Measuring Second Language Acquisition of Spanish Lenition

Rachel Meyer and **Ratree Wayland**
Department of Linguistics, University of Florida

Kevin Tang
Department of English Language and Linguistics, Heinrich Heine University Düsseldorf

Sophia Vellozzi and **Rahul Sengupta**
Department of Computer Science, University of Florida

Abstract

A deep learning model, Phonet, was used to examine the degrees of lenition (weakening) of Spanish stop consonants produced by native English speakers during their study abroad. Instead of traditional quantitative acoustic methods, recurrent networks were trained to identify the posterior probabilities of sonorant and continuant phonological features. The results confirmed the expected factors affecting lenition patterns in L1 Spanish. These findings support the effectiveness of this approach as an alternative or complement to quantitative acoustic measures for studying lenition.

1 Introduction

Lenition, or consonant weakening, is a prevalent phonological process in Spanish. Most non-word-initial voiced stops /b, d, g/ are lenited to approximants [β̞, ð̞, ɣ̞], respectively (Hualde, 2005). Voiced stop lenition is a *gradient* process, spanning from fricatives to full deletion. This range of realizations is influenced by environmental characteristics, such as syllable stress, adjacent vowel height, and place of articulation (Kingston, 2008).

Learning how and when to produce these voiced stop allophones presents a challenge for Spanish second language (L2) learners. Previous studies have found that native English speakers acquire voiced stop allophones more slowly than other new Spanish sounds (Face and Menke, 2009). Zampini (1994) noted that English speakers struggle with Spanish voiced stops /b, d, g/ due to the absence of an allophonic lenition rule in English and slower learning of the phone [ð] compared to [β] and [ɣ]. With more Spanish experience, L1 English speakers adopt more native-like phonetic cues in their stop pronunciations (Shea and Curtin, 2011). Learners lenite more strongly in word-medial and unstressed positions (Nagle, 2017). However, acquisition of lenition is subject to large individual differences, with Salinas (2015) finding both improvements and regressions among low-intermediate and advanced learners. Overall, these studies underscore the complexity of lenition for L2 learners.

1.1 Phonetic gradience and posterior probability

To fully capture the varied and gradual degrees of lenition, it is essential to go beyond the categorical manifestations of these changes. Computational methods have been used to capture gradient phonetic variation, like pronunciation changes (e.g., [dʒ]-[z] and [p^{h}]-[f] variations in Hindi English code-mixed speech (Pandey et al., 2020) and 'g'-dropping in English (Yuan and Liberman, 2011); (Kendall et al., 2021)). These studies often use forced alignment systems, which input word-level transcriptions and reference pronunciation dictionaries to suggest probable pronunciations based on acoustic properties. For instance, to capture 'th' fronting, words susceptible to this variation might be assigned two pronunciations ([θ] and [f]) in the dictionary. A trained forced aligner then selects the more probable pronunciation based on the acoustic evidence.

Yuan and Liberman (2009) expanded on this method by using log probability scores from forced alignments to measure variation in /l/-darkness in American English. This method highlighted both categorical distinctions and finer degrees of /l/-darkness depending on the phonetic contexts. Other approaches like Support Vector Machines (SVM) have been used to classify r-full and r-less tokens in English using Mel-Frequency Cepstral Coefficients (MFCCs) as the acoustic representation (McLarty et al., 2019). These models, trained on canonical pronunciations, estimate variable realizations based on acoustic similarities.

Priva and Gleason (2020) demonstrate the potential of different modeling methods to capture

Proceedings of the Society for Computation in Linguistics (SCiL) 2024, pages 297-301.
Irvine, California, June 27-29, 2024

lenition processes, suggesting that comparing surface forms can effectively represent various lenition processes, irrespective of their underlying forms. Specifically, they used three methods to model lenition in American English, (e.g. /t/ $\rightarrow$ [d]) . One method compared surface forms only (all [t] and [d]), the second focused on surface forms that share the same underlying form ([t] and [d] from underlying /t/), and the third evaluated only unchanged segments (/t/ $\rightarrow$ [t], /d/ $\rightarrow$ [d]). Remarkably, all three approaches produced consistent results, indicating that various acoustic manifestations of a lenition process, such as /t/ $\rightarrow$ [d], can be effectively captured by comparing relevant pairs of surface segments, without reference to their underlying forms.

