Tracking /r/ Deletion: Forced Alignment of Pronunciation Variants and Sociophonetic Insights into Post-Obstruent Final /r/ in French

Anisia Popescu[1], *Lori Lamel*[1], *Marc Evrard*[1], *Ioana Vasilescu*[1]

[1]LISN, CNRS, Université Paris-Saclay, France

anisia.popescu@lisn.fr

Abstract

This paper investigates post-obstruent final /r/ deletion in French (e.g., livre [liv(ʁ)], votre [vot(ʁ)]) using two large corpora and phoneme-level forced alignment to analyze both pronounced and deleted /r/ variants. The study examines factors influencing /r/ deletion, including phonetic context, speaker profession, gender, age, and speaking style. Results show that right-phone context, age, and speaking style significantly affect /r/ deletion, with specific patterns also observed for the preceding phone and speaker profession. Gender has no effect. With over 390 speakers producing 14,167 individual /r/ tokens, the study provides robust insights into phonetic and social factors shaping final post-obstruent /r/ deletion. The combination of speech technology, NLP methods, and linguistic analysis of large-scale naturalistic data improves our understanding of this phenomenon, challenging assumptions about gender's relevance in this phenomenon.

Index Terms: sociophonetics, corpus linguistics, forced alignment, pronunciation variants

1. Introduction

Final post-obstruent /r/ deletion is a well-known phonotactic phenomenon in French. Words ending in an obstruent (/p,t,k,b,d,g,f,v/) followed by a final rhotic /r/ (e.g., *votre* [vɔt(ʁ)] 'your', *livre* [liv (ʁ)] 'book', *concombre* [kɔkɔb (ʁ)] 'cucumber') violate the Sonority Sequencing Principle (SSP), which posits that there should be a sonority rise or plateau between the syllable peak (e.g., vowel) and any member of the syllable [1]. However, rhotics are more sonorant than obstruents, thus violating the SSP. Three adaptive strategies to this violation have been described [2]: (1) resyllabification, which occurs only if the next word has an initial vowel (e.g., *votre adresse* [vot.ʁa.dʁes]) ' your address), (2) the insertion of an epenthetic schwa after the final /r/ (e.g., *votre carte* [vo.tʁə.kaʁt] 'your card') and (3) the deletion of the final rhotic (e.g., *votre carte* [vo.kaʁt] 'your card'). The third strategy, final /r/ deletion, has attracted significant attention and is recognized as a sociolinguistic marker [3, 4, 5, 2]. This phenomenon is particularly interesting as it reflects not only phonotactic patterns but also social variation, making it an important marker of speech style and identity.

Several factors have been identified as influencing /r/ deletion, with some consistently recognized across studies. One commonly accepted factor is the right-phone context: /r/ is deleted at higher rates when followed by a consonant [2, 6, 5, 7, 8]. Listeners also compensate more than twice for deletion in pre-consonantal context [9]. Left-phone context has received less attention. Speech style and social class also play a significant role: conversational speech, as opposed to formal regis-

ters triggers more deletions [10, 6], and working-class or blue-collar speech exhibits higher deletion rates [5, 2]. Dialect and geographic origin further contribute to these patterns. Withing France, speakers from Nancy delete /r/ more frequently than those from Rennes [5]. Canadian French shows higher deletion rates than Metropolitan French [11], while Swiss French speakers surpass both French and Belgian French speakers in deletion frequency [2]. Similarly, Guadeloupean French exhibits higher deletion rates than Parisian French [7]. The impact of age and gender remains contested, with some studies suggesting they play a role, while others find no significant effect, or one that is modulated by other factors such as geographical origin.

The present paper investigates /r/ deletion in final post-obstruent position by utilizing forced alignment to analyze pronunciation variants within two large corpora. We focus on both phonetic factors, such as the left- and right-phone context, and social factors, including speaker identity, profession (as a proxy for social class), age, gender and speech style. The strength of this study lies in its use of a substantial dataset: approximately 150 hours of speech, originally recorded for broadcast news purposes, rather than linguistic analysis. The two corpora include both formal and less formal speech, offering a rich context for examining sociolinguistic patterns. The analysis covers 391 distinct speakers and over 14,000 tokens, providing a robust basis for our findings.

