PROBABILISTIC REDUCTION IN SPANISH-ENGLISH BILINGUAL SPEECH

Khia A. Johnson

University of British Columbia
khia.johnson@ubc.ca

ABSTRACT

Speakers reduce segments to a greater degree when they are more predictable and frequent. The outcome of this probabilistic reduction varies cross-linguistically—for example, /s/ is more predictable and likely to reduce in Spanish than in English [10]. If probabilistic reduction reflects a speaker's expectations about language, what happens when there is more than one language to contend with? This paper reports on a corpus study of consonant reduction in Spanish-English bilingual speech. Given that bilinguals' languages influence one another, are consonant duration patterns better accounted for when languages are pooled together or kept separate? In a comparison of two linear mixed effect models, fit for pooled- and separate-lexicon models does not significantly differ. More importantly, this study largely fails to find evidence of segmental probabilistic reduction, suggesting a fundamental difference in how probability operates in bilingual speech.

Keywords: Bilingualism, Speech production, Probabilistic reduction, Information theory

1. INTRODUCTION

In bilingual speech production, there is a broad consensus that languages influence each other regardless of dominance [7], and are activated regardless of which is in use [18]. To this end, researchers are concerned with how units (e.g. words, segments) are stored, how speakers mitigate interference from the language not in use [23], and how *mutual influence* impacts production [13, 15]. This study addresses segmental interference—specifically, how measures of information (e.g. frequency, predictability, and informativity) operate in probabilistic reduction for bilingual speech. Probabilistic reduction for segments is of particular interest, because while languages may share similar segments, they are not necessarily reduced in the same ways. In monolingual speech, these cross-linguistic differences have been accounted for with information content [9], which sheds light on speaker knowledge and expectations about language [16]. For bilinguals, it follows that probabilistic reduction would also provide insight on knowledge and expectations about language, as well as for how languages interact.

1.1. Probabilistic accounts of reduction

For various linguistic units, an increased probability of occurrence corresponds with increased articulatory reduction, and has been widely used to account for variation in speech production [16]. Studies exploring the relationship between probability and reduction are typically grounded in *Information Theory* [25]. Information-theoretic accounts have addressed variation in production for words [24], morphemes [22], and segments [9]. Regardless of the unit, the most common measure of reduction is duration, where shorter duration corresponds to increased reduction. While duration is not the only choice [1], it is the most well understood [8, 20]. In each of these cases, probability can be quantified in with information-theoretic measures like frequency, local predictability, and informativity [9, 16].

Frequency is a measure of how often a unit is encountered. Words and segments with higher frequencies are more likely to be reduced [9]. However, while word frequency typically predicts word duration [24], it is an inconsistent predictor of segment duration [9], as units within a word are not necessarily uniformly affected [22]. Frequency, defined in (1), is the probability of unit x.

(1) $P(x)$

Local predictability builds on frequency by including context. For segments, context typically comprises all (or some) of the preceding segments in the word [9]. While this definition often produces robust results [27], it is almost certainly inadequate (see [22]). Local predictability, defined in (2), is the negative log probability of unit x in context c.

(2) $-log_2 P(x|c)$

Informativity builds on local predictability, and has been used to account for why some units are more likely to reduce, even when they are locally unpredictable [9, 24]. Informativity, defined in (3), is a weighted average of local predictability for a unit x, taking all possible contexts, c, into account.

(3) $-\Sigma_c P(c|x) log_2 P(x|c)$

1.2. The present study

Information-theoretic measures are unique to specific languages. This point has been made effectively by many researchers (e.g. [9, 16]), but only for monolingual speech. Assuming a bilingual speech is subject to parallel activation and mutual influence, what happens to probabilistic reduction? Are predictability effects separate-lexicon in nature, and tied to specific languages? Or pooled-lexicon, with bilinguals drawing on all available source material? These competing hypotheses are considered here in the case of bilingual consonant reduction patterns.

2. METHODS

This corpus study comprises a comparison of two linear mixed effects models—separate-lexicon and pooled-lexicon. The models use the same sample, dependent variable, and structure, but vary with respect to how the values for the segmental information-theoretic measures are calculated.

