Phonological Processes in Code-Mixed Utterances: Tapping Across English-Spanish Language Boundaries

Hironori Katsuda, Gemma Repiso-Puigdelliura, Kie Zuraw*

University of Kansas; Universitat Autònoma de Barcelona; University of California, Los Angeles

1 Introduction

In a rule-based analysis (e.g., Chomsky & Halle 1968), phonological processes take the format A → B / X__Y, with the *structural description* XAY, composed of the *target* A and *environment* X__Y. For example, English cross-word tapping, roughly {t, d} → ɾ / V__#V, has the structural description V{t, d}#V, consisting of the target {t, d} and environment V__#V. In a constraint-based analysis (e.g., Prince & Smolensky 1993/2004), these processes can be captured by ranking a markedness constraint *XAY (*V{t, d}#V,) over a faithfulness constraint penalizing the change from A to B (IDENT(continuant), McCarthy & Prince 1995).

Phonological processes are language specific—cross-word tapping is a rule of English, and is not a rule of Spanish—but how exactly does this language-specificity work in the speech of bilinguals? For tapping to apply, does the target need to be part of an English word, or does the environment, or do they both?

Many bilinguals engage in code-switching, using more than one language within a single discourse or sentence (Poplack 1980). This can result in a match to a process's structural description that includes material from two languages. For example, in *He's go**t** abejas*, the target of tapping is in the /t/ in the English word *got*, and the environment for tapping is partially provided by the /a/ in the Spanish word *Abejas* 'bees'. Can a phonological process's structural description still match material from a different language, referred to here as the "incongruent" language? If so, does the process apply just as often? Does it matter whether the material from an incongruent language is the target or the environment?

To our knowledge, only the works of Macswan and Colina (2014), Olson (2019), and Henriksen et al. (2021) have addressed this issue. As reviewed below, these studies examine the same two Spanish phonological processes, *s*-voicing and spirantization (for Henriksen & al., only spirantization), which produce phonetically gradient outcomes even in monolingual utterances (Hualde, Simonet & Nadeu 2011; Garcia & Campos-Astorkiza 2015). This study aims to deepen our insight by examining and modeling English cross-word tapping, which typically results in acoustically categorical shifts from /t, d/ to [ɾ].[1]

In a production experiment, we show that tapping can apply in code-mixed utterances, at least when the target comes from the congruent language (English). However, for certain participants and conditions, the process applies less often than when all elements in the structural description are from English. Based on the effects we find of language dominance and experimental block, we argue that the lower rate of process application is not due to the phonological grammar itself, but rather to processing factors. We present a model that integrates processing factors to account for the differences across conditions and speaker groups.

1.1 *Previous studies* MacSwan and Colina (2014) examined Spanish s-voicing and Spanish spirantization in code-mixed sentences. S-voicing changes /s/ into [z] before a voiced consonant, while spirantization turns voiced stops into fricatives or approximants after a [+continuant] segment (1). Crucially, a word boundary may intervene between the target and the environment.

---

* We thank the audiences of AMP 2023, the North American Phonology Conference XII, and the 18th Conference on Laboratory Phonology, as well as the members of the UCLA Phonology Seminar, for their valuable feedback. We also thank our research assistants, Melissa Mendoza and Joshua Morales, for their support in running participant sessions.

[1] There have been conflicting findings on whether there is a smooth articulatory continuum from stops to taps that produces quantal acoustic outputs, or taps and stops can involve distinctive tongue movements (de Jong 1998; Fukaya & Byrd 2005; Derrick & Gick 2015; Derrick, Stavness & Gick 2015).

Proceedings of *AMP 2023*

(1) Spanish phonological processes examined in MacSwan and Colina (2014)
   a. *s*-voicing: /s/ → [z] / __ (#) [C, +voice]
   b. spirantization: {b, d, g} → fricatives or approximants / [+continuant] (#) __

In their study, five Spanish-English bilinguals read aloud code-mixed sentences containing sequences like *mis ghost* 'my ghosts', where a Spanish determiner ending in /s/ precedes an English noun starting with a voiced stop. For spirantization, target sentences contained sequences like *mi ghost* 'my ghost'. MacSwan and Colina found that /s/ changed to [z] in code-mixed sentences 20%-70% of the time, depending on the participant. This was slightly lower than the 45-90% rates seen in all-Spanish tokens (e.g., *hablamos de* 'we speak of'). Thus, the process applied even when its *environment* came from English, the incongruent language. For spirantization, fricatives or approximants appeared in just 1-7% of code-mixed tokens, as compared to 85-100% of all-Spanish tokens (e.g., *hablamos* 'we speak'). Thus, the process did not apply when its *target* came from the incongruent language.