2. Method

We analyzed two large corpora of French naturalistic speech: the ESTER corpus [12], which contains formal journalistic speech, and the ETAPE corpus [13], which includes less formal journalistic speech, such as conversations and debates. Together, these corpora comprise approximately 150 hours of speech (100 hours from ESTER and 50 hours from ETAPE).

We processed the data using an automatic speech recognition (ASR) system to force-align speech at the phoneme level. Manual orthographic transcriptions were matched with a pronunciation dictionary, which we enriched with pronunciation variants. Specifically, /r/ could be realized as an empty phone (deletion), /rə/, or /r/. Figure 1 shows three possible productions of the word *votre* 'your'. A small subset of the data, approximately 100 randomly selected tokens, was examined, confirming a classification accuracy above 96%.

From the alignment output, we extracted key information, including pronunciation type (canonical: /r/ present; non-canonical: /r/ deleted), left- and right-context phones, speaker identity, speaker gender, and year of broadcast. For each speaker, we sourced additional metadata for public figures, including their profession, nationality, and birth year. Only French native speakers with complete metadata and who pro-
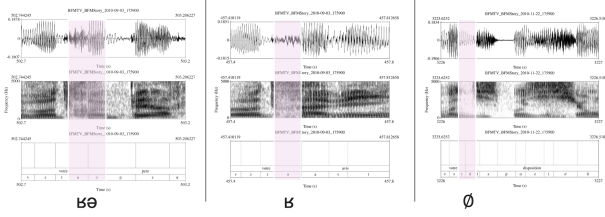
Figure 1: *Pronunciation variants of the word votre 'your' detected through phone-level forced alignment. Highlighted in pink are the regions of interest on the spectrogram*

duced a minimum of 5 observations were retained, resulting in a dataset of 391 speakers and 14,167 individual /r/ tokens.

Age was determined as the difference between the broadcast year and the speaker's birth year and was categorized into two groups: younger (≤40 years) and older (>40 years) speakers. To account for professional background, which has been used as an indicator for social class [14], we defined 18 occupational categories. Given the distribution of professions in our dataset, a strict classification into white- and blue-collar or working and middle-class categories, as used in previous studies [5, 2], was not feasible, as a majority of professions in our dataset fall within the white-collar domain. To ensure a more balanced dataset for analysis, we also included a broader "Other" category for professions represented by a maximum of four speakers (e.g., priests, architects, military officers). This structured approach allowed for more meaningful comparisons across professional backgrounds.

2.1. Statistics

Both exploratory and confirmatory analyses were conducted. In the exploratory we examined correlations between each predictor — previous phone, following phone, profession, age group, gender and corpus — and production type (canonical vs. non-canonical). For the confirmatory analysis, we fitted logistic mixed-effects models using the *lme4* package [15], incorporating the same predictors as in the exploratory analysis. Random intercepts for speaker ID and word were included as random factors. The *bobyqa* optimizer [16] was used to improve model convergence. Model selection was based on the Akaike Information Criterion (AIC) coefficient. To enhance interpretability, the model's output of log-odds was transformed into probabilities using the logistic function.

3. Results

The results will be presented in two stages: first, the explanatory analysis, followed by the confirmatory one.

3.1. Explanatory analysis

The distribution of /r/ tokens in the dataset, shown in Table 1, is notably unbalanced across phonetic context (left- and right-context), age groups, and corpora. For instance, in the left-phone context, the tokens are highly skewed, with alveolar stops /t/ (9,541 tokens) and /d/ (2,464 tokens) being much more frequent than velar stops /g/ (25 tokens) and /k/ (111 tokens), who are rare. Similarly the right context shows an imbalance with glides (674 tokens), and pauses (352 tokens) being less frequent than consonants, vowels or epenthetic schwas which all have over 3,000 occurrences. An age-related disparity can also be

observed with younger speakers being less represented. The imbalance between corpora is less pronounced with the ETAPE corpus (8878 tokens) having higher counts of /r/ tokens than the ESTER corpus (5796 tokens).