2.1. The database

This study uses the Bangor Miami corpus [12], which consists of conversational speech from highly-proficient Spanish-English bilinguals residing in Miami, USA. Speakers range in age from 9–78 years (median: 29.5 years), and typically knew each other. The corpus comprises 35 hours of recordings, with 34 percent in Spanish. Speakers conversed in various places, and many of the recordings are noisy. The corpus is transcribed orthographically by *utterance* (one main clause), with words tagged for language and part-of-speech. The transcriptions note unintelligible speech, disfluencies, and non-speech sounds, but do not contain sub-word annotations. As such, this study relies on transcriptions from the CALLHOME Spanish Lexicon [14] and Carnegie Mellon pronouncing dictionary [29].

2.2. Target segments

This study focuses on duration variation for word-medial, intervocalic /f/, /s/, and /ʧ/. If similar sounds are assumed to be linked (despite non-identical productions), Spanish and English substantially overlap. This assumption was not problematic in previous work on Spanish and English word-final sibilants [6] or word-initial stops [13]. Another reason for using these segments is that given their acoustic characteristics, they are robust under forced alignment. For a corpus without sub-word annotations, this increases tractability given time constraints.

2.3. Exclusionary criteria

Using the orthographic transcriptions and dictionary pronunciations, utterances were considered candidates if there was a target segment in either language—7896 utterances fit this search criteria. Utterances were excluded for code-switching (see [13]), disfluencies [4] and the presence of unintelligible speech. Targets were also excluded if they occurred in a function word [3], or a repeated word in the utterance [21]. For words with multiple target segments, all but one were randomly removed. This sample was further reduced to files where segment boundaries could be reliably marked, given the sub-optimal recording quality. As there were more Spanish than English targets at this point, a random sample was taken to match English, ensuring a minimum of five targets per speaker. The final sample comprised 2052 target segments.

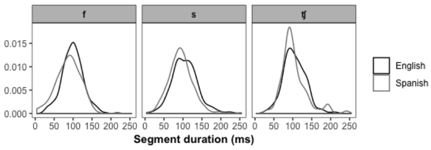
2.4. Duration measurements

The dependent variable was target segment duration. Target utterances were extracted from the corpus (stereo WAV files, sampling rate: 44.1 KHz), and automatically transcribed with the Montreal Forced Aligner, using pretrained acoustic models for both languages [19]. Target segment boundaries in the TextGrid annotations were hand-corrected in Praat [5] by inspecting the wideband spectogram and waveform, and supplemented by listening to the audio. Segment onset was defined as the point where high-frequency energy indicating frication first appeared for /f/ and /s/, and as the intensity minimum immediately following the preceding vowel for /ʧ/. Segment offset for all segments was defined as the intensity minimum immediately before the onset of periodicity for the following vowel. Figure 1 shows the distribution of target segment duration.

2.5. Control predictors

Each target segment was coded for six control fixed effect predictors. **Language** was a binary variable with possible values *Spanish* and *English*. **Mean syllable duration**—a gross measure of speech rate—was calculated over the utterance (including pauses), log-transformed, and centered. **Word posi-**

Figure 1: Duration is normally distributed, and is similar across languages for each segment.



tion was the distance from the end of the utterance in words, log-transformed, and centered. **Segment position** was the distance from the end of the word in segments, log-transformed, and centered. Diphthongs were always treated as two segments. **Stress precedes** and **stress follows** were binary variables with possible values *True* and *False*.

2.6. Information-theoretic predictors

Each target segment was coded for four information-theoretic fixed effect predictors. **Word frequency** is the number of times a word occurs in the corpus [9]. As Spanish and English do not make up equal halves of the corpus, relative word frequency for each language was log-transformed and centered. **Segment frequency** is the number of times a segment occurs in the corpus—relative values for each language were log-transformed and centered. **Local predictability** is the amount of information a particular segment has, given all preceding segments in the word. **Informativity** is the amount of information a segment typically has, given the above characterization of local predictability.

