FINAL DEVOICING OF FRICATIVES IN FRENCH: STUDYING VARIATION IN LARGE-SCALE CORPORA WITH AUTOMATIC ALIGNMENT

Adèle Jatteau[a], Ioana Vasilescu[a], Lori Lamel[a] & Martine Adda-Decker[a,b]

[a]LIMSI–CNRS, Université Paris-Saclay, Orsay, France
[b]LPP, CNRS-Université Sorbonne Nouvelle, Paris, France
{jatteau,vasilescu,lamel,madda}@limsi.fr

ABSTRACT

This study quantifies "final devoicing" (FD) in large-scale corpora of Standard French via automatic alignment with pronunciation variants. We use corpora of different speech styles, ESTER (journalistic speech) and NCCFr (conversation between friends), to compare the rates of devoicing and voicing of word-final fricatives as a function of the following context (voiceless obstruent, voiced obstruent, sonorant, vowel and pause). Three categories emerge: before obstruents, word-final fricatives undergo laryngeal assimilation (*arrive*[f] *tôt*); before vowels and sonorants, only little variation is found; before pause, there is a high rate of devoicing (*trouve*[f] ##) and only negligible voicing. This last point shows that French features FD. This finding, in accordance with the typology and phonetic studies, extends previous small-scale investigations on regional varieties of the language to large corpora of Standard French. The FD effect is reinforced in conversational speech, and is stronger in labial than in alveolar and post-alveolar fricatives.

Keywords: final devoicing, fricatives, large corpora, forced alignement, Standard French.

1. INTRODUCTION

Final devoicing (FD) is the process whereby (contrastively) voiced consonants are devoiced in domain-final position, as in Balearic Catalan *cav* [kaf] 'I dig'. The process is widely attested in the typology both as a sound change, which is believed to progress from larger to smaller prosodic domains, and as a static sound pattern [4]. It has been argued to result from aerodynamic constraints of the vocal tract: the falling of subglottal pressure in domain-final position quenches the vibration of the vocal folds [13, 20]. This contrasts with the opposite process, "final voicing", which is rare or unattested in the world's languages [10]. If "sound change is drawn from a pool of synchronic variation" [14], we should be able to find FD in the synchronic variation of languages for which the pattern has not been reported so far. The goal of this paper is to investigate this hypothesis in Standard French. Since such a pattern is likely to be peripheral, and better reflected in spontaneous speech, it is difficult to detect in laboratory settings. To be able to quantify the phonetic precursors of FD, we analyse large corpora of speech using automatic alignment with pronunciation variants.

2. FINAL DEVOICING IN FRENCH

French features a contrast between voiced and voiceless obstruents at three places of articulation. Since this contrast is maintained in word-final position, as in *case* /kaz/ 'compartment' ~ *casse* /kas/ 'breaks', *cage* /kaʒ/ 'cage' ~ *cache* [kaʃ] 'hides', it is a potential candidate for FD. FD has indeed been reported in a number of regional variants of French, especially in northern and eastern areas, where it is perceived as a marked vernacular feature and believed to result from contact with Flemish, Picard (northern) and Germanic (eastern) languages [15, 18]. FD is also a well-known feature of Belgian French [9], and has been occasionally found in other areas of France [17]. In these varieties, FD is not systematic, and it is word-final: it occurs regardless of the following context. To the best of our knowledge however, no large-scale study of FD has been undertaken for Standard French. The present paper proposes to enlarge this perspective by investigating the voicing alternation of word-final fricatives in large-scale corpora of Standard French. Fricatives are interesting because, since they devoice more easily than stops [13], we are likely to find more variation in fricatives than in stops; moreover, most of the works on the voicing alternations of word-final consonants in French (such as [16, 17, 18]) focus on stops only.