Table 1: *Distribution of /r/ tokens by phonetic context, age group, and corpus*

Left context	Count	Left context	Count
d	2,464	t	9,541
b	1,069	p	216
v	946	f	302
g	25	k	111
Right context	**Count**	**Age & Corpus**	**Count**
Consonant	4,871	Older	11,425
Vowel	3,149	Younger	3,249
Glide	674		
Pause	352	ESTER	5,796
Schwa	5,628	ETAPE	8878

The distribution of observations by speaker and profession is also unbalanced, with Journalists (5,298) and Politicians (3,550) making up the largest groups. TV/Radio personalities (2,171) also form a significant category. In contrast, professions related to Arts (65), Diplomacy (44), and Music (39) are much less represented. Other notable groups include Film Industry (605), Academics (350), Business Executives (300), and Lawyers (300). This imbalance is to be expected since the corpus consists of broadcast news data. These discrepancies point to the dataset's unbalanced nature, which may influence the analysis of /r/ token distribution.
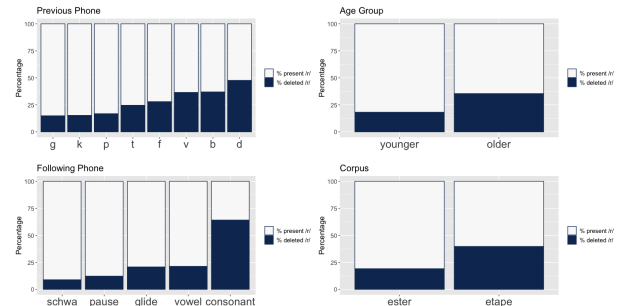

Figure 2: *Percentages of non-canonical (blue) vs. canonical (white) productions as a function of phonetic context, age group, and corpus.*

Figure 2 shows the percentage of non-canonical (blue) vs. canonical (white) /r/ productions for phonetic context, age group, and corpus. The observed patterns align with predictions, except for age, which shows the opposite trend. For right-context effects, /r/ deletion is highest before a consonant, likely to avoid a three-consonant sequence, and lowest before an epenthetic schwa, both reflecting known reparation strategies. For left-context effects, deletion increases when /r/ follows a consonant with a constriction in the front of the vocal tract and decreases with a back constriction, likely due to ease of articulation. Finally, as expected, a more informal speaking style (ETAPE corpus) corresponds to higher /r/ deletion rates. For age, however, we would have expected younger speakers to
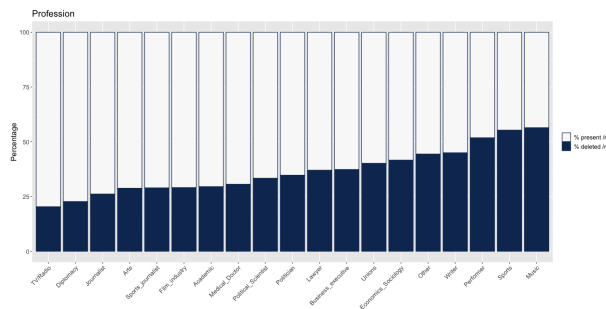
Figure 3: *Percentages of non-canonical (blue) vs. canonical (white) productions as a function of speakers' Profession.*

exhibit higher levels of /r/ deletion, the observed pattern is however reversed. For gender differences (not shown on Figure 2) the gap is smaller, with female speakers producing 27.5% of non-canonical forms (deleted /r/), and male speakers 31.8%.

Figure 3 illustrates /r/ deletion rates as a function of speakers' profession. The results reveal a continuum, with journalists exhibiting the lowest deletion rates and performers, athletes, and musicians showing the highest rates of deletion, reaching 50%. This pattern may reflect formality, with /r/ deletion serving as a marker of more informal speech. These results illustrate correlations that are found in the data. The next section will model the data with mixed-effects regression modeling.

