Extracting Linguistic Knowledge from Speech: A Study of Stop Realization in 5 Romance Languages

Yaru Wu, Mathilde Hutin, Ioana Vasilescu, Lori Lamel, Martine Adda-Decker

▶ **To cite this version:**

Yaru Wu, Mathilde Hutin, Ioana Vasilescu, Lori Lamel, Martine Adda-Decker. Extracting Linguistic Knowledge from Speech: A Study of Stop Realization in 5 Romance Languages. 13th Conference on Language Resources and Evaluation (LREC 2022), Jun 2022, Marseille, France. pp.3257-3263. hal-03706248

HAL Id: hal-03706248
https://hal.science/hal-03706248v1

Submitted on 27 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 3257–3263
Marseille, 20-25 June 2022
© European Language Resources Association (ELRA), licensed under CC-BY-NC-4.0

Extracting Linguistic Knowledge from Speech: A Study of Stop Realization in 5 Romance Languages

Yaru Wu[1,2,3], Mathilde Hutin[2], Ioana Vasilescu[2], Lori Lamel[2], Martine Adda-Decker[3]
[1]CRISCO/EA4255, Université de Caen Normandie, 14000 Caen, France,
[2]LISN, Univ.Paris-Saclay, 91405 Orsay cedex, France,
[3]Laboratoire de Phonétique et Phonologie (UMR7018, CNRS-Sorbonne Nouvelle), France
yaru.wu@unicaen.fr, {mathilde.hutin, ioana.vasilescu, lori.lamel}@lisn.upsaclay.fr,
martine.adda-decker@sorbonne-nouvelle.fr

Abstract

This paper builds upon recent work in leveraging the corpora and tools originally used to develop speech technologies for corpus-based linguistic studies. We address the non-canonical realization of consonants in connected speech and we focus on voicing alternation phenomena of stops in 5 standard varieties of Romance languages (French, Italian, Spanish, Portuguese, Romanian). For these languages, both large scale corpora and speech recognition systems were available for the study. We use forced alignment with pronunciation variants and machine learning techniques to examine to what extent such frequent phenomena characterize languages and what are the most triggering factors. The results confirm that voicing alternations occur in all Romance languages. Automatic classification underlines that surrounding contexts and segment duration are recurring contributing factors for modeling voicing alternation. The results of this study also demonstrate the new role that machine learning techniques such as classification algorithms can play in helping to extract linguistic knowledge from speech and to suggest interesting research directions.

Keywords: corpus-based linguistics, Romance languages, speech transcription and alignment, decision tree

1. Introduction

The "big data" revolution has affected many research domains due to both the amount of available data and to many methodological changes needed to efficiently exploit the data. As is the case for many domains, speech technologies and corpus-based linguistics have been impacted by the increasing amounts of data available for training automatic systems and for linguistic studies, and for novel machine learning techniques that can benefit from large corpora. The current scientific and technological state of the art reinforces the collaboration between the two communities, allowing them to jointly take advantage of the abundance of data and the emergence of new methods. A shared topic concerns the patterns of variation in connected speech: statistically supported modeling of phonetic variation is of high interest for linguists that aim to ground their hypotheses in "naturalistic" data and for speech technology specialists that aim to "master" the variation responsible for processing errors.

In fact, modeling variation in connected speech has been a long term research topic in linguistics and a challenge for speech technologists. For instance, phoneticians and phonologists are particularly interested in sound variation, whether it be synchronic, observed at the present moment, or diachronic, happening over time. It is now accepted by both communities, that studies of such phenomena can be facilitated by speech technologies, e.g. by analyzing large, varied corpora for linguistic exploration. We aim to shed a new light on the study of voicing alternation in connected speech by introducing an automatic classification step to determine which factors are most important to explain the observed data. By voicing alternation, we mean the non-canonical realization of voiced stops as voiceless stops (/bdg/ pronounced [ptk]), and, conversely, the voiceless realization of voiced stops (/ptk/ pronounced [bdg]). Over time the topic has been of interest for different linguistic communities (phoneticians, phonologists, historical linguists etc.) who addressed the issue in the broader framework of lenition (associated to weakening, that is a voiceless consonant such as /p/ becoming [b]) and fortition (associated to strengthening, that is a voiced consonant such as /b/ becoming [p]) (Honeybone, 2008; Gurevich, 2004), and for speech technologists who questioned these phenomena in order to overcome transcription errors dues to such non-canonical realizations (Vasilescu et al., 2018).

