POST-CONSONANTAL WORD-FINAL /K/ REALIZATION IN FRENCH: CONTRIBUTIONS OF LARGE CORPORA

Yaru Wu, Cédric Gendrot, Martine Adda-Decker, Cécile Fougeron

POST-CONSONANTAL WORD-FINAL /ʁ/ REALIZATION IN FRENCH: CONTRIBUTIONS OF LARGE CORPORA

Yaru WU[1,2], Cédric GENDROT[2], Martine ADDA-DECKER[2], Cécile FOUGERON[2]

[1]Modèles, Dynamiques, Corpus (MoDyCo), UMR 7114, CNRS, France
[2]Laboratoire de Phonétique et Phonologie (UMR7018, CNRS-Sorbonne Nouvelle), France
{yaru.wu, cedric.gendrot, martine.adda-decker, cecile.fougeron}@sorbonne-nouvelle.fr

ABSTRACT

French Cʁ#C sequences (eg. quatre sacs /katʁ#sak/, *four bags*) tend to violate the Sonority Sequencing Principle. This study investigates the realization of Cʁ#C sequences in continuous speech using large corpora: the formal journalistic speech corpus ESTER, the conversational journalistic speech corpus ETAPE and the casual speech corpus NCCFr. Results show that the absence/presence of /ʁ/ is dependent on the absence/presence of schwa: /ʁ/ is almost never absent when schwa is inserted (eg. [katʁə#sak]). Different speech styles trigger different strategies: the more formal the speech style is, the more we observe schwa insertion and the less we observe /ʁ/ deletion (ESTER: 65% schwa insertion, 15% /ʁ/ deletion; NCCFr: 13% schwa insertion, 69% /ʁ/ deletion). Interestingly, the absence/presence of /ʁ/ tends to be related to the durations of the surrounding consonants (Cs in [C(ʁ)#C]) while schwa is not inserted, namely the absence of /ʁ/ correlates with shorter duration.

Keywords: large corpora, continuous speech, French /ʁ/

1. INTRODUCTION

The realization of consonant clusters is known to be under the influence of the Sonority Sequencing Principle (SSP) [5, 4, 6]. In French, Cʁ# cluster is one of the few clusters that violate the SSP in the canonical form (eg. the cluster /tʁ/ in "quatre" /katʁ/). In speech production, the Cʁ# cluster can be followed by a word starting with a consonant or a vowel or by a pause. If the cluster is immediately followed by a word starting with a vowel, the SSP violation of the cluster no longer persists due to cross-word resyllabification (eg. quatre enfants /katʁ#ɑ̃fɑ̃/ [ka.tʁɑ̃.fɑ̃], *four children*). A study based on French syllable structures [1] showed that less than 20% of French word tokens start with a vowel. In contrast, when the following word starts with a consonant, the size of the consonant cluster grows across the word boundary, which makes the SSP violation problem even more challenging.

Hereafter, we focus on this latter situation, where the word-final Cʁ# cluster is followed by a word starting with a consonant (i.e. Cʁ#C). Our study aims at quantifying how these Cʁ#C clusters are actually produced using large speech corpora of continuous speech. To resolve the SSP violation of these cross-word 3-consonant sequences, speakers may delete the word-final post-consonantal /ʁ/ (eg. quatre /katʁ/ [kat]) or insert a schwa (eg. quatre /katʁ/ [katʁə]) thus creating an additional syllable. They might even delete /ʁ/ while inserting a schwa (eg. [katə]).

A recent study [3] on casual French showed that at least one of the segments in CLə# (L: liquids /l/ and /ʁ/) was absent in 80.7% of the word tokens. In their analyses [3], the researchers consider schwa as being part of the reference/canonical form (eg. quatre /katʁə/) and their study is carried out on all CLə# occurrences with all post-boundary/post-lexical contexts pooled (#C and #V, eg. "quatre sacs" /katʁə#s̲ak/ (*four bags*) & "quatre ans" /katʁə#ɑ̲̃/ (*four years*)).

In our study, schwa is not considered as part of the reference/canonical pronunciation of the Cʁ# cluster (eg. quatre /katʁ/, see also [12]). Rather, schwa insertion is considered as one of the speaker's strategies to resolve the violation of the Sonority Sequencing Principle (SSP), especially in a post-lexical consonantal context (/Cʁ#C̲/, eg. "quatre sacs" /katʁ#s̲ak/ (*four bags*)).

