Context, position in word and duration as predictors of voicing alternation of stops: a large-scale corpus-based study in 5 Romance Languages

Yaru Wu[1,2], Mathilde Hutin[1], Ioana Vasilescu[1], Lori Lamel[1], Martine Adda-Decker[1,2]
[1]*Université Paris-Saclay, CNRS, LISN, 91405 Orsay, France*
[2]*Laboratoire de Phonétique et Phonologie (UMR7018, CNRS-Sorbonne Nouvelle), France*

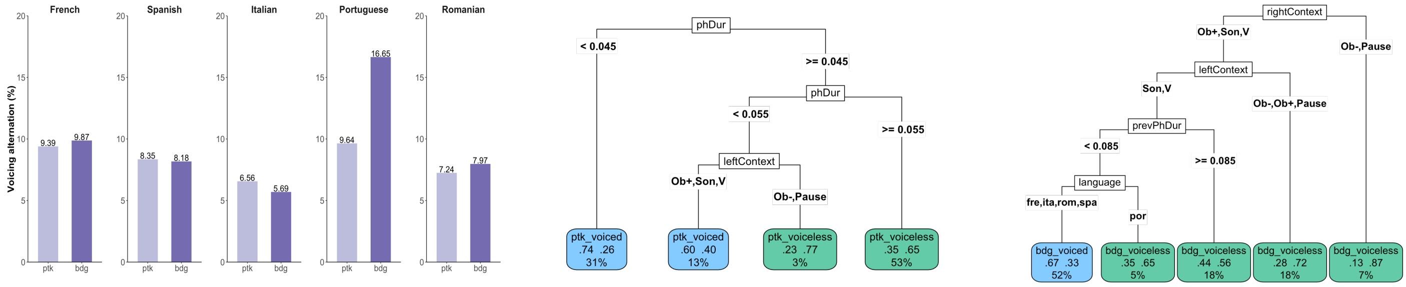
How synchronic phonetic variation interacts with diachronic change has been a key focus of variationist studies for over 50 years. Today, most linguists subscribe to Ohala's proposal that synchronic variation is one of the preconditions for diachronic change [1]. The "big data" revolution, which reached the humanities field in recent years, has added a new dimension to this research path. Large multilingual collections of spoken data covering various communication situations are now accessible for phonetic and laboratory phonology research, and can be explored with digital approaches [2,3]. In line with this new research direction, the present study explores two of the most common phonetic processes in the world's languages: voicing and devoicing of stop consonants (i.e. /ptk/ realized as [bdg] or /bdg/ realized as [ptk]). These processes are attested both as instances of synchronic variation and as processes provoking sound change. Using methods borrowed from speech technology, we seek to gain insight into the factors responsible for the phonetic precursors of these two phenomena.

Corpus and methodology. First, language-specific automatic speech recognition systems [4] in forced alignment mode with voicing variants (+/- voiced) were used to provide reliable accounts of the voicing alternation phenomena specific to connected speech [5,6,7]. Machine learning techniques are used to study variation in 5 Romance languages for which we have had abundant data (~ 1000 hours): French (176h), Italian (168h), Spanish (223h), Portuguese (114h) and Romanian (300h). Second, decision tree-based classification [8] allows the factors having the largest contribution to the prediction of how the voicing feature of stops is realized to be determined. Contextual, positional and durational factors were included in this study (see Table 1). For each experiment, 70% of the data set was randomly selected for training and the remaining 30% was used for test purposes in order to assess how well the tree generalizes to new data. Given that there are many fewer voicing alternations than unchanged observations, sampling techniques were used to balance the two groups (voiced vs voiceless) of each dataset.

Results. We present results from both the automatic alignment and automatic classification. Figure 1a shows the alternation rates obtained via our forced alignment allowing voiced and voiceless variants for both canonically voiceless and voiced stops. For instance, 9.4% of French voiceless stops /ptk/ were aligned with their corresponding non-canonical voiced variant [bdg] and 9.9% of voiced stops /bdg/ were aligned with a devoiced [ptk] variant. Portuguese exhibits a much higher rate of devoicing of voiced stops (/bdg/ realized as [ptk]) than the opposite phenomenon (voiced realizations of /ptk/, i.e., aligned with [bdg]). The rates of devoiced realization of voiced stops are also slightly higher than voiced realizations of voiceless stops for both French and Romanian, with the reverse tendency seen for Spanish and Italian. As for the respective roles of context, position in word and duration parameters correlated with voicing alternations, Figures 1b and 1c highlight the factors having the largest contributions to the prediction of voiced / voiceless segments (see texts in white rectangular boxes). The factors contributing the most to voicing (/ptk/ becomes [bdg]) are the duration of the segment in question and the left context, whereas devoicing is reliably predicted by the right context, followed by the left context, the duration of the preceding segment and the language. Interestingly, "language" is located towards the lower part of the tree in Figure 1c. This suggests that particular languages tend to subordinate to the historical movement of the language family. Nevertheless, the languages split into two groups with Portuguese in one branch and the other four languages in the other, which is in line with the higher devoicing rate seen for Portuguese in Figure 1a. The overall correct prediction of voicing alternation on unseen data (round-robin experiments with a random selection of data) are 73.6% for voiceless stops and 68.2% for voiced ones (mean of 10 experiments).