The purpose of this paper is to use a bottom-up approach to study voicing alternations in several languages and to compare the results with those of top-down studies. We hypothesize that bottom-up approaches, that specifically aim to extract linguistic knowledge from an extended array of local factors (acoustic, prosodic, contextual) can give unexpected results unforeseen by the top-down studies relying on *a priori* linguistic representations, and thus shake the current state of affairs (Meunier and Bigi, 2016; Wu and Adda-Decker, 2020; Wu and Adda-Decker, 2021). The underlying interest in applying speech technologies to corpus-based linguistic studies is two-fold: (i) such studies can improve our knowledge about speech variation in the targeted languages and allow linguis-

tic hypotheses to be tested and (in)validated against representative corpora; and (ii) the gained knowledge can hopefully serve to improve the acoustic, pronunciation and lexical models used in speech recognition and synthesis, and speech technologies in general. More specifically, in this paper we rely on large-scale corpora covering five languages and combine two automatic methods aimed to model the voicing alternations of stops: forced alignment of non-canonical variants with language-specific automatic speech recognition systems and machine learning techniques to prioritize the factors that trigger voicing alternation.

The paper is organized as follows. The Section 2 is dedicated to the methodology of the paper. The description of corpora, automatic alignment and the two methods used to model voicing alternation in the five Romance languages are detailed in the section. Section 3 focuses on the results: voicing alternation rates are described as a function of languages and subsequently, tested together with various contextual parameters to feed classification algorithms in order to highlight the factors that trigger voicing alternation the most. Section 4 provides a summary and conclusion.

2. Methodology

2.1. Corpora

Almost 1000 hours of spoken data were used in this study, covering five different languages: French, Spanish, Italian, Portuguese and Romanian. These five languages were selected for two reasons: (i) from a linguistic point of view, they share a common origin (i.e. Latin), and thus share some common features, but still demonstrate language specific patterns due to individual diachronic evolution; and (ii) from a practical perspective, given that all five languages are widely spoken, we were able to obtain access to numerous data and language specific speech recognition technology that allow us to automatically align the reference transcripts with the speech data.

These corpora were acquired from the *Linguistic Data Consortium* or *ELRA*, or developed in the framework of international research projects: IST ALERT for part of the data in French and Portuguese (da Silva et al., 2011; da Silva et al., 2013), IRST for part of the data in Italian (Marcello et al., 2000), and OSEO Quaero (Lamel et al., 2011) for all languages.

The data used in this study cover mainly the standard variety of each language, except for Spanish, for which a small amount of data of broadcast Latin American Spanish varieties were included, in addition to the recordings issued by European broadcast sources. More precisely, only broadcast-news, formal public discussions and debates were included in the database for all five languages.

Only 7 of the 300 hours of Romanian data was accompanied by manual transcriptions and the remainder was automatically transcribed with a system built for this language (Vasilescu et al., 2014). All of the spoken data was manually transcribed for the four other languages. Table 1 shows the amount of data for each language, along with the quantification of word-tokens (M), word-types (k), and the mean of the number of variants for each word.