Little is known about the distribution of the different realizations of Cʁ#C sequences in continuous speech and, in particular, whether and how these realizations vary across different speech styles. Furthermore, in the challenging case where schwa is not inserted ([C(ʁ)#C]), we propose to study whether the absence/presence of /ʁ/ correlates with the durations of its surrounding segments which may be seen as indicative of local speech rate.

2. CORPORA AND ALIGNMENT

Three large corpora of journalistic and casual French, were used for this study: ESTER [7],

ETAPE [10] and NCCFr [14]. All selected material is representative of "standard" French (as spoken in France) with no marked regional accent. The ESTER corpus [7] consists of 100h of radio broadcast news shows in French (only broadcast news from France are included). The ETAPE corpus contains 42.5 hours of conversational journalistic speech in French (13.5 hours of radio data and 29 hours of TV data, including debates and free conversations). The Nijmegen Corpus of Casual French (NCCFr) is consist of 35 hours of conversations between friends.

Both automatic and manual alignments, i.e. segmentations and labellings, were used for this study. Automatic alignment is available for all three corpora (ESTER, ETAPE and NCCFr) and was carried out using the LIMSI speech transcription system [9] in forced alignment mode. Starting with manual orthographic transcriptions of the speech files, forced alignment automatically matches speech segments with optimal phonemic transcriptions, given the system's acoustic phone models and pronunciation lexicon (which includes variants with respect to presence/absence of word-final /ʁ/ and schwa). Thus, forced alignment generates word and phone boundaries using the available orthographic transcriptions and chooses the most suitable pronunciation for each word of the dictionary given the specific pronunciations produced by the speakers. The manual alignment was carried out by an experienced phonetician who makes corrections/adjustments on the existing automatic alignments.

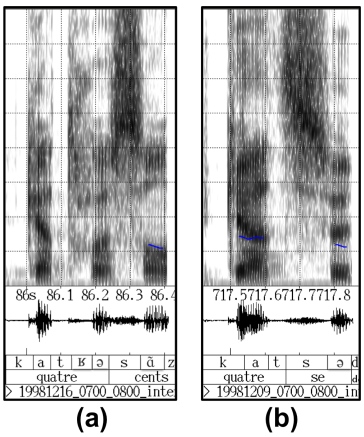
3. METHOD

This study examines the Cʁ#C sequences (eg. quatre sacs /ka**tʁ#s**ak/, *four bags*), which provoke potential SSP violation, using both manual and automatic alignments on large corpora. Note that obstruent + ʁ clusters (Cʁ#) are the only Cʁ# clusters allowed in French.

Optional /ʁ/ and schwa were included as variants in the system's pronunciation dictionary in this study. The absence/presence of /ʁ/ or schwa was thus automatically decided using forced alignment. Given that the alignment system selects the best matching variants, the aligned pronunciation tends to reflect the actual pronunciation of the speaker. Figure 1 illustrates spectrograms of two different automatically aligned productions of the word "quatre" (/katʁ/, *four*) with (Figure 1a) and without (Figure 1b) /ʁ/ aligned.

With regard to manual alignment, beyond listening, the following acoustic features were taken into account to decide on the absence/presence of /ʁ/ and schwa: periodicity and intensity of the speech signal, as well as movements of the second formant in the spectrogram. Manual alignment was produced on a subset of the ESTER corpus (30h of speech).

Figure 1: The word "quatre" (/katʁ/, *four*) with (a) and without (b) /ʁ/ aligned by the LIMSI speech transcription system.



Manual alignment was used for strategy and duration analyses. Speech style was analyzed using automatic alignment. Cohen's kappa coefficient [8] shows that manual and automatic alignments have "almost perfect agreement" (kappa = 0.832) on /ʁ/ realization and "substantial agreement" (kappa = 0.704) considering both the absence/presence of /ʁ/ and the absence/presence of schwa.

