Synchronic variation and sound change in Romance languages: corpus-based lenition phenomena in Romanian and Spanish

Ioana Vasilescu, Lori Lamel

Many diachronic phonological changes mirror synchronic phonetic variations, and as J. Ohala proposed, some historical processes can be "duplicated" in laboratory conditions through acoustic and perceptual experiments [1]. In this presentation we compare two **lenition phenomena** in Romanian and Spanish through corpus-based analyses, both being linked to phonological changes from Latin to modern Romance languages.

In Romanian we discuss the deletion of the masculine definite article -*l* (pom 'apple' – pomul 'the apple') whose function is taken by the desinence -*u*-, in connected speech, and we compare -*l* deletion in all attested contexts (word final +/-definite article, word initial and internal). Derived from the Latin personal pronoun *illu(m)*, the definite article is attested since Middle Ages, but a unsettled orthography with and without -*l* persists until the 20th century when the orthographic norm stabilizes and imposes a written -*l*. As for Spanish, we address the issue of the intervocalic voiced stops /bdg/ lenition in varieties from Spain and Latin America (Caribbean). The sound change is no longer active, however nowadays Spanish attests this phenomenon as synchronic variation. For both Romanian and Spanish, our purpose is to ground the phonological constraints and sound change patterns in the language specific and contemporary phonetic variation.

Our **approach** is in line with the methodological trend observed during the last two decades in phonetic studies which consists in increasingly using speech data collections gathered not only in controlled but also in natural settings. In this research framework, corpora evolved from small, task dependent, to large-scale containing dozens to thousands of hours, collected for various purposes, not necessarily linguistic, and heterogeneous in terms of recording conditions and speaker profile [2]. Corpora of this kind provide ideal conditions to study the patterns of phonetic variation, that are often linked to reduction phenomena [3]. Such patterns can be strictly speaking reduction, that is deletion of segments, but also can cover a wide variety of coarticulatory events, some of which refer to language specific phonological rules. Large corpora can be also beyond the human capacity of processing for the purpose of linguistic studies and require help from automatic processing. This assistance is provided by tools coming from speech technology developments, and specifically speech recognition, that through techniques such as forced alignment of the speech signal with the manual transcriptions, are able to convert virtually unlimited quantities of spoken data in material for corpus-based phonetic research.

The **corpora** used for this study are mainly coming from research projects dedicated to the development of speech-to-text systems for the two languages [4,5]. For Romanian we are using both semi-prepared and spontaneous speech, consisting in semi-prepared news, televised interviews, dialogues and read speech, and totaling 10.5 hours [6]. For Spanish we are using semi-prepared and spontaneous news in Peninsular (13 hours) and Latin American Spanish (Carribean variety, almost 4 hours). Those data are contrasted with a 5 hours corpus of telephonic conversational speech in Peninsular Spanish [5]. All data are manually transcribed.

On the **methodological** side, we use speech-to-text systems to estimate the frequency and triggering contexts of the two lenition phenomena. Specifically, we make use of the method which consists of the alignment with the speech signal of both canonical (phonological, here non lenited variants) transcriptions of words presents in dictionary and of their reduced or modified variants (lenited variants, e.g. *pomu* instead of *pomul* for Romanian, and *via* instead of *vida* for Spanish). This method has been already used in a wide range of phonetic and laboratory phonology studies, from broad quantification of reduction phenomena as function of the language and speaking style [7,8], to in depth analyses of the relation that can be established between variation and underlying phonological rules (e.g. schwa and liaison in French [9,10,11], dark vs light -*l* occurrence in English [12] ; minimal phonological contrast in the Romanian vocalic system [13]).