Lang	#hours	word tokens (M)	word types (k)	#variants
Fre	176	2.4	55.7	6.8
Spa	223	2,6	61.9	4.4
Ita	168	1.8	57.0	5.3
Por	114	1.0	40.0	3.7
Rom	300	3.6	48.0	3.7

Table 1: Data description: language, duration of the corpus, number of word tokens (in millions, M), number of word types (in thousands, k), mean number of variants/word when allowing voicing alternation for each stop occurrence.

2.2. Forced alignment with pronunciation variants

We adopt the method proposed in (Adda-Decker and Hallé, 2007; Hallé and Adda-Decker, 2007; Jatteau et al., 2019a; Jatteau et al., 2019b) to study voicing alternations of Romance stop consonants /ptkbdg/ through automatic forced alignment with specific variants in the pronunciation dictionaries. In this study, we included in the pronunciation dictionary both the canonical form and the non-canonical forms, i.e. possible pronunciations of word with systematic pronunciation variants (Adda-Decker and Lamel, 2017). The baseline pronunciation dictionaries are those used by the LISN (former LIMSI) speech transcription systems.

The method consists of providing non-canonical pronunciation variants in pronunciation dictionaries and allowing the speech recognition systems to select the best matching one during the forced alignment process. The described in-house speech recognition systems had been previously trained on the same type of data selected for the study (Vasilescu et al., 2020). The architecture of the systems are therefore comparable to that used in (Vasilescu et al., 2020). Systems exist for all the languages selected, namely French (Gauvain et al., 2002), Spanish (Vasilescu et al., 2018), Italian (Després et al., 2013), Portuguese (da Silva et al., 2011; da Silva et al., 2013) and Romanian (Vasilescu et al., 2014). Automatic word and phone level alignments of the speech data with their manual orthographic transcriptions were produced using a system derived from that described in (Vasilescu et al., 2014). As described in (Gauvain et al., 2005), the system selects the most probable variant given the actual acoustic realization.

During alignment, voicing and devoicing are decided if the best matching phone model corresponds to the voiced or voiceless variant respectively. Hence, the system can select (1) for any segment of /ptk/, its as-

sociated voiceless acoustic model or its voiced counterpart [bdg] and (2) for any segment of /bdg/, its originally voiced acoustic model or its voiceless counterpart [ptk]. For instance, the Romanian word *grup*, /grup/ could be transcribed either as [grup], [grub], [krup] or [krub] depending on whether the system considered the first and last consonants to best correspond to the voiceless or voiced realization.

Several recent works took advantage of large corpora and automatic alignment with pronunciation variants to investigate the voicing alternation processes. This variant-based approach has given reliable accounts of voicing variants (voiced *vs* voiceless) of consonants for Spanish (Vasilescu et al., 2018; Ryant and Liberman, 2016), French (Jatteau et al., 2019a; Jatteau et al., 2019b; Hutin et al., 2020a) and Romanian (Hutin et al., 2020a; Hutin et al., 2020b). In (Vasilescu et al., 2020) this method was applied to a corpus of nearly 1000 hours of speech to compare alternation patterns in the same five Romance languages as studied here. The main result reported in (Vasilescu et al., 2020) was that the stops' realizations are influenced by both the stops' own position in the word and its adjacent (left and right) segments. In this paper, we extend the work of (Vasilescu et al., 2020) and make use of the same corpora to gain insight in the factors relevant to voicing alternation via automatic classification. The role of positional and contextual information in detecting voicing alternation will be assessed here by decision trees.

2.3. Decision tree-based classification

Over the last decade, technologies for browsing or mining content from large collections of textual material have been extended to the exploration of audio material. The large-scale data mining on text had helped transform the relevant disciplines and it was expected that the disciplines dealing with spoken language would reap similar benefits from accessible, searchable, large corpora[1]. Data mining phonetics has proven useful to investigate various phonetic questions such as foreign accents in French (Boula de Mareüil et al.,), tone contour shapes in Mandarin Chinese (Zhang, 2019) or, more generally, surface pronunciations of a word in conversational speech (Bowman and Livescu,).

