SEMANTIC PROCESSING TRIGGERS CROSS-LINGUISTIC INTERFERENCE DURING EARLY PHONETIC CATEGORY DEVELOPMENT

Joseph V. Casillas

Rutgers University
joseph.casillas@rutgers.edu

ABSTRACT

The present work explored the initial stages of adult L2 learning with a focus on the acquisition of the target language sound system. Specifically, the study analyzed the development of Spanish stop voicing in an immersion learning context in which L1 use was minimal and L2 input was maximized. Native English speaking learners of Spanish and bilinguals provided production data that were analyzed using Bayesian multilevel models. The analyses included a pre/post program evaluation of stop voicing in a delayed shadowing task, as well as a comparison with the participants' post-program production in a task that required semantic processing. The analyses revealed phonetic learning for stop voicing over the course of the program, though production gains diminished when the task required semantic processing. The results suggest L2 phonetic category formation can occur at an early stage of development, but L2 phonetic representations are unstable during initial stages of learning.

Keywords: Semantic processing, SLA, VOT

1. INTRODUCTION

Recent investigations on second language (L2) speech production have focused on the nature of the representations acquired by L2 learners at different stages of learning. One question posed by this line of research asks whether the distinct types of cognitive processes utilized during speech production influence L2 planning, programming, and execution of articulation, i.e., phonetic processing [5]. Recent studies show semantic processing affects speech production in monolingual [10, 13] and high-intermediate L2 speech [5]. The present investigation extends this work by exploring semantic processing effects (SPE) on L2 production in pure beginners and native bilinguals.

Research on SPE in L2 speech is motivated by investigations analyzing pathological speech. These studies showed that cognitive processes can affect phonetic processing in monolinguals suffering from Parkinson's [13] and aphasia [10]. SPE are believed to occur because semantic processing requires the speaker to first access meaning in order to retrieve the segments needed to produce the relevant lexical item. In a picture naming task speakers must access the meaning of the object in the picture before naming it. Reading and repetition tasks obviate this additional step, and thus avoid the need for semantic processing because speech segments can be accessed via direct mapping with orthography [13] and sound representations can be directly accessed from input phonology [10]. This raises important questions regarding L2 and bilingual speech. For example, what happens when the target language (TL) phonology is still developing, or the languages have phonologically similar, but acoustically different contrasts?

Gustafson et al. [5] showed SPE in L2 speech in their analysis of voice-onset time (VOT) of voiced bilabial and coronal stops in L2 learners of French. VOT refers to the time interval between the stop burst and the onset of voicing [8]. Similar to Spanish, French voiced stops are realized with lead (negative) VOT, and voiceless stops are realized with short-lag (positive) VOT. English, on the other hand, contrasts short-lag (positive) phonologically voiced segments with long-lag (positive), aspirated voiceless segments. Thus, for English speakers, learning to accurately produce French (or Spanish) stops implies reducing VOT. Implicitly this includes (1) developing a new phonetic category and (2) reducing cross-linguistic interference from the phonetically and perceptually similar L1 segments. For their analysis, [5] dichotomized VOT for voiced stops (lead vs. short-lag) and found picture naming increased the proportion of short-lag stops in high-intermediate learners.

The goal of the present study was to explore the ongoing acquisition of voice-timing in beginning adult L2 learners. Specifically, this work aimed to test for SPE on phonetic processing of Spanish stops, thus partially replicating [5] and extending it to a different language (Spanish), to more speech segments (voiced/voiceless bilabial, coronal and velar stops), and to adult L2 learners during the initial stages of phonetic category development.

2. METHOD

2.1. Participants

Ten adult native English speaking L2 learners of Spanish participated in the present study. They were enrolled in a Spanish immersion program and had a mean age of 23.7 (*SD* = 5.27) at the time of testing. To participate they completed a language background questionnaire to ensure (1) they did not have knowledge of any languages besides English and (2) had not spent more than 3 weeks in a Spanish speaking country. The program required all students sign a formal pledge by which they promised to speak only Spanish for 7 weeks (the duration of the program). Thus the learners received maximal TL input and used their L1 minimally.