The absence/presence of /ʁ/ and that of schwa were determined by comparing the aligned production of the speakers (i.e. surface pronunciation) with the phonological transcription of Lexique380 [12], which is considered as the reference (or full) pronunciation for the analyses in this study. Doing so, aligned words can be classified into four categories: *Full*+ə, *Full*, *Full*-ʁ, *Full*-ʁ+ə (see Table 1). The *Full*+ə and the *Full*-ʁ classes are supposed to collect most of the examined items as they resolve the SSP violation. The *Full* class will gather words that are problematic with respect to SSP. The *Full*-ʁ+ə is expected to be almost empty, at least for the formal speech styles, as this pronunciation engages two strategies at the same time.

4. ANALYSES AND RESULTS

Hereafter, we analyze the realizations of Cʁ#C sequences, first on the manual alignments and then on the three corpora using automatic alignments. Generalized linear mixed models (GLMM) [11] were carried out in R [13] (package lme4 [2]) for the statistical analyses of this study.

Table 1: Comparison between the reference and the potential surface pronunciations for the word “quatre” (*four*) and the word “nombre" (*number*).

Spelling	Ref. pron.	Surface pron.	ʁ	Schwa	Surface class
quatre (*four*)	katʁ	katʁə	present	present	Full+ə
		katʁ	present	absent	Full
		kat	absent	absent	Full-ʁ
		katə	absent	present	Full-ʁ+ə
nombre (*number*)	nɔ̃bʁ	nɔ̃bʁə	present	present	Full+ə
		nɔ̃bʁ	present	absent	Full
		nɔ̃b	absent	absent	Full-ʁ
		nɔ̃bə	absent	present	Full-ʁ+ə

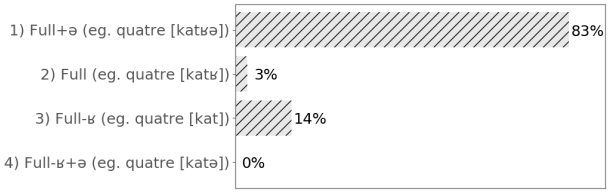
4.1. Manual alignment on the ESTER corpus

In this section, we examine the observed pronunciations using the manually checked alignments and we question a potential link between the duration of the surrounding consonantal segments of /ʁ/ and the absence/presence of /ʁ/ (when schwa is not inserted).

4.1.1. Strategies for Cʁ#C realization

Figure 2 presents different realization rates of the Cʁ# cluster while it is followed by a word starting with a consonant using manual alignment on a subset of the ESTER corpus. The most adopted

Figure 2: Realization rates of Cʁ# followed by #C on a subset of the formal journalistic corpus ESTER using 4 classes: *Full*+ə, *Full*, *Full*-ʁ, *Full*-ʁ+ə.



strategy (83%) is to add a schwa to the canonical form (eg. /katʁ/ [katʁə̲], *four*) in order to resolve the SSP violation. The second most popular (14%) strategy observed is word-final /ʁ/ deletion (eg. /katʁ/ [kat]). Interestingly, only 3% of the Cʁ# clusters are produced using the problematic form "*Full*" (eg. /katʁ/ [katʁ]) followed by a word starting with a consonant. Moreover, Cʁ# clusters are almost never pronounced using the form *Full*-ʁ+ə (eg. /katʁ/ [katə̲]).

4.1.2. Duration of the surrounding consonants and absence/presence of /ʁ/

In what follows, we focus on the subset of Cʁ#C productions where no schwa is realized. We examine the segmental duration of the consonants surrounding the /ʁ/ position.

Figure 3 (left) illustrates the duration of the preceding and the following consonants of potential /ʁ/ for Cʁ#C sequences with and without /ʁ/ realized, while schwa is not inserted. Raw duration measurements in millisecond are used in this figure, giving us a general view of how the segmental durations are distributed before any data transformation. Results show that when /ʁ/ is absent, the durations of the surrounding consonants tend to be shorter (as a more general observation, one can also note that the left (coda) consonant tends to be shorter than the right (onset) consonant).

Additional duration analyses were carried out using normalized data (Figure 3 center and right) The central part of Figure 3 shows the distribution of normalized duration of the preceding consonant. Results show that the absence of /ʁ/ corresponds to shorter durations of the preceding consonant. Likewise, the right part of Figure 3 shows the distribution of normalized duration of the following consonant. As for the preceding consonant (central figure), the absence of /ʁ/ corresponds to shorter durations of the following consonant.