1.2 Phonological features and lenition

Lenition can be conceptualized abstractly as a change in certain phonological feature classes. Phonological features are abstract categories that group phonemes based on shared phonetic traits. The [continuant] class involves sustained airflow: [+continuant] includes fricatives, approximants, and vowels, allowing ongoing airflow despite partial closure, while [-continuant] sounds like stops and nasals, block airflow completely. In this study, two major phonological features classes are considered. The [+sonorant] class comprises phonemes like nasals, approximants, and vowels that allow relatively free airflow and resonance. Phonemes that are [-sonorant], such as stops and fricatives, which are produced with a substantial or complete obstruction of airflow through the vocal tract (see Hayes, 2008 for further detail on phonological features).

This study broadens examination of lenition beyond individual segment comparisons to encompass entire classes of segments defined by specific phonological features. Unlike the method used by Priva and Gleason (2020), which focuses on segment pairs, this approach categorizes segments into groups. Specifically, it assesses the probability of the [continuant] feature, which distinguishes stops from non-stops, and the [sonorant] feature, which separates stops and fricatives from non-stops and non-fricatives. These features encapsulate the primary categorical outcomes of stop lenition in Spanish. A high [continuant] probability coupled with a low [sonorant] probability indicates a fricative-like transformation, while both high [continuant] and [sonorant] probabilities suggest an approximant-like realization (Table 1). This approach diverges from Yuan and Liberman's (2009) method, which calculated phonetic variation from log probability differences between two alignments (dark /l/ vs. light /l/). Instead, it reflects degrees of lenition through the probabilities of phonological features derived from the acoustic properties of the signals.

To summarize, this study quantifies the lenition of Spanish stop sounds among native speakers of English over the course of a study-abroad program using a deep learning model, Phonet, that calculates posterior probabilities of phonological features to measure lenition continuously (Vásquez-Correa et al., 2019). This gradient approach can provide more detail on both the frequency and degree of lenition, and thereby a more nuanced view of the learning process. In addition, the measurement method is largely automatic, and can be tailored to any language and phonological feature set.

	Stop	**Fricative**	**Approx.**
Continuant	0	1	1
Sonorant	0	0	1

Table 1: Voiced stop allophone feature values.

2 Methods

2.1 Data

Data comes from the LANGSNAP project (MacWhinney, 2000; Mitchell et al., 2014). Twenty-seven native English speakers studying abroad in either Spain (n = 18) or Mexico (n = 9) participated in the project. All were students in a four-year Spanish program at a British university. The third year was comprised of a study-abroad for the entire the academic year. Comparison data was provided by ten native Spanish speakers from Spain (n = 8) and Mexico (n = 2) studying in the UK.

Participants completed three picture description tasks. After viewing a series of pictures, they retold the story depicted in their own words. The native Spanish-speakers completed the task once for each of the three sets of pictures, while the L2 learners completed the task for each set of pictures twice over the course of the project, for a total of six test times: pretest, three times while abroad, immediately after returning (posttest), and one year after returning (delayed posttest). The number of tokens for each target phoneme in word-initial and word-medial positions is displayed in (Table 2).

Phoneme	Word-Initial	Word-Medial
/p/	2617	394
/b/	1095	3644
/t/	1002	1312
/d/	2277	2128
/k/	2195	865
/g/	363	155

Table 2: Counts of target phonemes by word position.