Ten simultaneous (native) bilinguals, born and raised in Arizona, U.S.A., in Spanish speaking households, also participated. Members of this group had an average age of 23.27 (*SD* = 2.76) at the time of testing and reported using Spanish and English on a daily basis for as long as they could remember. They completed the Bilingual Language Profile (BLP) [1], which provided each participant with a language dominance score ranging from -218 to 218. Negative values implied dominance in English, positive values implied dominance in Spanish. The bilingual group had an average BLP score of -2.76 (*SD* = 38.93), thus they were considered balanced bilinguals for the purposes of this study.

2.2. Materials and Procedure

A delayed shadowing task in which speakers heard and repeated target phrases was administered first. The purpose of this task was to elicit production without requiring semantic processing. A picture naming task in which speakers said out loud the name of an object illustrated in an image was administered second. The purpose of this task was to elicit production that required semantic processing. Both tasks included 12 critical items containing the 6 Spanish stops in word initial position. Each word had a CV.CV syllable structure with paroxytonic stress. The stops were followed by either /a/ or /o/, forming real words that could easily be represented in an image. Both tasks utilized the same stimuli, but in different modalities. Table 1 lists the target items.

For delayed shadowing, the 12 items were embedded in the carrier phrase "_ es la palabra" (Eng. _ *is the word*). A 29 year old native Spanish speaker from Cádiz, Spain, provided the audio stimuli, which was recorded in a sound attenuated booth. The stimuli were presented randomly using PsychoPy2 [11], along with 38 distractor items. The distractors consisted of 30 CV.CV nonse words and 8 V.CV pseudowords, which were stimuli for other experiments. The 12 items were repeated 4 times by each participant, for a total of 960 tokens.

Table 1: Target items utilized across tasks.

	Bilabial	Dental	Velar
Voiced	bala	dado	gato
	bola	dodo	gota
Voiceless	palo	taco	caca
	polo	topo	coco

Of the 50 items used in the delayed repetition task, the 12 real words, along with 8 novel words, were selected for the picture naming task. The novel words were distractors and were selected from a list of vocabulary the learners knew. Thus, the distractors were only novel in that they had not previously been used in any of the tasks. Images of each of the 20 items were retrieved from the internet or created in Photoshop. PsychoPy2 [11] presented the images randomly on a laptop computer. The 12 items were repeated 4 times by each participant, for a total of 960 tokens.

A Shure SM10A dynamic head-mounted microphone recorded both groups, as well as the experimental stimuli used in the delayed shadowing task. A Sound Devices MM-1 pre-amplifier boosted the signal and sent it to a laptop computer were it was recorded using Praat at a 44.1 kHz sample rate with 16-bit quantization [2]. The learners completed all tasks in a quiet room on site at the immersion program. The delayed shadowing task was administered on a weekly basis throughout the course of the program, though only data from the first session (day 1 of the program) and the final session (week 7) are reported. The picture naming task was completed during the final week of the program, at least 1 day after the final session of the delayed shadowing task. The bilinguals completed the production tasks in distinct sessions separated by a minimum of 24 hours. Their productions were recorded in an isolated sound booth at a university in the U.S. southwest.

Recorded productions were segmented in Praat following standard procedures. Synchronized waveform and spectrographic displays served to hand-mark the onset of voicing and the burst of each stop. The voicing onset was considered the first periodic pattern found in the waveform. The criterion for bursts was the onset of broad-band sudden noise in the spectrogram. VOT was was calculated as the difference (in ms) between the onset of voicing and the burst.