In the following, we use a decision tree paradigm with multiple contextual and durational factors which are also available to automatic speech recognition processing. We want to explore to what extent the following features allow for classifying data that behave canonically or non-canonically with respect to their voicing feature[2].

The selected features that are modeled are:

[1] `http://www.phon.ox.ac.uk/mining_speech/`

[2] In future work we plan to add more specific features such as the acoustic voicing ratio (vocal fold vibration rate).

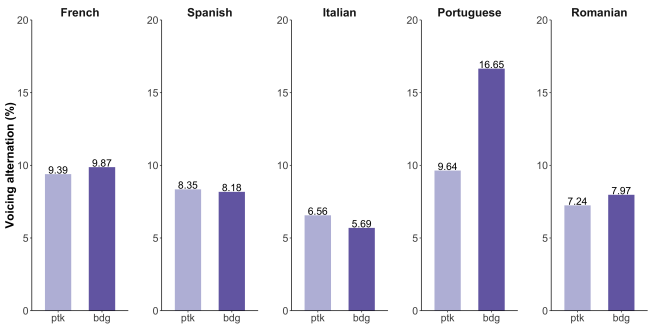

Figure 1: Voicing alternation rates of voiceless (ptk) and voiced (bdg) stops in 5 Romance languages, according to the automatic labeling via forced alignment with pronunciation variants (i.e., bdg $\Longleftrightarrow$ ptk).

- language: one of French (fre), Spanish (spa), Italian (ita), Portuguese (por) or Romanian (rom)
- wordPosition: whether the stop is word initial (wInitial), medial (wMedial) or final (wFinal) (Vasilescu et al., 2020; Ségéral and Scheer, 2008)
- leftContext: whether the stop is preceded by a pause (hesitation, breath or silence), a vowel (V), a sonorant (Son), a voiced obstruent (Ob+) or a voiceless obstruent (Ob-) (Vasilescu et al., 2020; Niebuhr and Meunier, 2011; Meunier and Espesser, 2011)
- rightContext: whether the stop is followed by a pause (hesitation, breath or silence), a vowel (V), a sonorant (Son), a voiced obstruent (Ob+) or a voiceless obstruent (Ob-) (Jatteau et al., 2019b; Hutin et al., 2020b)
- segment/word duration: how long the segment/word is (Vasilescu et al., 2014; Ryant and Liberman, 2016; Snoeren et al., 2006) (duration of the stop (phDur), durations of the preceding (prevPhDur) and following phones (nextPhDur) and duration of the word in question).

All of these features were chosen based on linguistic literature and the results of our investigations will therefore reflect the underlying linguistic parameters.

3. Results

In this section, we present results from both the automatic alignment and automatic classification.

Figure 1 shows the alternation rates obtained via the forced alignment with voiceless and voiced stops variants for both canonically voiceless (/ptk/) and voiced stops (/bdg/). The results show that the voicing alternation rates of voiceless stops tend to be lower than that of voiced stops for Portuguese and Romanian, followed by French, for which 9.4% of the voiceless stops (/ptk/) were aligned with their non-canonical voiced

variant [bdg] and 9.9% of French voiced stops /bdg/ were aligned with the devoiced [ptk] variant. The voicing alternation rates of voiceless stops tend to be higher than that of voiced stops for Italian, while similar voicing alternation rates are found for voiceless and voiced stops in Spanish. Portuguese demonstrates a much higher voicing alternation rate for voiced stops (/bdg/ realized as [ptk]) than for the voiceless stops (/ptk/ produced as [bdg]). Although the devoicing rate of /bdg/ in Portuguese is surprisingly high, the outcome is in line with the special pattern found for Portuguese in (Pape and Jesus, 2015) and concluded that Portuguese stops behave more similarly to German ones than to Italian ones.