2.2 Phonet

To derive gradient measures for [continuant] and [sonorant], we employ Phonet (Vásquez-Correa et al., 2019), a neural network model that predicts discrete phonological features from acoustic data. Phonet was designed for use with pathological speech, but this study expands it to non-pathological L2 speech. Phonet chunks the input signal into half-second segments, then computes the log energy signal distributed across 33 triangular filters along the Mel scale. This calculation is done for each 25-ms window in the chunk. Using this acoustic data and the force-aligned transcripts, Phonet learns the typical acoustic energy patterns of each phoneme. It uses two bidirectional GRUs (gated recurrent units) to account for coarticulation by considering the acoustic energy of the previous phone and following phones as well. This method is also used to learn the typical acoustic energy patterns for individual features.

We first force-aligned the audio and transcripts from LANGSNAP using the Spanish Montreal Forced Aligner (McAuliffe et al., 2017). These became the model input. We then trained a Phonet model on a corpus of Argentinian Spanish (Guevara-Rukoz et al., 2020), maintaining default parameters.[1]. The phonemes /b, d, g/ were excluded from the initial training process, since they would be featurally ambiguous. After training, the model was applied to all stop phonemes: /b, d, g, p, t, k/ in the LANGSNAP dataset. The model performed binary classification for each feature (see Table 1), and generated a corresponding posterior probability. Predictions were computed over 10ms windows. The model was quite accurate at predicting continuant (91%) and sonorant (92%) features.

2.3 Statistical Analysis

Linear mixed effect regression models were run in R with sonorant and continuant posterior probabilities as dependent variables. Following Kingston (2008), predictor variables were session (the six study-abroad timepoints, plus native speakers), syllable stress, stop voicing, stop place of articulation, word position (word-initial, word-medial), preceding vowel height (high, mid, low), and following vowel height. Session, place of articulation, and vowel heights were forward-difference coded, and the other variables were contrast-coded. To minimize coarticulatory effects, only intervocalic word-initial and word-medial stops were included in the regression analysis. Phones longer than two windows (20 ms) were filtered to only the middle third of the phone. Then posterior probabilities for these windows were averaged to give one posterior probability per phone.

3 Results

3.1 Continuant Posterior Probability

As expected, voiceless stops had a lower continuant posterior probability than voiced stops ($b = -.262, t = -48.867, p < .001$), and word-initial stops had a lower continuant posterior probability than word-medial stops ($b = -.079, t = 14.398, p < .001$) 1. Stops in stressed syllables had a lower continuant posterior probability than stops in unstressed syllables ($b = -.015, t = -2.863, p < .001$).

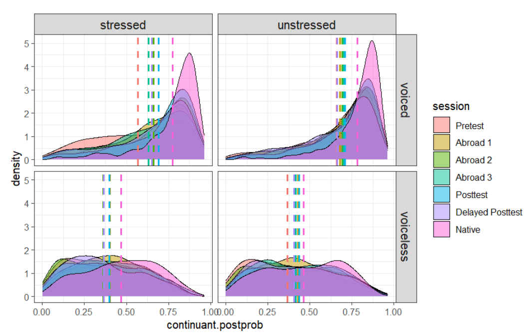

Figure 1: Continuant Posterior Probability by test time, voicing, and word position.

Dental stops had a higher continuant posterior probability than velar stops($b = .032, t = 4.483, p < .001$). Stops followed by a mid vowel had higher continuant posterior probabilities than stops followed by either high vowels ($b = -.016, t = 2.192, p = .028$) or low vowels ($b = .020, t = 3.405, p = .001$).

Mean posterior probabilities across all target phonemes by test time and word position for the learners compared to those of the native speak-

[1] https://phonet.readthedocs.io/en/latest/

ers are shown in Figure 2. The results of the regression models indicated that there was a significant increase in continuant posterior probability between pretest and the first abroad test time ($b = -.049, t = -2.931, p = .003$) but a significant decrease between the posttest and delayed posttest ($b = .042, t = 2.496, p = .014$). Native Spanish speakers also had a larger continuant posterior probability than L2 learners at delayed posttest ($b = -.100, t = -6.025, p < .001$). These results suggest that L2 learners learn to lenite voiced stops to some degree while abroad, but regress after leaving the immersion environment.