3.2. Confirmatory analysis

The reference values for the logistic regression model were set to correspond to the levels associated with the highest probabilities of non-canonical productions observed in the explanatory analysis (i.e., left context /d/, right context "consonant", age — "older" corpus "etape", gender — "male" and Profession "Music"). The model has a marginal effect size of (delta) R^2= 0.27, representing the variation explained by the fixed effects alone and a conditional effect size of (delta) R^2=0.46, which accounts for the combined influence of both fixed and random effects. The resulting intercept (Est. 2.82599, z-value = 3.396, p<0.01) indicates that when all predictors are set to their reference levels, the probability of non-canonical pronunciation is 94%. Figures 4 and 5 show the predicted probabilities for phonetic context (left and right), age, and corpus. While results differ from those presented in the previous section, the overall patterns remain similar. For the left-phone context, significant drops in probability (P) of non-canonical productions are observed for the obstruent /f/ (P $\sim$ 86%, p<0.05), /t/ (P $\sim$ 89%, p<0.001) and /k/ (P $\sim$ 74%, p<0.001). In the right-phone context, all conditions (vowel, glide, pause, schwa) exhibit highly significant drops (p<0.001) in the probability of non-canonical pronunciation compared to the consonant case. For the case of the epenthetic schwa (with all other variables held constant), the probability of /r/ deletion is half that of when the next word starts with a consonant (P $\sim$ 47%, p<0.001). Regarding speaking style, as defined by the corpus, non-canonical productions are less probable in more formal speech (ESTER corpus) with probabilities dropping to P $\sim$ 89% (p<0.001) compared to the more informal ETAPE corpus, which aligns with expectations. The confirmatory analysis also supports the opposite trend for age, with younger speakers showing significantly lower probability of /r/ deletion (P $\sim$ 91%, p<0.01) compared to older speakers (P $\sim$ 94%), with all other variables held constant, con-

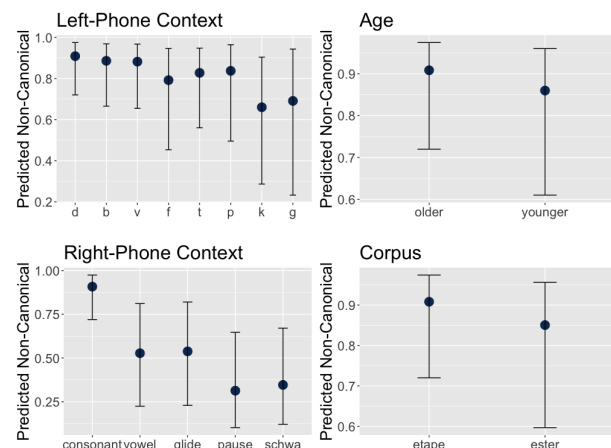trary to the predicted pattern. No significant effect of gender was found (p $\sim$ 0.39).

Figure 4: *Predicted probabilities of non-canonical production for different predictors (phonetic context, age, corpus) based on the model output.*

Regarding professions, Table 2 contains the output of the logistic regression mixed effects model. Probability drops relative to the reference-based levels are also given for each variable. Figure 5 illustrates the predicted probabilities of non-canonical production for different professional categories relative to the base reference variables.

Table 2: *Transformed probability drop (P drop) based on model output log-odds for profession effects on /r/ deletion, relative to the intercept of the base reference levels = Music*

Profession	P Drop	Log-Odds	p-value
Sports	17%	-1.576	0.086
Performer	26%	-1.028	0.196
Writer	19%	-1.391	0.079
Other	16%	-1.625	0.044 *
Economics/Sociology	15%	-1.708	0.036 *
Lawyer	10%	-2.116	0.008 **
Politician	13%	-1.883	0.011 *
Political Scientist	14%	-1.761	0.033 *
Medical Doctor	16%	-1.624	0.047 *
Academic	14%	-1.781	0.026 *
Film Industry	14%	-1.749	0.026 *
Arts	17%	-1.541	0.085
Journalist	11%	-2.072	0.005 **
Diplomacy	9%	-2.225	0.030 *
TV/Radio	11%	-1.999	0.009 **