As for decision trees (Quinlan, 1986), they allow the explicit inclusion of linguistic features and perform a relevance analysis which provides information about their contribution to the prediction of the targeted variable. Here, the classification examines the contribution of investigated factors to the differentiation of the voicing pattern (voiceless *vs* voiced) of canonical stops.

The features presented above were included as predictors and the voicing labels resulting from the forced alignment were included as target variables in the analyses using the rpart package (Therneau et al., 2010) in R (R Core Team, 2013). For each experiment, 70% of the data set was randomly selected for training and the remaining 30% was used as test data to assess how well the tree generalizes to new data. Since there are much fewer voicing alternation observations than unchanged observations, sampling techniques were used to balance the two groups of each data set (voiced *vs* voiceless). We used the number of observations for voicing alternation as a baseline and randomly extracted the same number of unchanged observations for our analyses. Each classification tree was pruned using the lowest cross validation error.

The classification trees show the features that had the largest contributions to the prediction of voiced / voiceless segments (see texts in white rectangular boxes). Conditions and thresholds are specified on the branches of the tree. Each leaf node shows (1) the predicted class (ptk_voiced *vs* ptk_voiceless; bdg_voiced *vs* bdg_voiceless); (2) the predicted probability of each class; and (3) the percentage of observations in the node. More precisely, the leaves in the blue and green frames indicate the target variable (i.e. voiced *vs* voiceless for canonically voiced / voiceless stops). The two numbers in the second row of each leaf indicate the performance rate of each target value, with the predicted token for voiced segments on the left and for voiceless segments on the right. Here, the correctly predicted rate for each blue leaf is the first number on the second row of the leaf; the correctly predicted rate for each green leaf is the second number on the second row of the leaf. The percentage in the leaf suggests what percentage of the training samples ended up in each leaf. The sum of the percentages in all leaves is 100%.

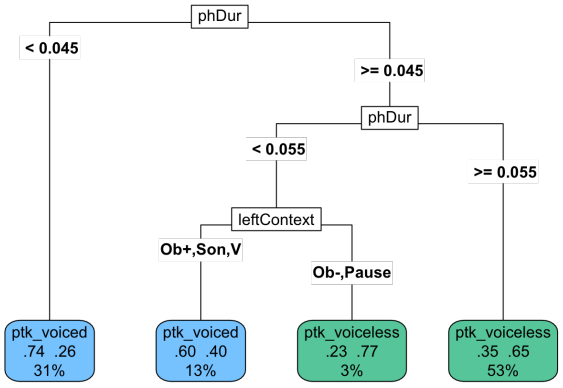

Figure 2: Classification tree for voicing pattern of the canonically voiceless stops (ptk_voiced *vs* ptk_voiceless), all languages pooled.

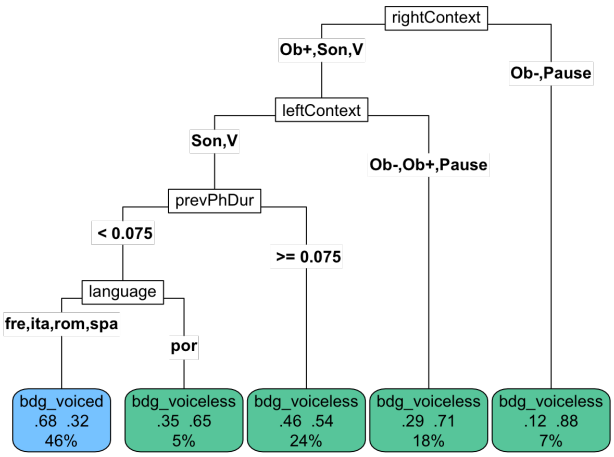

Figure 3: Classification tree for voicing patterns of canonically voiced stops (bdg_voiced *vs* bdg_voiceless), all languages pooled.