For each model, the values for the segmental information-theoretic predictors were calculated in different ways. In the separate-lexicon model, each target segment has a separate frequency and informativity value for each language, and local predictability was determined for each language, such that even if the same context occurs in the other language, the cross-language context was not considered. In the pooled-lexicon model, the converse was true. Target segments shared frequency and informativity values across languages, and local predictability included cross-language contexts.

2.7. Regression analysis

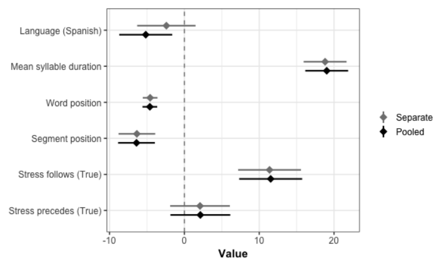
The goal of this analysis is to assess whether probabilistic reduction operates according to a separate- or pooled-lexicon model. Two linear mixed effects models were fit to the same data with the *lme4* R package [2]. The models share the same structure, dependent variable, and fixed effects outlined above. Also included were random intercepts for speaker, and by-speaker random slopes for segment frequency, local predictability, and informativity (as in [9]). Given the number of predictors, and need to limit model complexity, interaction terms were not included. As it stands, the number of observations in this study ($n = 2052$) is sufficient, following a conservative minimum of 50–100 per predictor [17].

3. RESULTS

Relative model fit can be assessed with the Akaike Information Criterion (AIC). While the pooled-lexicon model has a lower AIC value (separate: 19512.11, pooled: 19509.51, $\Delta = 2.60$), the difference is not significant. Note that in reporting overall model fit, collinearity is not problematic [30].

The output of pooled-lexicon model is reported here, as it fits the data better. The intercept was significant ($\beta = 125.76$, $SE = 7.84$, $t = 16.03$, $p < 2 \times 10^{-16}$), and the control fixed effects followed expected patterns. Increased mean syllable duration ($\beta = 19.03$, $SE = 1.46$, $t = 13.01$, $p < 2 \times 10^{-16}$) and following stress ($\beta = 11.54$, $SE = 2.15$, $t = 5.37$, $p < 8.63 \times 10^{-8}$) led to increased target segment duration. Preceding stress trended in the same direction, but was not significant ($\beta = 2.14$, $SE = 2.04$, $t = 1.05$, $p = 0.30$). Word position ($\beta = -4.61$, $SE = 0.50$, $t = -9.19$, $p < 2 \times 10^{-16}$) and segment position ($\beta = -6.40$, $SE = 1.25$, $t = -5.10$, $p < 3.7 \times 10^{-7}$) both led to decreased target segment duration, as expected due to final lengthening effects. The only control fixed effect that qualitatively differed across models was language. Spanish indicates shorter segment duration in the pooled-lexicon model ($\beta = -5.16$, $SE = 1.81$, $t = -2.86$, $p < 0.004$), but not in the separate-lexicon model ($\beta = -2.40$, $SE = 1.99$, $t = -1.21$, $p = 0.23$). Control fixed effect coefficients are depicted in Figure 2.

Figure 2: Control fixed effect coefficients and 95% CIs for both models. Positive estimates indicate increased duration, and negative decreased.

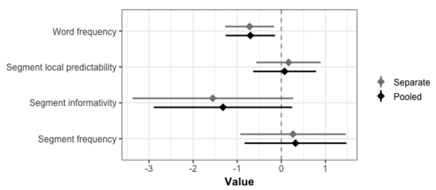


Of the information-theoretic fixed effects, only word frequency patterned as expected—an increase in word frequency ($\beta = -0.70$, $SE = 0.29$, $t =$

-2.46, $p < 0.02$) led to a small decrease in segment duration. Segment frequency, local predictability, and informativity did not follow expected patterns.