Olson (2019) examined the same two phonological processes in 49 Spanish-English bilinguals. In addition to the all-Spanish condition (2a) and the Spanish-target/English-environment condition (2b) that MacSwan and Colina (2014) examined, Olson introduced an all-English condition (2c) and an English-target/Spanish-environment condition (2d).

(2) Sample stimuli in Olson (2019)

a.	All-Spanish: process applied	…*escuchas niños*… 'hear children' (s-voicing)
		…*exige guerras*… 'demand wars' (spirantization)
b.	Spanish-target/English-trigger: applied, but less than in All-Spanish	
		…*escuchas noises*… 'hear noises'
		…*flee guerra*… 'flee wars'
c.	All-English: process did not apply	…*eats nuts*…
		…*flee guns*…
d.	English-target/Spanish-trigger: did not apply	…*eats naranjas*… 'eats oranges'
		…*lleva guns*… 'carries guns'

Like MacSwan and Colina (2014), Olson found that the two processes applied when the target was congruent with the rule, and did not apply when it was incongruent.[2] Olson further found that both processes applied more strongly in all-Spanish sentences than in Spanish-target/English-environment sentences.

Henriksen et al. (2021) examined the Spanish spirantization rule in Afrikaans-Spanish code-switching. Consistently with the previous studies, consonants at the beginnings of Afrikaans words (that is, incongruent-language targets) did not undergo spirantization. Consonants at the beginnings of Spanish words did spirantize when the preceding word was Afrikaans, but less than when the preceding word was Spanish.

In summary, previous studies indicate that a phonological process can apply when the target comes from the congruent language and the environment is from an incongruent language. However, application is reduced compared to instances where all of the structural description is from the congruent language.

2 Experiment

2.1 *Materials* We conducted a production task with two language conditions. In the All-English condition, sentences were solely in English, like *The boy has got asparagus*. In the Code-Mixed condition, sentences began with English words and ended with a Spanish word, such as *The boy has got abejas*.

We formulated 30 target sentences for each condition, with the context for cross-word tapping occurring across the penultimate (Word1) and final word of the sentence (Word2). Word1 was English in both conditions (e.g., *got*), ending in /t/ in half of sentences and /d/ in the other half (e.g. *need*). Word2 was English for the all-English condition and Spanish for the Code-Mixed condition (e.g., *asparagus* vs. *abejas*). Word1

[2] The lack of *s*-voicing in *eats naranjas*-type sentences should be interpreted with caution, since Olson's stimuli involved a competing English process that makes *s* voiceless because of the preceding voiceless consonant.

could be a verb (e.g., *go**t***), an adjective (e.g., *grea**t***), or a preposition (e.g., *abou**t***), but Word2 was always a noun. We included 30 filler sentences in both conditions where Word2 began with a consonant, and thus did not create the context for tapping. The complete list of sentences is available on the project OSF page.

2.2 *Procedure* The experiment was administered online using PowerPoint slides, with Spanish-English bilingual research assistants proctoring. Participants were instructed to read the sentences as naturally as possible. The stimuli were grouped by condition (all-English or Code-Mixed), and the sequence of conditions was counterbalanced across participants. Within each condition, sentence order was randomized. The production task was repeated twice per participant, yielding two experimental blocks. Participants recorded their productions themselves using the smartphone app ShurePlus MOTIV®, which records uncompressed audio files with a sampling rate of 48.1 kHz and a sample size of 16 bits. Participants were coached to hold their phone's microphone directly facing their mouth at a distance of approximately four inches. After the experiment, participants filled out the Bilingual Language Profile (Birdsong, Gertken & Amengual 2012).