Figure 2 shows the classification outcome (ptk_voiceless, ptk_voiced) for the canonically voiceless stops (/ptk/). The features that contribute the most to the classification of the voicing feature of /ptk/ (/ptk/ realized as voiceless [ptk] or voiced [bdg]) are the duration of the segment in question and the left context. When the voiceless stop is shorter than 0.045s, it is more likely to observe voicing alternation (/ptk/ produced as [bdg], 74% correctly predicted). If the segment duration is greater than 0.055s, voiceless stops tend to stay voiceless, with 65% correct prediction. If the segment duration is between 0.045 and 0.055s, the left context plays an important role in the prediction of the voicing pattern of /ptk/: when the voiceless stops are preceded by a voiced context (a voiced obstruent, a sonorant or a vowel), they are more likely to undergo voicing alternation (60% correctly predicted) and they tend to stay voiceless when preceded by a voiceless obstruent or a pause (77% correct prediction).

Figure 3 shows the classification outcome (bdg_voiced, bdg_voiceless) for the canonically voiced stops (/bdg/). The feature that contributes the most to the classification of the voicing feature of voiced stops is the right

context, followed by the left context, the duration of the preceding segment and the language. When the voiced stop is followed by a voiceless obstruent or a pause, it is predicted to undergo voicing alternation (88% correctly predicted). When the voiced stop is followed by a voiced obstruent, a sonorant or a vowel but preceded by an obstruent (voiced or voiceless) or a pause, voicing alternation is predicted for the segment (71% correctly predicted). Interestingly, among the contributing features, "language" is located towards the lower part of the tree. This suggests that particular languages tend to be 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 alternation rate seen for Portuguese in Figure 1.

	Factor \ Language	Fre.	Spa.	Ita.	Por.	Rom.
ptk	Left context	✔	✔	✔		✔
	Right context	✔		✔		
	Phone duration	✔	✔	✔	✔	✔
	Previous phone duration	✔				
bdg	Left context	✔	✔			✔
	Right context	✔	✔	✔		✔
	Phone duration				✔	✔
	Previous phone duration		✔	✔		✔
	Next phone duration	✔				✔
	Word position		✔		✔	
	Word duration				✔	

Table 2: Contributing factors for predicting /ptk/ (ptk_voiced *vs* ptk_voiceless) and /bdg/ (bdg_voiced *vs* bdg_voiceless) voicing patterns for each language.

The contributing features for predicting the voicing patterns of stops in each language are presented in Table 2. With respect to the voicing patterns for /ptk/ (ptk_voiced *vs* ptk_voiceless), phone duration is an influential feature for predicting voicing alternations in all five languages. More specifically, shorter phone durations correspond to a higher tendency towards voicing of /ptk/. Since the realization of voiceless stops as voiced stops in speech production is essentially a phenomenon of reduction and weakening, the results based on duration are consistent with linguistic theory. The surrounding context also plays a role, with voiced contexts favoring voicing alternation of the voiceless stops. With regard to voicing patterns for /bdg/ (bdg_voiced *vs* bdg_voiceless), the duration (of the phone in question, of surrounding phones or of the word containing the phone in question) is a recurring feature for predicting voicing alternations. In addition, voiceless contexts also generally predict voicing alternation of /bdg/.

In order to find out how well each algorithm generalizes, we made a prediction using the test data sets.

ptk			
Cat. \ Predict.		voiced	voiceless
All 5lang	voiced	**62**	38
	voiceless	27	**73**
Fre.	voiced	**61**	39
	voiceless	20	**80**
Spa.	voiced	**68**	32
	voiceless	27	**73**
Ita.	voiced	**65**	35
	voiceless	28	**72**
Por.	voiced	**67**	33
	voiceless	26	**74**
Rom.	voiced	**59**	41
	voiceless	24	**76**

bdg			
Cat. \ Predict.		voiced	voiceless
All 5lang	voiced	**69**	31
	voiceless	35	**65**
Fre.	voiced	**78**	22
	voiceless	46	**54**
Spa.	voiced	**78**	22
	voiceless	40	**60**
Ita.	voiced	**58**	42
	voiceless	31	**69**
Por.	voiced	**67**	33
	voiceless	41	**59**
Rom.	voiced	**75**	25
	voiceless	37	**63**

Table 3: Confusion matrices (left panel /ptk/; right panel /bdg/) obtained for each language. The correct prediction rates (%) on unseen test data are shown in bold.