3. CORPORA

Two manually transcribed corpora of French were used. The first is a subset of the ESTER corpus of radio and television news broadcast [6]. In order to focus on Standard French, we filtered out the data coming from RTM (*Radio Television Maroc*) and RFI (*Radio France International*) which contain many portions of non-metropolitan French. The remaining corpus comprises 43 hours of speech. The second corpus is NCCFr, which comprises 36 hours of conversations between friends [19]. Both corpora were cleaned of monosyllabic clitics such as *je, ce, se*, interjections (eg. *pff, oups*) and loanwords for which more than one pronunciation may be possible (eg. *Gomez, Museuw, Zinoviev*). Although we cannot precisely control for the percentage of marked regional variants and foreign accent in the corpora, our listening to a hundred samples in each of them suggests a globally high homogeneity. The comparison of ESTER and NCCFr therefore allows us to measure the trend for FD across two distinct speech styles, formal (ESTER) and casual (NCCFr).

4. METHODOLOGY

4.1. Automatic alignment with variants

To measure the voicing alternations in fricatives, these corpora were segmented using the LIMSI's automatic speech recognition (ASR) system to carry out the forced alignments [7]. Forced alignment matches the corpora's speech segments with their orthographic transcription, using acoustic models and a pronunciation dictionary. The pronunciation dictionary provides, for each orthographic word, a short list of its standard broad phonetic transcriptions (see Table 1). For each word, the system chooses the most adapted pronunciation variant listed in the dictionary, and returns the word and phone boundaries.

In order to detect voicing alternation in fricatives, we introduced additional pronunciation variants in the dictionary by systematically applying voicing and devoicing rules to the voiceless and voiced fricatives. For instance, for the word *grève* the system could choose between the pronunciation variants [gʁɛv] and [gʁɛf] (plus optional schwa) (cf. Table 1). This decision is made based on the system's acoustic models: in each case, the system decides whether the fricative best matches the acoustic model of the voiced or voiceless member of the pair at each place of articulation. Adding variants to the pronunciation dictionary is a method used to improve the system's performance [12, 2, 3]. It is here leveraged to assist the linguistic investigation, by allowing us to detect general trends in the phonology at a large scale in spontaneous speech.

Base dic.	*grève*	grɛv, grɛvə
Var. dic.	*grève*	grɛv, grɛvə, grɛf, grɛfə

Table 1: Sample of the base dictionary vs. dictionary with added variants.

The acoustic models were previously trained on about 250 hours of transcribed speech from a variety of broadcast radio and TV sources. As such, the models inherently capture a degree of variation in the signal, introducing some noise in the results. To limit this effect, we used context-independent phone models [2].

4.2. Comparisons

In order to quantify FD, we proceed from larger to more precise comparisons. To control for the general variation found in spontaneous speech, we first compare the rates of alternations in word-final vs. word-internal positions (§5.1). A total of 247 972 word-internal fricatives and 27 424 word-final fricatives were extracted. The latter were then restricted to words where the fricative is in absolute final position (so [gʁɛvə], [gʁɛfə] were set apart), reducing the number of tokens to 24 024 (Table 2).

We then focus on the word-final position, and compare the rates of /f, s, ʃ/ voicing and /v, z, ʒ/ devoicing (§5.2). Word-final tokens were extracted together with the following context, which was sorted in five classes: voiceless obstruent, voiced obstruent, sonorant, vowel and pause. The "pause" category includes silences and breaths. Note that these do not necessarily occur at the end of sentences or syntactic groups: the context we are looking at for FD is hence purely phonological. Examination of voicing and devoicing alternations in each of these five categories reveals two types of processes: before obstruents, the potential effect of FD is masked by a laryngeal assimilation process (§5.3); in the remaining contexts, we find a relatively high level of devoicing before pause (§5.4).