2.3. Statistical analyses

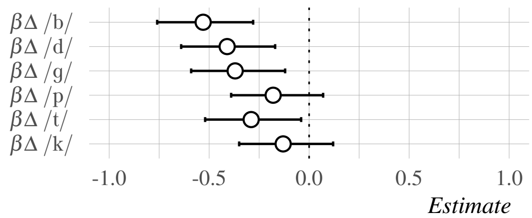
Three principle analyses are reported. For all models, VOT was standardized by phoneme and analyzed using Bayesian multilevel models fitted in `stan` via the R package `brms` [3].[1] VOT, the criterion variable for each of the 3 analyses, was modeled as a function of the fixed effects *exposure time* (analysis 1, start/end of program), *group* (analysis 2, learners/bilinguals), or *task* (analysis 3, non-semantic/semantic) and *item repetition*. In each case, the categorical predictor was sum coded (-1, 1), thus the posterior distribution of parameter estimates provided a sampling distribution of plausible values for the difference between each level of the fixed effect in standard units. The random effects structure included by-subject and by-item intercepts with random slopes for *exposure time* or *task* and *item repetition*. The same model specification was used to fit the data for each stop segment with 2000 iterations (1000 warm-up). Hamiltonian Monte-Carlo sampling was carried out with 4 chains distributed between 4 processing cores in order to draw samples from the posterior predictive distribution. The models included regularizing, weakly informative priors [4, 14]. Specifically, all parameters were assumed to be distributed as normal with a standard deviation of 1. The standard deviation parameters for random effects and the model residual error (sigma) were truncated to exclude negative values. The correlation parameter used an LKJ-correlation prior with regularization set at 2.

3. RESULTS

3.1. Phonetic category development

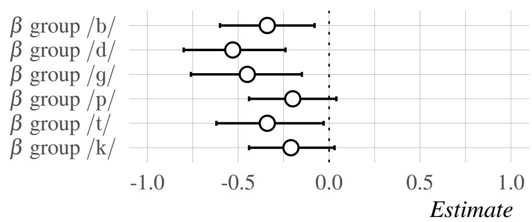
The 1st analysis examined the learners' pre- and post-program stop production in the delayed shadowing task. Figure 1 plots the point estimates and 95% credible intervals of change in VOT over time for each segment.[2] All point estimates were negative at the end of the program, which is the expected direction for a decrease in VOT values. The 95% credible intervals were rather wide and, in all cases except for /p/ and /k/, no longer contained 0, giving moderate evidence to support the notion that VOT values reduced for the voiced segments. There was weak evidence supporting the notion that the late learners, as a group, reduced aspiration in voiceless stops at the end of the program. An individual difference analysis showed that, by the end of the program, all participants produced pre-voiced stops some of the time. Furthermore, point estimates for change over time were negative with 95% credible intervals that did not encompass 0 for 6 participants for /b/, 5 participants for /d/, 5 participants for /g/, 3 participants for /p/, 2 participants for /t/, and 1 participant for /k/. Though production gains were unstable, the analysis suggests that learners' stops at the end of the program had become more native-like (i.e., overall reduction in VOT), with most change occurring in the voiced stops.

Figure 1: Parameter point estimates ± 95% credible intervals of Δ VOT over time.



The 2nd analysis compared the learners' post-program production in the delayed shadowing task with that of the control group of native bilinguals. Figure 2 plots the point estimates and 95% credible intervals of the *group* effect. There is a clear pattern for all segments: bilingual VOT production is lower in standard units when compared with the learners' productions. This suggests that the learners produced the voiced and voiceless stops differently from the native bilinguals. Specifically, their productions had overall longer VOT. This is apparent when sampling the posterior distributions, as the point estimates for β group were negative for all segments, though within native range for the voiceless coronal and velar segments. Taken together, the analyses showed that, overall, the learners did become more native-like at the end of the program, though their production of Spanish stops, by and large, still differed from that of the native bilinguals.

