An automatic study of lenition of intra-lexical intervocalic /bdg/ and coda -s in Peninsular vs Latin America Spanish

N Hernandez I Vasilescu B Vieru L Lamel

We focus on two common lenition processes in Peninsular and Latin America varieties of Spanish : the weakening and/or suppression of the intervocalic /bdg/ and of the coda -s. They benefit of a long linguistic and sociolinguistic tradition, resulting in extensive studies of the historical mechanisms that triggered lenition (Hualde & Chitoran, 2016), the acoustic and prosodic features underlying the weakening processes (Hualde & Prieto, 2014; Torreira & Ernestus, 2012), or the dialectal and/or social patterns of lenition (Carasco et al., 2012, Ryant & Liberman, 2016). The present study examines the lenition from the perspective of the geographical and stylistic frequency of the phenomenon in large scale corpora. We make use of a large spoken dataset and of automated tools to quantify the incidence of lenition in Peninsular and Latin America varieties.

Corpora. The study is based on corpora totaling up to 22h of manually transcribed speech from Spain (17h) and Latin America (5h). Corpora cover different speaking styles and consist of both carefully articulated and conversational Spanish. Peninsular Spanish is illustrated by 11h hours of broadcast news in Castilian Spanish (BN_Spa), 44’ of broadcast conversations (BC_Spa) and 5h of telephonic conversations (TC_Spa). Latin America data consist of 3h20 of broadcast news (BN_LA) and of 5h of telephonic conversations (TC_LA) both illustrating the Caribbean variety, as well as by a corpus of monologues selected among recordings of the Catalogo de Voces Hispanicas (CVC_LA)[1]. 31 samples totaling 1h40' from Argentina (5 samples), Mexico (4), Bolivia (2), Colombia and Venezuela (3), Peru and Chile (2) and one for each of the remaining countries were included in the analysis. The samples have been merged according to the classification proposed by (Quesada Pacheco, 2014) in the following main varieties : Andean, Caribbean, Mexican, Rioplatense, Chilean.

Method. Speech-to-text systems are increasingly used by linguists as tools enabling the study of variation in large scale corpora (McAuliffe et al.,2015). Using as input material wav files and orthographic transcriptions, they can provide a sound-to-phoneme alignment of virtually unlimited amount of data. Through experiments that combine acoustic decoding and analysis of contextual distribution of pronunciation variants, forced alignment of big spoken corpora can also ground phonological representations in the everyday spoken realizations (Renwick et al., 2016), and estimate potential sound changes (Chitoran et al., 2017). In this study forced alignment of pronunciation variants is used to estimate the geographical and stylistic incidence of lenition of intra-lexical intervocalic /bdg/ and coda -s. The speech-to-text system designed for Peninsular Spanish (REF) is provided with speech files, orthographic transcriptions and a pronunciation lexicon linking the orthographic transcriptions with phonemic representations, and acoustic phone models. The lexicon based on canonical and frequent pronunciations of Spanish words is modified to encode the potential variation due to lenition. For each concerned word new pronunciation variants are defined that enable the selection of weakened pronunciations of the intra-lexical intervocalic voiced stops /bdg/ (e.g. pronunciation variants for “adulto” are /aðulto/ and /aulto/, the latter corresponding to a weakened articulation or deletion of /ð/), and of the coda -s (e.g. pronunciation variants for “estos” are /etos/, /eto/, /esto/ or /estos/). The system will select the variant that matches best the speech signal and the pronunciation, given the acoustic models. In the specific case of lenition, the system will select a reduced variant if a deletion or a phonetic undershoot of the target consonant occur. In this case the expected maximum span is not achieved and it will leave one segment out.

1Available at Centro Virtual Cervantes (http://cvc.cervantes.es/lengua/voces_hispanicas).

Results. Lenition rate is computed per item (VbV, VdV, VgV, coda -s) and per variety, given the relative frequency of each sequence in the data. Results show that lenition concerns all sequences and varieties of both Peninsular and Latin America Spanish, and all speaking styles at varying degrees (Table 1). The rate of lenition for both intervocalic /bdg/ and coda -s is higher for Latin America varieties compared to Peninsular Spanish. The speaking style factor plays an equally important role for all varieties: rate of lenition increases with less carefully articulated speech. However, spontaneous speech from both Peninsular and Latin American is equally and strongly affected by lenition of coda -s, as 62% of the -s are weakened or deleted. Inside the Latin America varieties, Caribbean, Chilean and Rioplatense known as "Low lands" (Quesada Pacheco, 2014) show a stronger tendency to both intervocalic /bdg/ and coda -s lenition compared to Andean and Mexican varieties ("High lands"). Finally, among the items, coda -s and intervocalic /d/ are the most affected by lenition, with 27.8% and 33.7% weakened realizations across corpora.