5. RESULTS

5.1. Word-internal vs. word-final alternations

Voicing alternations were quantified in word-medial vs. word-final position. As expected, a much lower rate of variation is found in medial than in final position (Table 3). These low level voicing alternations

	ESTER	NCCFr	Total
N. fric.	17 481	9 543	27 024
/f/	1 260	826	2 086
/s/	7 376	3 765	11 141
/ʃ/	1 399	631	2 030
/v/	1 301	1 673	2 974
/z/	3 969	1 664	5 633
/ʒ/	2 176	984	3 160

Table 2: Number of extracted word-final fricatives.

roughly reflect unpredictable variation, as when a whole word is voiced or devoiced. The rate of word-internal variation can then be used as an indication for the overall rate of variation in our corpora.

	Medial	Final
/f, s, ʃ/ voicing	6	17
/v, z, ʒ/ devoicing	5	27

Table 3: Rate (in %) of voicing alternation in ESTER + NCCFr as a function of the position in the word.

The situation where the consonant of interest is in absolute final position can be compared to the situation in which it is followed by a schwa and/or a liaison consonant (eg. [gʁɛv] vs. [gʁɛvə], [gʁɛvz], [gʁɛvəz]), representing 3400 tokens. As expected, we find less variation in this case than in the absolute final position, with a voicing rate of 10% and a devoicing rate of 9% across all places of articulation and both corpora. Interestingly however, these rates are still higher than the word-internal ones in Table 3. They reflect cases where the whole word-final sequence is voiced or devoiced depending on the following context.

5.2. Final position: voicing vs. devoicing

Focusing now on the word-final position, we compare the rates of voicing and devoicing depending on the following contexts. The comparison of Fig. 1a (voicing) and 1b (devoicing) shows that three groups of contexts emerge: before obstruents, laryngeal assimilation takes place; before vowels and sonorants, only small amounts of variation are observed; before pause, there is a high rate of devoicing but no voicing. This is evidence for two processes: assimilation (§5.3) and FD (§5.4).

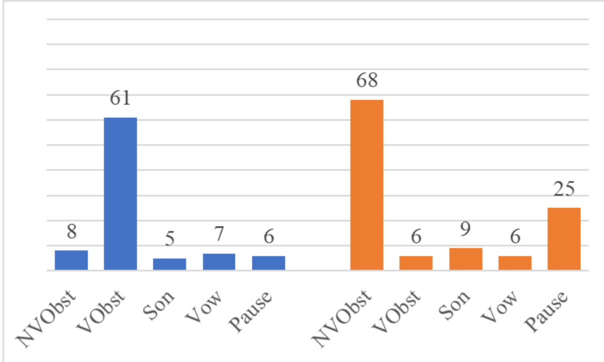

Fig. 1a. Fig. 1b.

Figure 1: Rate (in %) of word-final /f, s, ʃ/ voicing (Fig. 1a) and /v, z, ʒ/ devoicing (Fig. 1b) in ESTER + NCCFr as a function of the following context. NVObst = voiceless obstruent, VObst = voiced obstruent, Son = sonorant, Vow = vowel.

5.3. Laryngeal assimilation

Both Fig. 1a and Fig. 1b show a high amount of variation before obstruents. When the following context is a voiced obstruent, word-final /f, s, ʃ/ are voiced 61% of the time (eg. *grosse*[z] *différence*). When it is a voiceless obstruent, word-final /v, z, ʒ/ are devoiced 68% of the time (eg. *mobilise*[s] *pour*). NCCFr shows overall more variation than ESTER. This confirms that regressive laryngeal assimilation at word-boundary is a productive process in Standard French, in line with previous studies on the subject [5, 1, 8]. A difference is that we find a preference for devoicing assimilation, while Hallé & Adda-Decker found a symmetric effect for fricatives [8].

5.4. Final devoicing

The second process which emerges from the comparison of Fig. 1a and Fig. 1b is FD. Comparing the rates of /v, z, ʒ/ devoicing before sonorant, vowel and pause in Fig. 1b, we find a relatively high rate before pause (25%), but only marginal variation before sonorants (9%) and vowels (6%). Our study therefore shows that there is pre-pausal FD in Standard French: the process does not target all word-final positions, but only the end of chunks which can be analysed as utterances. Comparing this to /f, s, ʃ/ voicing in Fig. 1a, we can see that there is only marginal variation in all three cases (5% before sonorants, 7% before vowels, 6% before pause). As expected, there is no final voicing in French.