2.3 *Participants* A total of 54 Spanish-English bilinguals were recruited through a student participant pool at the University of California, Los Angeles. They obtained course credit for their participation. Of these, 21 were excluded because of noisy or missing audio, both primary caregivers during the participant's childhood being born in the US or having migrated from a non-Spanish-speaking country, and unexpected target productions in the baseline English condition. Of the final 33 participants (32 females, 1 male; mean age = 20.09 years, SD = 1.7 years), most had learned Spanish first (mean age of acquisition of Spanish = 0.30 years, SD = 0.95 years; mean age of acquisition of English = 2.87 years, SD = 2.70 years). All had at least one primary caregiver raised in a Spanish-speaking country. The majority (N = 23) had both caregivers from Mexico. Other combinations were both caregivers from Central America (specifically Guatemala and El Salvador, N = 4), one each from Mexico and the US (N = 3), one each from Mexico and Central America (N = 2), and one each from Central America and the US (N = 1).

The Bilingual Language Profile (BLP) (Birdsong, Gertken & Amengual 2012) was used to determine participants' dominance in either English or Spanish, based on four self-reported factors: Language History, Language Use, Language Proficiency, and Language Attitudes. Scores can range from -218 to 218, with negative values indicating a stronger dominance in Spanish and positive values in English. Scores near 0 indicate balanced bilingualism. Language dominance of our participants varied from -55 to 99.35, with a mean of 33.93 (SD = 33.34), meaning that participants were mostly English-dominant or balanced.

2.4 *Annotation* The target segments were narrowly annotated by the first two authors, using the categories described in Table 1. Discrepancies in annotation were resolved through discussion between them. The original annotations were then collapsed into the four groups shown. We grouped categories (h) and (i), which were unexpected and rarely observed, together with (f) and (g), since they all suggest a syllable boundary after Word1. Tokens were eliminated if we observed a pause longer than 150 msec. (n=140), disfluencies within the sentence (72), intonational focus on Word1 (23), the wrong word read (12), a word added (9), a word omitted (8), the whole sentence omitted (7), background noise (6), or strong creakiness extending beyond the target segment (4). This left 3,679 tokens (out of the original 3,960) for analysis.

<table>
<tr><th>narrow annotation</th><th>groupings for figures</th><th>grouping for statistics</th></tr>
<tr><td>a. [ɾ]: tap</td><td>[ɾ]</td><td>[ɾ]</td></tr>
<tr><td>b. [t/d]: [t] or [d]</td><td>[t/d]</td><td rowspan="8">non-[ɾ]</td></tr>
<tr><td>c. [Ø]: complete deletion</td><td rowspan="3">[ʔ]-variants (with [V̰] and [Ø] being viewed as reduced forms of [ʔ])</td></tr>
<tr><td>d. [V̰]: creaky vowel</td></tr>
<tr><td>e. [ʔ]: glottal stop</td></tr>
<tr><td>f. [t/d]+[V̰]: [t/d] followed by creaky vowel</td><td rowspan="4">[t/d]+[ʔ]-variants</td></tr>
<tr><td>g. [t/d]+[ʔ]: [t/d] followed by glottal stop</td></tr>
<tr><td>h. [ɾ]+[V̰]: [ɾ] followed by creaky vowel</td></tr>
<tr><td>i. [ɾ]+[ʔ]: [ɾ] followed by glottal stop</td></tr>
</table>

Table 1: Annotation categories

2.5 *Statistical analysis* We employed Bayesian mixed effects models using the *brms* package (Bürkner 2017) in R, with a binary dependent variable of [ɾ] (coded as 1) vs. non-[ɾ] (0), using a logistic link function. To avoid excessively complex models, we fitted separate models for underlying /t/s and /d/s.

Fixed effects were condition (All-English or Code-Mixed), language dominance (continuous), block (1st or 2nd), syntactic boundary (verb-noun, adjective-noun, or preposition-noun), and the interactions among condition, language dominance, and block. We included random intercepts for participants to capture baseline variations in tapping rate, and by-participant slopes for condition, block, and their interaction.[3]

The models were fitted with weakly informative priors, indicating no prior expectation for baseline realization or the effect of each factor. The intercept was specified as normal, with log-odds of producing a tap for /t/ or /d/ centered at 0, with standard deviation of 1. Priors for fixed effects were likewise specified as normal (0, 1) in log-odds space. The model was fitted by drawing 4000 samples in each of four Markov chains, discarding the first 1000 samples from each chain and analyzing 75% of the remaining samples.