Table 1. Variation factors included in the automatic classification.

Factors	Details
Language	Spanish (spa), French (fre), Italian (ita), Portuguese (por) or Romanian (rom)
Position in word [9,10]	Position of the stop in the word: word initial position (wInitial), word medial position (wMedial) or word final position (wFinal)
Left context [10,11,12]	Nature of the segment preceding the stop in question: pause including hesitation, breath and silence;vowel (V); sonorant (Son); voiced obstruent (Ob+); voiceless obstruent (Ob-)
Right context [7,10]	Nature of the segment following the stop in question: pause including hesitation, breath and silence; vowel (V); sonorant (Son), voiced obstruent (Ob+); voiceless obstruent (Ob-)
Segment/Word duration [3,13,14]	Duration of the stop / of the segment preceding the stop / of the segment following the stop / of the word



(a) **(b)** **(c)**

Figure 1. *Voicing alternation rates of voiceless (/ptk/) and voiced (/bdg/) stops (a) , and classification tree on the voicing of canonically voiceless (b) and voiced (c) stops.*

[1] 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.

[2] Adda-Decker, M., & Lamel L. (2018). Discovering speech reductions across speaking styles and languages, in *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*, F. Cangemi, M. Clayards, O. Niebuhr, B. Schup-pler, and M. Zellers, Eds. De Mouton Gruyter.

[3] Ryant, N., & Liberman, M. (2016, September). Large-scale analysis of Spanish/s/-lenition using audiobooks. In *Proceedings of Meetings on Acoustics 22ICA* (Vol. 28, No. 1, p. 060005). Acoustical Society of America.

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

[5] Adda-Decker, M., & Hallé, P. (2007). Bayesian framework for voicing alternation and assimilation studies on large corpora in French. In *16th Congress of Phonetic Sciences* (pp. 613-616).

[6] Jatteau, A., Vasilescu, I., Lamel, L., Adda-Decker, M., & Audibert, N. (2019, September). " Gra [f] e!" Word-Final Devoicing of Obstruents in Standard French: An Acoustic Study Based on Large Corpora. In *INTERSPEECH* (pp. 1726-1730).

[7] Hutin, M., Jatteau, A., Vasilescu, I., Lamel, L., & Adda-Decker, M. (2020, October). Ongoing phonologization of word-final voicing alternations in two Romance languages: Romanian and French. In *Interspeech 2020*.

[8] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*(1), 81-106.

[9] Ségéral, P., & Scheer, T. (2008). Positional factors in lenition and fortition. *Lenition & fortition, ed. J. Brandao de Carvalho, T. Scheer, & P. Ségéral*, 131-172.

[10] Vasilescu I., Wu Y., Jatteau A., Adda-Decker M., & Lamel L. (à paraître). Alternances de voisement et processus de lénition et de fortition: une étude automatisée de grands corpus en cinq langues romanes. In *Revue TAL* (Volume 61, Numéro 1). www.atala.org/content/varia-12.

[11] Niebuhr, O., & Meunier, C. (2011). The phonetic manifestation of French /s#ʃ/ and /ʃ#s/ sequences in different vowel contexts: On the occurrence and the domain of sibilant assimilation. *Phonetica*, *68*(3), 133-160.

[12] Meunier, C., & Espesser, R. (2011). Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics*, *39*(3), 271-278.

[13] Vasilescu, I., Vieru, B., & Lamel, L. (2014). Exploring pronunciation variants for Romanian speech-to-text transcription. In *Spoken Language Technologies for Under-Resourced Languages*.

[14] Snoeren, N. D., Hallé, P. A., & Segui, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics*, *34*(2), 241-268.