Using this method, we demonstrate that in Romanian, although in all the lexical contexts but the definite article the deletion of the liquid occurs randomly, due to connected (fast speech)

constraints, the definite article position undergo two constraints: (i) the right edge consonantal contexts triggers more often deletion and (2) the trend is reinforced by the speaking style (the spontaneous speech triggers deletion more often). Data suggest that -*l* deletion is not a sound change in progress, but rather a segment insertion triggered by orthography. It should be noted, however, that traces of -*l* definite article still persist in the nominal inflection (ie. genitive sg. pomu**l**ui 'of the apple', pl. pomi**l**or "of the apples"). In Spanish we measure the extent of the phenomenon as a function of geographical and stylistic repartition. Data show that the distribution of pronunciation variants across Peninsular and Latin American Spanish varieties is consistent with trends depicted by classical linguistic studies. More lenited variants are thus selected by the speech recognition system for Caribbean speech compared to Peninsular Spanish, and for telephonic conversational recordings compared to news. In an on-going study we also contrast intervocalic /bdg/ lenition with word initial and other lexical positions contexts. In comparison with Romanian, Spanish data clearly support the hypothesis that traces of sound change can be find in systematic synchronic variation.

[1] Ohala, J. J. 1989. Sound change is drawn from a pool of synchronic variation. L. E. Breivik & E. H. Jahr (eds.), Language Change: Contributions to the study of its causes. [Series: Trends in Linguistics, Studies and Monographs No. 43]. Berlin: Mouton de Gruyter, p. 173-198.
[2] Mirjam Ernestus. 2011. An introduction to reduced pronunciation variants. Journal of Phonetics, 3 (39):253–260, 2011.
[3] Adda-Decker, M., Lamel, L.. 2017. Discovering speech reductions across speaking styles and langages, in Cangemi, F., Clayards M., Niebuhr O., Schuppler B., & Zellers M. *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*, De Gruyter Mouton.
[4] Vasilescu, I., Vieru, B., Lamel, L. Exploring pronunciation variants for Romanian speech-to-text transcription. In SLTU-2014 – International Workshop on Spoken Language Technologies for Under-Resourced Languages, pages 161–168, 2014.
[5] Vasilescu, I., Hernandez, N., Vieru, B., Lamel, L. 2018. Exploring Temporal Reduction in Dialectal Spanish : A Large-scale Study of Lenition of Voiced Stops and Coda-s In INTERSPEECH 2018, 2728-2732, DOI : 10.21437/Interspeech, Hyderabad, India..
[6] Vasilescu, I., Chitoran, I., Vieru, B., Adda-Decker, M., Candea, M., Lamel, L., Niculescu, O. 2017. Studying variation in Romanian : deletion of the definite article -l in continuous speech. Linguistic Vanguard. Accepted.
[7] Martine Adda-Decker and Natalie Snoeren. 2011. Quantifying temporal speech reduction in French using forced speech alignment. Journal of Phonetics, 3(39):261–270.
[8] B. Schuppler, M. Ernestus, O. Scharenborg, & L. Boves (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics* 39, 96-109.
[9] Adda-Decker, M., Boula de Mareüil, P. & Lamel, L. 1999. Pronunciation variants in French: schwa & liaison", ICPhS, San Francisco (pp. 2239-2242).
[1à] Cécile Fougeron, Cédric Gendrot, Audrey Bürki. On the phonetic identity of French schwa compared to /ø/ and /oe/. *Actes des 5èmes Journées d'Etudes Linguistiques.*, 2007, Nantes, France. pp.191-198, 2007.
[11] Audrey Bürki, Mirjiam Ernestus, Cécile Fougeron, Cedric Gendrot, Ulrich Frauenfelder. What affects the presence versus absence of schwa and its duration : A corpus analysis of French connected speech. *Journal of the Acoustical Society of America*, Acoustical Society of America, 2011, 130 (6), pp.3980-3991.
[12] Yuan, Jiahong, and Mark Liberman. 2009. Investigating /l/ variation in English through forced alignment. In Proceedings of Interspeech.
[13] Margaret Renwick, Ioana Vasilescu, Camille Dutrey, Lori Lamel, and Bianca Vieru. 2016. Marginal contrast among Romanian vowels: Evidence from ASR and functional load. In Interspeech 2016, San Francisco, p. 2433–2436.