We present the median estimate along with the 95% credible intervals (CrI) for an effect. In the Bayesian framework, a 95% CrI that excludes zero indicates "credible" evidence for an effect—that is, the model consistently estimates the same direction, with a distribution that is unlikely to include zero. We provide the percentage of the posterior distribution for an estimate that exhibits a specific direction, expressed as "pd" (probability of direction). This is calculated using the p_direction function in bayestestR (Makowski, Ben-Shachar & Lüdecke 2019). We regard values exceeding 95% as reliable evidence for an effect.

2.6 *Results* As shown in Figure 1, substantial tapping occurs even in the Code-Mixed condition, for both /t/ and /d/. In line with previous studies' results for Spanish spirantization and *s*-voicing, the rule can apply when its environment comes from a non-congruent language.

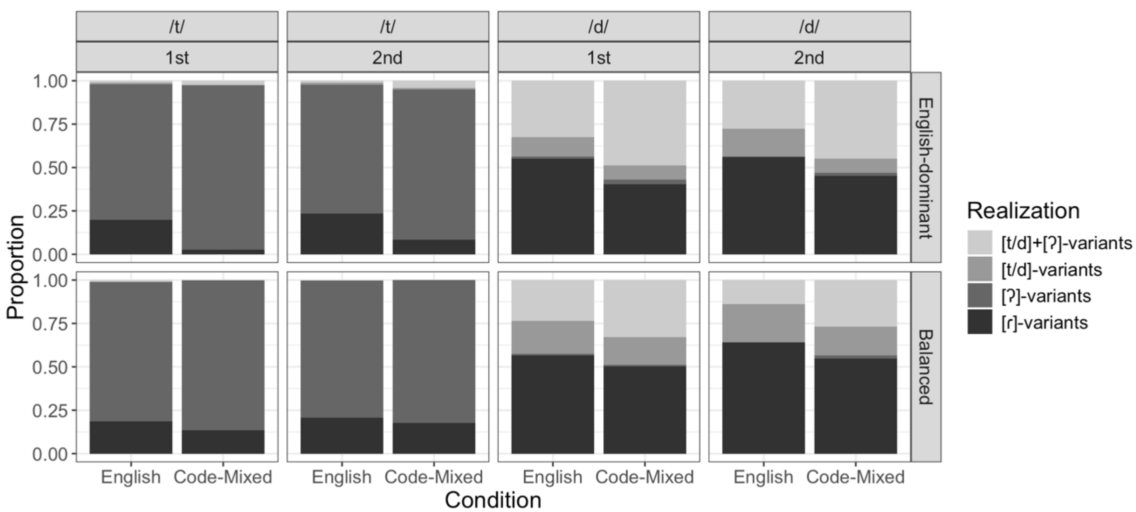

Figure 1: Distribution of realizations separated by block, language dominance and condition

The figure shows that the rate of tapping is consistently lower in the Code-Mixed condition. To understand this, we examined the effects of language dominance and experimental block. The figure splits participants into those who are more English-dominant and those who are more balanced bilinguals (with a cut-off at 40.78, the median value; the statistical analysis treats language dominance as a continuous variable), showing that the difference between the All-English and Code-Mixed conditions is driven by the English-dominant participants, particularly for /t/. The Bayesian analysis revealed credible effects of condition for both /t/ (β=-1.73, 95% Crl [-2.65,-0.85], pd=100%) and /d/ (β=-0.69, 95% Crl [-1.19,-0.20], pd=100%), indicating reduced tapping in the Code-Mixed condition. The model for /t/ also showed a credible

[3] Model specifications: brm(realization~condition*language.dominance*block+syntactic.boundary+(1+condition*block |participant)), data=/t/ and brm(realization~condition*language.dominance*block + syntactic.boundary+(1+condition*block|participant)), data=/d/

interaction between condition and language dominance (β=-0.71, 95% Crl [-1.36,-0.09], pd=99%): the more the participant is English-dominant, the greater the difference between the All-English and Code-Mixed conditions. For /d/, where the difference in tapping rates between conditions is smaller, there was no credible interaction between condition and language dominance.