Local predictability has a clear interpretation—there is no effect on segment duration ($\beta = 0.07$, $SE = 0.36$, $t = 0.20$, $p = 0.84$). The results for segment frequency ($\beta = -4.24$, $SE = 1.31$, $t = -3.25$, $p < 0.001$) and informativity ($\beta = -7.12$, $SE = 1.83$, $t = -3.90$, $p < 0.0002$) exhibit strong collinearity in both models ($VIF > 3$). While this was not a problem for interpreting overall model fit [30], it renders the collinear fixed effects uninterpretable. To assess if the estimates were biased by collinearity, two checks were performed. First, the zero-order correlation for segment frequency and informativity with the dependent variable was compared against the model estimates. There was a change in sign for informativity ($R = 0.018$, $\beta = -7.12$), which suggests bias. There was no sign change for segment frequency ($R = -0.035$, $\beta = -4.24$). This was consistent for both models. Second, the model was refit twice, leaving out each collinear predictor in turn. With segment frequency removed, the effect size for informativity was reduced and no longer significant ($\beta = -1.32$, $SE = 0.80$, $t = -1.65$, $p = 0.10$). With informativity removed, the effect for segment frequency changed directions and was no longer significant ($\beta = 0.32$, $SE = 0.59$, $t = 0.54$, $p = 0.59$). All other fixed effects were stable. This finding was consistent across both models. These checks strongly suggest that the significant negative effects for segment frequency and informativity were artifacts of collinearity. The estimates from the refit models likely better reflect the actual effect, and as such, are depicted with word frequency and local predictability in Figure 3.

Figure 3: Information-theoretic fixed effect coefficients and 95% CIs for both models. Increased word frequency corresponds to decreased duration. All other factors are not significant.



4. DISCUSSION

This study addresses how segmental probabilistic reduction operates in the case of bilingual speech. Two linear mixed effects models were fit to a subset of utterances in the Bangor Miami corpus [12] and compared against each other. Each model represents a competing hypothesis about what information-theoretic measures draw on—a separate- or pooled-lexicon. As neither model gives a better fit, the far more interesting outcome is that this study largely fails to replicate basic monolingual findings for segmental probabilistic reduction. This differs from previous work showing effects for segment frequency, local predictability [1, 27], and informativity [9]. It is important to highlight that previous work has modeled segment duration for a much larger set of consonants [9], and this study focuses on just three. While Cohen Priva [9] does not report unique behavior for /f/, /s/, and /ʧ/, it is nonetheless possible that the results of this study follow from this particular subset of consonants and their phonological patterns in two languages.

Sample size ($n = 2052$) may be a limiting factor in this study, though given the model specified and result, it does not seem likely. While Cohen Priva and Jaeger [11] find an increased risk of spurious effects for small sample sizes, the main risk they report is finding a significant effect of segment frequency while failing to control for local predictability and informativity. They don't consider the outcome where all are included, but no effect is found. In this light, and because the sample size large enough for the complexity of the model [17], it is worth entertaining the more interesting possibility—bilingual speech is fundamentally different. This could be an artifact of bilingualism, the specific languages involved, or a combination of both. Recent work shows that the relative importance of probabilistic measures varies by language [28], and that defining context remains a major issue [26]. In addition to the strictly local context used here, researchers have observed effects from neighboring words [22, 24], and farther [21]. A major challenge going forward is to determine what aspects of context affect bilingual speech. It is possible that context dynamically shifts according the languages in use and code-switching (see [13]), or as an effect of recency and cumulative experience (see [7]). If probabilistic measures vary by language (as in [28]), context may serve as a modulating force.

While this study presents a null result, it highlights an important point—monolingual findings cannot be simply applied to bilingual speech. This study provides an important contribution towards understanding how probabilistic reduction operates in a new population, and in the process, treats bilingualism as an interesting question rather than as a complicating factor.

5. REFERENCES

[1] Aylett, M., Turk, A. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America* 119(5), 3048–3058.

[2] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.

[3] Bell, A., Brenier, J. M., Gregory, M., Girand, C., Jurafsky, D. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1), 92–111.

[4] Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., Gildea, D. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America* 113(2), 1001–1024.

[5] Boersma, P. 2002. Praat, a system for doing phonetics by computer. *Glot international* 5(9/10), 341–345.