Furthermore, to enable comprehensive analyses across different professional categories, we redefined the reference level, which was initially set to musicians. Consistent with the findings from the explanatory analysis, the results indicate that professional categories associated with less formality — such as musicians, performers (actors, dancers, comedians), and athletes — exhibit significantly higher probabilities of non-canonical productions compared to journalists, politicians, diplomats and TV/Radio hosts.
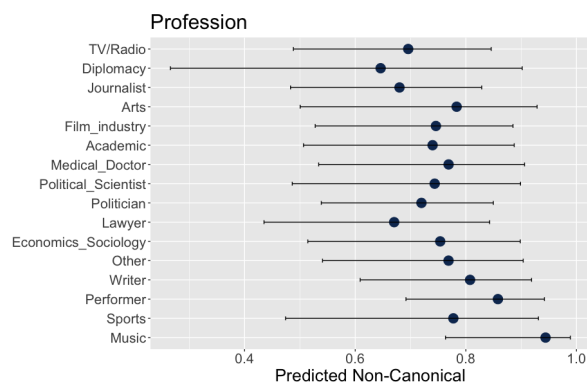
Figure 5: *Predicted probabilities of non-canonical production for different Profession based on the model output.*

4. Discussion

The present study investigated final post-obstruent /r/ deletion in a large corpus comprising over 14,000 tokens from 391 different speakers. We considered several factors, including phonetic context (both left- and right-phone contexts), speech style, and speaker characteristics such as age, gender, and professional background.

The results are consistent with patterns observed in previous research, particularly with regard to the right-phone context. In line with prior studies, /r/ deletion was more common when the following word began with a consonant. This finding aligns with the third strategy outlined in the introduction to address the SSP violation. Additionally, the second strategy — introduction of an epenthetic schwa — was also observed in 38% of cases. We were unable to test the remaining strategy, resyllabification, due to the lack of syllabification and prosodic annotations in the corpus. The left-phone context revealed some significant trends as well, with /r/ deletion being more likely following a velar consonant. The most deletions were observed after the voiced stops /d/ and /b/, followed by the fricatives /f/ and /v/. This pattern contradicts previous findings, which suggest that deletions are more common with fricatives than stops [17]. However, these results were recorded for a specific dialect, Vimeu French, which may account for the difference. As expected, speech style — represented by the corpus from which the tokens originated — played a significant role, with /r/ deletion being more likely in conversational speech contexts. In contrast, gender did not have a significant impact on the likelihood of /r/ deletion in our dataset. Age, however, was a significant factor, with speakers over the age of 40 being more likely to delete final /r/. Previous studies have reported conflicting findings regarding the role of age in /r/ deletion, and our study's different classification system and data distribution may have influenced these results. Future analyses will explore potential interactions between age and other factors.

Another factor we examined in this study was professional background. While our data could not permit a binary categorization of professions, such as working versus middle class, as done in previous studies [5, 2], our findings suggest that professional backgrounds exist along a continuum. Even within this more nuanced framework, we observe marked distinctions between categories. One explanation for these patterns is that individuals in professions such as broadcast journalism, diplomacy, and politics tend to produce more formal speech due to the structured and professional nature of their communication

contexts. In contrast, performers and athletes, who often speak in a more informal, personal capacity outside of strictly professional settings, exhibit less formal language use and produce more non-canonical tokens. This distinction highlights the role of professional background in shaping language patterns. However, profession alone may not be sufficient to fully explain variation. To better understand these differences, additional speaker-specific information and metadata, such as place of origin, upbringing, and level of education, are needed to provide a more comprehensive picture of the factors influencing sociolinguistic variation. Expanding the metadata available for speakers is crucial, though this remains a challenge. From our original dataset of over 1900 speakers, we were left with 391 after excluding those whose identities were unknown, lacked metadata, or did not provide sufficient observations. Furthermore, the unbalanced nature of our corpus — stemming from the large and heterogeneous dataset — poses a limitation, yet this is a common disadvantage when working with such extensive data.

Further analyses of the data should incorporate additional factors that can be extracted from the acoustic signal, such as speech rate and the duration of phones. The type of /r/, which exhibits considerable variation in French [18, 19, 20], could also be a contributing factor influencing /r/ deletion.