The pre-pausal FD process is reinforced in casual speech, as opposed to prepared speech: as shown by Fig. 2, there is a higher rate of pre-pausal devoicing in NCCFr (32%) than in ESTER (21%) ($\chi^2(1)$ =

17.45, $p < .001$).

Next, our results show an effect of place of articulation (Table 4): labial fricatives devoice more frequently than alveolar and post-alveolar fricatives. When both corpora are pooled, the differences between /v/ and /z/, as well as /v/ and /ʒ/ are significant ($\chi^2(1) = 13.37, p < .001$ and $\chi^2(1) = 9.21, p < .05$ respectively); the one between /z/ and /ʒ/ is not ($\chi^2(1) = .10, p = .74$). This result is due to the influence of the ESTER corpus: in NCCFr alone, none of the comparisons are significant.

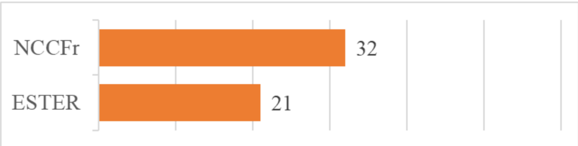

Figure 2: Rate (in %) of /v, z, ʒ/ devoicing before pause in NCCFr (top) vs. ESTER (bottom).

	/v/	/z/	/ʒ/
ESTER	32	19	20
NCCFr	35	32	28
Both corp.	34	22	23

Table 4: Rate (in %) of devoicing before pause in ESTER vs. NCCFr.

6. DISCUSSION

This study explored the voicing alternations of fricatives in two large corpora of Standard French. We report two processes responsible for these alternations in word-final position. The first is regressive laryngeal assimilation before obstruents: our findings corroborate previous studies on the topic. The second is final devoicing before pause, that is, at the end of utterances. The comparison of voicing and devoicing patterns in non-assimilatory contexts, that is, before sonorant, vowel and pause, shows a relatively high rate of devoicing before pause, and no special effect in the other contexts. Our study thus confirms the existence of FD in the "pool of variation" in Standard French, before the process may take on and become a sound change. This situation differs from the one reported in regional varieties of the language: in our corpora, FD does not target the end of the word, but the end of larger prosodic groups. It is not systematic in this position. Standard French thus shows a preliminary stage of the process. Furthermore, we find an effect of speech style (there is more FD in casual speech) and an effect of place of articulation: in ESTER, labial fricatives devoice more often than alveolar and post-alveolar fricatives. These are new results, since previous works targetted only one speech style, and either focused on stops [17, 18] or found a great variety in the selection of devoicing fricatives depending on the dialects [15]. In future work we plan to extend the study to stops, in order to allow for the comparison and to provide a full picture of the process in Standard French.

Finally, a word on the methodology is in order. A major difficulty in studying neutralisation patterns is how to label the categories, in our case "voice" and "voiceless". What is quantified in this paper as a binary decision most probably corresponds in the data to a gradient continuum along the acoustic parameters cuing the voice contrast. Telling apart the canonical vs. alternating tokens is all the more difficult before pause, where the cues for the contrast are weakened anyway; even in languages where FD has been reported as a systematic rule, numerous studies have shown that the neutralisation is acoustically and even perceptually incomplete [11]. To capture this gradiency, we are planning in future work to study in more detail the acoustic parameters at stake in our corpora. Yet ASR systems are interesting in that they approach exemplar-based modelisations of the listener. In real life, listeners are able to compensate for the variation in the signal and to correctly identify the phoneme categories by relying, among other tools, on subtle acoustic cues and their lexicon. In a similar way, the acoustic models of ASR systems are trained on hundreds of hours of speech, including word-final tokens, and may therefore compensate for some of the variation in the signal. An important difference in the method used in this paper is that the system cannot rely on its lexicon to identify the phoneme, since the lexicon was modified with additional variants. It is thus forced to make the decision based on its acoustic models only. In that sense, this method approximates perceptual experiments at a very large scale, and stands as a useful complement to classical acoustic and perceptual studies.