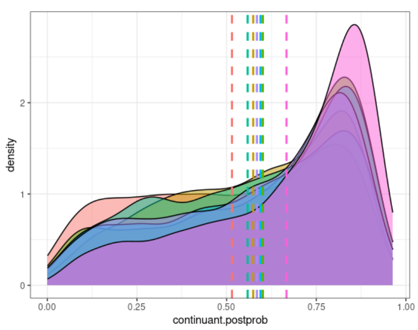

Figure 2: Continuant Posterior Probability by Test Time.

3.2 Sonorant Posterior Probability

As expected, voiceless stops had a lower sonorant posterior probability than voiced stops ($b = -.229, t = -43.655, p < .001$), and word-initial stops had a lower posterior probability than word-medial stops ($b = .047, t = 8.784, p < .001$) (Figure 3). There was no effect of stress.

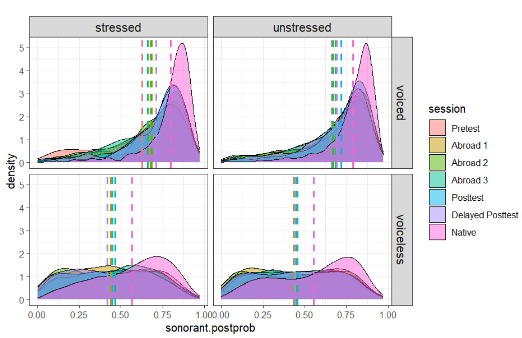

Figure 3: Sonorant Posterior Probability by test time, voicing, and word position.

Bilabial stops had a higher continuant posterior probability than dental stops ($b = .014, t = 2.567, p = .01$), and stops preceded by a mid vowel had higher continuant posterior probabilities than stops preceded by a high vowel ($b = -.026, t = -5.066, p < .001$).

There were no significant effects of session across any of the study-abroad timepoints. There was a significant difference between the delayed posttest and native speakers ($b = -.121, t = -6.419, p < .001$), suggesting that L2 speakers never demonstrated approximant-like productions (Figure 4.

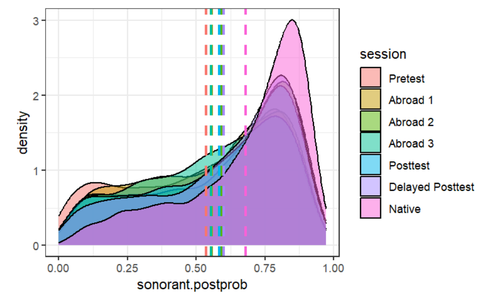

Figure 4: Sonorant Posterior Probability by Test Time

4 Discussion

These results show some acquisition of Spanish voiced stop lenition by L2 learners studying abroad. While their voiced stops became more continuant while abroad, L2 learners' voiced stops never increased in sonorance. The combination of high [continuant] and low [sonorant] posterior probabilities indicates a fricative production, signally that advanced learners learn *when* to lenite, but not *how strongly*. Furthermore, Phonet's posterior probability estimates of [continuant] and [sonorant] reliably measure lenition. These results show expected patterns of stronger lenition in word-medial, unstressed, and voiced environments, as in Nagel (2017). The increased gradience of our approach enabled us to see that even though learners made progress towards native-like acoustic cues, they never produced them in a native-like way.

Future work will also consider post-nasal stops, which do not undergo lenition, and post-lateral stops, an environment where /b, g/ lenite, but not /d/ (Hualde, 2005). Examining these post-nasal and post-lateral stops can show if learners overgeneralize lenition to all non-word-initial environments.