The influence of the duration of the preceding and that of the following consonant on the absence/presence of /ʁ/ were taken into consideration in a GLMM model. We considered the duration of the preceding consonant and the duration of the following consonant as fixed effects. A random intercept per speaker and one per word were included as random terms in the model. Results of the model suggest that the duration of the preceding consonant has a significant effect on the absence/presence of /ʁ/ while schwa is not inserted (log odds ratio = 0.6346, $|Z| = 2.251$, $p < 0.05$). The effect of the duration of the following consonant is also found in our data (log odds ratio = 0.6120, $|Z| = 3.304$, $p < 0.001$).

4.2. Automatic alignment on the three corpora: strategy changes for Cʁ# realization?

Hereafter, we analyze the influence of speech style on the different realization strategies using automatic alignment, as the agreement between the automatic and the manual alignment is high according to Cohen’s Kappa (see section 3). We focus only on the two most interesting pronunciation classes with respect to the SSP violation. These two classes captured the highest rates in the previous section using the manually checked alignments, namely the *Full*+ə and the *Full*-ʁ classes (eg. quatre /katʁ/ with schwa insertion [katʁə̲] and with /ʁ/ deletion [kat]).

Figure 3: Analyses of the duration of the surrounding consonants of /ʁ/ while schwa is not inserted. **Left**: Box-plot (raw data in ms). **Center**: Density plot of the duration of the consonant preceding /ʁ/ (normalized in z-score). **Right**: Density plot of the duration of the consonant following /ʁ/ (normalized in z-score).

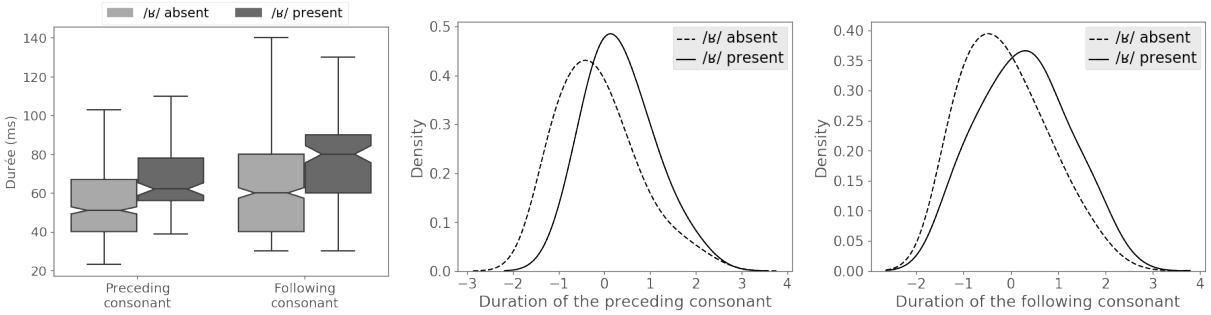
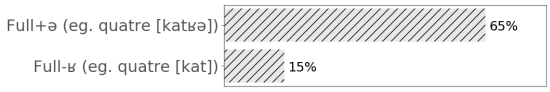
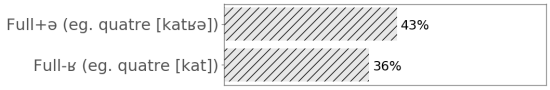


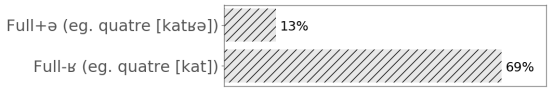
Figure 4: Realization of Cʁ# in #C context for 3 speech styles. *Full*+ə corresponds to words with final schwa, *Full*-ʁ to those with /ʁ/ deleted.



(a) ESTER: formal news



(b) ETAPE: broadcast conversations



(c) NCCFr: casual conversations

Figure 4 illustrates the realization of Cʁ#C (schwa insertion vs. /ʁ/ deletion) according to different speech styles. Schwa insertion tends to be the preferred strategy for speakers in the formal journalistic corpus ESTER (65% schwa insertion vs. 15% /ʁ/ deletion) whereas /ʁ/ deletion tends to be the privileged strategy for speakers in the casual speech corpus NCCFr (13% schwa insertion vs. 69% /ʁ/ deletion). Schwa insertion is 7% ahead of /ʁ/ deletion for the conversational journalistic corpus ETAPE. Results show that speech style has a strong impact on the realization of Cʁ#. The less formal the speech style is, the less we observe schwa insertion and the more we observe /ʁ/ deletion.