[6] Brown, E. L., Harper, D. 2009. Phonological evidence of interlingual exemplar connections. *Studies in Hispanic and Lusophone Linguistics* 2(2), 257–274.

[7] Chang, C. B. 2013. A novelty effect in phonetic drift of the native language. *Journal of Phonetics* 41(6), 520–533.

[8] Clopper, C. G., Pierrehumbert, J. B. 2008. Effects of semantic predictability and regional dialect on vowel space reduction. *The Journal of the Acoustical Society of America* 124(3), 1682–1688.

[9] Cohen Priva, U. 2015. Informativity affects consonant duration and deletion rates. *Laboratory Phonology* 6(2), 243–278.

[10] Cohen Priva, U. 2017. Informativity and the actuation of lenition. *Language* 93(3), 569–597.

[11] Cohen Priva, U., Jaeger, T. F. 2018. The interdependence of frequency, predictability, and informativity in the segmental domain. *Linguistics Vanguard* 4(s2).

[12] Deuchar, M., Davies, P., Herring, J. R., Couto, M. C. P., Carter, D. 2014. Building bilingual corpora. In: Thomas, E. M., Mennen, I., (eds), *Advances in the Study of Bilingualism*. Multilingual Matters 93–110.

[13] Fricke, M., Kroll, J. F., Dussias, P. E. 2016. Phonetic variation in bilingual speech: A lens for studying the production–comprehension link. *Journal of Memory and Language* 89, 110–137.

[14] Garrett, S., Morton, T., McLemore, C. 1996. CALLHOME Spanish lexicon, LDC96L16. *Linguistic Data Consortium, Philadelpha, PA*.

[15] Goldrick, M., Runnqvist, E., Costa, A. 2014. Language switching makes pronunciation less nativelike. *Psychological Science* 25(4), 1031–1036.

[16] Hall, K. C., Hume, E., Jaeger, T. F., Wedel, A. 2018. The role of predictability in shaping phonological patterns. *Linguistics Vanguard* 4(s2).

[17] Jaeger, T. F. 2011. Corpus-based research on language production: Information density and reducible subject relatives. In: Bender, E. M., Arnold, J. E., (eds), *Language from a cognitive perspective: Grammar, usage, and processing. Studies in honor of Tom Wasow*. Stanford, CA: CSLI Publications 161–197.

[18] Kroll, J. F., Dussias, P. E., Bice, K., Perrotti, L. 2015. Bilingualism, mind, and brain. *Annual Review of Linguistics* 1(1), 377–394.

[19] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., Sonderegger, M. 2017. Montreal forced aligner: Trainable text-speech alignment using Kaldi. *Interspeech 2017*. ISCA.

[20] Munson, B., Solomon, N. P. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech Language and Hearing Research* 47(5), 1048–1058.

[21] Pluymaekers, M., Ernestus, M., Baayen, R. H. 2005. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62(2-4), 146–159.

[22] Rose, D. E. 2017. *Predicting plurality: An examination of the effects of morphological predictability on the learning and realization of bound morphemes*. PhD thesis University of Canterbury Christchurch, New Zealand.

[23] Runnqvist, E., Strijkers, K., Costa, A. 2014. Bilingual word access. In: Goldrick, M., Ferreira, V. S., Miozzo, M., (eds), *The Oxford Handbook of Language Production*. Oxford University Press –.

[24] Seyfarth, S. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133(1), 140–155.

[25] Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27(3), 379–423.

[26] Shaw, J., Kawahara, S. 2018. Predictability and phonology: Past, present and future. *Linguistics Vanguard* 4(s2).

[27] van Son, R., van Santen, J. P. 2005. Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication* 47(1-2), 100–123.

[28] Turnbull, R. 2018. Patterns of probabilistic segment deletion/reduction in English and Japanese. *Linguistics Vanguard* 4(s2).

[29] Weide, R. 2007. The Carnegie Mellon pronouncing dictionary [cmudict. 0.7b].

[30] Wurm, L. H., Fisicaro, S. A. 2014. What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language* 72, 37–48.