7. ACKNOWLEDGEMENTS

This work was partially supported by the *Maison des Sciences de l'Homme* Paris-Saclay through a Maturation grant 18-MA-01, and by the French National Agency for Research as part of the SALSA project (Speech and Language technologies for Security Applications) under grant ANR-14-CE28-0021.

8. REFERENCES

[1] Adda-Decker, M., Hallé, P. 2007. Bayesian framework for voicing alternation and assimilation stud-

ies on large corpora in French. *16th International Congress of Phonetic Sciences* 613–616.
[2] Adda-Decker, M., Lamel, L. 1999. Pronunciation variants across system configuration, language and speaking style. *Speech Communication* 29(2-4), 83–98.
[3] Adda-Decker, M., Lamel, L. 2000. The use of lexica in automatic speech recognition. In: Van Eynde, F., Gibbon, D., (eds), *Lexicon Development for Speech and Language Processing*. Springer 235–266.
[4] Blevins, J. 2006. Theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32(2), 117–166.
[5] Féry, C. 2003. Gradient prosodic correlates of phrasing in French. In: Meisenburg, T., Selig, M., (eds), *Nouveaux Départs en Phonologie*. Niemeyer.
[6] 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. *Ninth European Conference on Speech Communication and Technology*.
[7] Gauvain, J.-L., Lamel, L., Adda, G. 2002. The LIMSI broadcast news transcription system. *Speech communication* 37(1-2), 89–108.
[8] Hallé, P., Adda-Decker, M. 2011. Voice assimilation in French obstruents: A gradient or a categorical process? In: *Tones and features: A festschrift for Nick Clements*. De Gruyter 149–175.
[9] Hambye, P. 2005. *La prononciation du français contemporain en Belgique : variation, normes et identités*. PhD thesis Université Catholique de Louvain.
[10] Kiparsky, P. 2008. Universals constrain change, change results in typological generalizations. In: Good, J., (ed), *Linguistic Universals and Language Change*. Oxford University Press.
[11] Kleber, F., John, T., Harrington, J. 2010. The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics* 38, 185–196.
[12] Lamel, L., Adda, G. 1996. On designing pronunciation lexicons for large vocabulary continuous speech recognition. *The 4th International Conference on Spoken Language Processing, Philadelphia, PA, USA, October 3-6, 1996*.
[13] Ohala, J. J. 1983. The origin of sound patterns in vocal tract constraints. In: MacNeilage, P., (ed), *The production of speech*. New York: Springer-Verlag 189–216.
[14] Ohala, J. J. 1989. Sound change is drawn from a pool of synchronic variation. *Language change: Contributions to the study of its causes* 173–198.
[15] Pooley, T. 1994. Word-final consonant devoicing in a variety of working-class French - a case of language contact? *Journal of French Language Studies* 4(2), 215–233.
[16] Snoeren, N., Hallé, P., Segui, J. 2006. A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics* 34, 241–268.
[17] Temple, R. 1999. Phonetic and sociophonetic conditioning of voicing patterns in the stop consonants of French. *Proceedings of the 14th International Congress of Phonetic Sciences* 1409–1412.
[18] Temple, R. 2000. Old wine into new wineskins. A variationist investigation into patterns of devoicing in plosives in the Atlas linguistique de la France. *Transactions of the Philological Societ* 98, 353–294.
[19] Torreira, F., Adda-Decker, M., Ernestus, M. 2010. The Nijmegen corpus of casual French. *Speech Communication* 52(3), 201–212.
[20] Westbury, J., Keating, P. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22, 145–166.