We also noted a difference between experimental blocks. The second experimental block repeats the stimuli from the first block, so items are more familiar to participants in the second block. As the figure shows, the proportion of taps slightly increases in the second block compared to the first block, and the very lowest rate of tapping is for English-dominant participants in the first block. The statistical analysis revealed a credible effect of block for /d/ (β=0.26, 95% Crl [-0.00,0.52], pd=97%). Furthermore, the interaction between language dominance and block was credible for /t/ (β=0.36, 95% Crl [-0.04,0.75], pd=96%): the increase in taps during the second block is greater among participants who were more English-dominant. The three-way interaction among condition, language dominance, and block was nearly credible for /d/ (β=0.35, 95% Crl [-0.08,0.78], pd=94%), suggesting an increase in taps during the second block among participants with a stronger English dominance in the Code-Mixed condition.

3 Analysis

Across previous studies and our own, we have the following findings to account for. First, a phonological process cannot apply if its target is from the incongruent language, as in *lleva **g**uns*, but can when some or all of its environment comes from the incongruent language, as in *escuchas **n**oises*, *go**t** abejas*, *flee **g**uerras*. We present a model to account for this difference in section 3.1. Second, when environment material does come from the incongruent language, in our study balanced bilinguals apply the rule nearly as often (as if the environment all came from the congruent language), but English-dominant bilinguals apply the rule less often, especially when the stimuli are less familiar (first experimental block). We account for this in section 3.2, and discuss Olson's (2019) findings on language dominance.

3.1 *Target vs. environment* To explain the basic difference between incongruent target (process can't apply[4]) and incongruent environment (process can apply), we adopt the same assumption as MacSwan and Colina (2014): each word, or same-language sequence of words, is evaluated by its own language's phonological grammar, but neighboring words from the other language can also be visible in that evaluation. Like MacSwan and Colina, we implement this idea in Optimality Theory (Prince & Smolensky 1993/2004). (Olson 2019 takes a similar view, informally and using rules.) In *flee guerras*, for example, a tableau with the English constraint ranking evaluates *flee*; Spanish *guerras* is also visible to constraints, but Gen(), the function that generates output candidates, can change only the congruent-language material in each tableau, so the *guerras* portion is the same in every candidate. A tableau with the Spanish constraint ranking evaluates *guerras*, with English *flee* similarly present but not modifiable. The results are then spliced together.

(3)

English			Spanish		
/**fli**/ (+geras)	IDENT (continuant)	AGREE (stricture)	(fli+) /**geras**/	AGREE (stricture)	IDENT (continuant)
a. ☞ **fli** (geras)		** (ig, ge)	c. (fli) **geras**	*!* (ig, ge)	
b. **flɪt** (geras)	*!	*	d. ☞ (fli) **ɰeras**		*

Final output: [fli] + [ɰeras] - spirantization

MacSwan and Colina's AGREE(stricture) requires that stops/nasals/laterals be adjacent only to other stops/nasals/laterals, fricatives to other fricatives, and approximants/vowels to other approximants/vowels.

[4] MacSwan and Colina exclude phonetic effects, such as the influence of code-switching on Voice Onset Time (e.g., Toribio et al. 2005); presumably they would result from phonetic grammars that can handle multi-language sequences.

(The sequences [fl], [li], [er], [ra], [as], and [lt] also incur violations, but we show only those for [ig] and [ge]). In the English tableau, IDENT(continuant) prevents repairing the disagreement. In the Spanish tableau, IDENT is ranked lower, so spirantization applies, eliminating both AGREE violations. In brief, *guerras* can spirantize because in the Spanish tableau, where Gen() allows it to be modified, the constraint ranking allows spirantization, and the environment needed to trigger spirantization is present.

In *tiene guns*,[5] there is no way to repair the AGREE violations in the Spanish tableau for *tiene* without violating even higher-ranking constraints that are not shown (say, by deleting the word-final /e/ or turning it into a stop); the English portion of the candidate, [gʌnz], remains constant across all the candidates, so spirantizing English *g* is not an option in the Spanish tableau. In the English tableau, high-ranking IDENT prevents spirantization. In brief, *guns* cannot spirantize because the tableau where Gen() allows it to be modified is English, and even though the environment needed to trigger spirantization is present, the English grammar does not allow spirantization.

(4)

Spanish			English		
/**tjene**/(+gʌnz)	AGREE (stricture)	IDENT (continuant)	(tjene+)/**gʌnz**/	IDENT (continuant)	AGREE (stricture)
a. ☞ **tjene** (gʌnz)	**		c. ☞ (tjene) **gʌnz**		**
			d. (tjene) **ɰʌnz**	*!	