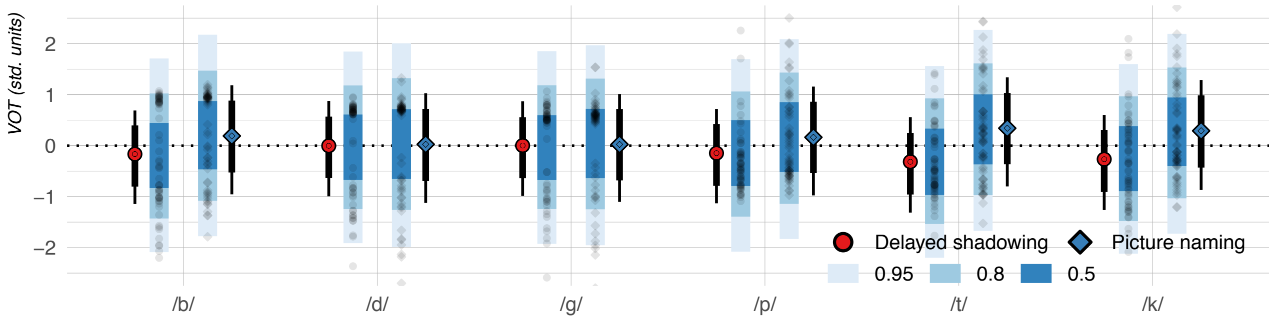
Figure 2: Parameter point estimates ± 95% credible intervals of group effect for each segment.



3.2. Semantic processing effect

The final analysis examined the SPE in the learner and bilingual groups. Specifically, we tested for the SPE on the phonetic processing of Spanish stops by

Figure 3: Raw data, posterior predictions, and posterior distributions of the means for VOT for each segment as a function of task.



directly comparing data from the delayed shadowing task with that of the picture naming task. Figure 3 plots the learners' standardized VOT data, posterior predictions, and posterior distributions of the means for each segment as a function of task. There was a non-zero, positive effect for all segments except /d/ and /g/, suggesting that picture naming increased VOT in the remaining stops. Specifically, there is strongest evidence for the SPE in the voiceless coronal and velar segments, as these posterior point estimates, which are interpreted as the SPE for each group segment, are largest with 95% credible intervals that are furthest from including 0 (See Table 2). With regard to the bilingual group, the SPE was practically non-existent for all segments. Furthermore, the posterior point estimates that were non-zero were small, and in the opposite direction with credible intervals that included 0 (see Table 2). In sum, the analysis suggests that (1) semantic processing affects phonetic processing of stops for learners, though the effect is not uniform across all segments, and that (2) native bilingual speech is unaffected.

Table 2: Parameter point estimates ± 95% credible intervals of the semantic processing effect as a function of segment and group.

Parameters	Estimate	Est. Error	l-95% CI	u-95% CI
Learners				
β SPE /b/	0.19	0.10	-0.02	0.39
β SPE /d/	0.00	0.10	-0.20	0.20
β SPE /g/	0.01	0.10	-0.19	0.21
β SPE /p/	0.17	0.10	-0.04	0.37
β SPE /t/	0.33	0.11	0.12	0.53
β SPE /k/	0.27	0.10	0.07	0.47
Bilinguals				
β SPE /b/	-0.06	0.10	-0.25	0.13
β SPE /d/	0.00	0.10	-0.19	0.19
β SPE /g/	-0.05	0.10	-0.24	0.15
β SPE /p/	-0.11	0.10	-0.30	0.08
β SPE /t/	-0.13	0.10	-0.31	0.06
β SPE /k/	0.00	0.10	-0.20	0.18

4. DISCUSSION AND CONCLUSION

The results of the analyses provided evidence for phonetic learning in adults. Specifically, L2 learners of Spanish improved their production of stop voicing after 7 weeks of TL exposure. Despite the general decrease in VOT, the learners still produced stops that overall fell outside the range of the bilingual group. Importantly, the analyses revealed that production gains were negated when the experimental task required semantic processing. That is, the late learners were particularly susceptible to SPE, which caused cross-linguistic interference. The same was not true for the bilingual group, whose stop production was unaffected by the nature of the experimental task.

