveolar rhotics in catalan," *Phonetica*, vol. 64, pp. 1 – 28, 2007.

[4] A. H. P. Silva and E. C. Albano, "Brazilian Portuguese rhotics and the phonetics/phonology boundary," in *ICPhS-14*, 1999, pp. 2211–2214.

[5] A. H. P. Silva, "Towards a dynamical representation for gradient allophony of Brazilian Portuguese rhotics," in *ICPhS-15*, 2003, pp. 1863–1866.

[6] I. Rennicke, "Variation and change in the rhotics of brazilian portuguese," Ph.D. dissertation, 2015.

[7] P. Howson and A. Kochetov, "Gestural Lenition of Rhotics Captures Variation in Brazilian Portuguese," in *Proceedings of Interspeech 2018*, 2018, pp. 182–186.

[8] E. R. Webb, "Minimalism and French /R/: Phonological representations in phonetically based phonology," *Journal of French Language Studies*, vol. 19, no. 1, pp. 87–115, 2009.

[9] G. Dulong and G. Bergeron, "Le parler populaire du Québec et de ses régions voisines," *Atlas linguistique de l'Est du Canada. Collection Études et dossiers Québec*.

[10] C. Verreault and T. Lavoie, "«La langue de nos gens» du Père Laurent Tremblay: Une première synthèse sur la variation géolinguistique du français parlé au Québec au début des années 1940," *Langues et linguistique*, no. 25, pp. 145–212, 1999.

[11] P. Balcom, L. Beaulieu, G. R. Butler, W. Cichocki, and R. King, "The Linguistic Study of Acadian French," *Canadian Journal of Linguistics/Revue canadienne de linguistique*, vol. 53, no. 1, pp. 1–5, 2008.

[12] "La situation démolinguistique du Québec," https://www.axl.cefan.ulaval.ca/amnord/Quebec-2demo.htm.

[13] J. Clermont and H. Cedergren, "Les "R" de ma mère sont perdus dans l'air," *Le français parlé. Études socio-linguistiques*, pp. 13–28, 1979.

[14] L. Santerre, "Les (r) montréalais en régression rapide," in *Les français régionaux du Québec, Numéro spécial de Protée VII (2)*, L. Lavoie, Ed., 1979, pp. 117–132.

[15] ——, "Des [r] montréalais imprévisibles et inouïs," *Revue québecoise de linguistique*, vol. 12, pp. 77–96, 1982.

[16] G. Sankoff and H. Blondeau, "Language change across the lifespan: /r/ in Montreal French," *Language*, pp. 560–588, 2007.

[17] M. Lancien, "More on the transition from apical to dorsal /R/ in Quebec," in *R-atics'7, International Congress on Rhotic Sounds*, 2021.

[18] J. Durand, B. Laks, and C. Lyche, "La Phonologie du français contemporain (PFC): usages, variétés et structure," C. Pusch and W. Raible, Eds. Günther Narr Verlag Tübingen, Germany, 2002, pp. 93–106.

[19] P. Boersma, "Praat: Doing phonetics by computer version 5.2. 03," *http://www. praat. org/*, 2015.

[20] E. R. Webb, "From r to ʁ via z and l: Rethinking the history of french/r/," *High Desert Linguistics Societyâ¦ Albuquerque, New Mexico*, p. 55, 2002.

[21] M. Elff, M. M. Elff, M. Suggests, and L. Yes, "Package 'mclogit'," 2022.

[22] D. Steriade, *Phonetics in Phonology: The Case of Laryngeal Neutralization*. UCLA, 1997.

[23] J. Stuart-Smith, "A sociophonetic investigation of postvocalic/r/in glaswegian adolescents," in *Proceedings of the 16th International Congress of Phonetic Sciences*, 2007, pp. 1307–10.

[24] M. Pukli, "L'effet de la fréquence lexicale sur les réalisations des rhotiques en Ecosse," in *XXXIIe Journées d'Études sur la Parole*, Aix en Provence, France, 2018. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02530429

[25] Y. Wu, M. Hutin, I. Vasilescu, L. Lamel, and M. Adda-Decker, "Extracting linguistic knowledge from speech: A study of stop realization in 5 Romance languages," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, Jun. 2022, pp. 3257–3263. [Online]. Available: https://aclanthology.org/2022.lrec-1.348

---

[1] In French, /R/ is almost never an internal onset since sequences of an obstruent followed by /R/ usually result in a so-called *muta cum liquida*, or CR-cluster.

[2] Throughout the paper, we use a capital R to note the rhotic of French without posing an underlying representation. This notation is not the same as the one for the uvular trill ʀ, and is merely a way to refer to some sort of rhotic archiphoneme.

[3] Full lists can be found on the project's webpage https://public.projet-pfc.net

[4] Apical taps and trills are usually confused and merged as a single "apical" macro-category in classical studies on Quebec French, thus we confuse them in this paper but will address the possibility of the tap being a lenited variant of the trill in further analyses.

[5] A threshold was set to 5% apical realizations (for Apical speakers) and 5% uvular realizations (for Uvular speakers) to ensure that the use of these /R/ variants was not an articulatory "accident" but an actual use of the variant. The groups of Fricative, Apical and Uvular speakers are mutually exclusive.