Acknowledgements

This research was funded by NSF-National Science Foundation (SenSE). Award No. 852 2037266

– SenSE

References

Timothy Face and Mandy Menke. 2009. Acquisition of the Spanish voiced spirants by second language learners. In *Selected Proc of the 11th Hispanic Linguistics Symposium*, pages 39–52.

Adriana Guevara-Rukoz, Isin Demirsahin, Fe He Fei, Shan Hui Cathy Chu Sarin Supheakmungkol, Knot Pipatsrisawat, Alexander Gutkin, Alena Butryna, and Oddur Kjartansson. 2020. Crowdsourcing Latin American Spanish for low-resource text-to-speech. In *Proc of the 12th LREC*, pages 6504–6513.

Bruce Hayes. 2008. *Introductory phonology*, volume 7. John Wiley & Sons.

José Ignacio Hualde. 2005. *The Sounds of Spanish*. Cambridge University Press, Cambridge, UK.

Tyler Kendall, Charlotte Vaughn, Charlie Farrington, Kaylynn Gunter, Jaidan McLean, Chloe Tacate, and Shelby Arnson. 2021. Considering performance in the automated and manual coding of sociolinguistic variables: Lessons from variable (ING). *Frontiers in Artificial Intelligence*.

John Kingston. 2008. Lenition. In *Selected Proc of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, pages 1–31.

Brian MacWhinney. 2000. *The CHILDES Project: Tools for Analyzing Talk*, 3rd edition. Lawrence Erhlbaum Associates, Mahwah, New Jersey. CHILDES-NICHD HD082736.

Michael McAuliffe, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. 2017. Montreal forced aligner: Trainable text-speech alignment using Kaldi. In *Proc Interspeech 2017*, pages 498–502.

Jason McLarty, Taylor Jones, and Christopher Hall. 2019. Corpus-based sociophonetic approaches to postvocalic r-lessness in African American Language. *American Speech: A Quarterly of Linguistic Usage*, 94(1):91–109.

Rosamund Mitchell, Patricia Romero, and Laurence Richard. 2014. Social networks, target language interaction and second language acquisition during the year abroad: A longitudinal study. *UK Data Archive*.

Charles L. Nagle. 2017. Individual developmental trajectories in the L2 acquisition of Spanish spirantization. *Journal of Second Language Pronunciation*, 3:218–241.

Ayushi Pandey, Pamir Gogoi, and Kevin Tang. 2020. Understanding forced alignment errors in Hindi-English code-mixed speech–A feature analysis. In *Proceedings of First Workshop on Speech Technologies for Codeswitching in Multilingual Communities*, pages 13–17.

Uriel Cohen Priva and Emily Gleason. 2020. The causal structure of lenition: A case for the causal precedence of durational shortening. *Language*, 96(2):413–448.

Lorena M. Botella Salinas. 2015. Lenition of voiced stops in L2 spanish speakers: Going from [bdg] to [βðɣ]. *LSO Working Papers in Linguistics*, 10.

Christine E. Shea and Suzanne Curtin. 2011. Experience, representations and the production of second language allophones. *Second Language Research*, 27:229–250.

Juan Camilo Vásquez-Correa, Philipp Klumpp, Juan Rafael Orozco-Arroyave, and Elmar Nöth. 2019. Phonet: A tool based on gated recurrent neural networks to extract phonological posteriors from speech. In *Proc Interspeech 2019*, pages 549–553.

Jiahong Yuan and Mark Liberman. 2009. Investigating/l/variation in English through forced alignment. In *Tenth Annual Conference of the International Speech Communication Association*.

Jiahong Yuan and Mark Liberman. 2011. /l/ variation in American English: A corpus approach. *Journal of Speech Sciences*, 1:35–46.

Mary L. Zampini. 1994. The role of native language transfer and task formality in the acquisition of spanish spirantization. *Hispania*, pages 470–481.