5. Conclusion

In summary, this study provides valuable insights into the factors influencing final post-obstruent /r/ deletion. Our findings highlight the importance of phonetic context, speech style, and speaker characteristics, such as age and professional background, in shaping this linguistic phenomenon. While certain patterns align with previous research, others present novel insights, particularly with regard to the left-phone context and the role of professional background. Further research incorporating additional variables and refining speaker metadata is needed to fully capture the complexities of sociolinguistic variation in French /r/ deletion.

6. Acknowledgments

7. References

[1] J. Blevins, *Handbook of phonologica theory*. London: Blackwell, 1995, ch. The syllable in phonological theory, pp. 206–244.

[2] M. Avanzi, "Vot' artic' est formidab': une étude multifactorielle de la chute des liquides post-obstruantes finales de mot en francais," *Journal of French Language Studies*, pp. 1–31, 2023.

[3] B. Laks, "Contribution empirique à l'analyse socio-différentielle de la chute des /r/ dans les groupes consonantiques finals," *Langue Française*, vol. 34, pp. 109–125, 1977.

[4] P. Guiraud, *Le français populaire*. Paris: PUF, 1965.

[5] Z. Boughton, "Social class, cluster simplification and following context: Sociolinguistic variation in word-final post-obstruent liquid deletion in french," *French Language Studies*, vol. 25, pp. 257–286, 2015.

[6] Y. Wu and M. Adda-Decker, "Dstribution and deletion of /ʁ/ in fluent speech," *Studii de lingvistica*, vol. 11, pp. 39–53, 2021.

[7] E. Pustka, "Le conditionnement lexical de l'élision des liquides en contexte post-consonantique final," *Langue française*, no. 169, pp. 257–286, 2011.

[8] T. Pooley, *Chtimi: The Urban Vernaculars of Northern France.* Clevedon: Multilingual Matters, 1996.

[9] S. Peperkamp and A. Zurita, "Compensation for french liquid deletion during auditory sentence processing," in *Proc. INTER-SPEECH 2019 – 20th Annual Conference of the International Speech Communication Association*, Graz, Austria, 2019, pp. 1951–1955.

[10] Y. Wu, C. Gendrot, M. Adda-Decker, and C. Fougeron, "Post-consonantal word-final /r/ realization in french: contributions of large corpora," in *Proc. of the International Congress of Phonetic Sciences ICPhS219*, Melbourne, Australia, 2019.

[11] P. Milne, "The variable pronunciations of word-final consonant clusters in a force aligned corpus of spoken french," Ph.D. dissertation, University of Ottawa, 2014.

[12] S. Galliano, E. Geoffrois, G. Gravier, J. Bonastre, D. Mostefa, and K. Choukri, "Corpus description of the ester evaluation campaign for the rich transcription of french broadcast news," in *Proc. LREC 2006*, vol. 6, 2006, pp. 315–320.

[13] G. Gravier, G. Adda, N. Paulson, M. Carré, A. Giraudel, and O. Galibert, "The etape corpus for the evaluation of speech-based tv content processing in the french language," in *Proc. LREC 2012*, 2012.

[14] J. Chambers, *Sociolinguistic Theory: Variation and its Social Signficance*. Oxford: Wiley-Blackwell, 2009.

[15] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[16] M. Powell, "The bobyqa algorithm for bound constrained optimization without derivatives," *Technical Report, Department of Applied Mathematics and Theoretical Physics*, 2009.

[17] A. Villeneuve, "Word-final cluster simplification in vimeu french," *University of Pennsylvania Working Papers in Papers in Linguistics*, vol. 15, no. 2, pp. 133–144, 2010.

[18] C. Gendrot, B. Kuehnert, and D. Demolin, "Aerodynamic, articulaotry and acoustic realization of french /r/," in *Proc. of the 18th International Congress of Phonetic Sciences (ICPhS2015)*, 2015.

[19] C. Fougeron, "Word boundaries and contrast neutralization in the case of enchainement in french," *Papers in Laboratory Phonology IX: Change in Phonology*, vol. 9, pp. 609–642, 2007.

[20] M. Chafcouloff, "Les caractéristiques acoustiques de /j,y,w,l,r/ en français," *Travaux de l'Institut de Phonétique d'Aix*, vol. 7, pp. 7–56, 1980.