The findings corroborate those of [5]. As is the case with the varieties of Canadian English and French investigated in [5], the two varieties of English and Spanish investigated here differ in the realization of voice-timing for stop contrasts. In our data, similar to [5], picture naming incited SPE in L2 learners. The absence of a task effect for bilinguals suggests that phonetic processing may be less susceptible to SPE as L2 proficiency increases. This hypothesis was also put forth in [5] based on self-reported speaking proficiency data. Future research could shed more light on this issue by including an objective measure of proficiency as a predictor variable.

The present study contributes to our knowledge on the relationship between cognitive processes and phonetic processing by extending the work of [5] to include (1) a different language (Spanish), (2) more L2 segments (voiceless stops and voiced velars), and (3) native bilinguals and beginning L2 learners. The results suggest that L2 phonetic category formation can occur at an early stage of development, but L2 phonetic representations are unstable during initial stages of learning and susceptible to semantic processing effects.

5. REFERENCES

[1] Birdsong, D., Gertken, L. M., Amengual, M. 2012. Bilingual language profile: an easy-to-use instrument to assess bilingualism.

[2] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [computer program].

[3] Bürkner, P.-C. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1), 1–28.

[4] Gelman, A., Simpson, D., Betancourt, M. Oct 2017. The prior can often only be understood in the context of the likelihood. *Entropy* 19(10), 1–13.

[5] Gustafson, E., Engstler, C., Goldrick, M. 2013. Phonetic processing of non-native speech in semantic vs non-semantic tasks. *The Journal of the Acoustical Society of America* 134(6), EL506–EL512.

[6] Kimball, A., Shantz, K., Eager, C., Roy, J. 2016. Beyond maximal random effects for logistic regression: Moving past convergence problems. *ArXiv e-prints*.

[7] Kruschke, J. K., Liddell, T. M. 2018. The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. *Psychonomic Bulletin & Review* 25(1), 178–206.

[8] Lisker, L., Abramson, A. S. 1964. A Cross-language Study of Voicing in Initial Stops: Acoustical Measurements. *Word* 20.3, 384–422.

[9] Nicenboim, B., Vasishth, S. 2016. Statistical methods for linguistic research: Foundational ideas–part ii. *Language and Linguistics Compass* 10(11), 591–613.

[10] Nozari, N., Kittredge, A. K., Dell, G. S., Schwartz, M. F. 2010. Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of memory and language* 63(4), 541–559.

[11] Peirce, J. W. 2008. Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics* 2(10), 1–8.

[12] van de Schoot, R., Depaoli, S. 2014. Bayesian analyses: Where to start and what to report. *European Health Psychologist* 16(2), 75–84.

[13] Van Lancker Sidtis, D., Cameron, K., Sidtis, J. J. 2012. Dramatic effects of speech task on motor and linguistic planning in severely dysfluent parkinsonian speech. *Clinical linguistics & phonetics* 26(8), 695–711.

[14] Vasishth, S., Beckman, M., Nicenboim, B., Li, F., Kong, E. J. 2018, in press. Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*.

[15] Vasishth, S., Nicenboim, B. 2016. Statistical methods for linguistic research: Foundational ideas–part i. *Language and Linguistics Compass* 10(8), 349–369.

---

[1] Bayesian data analysis (BDA) has emerged as an alternative to frequentist data analysis (see [15, 6]). The details behind how BDA works are beyond the scope of this work. The interested reader is encouraged to consult [7], [9], and [14] for descriptions and tutorials designed for linguists, as well as [12] for a more general presentation of BDA methods and reporting results for the Social and Behavioral Sciences.

[2] Model plots only include parameter estimates of the relevant predictor (i.e., exposure time, group, or task). Given that VOT was standardized as a function of phoneme, intercept terms were always 0 (the grand mean) and uninformative. Due to space limitations, parameter estimates for *item repetition* were also excluded. Repetitions had no effect on VOT in any of the models.