The generated confusion matrices are shown in Table 3, based on the mean of 10 round-robin experiments with a random selection of data. The overall correct predictions of voicing alternation (voiceless stops /ptk/, voiced stops /bdg/) on unseen data are 73.6% for voiceless stops and 68.2% for voiced ones.

4. Summary and conclusions

The aim of this study was to apply a bottom-up method to extract linguistic knowledge from speech using machine learning techniques, i.e. decision tree-based classification, to investigate features that could help model voicing patterns of canonically voiceless (/ptk/) and voiced (/bdg/) stops. A total of $\sim$1000 hours of speech data in French, Spanish, Italian, Portuguese and Romanian were analyzed. We were able to isolate the most contributing features for predicting the observed voicing alternation patterns for all languages individually and combined. The decision tree classification results confirm that the surrounding (left and right) contexts and the segment duration are recurring contributing factors for predicting voicing alternation (both voiceless stops become voiced ones and voiced stops become voiceless ones). The trained algorithms exhibit satisfactory results when generalizing to new data sets. Our results suggest that machine learning techniques could help extract important linguistic knowledge from speech and open interesting research directions for the linguistic community.

5. Acknowledgments

This research was partially supported by DigiCosme (project ANR-11-LABEX-0045-DIGICOSME) and DATAIA / MSH Paris-Saclay "Excellence" grant.

6. Bibliographical References

Adda-Decker, M. and Hallé, P.-A. (2007). Bayesian framework for voicing alternation and assimilation studies on large corpora in french. In *Proc. ICPhS*, pages 613–616.

Adda-Decker, M. and Lamel, L. (2017). Discovering speech reductions across speaking styles and languages. In Francesco Cangemi, et al., editors, *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*. De Mouton Gruyter.

Boula de Mareüil, P., Vieru-Dimulescu, B., Woehrling, C., and Adda-Decker, M.). Accents étrangers et régionaux en français. caractérisation et identification. *Traitement Automatique des Langues*, 49.

Bowman, S. and Livescu, K.). Modeling pronunciation variation with context-dependent articulatory feature decision trees. In *Proc. ISCA Interspeech*.

da Silva, T. F., Gauvain, J.-L., and Lamel, L. (2011). Lattice-based unsupervised acoustic model training. In *Proc. IEEE ICASSP*, pages 4656–4659.

da Silva, T. F., Gauvain, J.-L., and Lamel, L. (2013). Interpolation of acoustic models for speech recognition. In *Proc. ISCA Interspeech*, pages 3347–3351.

Després, J., Lamel, L., Gauvain, J.-L., Vieru-Dimulescu, B., Woehrling, C., Le, V. B., and Oparin, I. (2013). The Vocapia Research ASR systems for Evalita 2011. In Cristina Bosco et al., editors, *Lecture Notes in Computer Science*, volume 7689, pages 286–294. Springer, Berlin Heidelberg.

Gauvain, J.-L., Lamel, L., and Adda, G. (2002). The LIMSI Broadcast News transcription system. *Speech Communication*, 37(1-2):89–108.

Gauvain, J.-L., Adda, G., Adda-Decker, M., Allauzen, A., Gendner, V., Lamel, L., and Schwenk, H. (2005). Where are we in transcribing french broadcast news? In *Ninth European conference on speech communication and technology*.

Naomi Gurevich, editor. (2004). *Lenition and Contrast. The functional consequences of certain phonetically motivated sound changes*. Routledge, New York and Londres.