Table 1a & b. Lenition (in %) of intra-lexical intervocalic /bdg/ and coda -s across corpora.

Corpus	Lenition (%)			
	VbV	VdV	VgV	Coda -s
BN_LA	34	46.8	34.6	27.2
BN_Spa	19.8	22.8	26.8	15.3
CVC_LA	30.2	43.6	21.4	34.2
BC_Spa	31	40.2	31.3	25.3
TC_LA	60	75.8	71.4	62.1
TC_Spa	46.4	58.7	49.7	62.9

Variety	Lenition (%)			
	VbV	VdV	VgV	Coda -s
Andean	17.6	37.1	8.1	16
Mexican	26.6	53	29	20.4
Caribbean	38.3	58.7	24.1	45.3
Chilean	82.7	71.5	11.1	68.2
Rioplaten	34.7	42.2	23.8	47.8

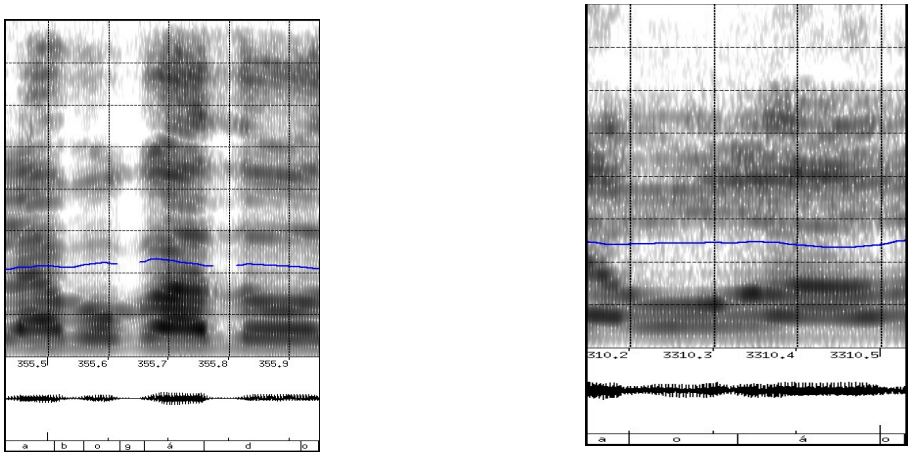

Figure 1. Spectrograms of abogado and aoao.

[1] Hualde, I., Chitoran, I.. 2016. Surface sound and underlying structure: The phonetics-phonology interface in Romance languages. Fischer and C. Gabriel. Manual of grammatical interfaces in Romance, 10, Mouton de Gruyter, 23-40, Manuals of Romance Linguistics; [2] Hualde, I. & Prieto, P. 2014. Lenition of intervocalic alveolar fricatives in Catalan and Spanish. *Phonetica* 71, 109–127; [3] Torreira, F. & Ernestus, M. 212. Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish. *Phonetica* 69, 124–148; [4] Carrasco, P., Hualde, J.,& Simonet, M. 2012. *Dialectal differences in Spanish voiced obstruent allophony. Costa Rican vs. Iberian Spanish*. Phonetica 69, 149–179 ; [5] Ryan, N. & Liberman, M. 2016. Large-scale analysis of Spanish /s/-lenition using audiobooks. *Proceedings of the 22d International Congress on Acoustics*, Buenos Aires, 2016 ; [6] Renwick, M., Vasilescu, I., Vieru, B., C. Dutrey, Lamel, L. 2016. *Marginal contrast among Romanian vowels: evidence from ASR and functional load*. Proceedings of Interspeech (San Francisco) ; [7] Chitoran, I., Vasilescu, I., Vieru, B. & Lamel, L. 2017 Connected speech in Romanian: Exploring sound change through an ASR system. Daniel Recasens & Fernando

Sánchez Miret (eds.), *Production and Perception Mechanisms of Sound Change*, Munich, Lincom Europa.