The influence of speech style on the strategies between schwa insertion (eg. quatre /katʁ/ [katʁə̲]) and /ʁ/ deletion (eg. quatre /katʁ/ [kat]) was also tested using GLMM. The fixed factor considered was speech style (ETAPE, ESTER or NCCFr, reference: ESTER) and sex (male or female, reference: female) was considered as covariate. The following random terms were included in the model: a random intercept per speaker and one per word, a by-speaker slope for the effect of speech style and a by-word slope for the effect of speech style. Post-hoc analyses based on the model were performed for the fixed effect to obtain information on each level of the fixed effect. Results show that the probability of applying the schwa insertion strategy decreases significantly both in ETAPE (log odds ratio = -1.2960, |Z|= 10.129, $p < 0.001$), and in NCCFr (log odds ratio = -3.6069, |Z|= 17.749, $p < 0.001$) with respect to that observed in ESTER. The post-hoc test based on our model suggests that the corpora, which stand for different speech styles, are significantly different from each other concerning strategies for resolving SSP violation (ESTER vs. ETAPE: $p < 0.001$; ETAPE vs. NCCFr: $p < 0.001$; ESTER vs. NCCFr: $p < 0.001$).

5. CONCLUSION

Our study on Cʁ#C sequences allows us to better understand the production of word-final /Cʁ#/ followed by #C in continuous speech, thanks to large corpora. Results on Cʁ#C realization show that speakers may apply different strategies to avoid the violation of the Sonority Sequencing Principle (SSP). According to the manual alignment on the formal journalistic speech ESTER, the preferred strategies are mostly schwa insertion followed by /ʁ/ deletion. Duration analyses show that the absence/presence of /ʁ/ is also linked to the durations of the preceding and following consonants of the word-final /ʁ/ position. This result suggests a potential impact of local speech rate. We found interesting tendencies on different speech styles: the less formal the speech style is, the less we observe schwa insertion and the more we observe /ʁ/ deletion.

6. ACKNOWLEDGEMENTS

The present study was funded by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083) and by SALSA project (ANR 14-CE28-0021).

7. REFERENCES

[1] Adda-Decker, M., de Mareüil, P. B., Adda, G., Lamel, L. 2005. Investigating syllabic structures and their variation in spontaneous French. *Speech Communication* 46(2), 119–139.

[2] Bates, D., Mächler, M., Bolker, B., Walker, S. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

[3] Brand, S., Ernestus, M. 2018. Reduction of word-final obstruent-liquid-schwa clusters in Parisian French. *Corpus Linguistics and Linguistic Theory*.

[4] Clements, G. 1990. The role of the sonority cycle in core syllabification. Papers in Laboratory Phonology I, edited by John Kingston & Mary Beckmann, 283-333.

[5] Clements, G. N. 1988. The sonority cycle and syllable organization. *Phonologica* 63–76.

[6] Côté, M.-H. 2004. Consonant cluster simplification in Québec French. *Probus* 16(2), 151–201.

[7] Galliano, S., Geoffrois, E., Gravier, G., Bonastre, J.-F., Mostefa, D., Choukri, K. 2006. Corpus description of the ESTER evaluation campaign for the rich transcription of french broadcast news. *Proceedings of LREC* volume 6 315–320.

[8] Gamer, M., Lemon, J., Fellows, I., Singh, P. 2012. irr: Various coefficients of interrater reliability and agreement. R package v. 0.84.

[9] Gauvain, J.-L., Lamel, L., Adda, G. 2002. The LIMSI broadcast news transcription system. *Speech communication* 37(1-2), 89–108.

[10] 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-Eighth international conference on Language Resources and Evaluation*.

[11] McCulloch, C. E., Neuhaus, J. M. 2001. *Generalized linear mixed models*. Wiley Online Library.

[12] New, B., Brysbaert, M., Veronis, J., Pallier, C. 2007. The use of film subtitles to estimate word frequencies. *Applied psycholinguistics* 28(4), 661–677.

[13] Team, R. C. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.

[14] Torreira, F., Adda-Decker, M., Ernestus, M. 2010. The Nijmegen Corpus of Casual French. *Speech Communication* 52(3), 201–212.