Hallé, P.-A. and Adda-Decker, M. (2007). Voicing assimilation in journalistic speech. In *ICPhS*, pages 493–496.

Honeybone, P. (2008). Lenition, weakening and consonantal strength: tracing concepts through the history of phonology. In Joaquim Brandao de Carvalho, et al., editors, *Lenition and Fortition*, pages 9–92. Mouton de Gruyter, Berlin.

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

Hutin, M., Niculescu, O., Vasilescu, I., Lamel, L., and Adda-Decker, M. (2020b). Lenition and fortition of stop codas in romanian. In *Proc. SLTU-CCURL*.

Jatteau, A., Vasilescu, I., Lamel, L., and Adda-Decker, M. (2019a). Final devoicing of fricatives in French: Studying variation in large-scale corpora with automatic alignment. In *Proc. ICPhS*, pages 295–299, Melbourne, Australia.

Jatteau, A., Vasilescu, I., Lamel, L., Adda-Decker, M., and Audibert, N. (2019b). " gra[f] e!" word-final devoicing of obstruents in standard french: An acoustic study based on large corpora. In *Proc. ISCA Interspeech*, pages 1726–1730.

Lamel, L., Courcinous, S., and al. (2011). Speech Recognition for Machine Translation in Quaero. In *IWSLT*.

Marcello, F., Giordani, D., and Coletti, P. (2000). Development and Evaluation of an Italian Broadcast News Corpus. In *Proc. LREC*.

Meunier, C. and Bigi, B. (2016). Répartition des phonèmes réduits en parole conversationnelle. approche quantitative par extraction automatique. In *Journées d'Études sur la Parole*, number 31, page 9.

Meunier, C. and Espesser, R. (2011). Vowel reduction in conversational speech in French: The role of lexical factors. *Phonetics*, 39(3):271–278.

Niebuhr, O. and Meunier, C. (2011). The Phonetic Manifestation of French /sS/ and /Ss/ Sequences in Different Vowel Contexts: On the Occurrence and the Domain of Sibilant Assimilation. *Phonetica*, 68.

Pape, D. and Jesus, L. M. (2015). Stop and fricative devoicing in european portuguese, italian and german. *Language and speech*, 58(2):224–246.

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1):81–106.

R Core Team, (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ryant, N. and Liberman, M. (2016). Large-scale analysis of spanish /s/-lenition using audiobooks. In *Proceedings of the 22nd International Congress on Acoustics*.

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

Snoeren, N., Hallé, P.-A., and Segui, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Phonetics*, 34:241–268.

Therneau, T. M., Atkinson, B., and Ripley, M. B. (2010). The rpart package.

Vasilescu, I., Vieru, B., and Lamel, L. (2014). Exploring pronunciation variants for Romanian speech-to-text transcription. In *Proc. SLTU*, pages 161–168.

Vasilescu, I., Hernandez, N., Vieru, B., and Lamel, L. (2018). Exploring temporal reduction in dialectal spanish: A large-scale study of lenition of voiced stops and coda-s. In *Proc. ISCA Interspeech*, pages 2728–2732.

Vasilescu, I., Wu, Y., Jatteau, A., Adda-Decker, M., and Lamel, L. (2020). Alternances de voisement et processus de lénition et de fortition: une étude automatisée de grands corpus en cinq langues romanes. *Traitement Automatique des Langues*, 61(1).

Wu, Y. and Adda-Decker, M. (2020). Réduction temporelle en français spontané: où se cache-t-elle?

une étude des segments, des mots et séquences de mots fréquemment réduits (). In *Actes des 33èmes Journées d'Etude sur la Parole - JEP2020*, pages 627–635.

Wu, Y. and Adda-Decker, M. (2021). Réduction des segments en français spontané: apports des grands corpus et du traitement automatique de la parole. *Corpus*, (22).

Zhang, S. (2019). Data mining mandarin tone contour shapes. *arXiv preprint arXiv:1907.01668*.