Final output: [tjene] + [gʌnz] –no spirantization

Why should each word be subject only to its own language's constraint ranking? MacSwan and Colina propose that the restriction follows from independent assumptions, including that constraint ranking is invariable within a language, so that two grammars can't be combined without producing ranking paradoxes. But this is not a universal assumption in the constraint-grammar world. For example, Daland and Norrmann-Vigil (2015) employ grammars in which each constraint is associated with a numerical weight (Goldwater & Johnson 2003), and each candidate's probability of being selected is a function of how many times it violates each constraint and what weights those constraints have. Daland and Norrmann-Vigil combine English and Spanish phonological grammars by interpolating between the two languages' weights for each constraint. We also employ weighted constraints below, so we must simply stipulate that the two phonological grammars apply separately; it is not a necessary outcome of independent assumptions about grammar architecture.

Why is material from the other language present in the candidates? Empirically, it would be incorrect to have a system in which only material from the congruent language is present in each tableau—one tableau for just *got* and one for just *abejas*—because it would prevent any cross-word rules from applying across a language boundary. We therefore accept as a necessary stipulation, for both monolingual and code-mixed utterances, that material beyond that being evaluated can be present in the tableau. In sub-section 4.2 below, we will make the "can be present" part of this stipulation more precise, by arguing that processing factors determine whether that other material actually *is* present in the tableau.[6]

3.2 *Lower rates of process application with an incongruent-language environment* Balanced bilinguals in our study tapped almost as often in mixed utterances (*got abejas*) as in all-English utterances (*got asparagus*), consistent with the idea presented above that the English tableau for *got* can include the following word, with markedness constraints insensitive to whether that word is English or Spanish.

By contrast, for English-dominant bilinguals, especially with /t/, there was sharply less tapping in the mixed utterances, especially in the first experimental block, where the sentences were less familiar. To explain this, we propose that the lower rate of tapping in *got abejas* is caused by processing factors. For more

[5] Changed from Olson's *lleva guns* for simplicity of constraint evaluation

[6] We leave open questions about the model that our experiment cannot address, including whether the non-congruent-language material in a tableau should be underlying or surface forms, and what happens to inserted material that does not belong to either word.

English-dominant bilinguals especially, retrieving a Spanish word could cause a slowdown in speech planning that prevents *abejas* from being accessed soon enough for it to be present in the tableau. For more balanced bilinguals, this effect would be smaller.

Readers might object that in picture-naming studies that cue participants which language to use on which trial, it is a switch into the *dominant* language that tends to cause the slowest reaction times (e.g., Meuter & Allport 1999, Costa & Santesteban 2004). However, other tasks have produced different findings. Out of various tasks that collect reaction times, self-paced reading and "word naming" (reading aloud a single word) are closer to what we asked participants to do, and there the results seem more consistent with our proposal. Bultena, Dijkstra and van Hell (2015), in a self-paced reading study, found that when a sentence switched from participants' L1 into L2, reading times slowed down, with more of a slowdown the lower the participant's L2 proficiency. Word-naming studies have tended to find slower reaction times for words in participants' L2 (e.g., de Groot et al. 2002, Friesen, Jared & Haigh 2014). In addition, some of the studies using picture-naming tasks (e.g., Costa & Santesteban 2004) also show overall slower reaction times in the non-dominant language than in the dominant language. We speculate that, in our reading task, participants planned the switch and recruited inhibitory skills prior to the code-switch location, thereby mitigating switch cost effects found in some picture-naming studies.

The idea that speech-production planning can affect cross-word phonology has been argued for in several studies, including some of English tapping (Wagner 2012; Tilsen 2012; Kilbourn-Ceron 2017a; Kilbourn-Ceron 2017b; Tanner, Sonderegger & Wagner 2017; Kilbourn-Ceron & Sonderegger 2018; Tamminga 2018; Kilbourn-Ceron, Clayards & Wagner 2020; Kilbourn-Ceron & Goldrick 2021). Kilbourn-Ceron et al. (2020), for instance, argue that producing a tap in *rabbi**t** ears* is challenging because the speaker has to retrieve the fact that *ears* begins with a vowel, and do so early enough to plan the tap. If *ears* is not retrieved in time, the phonological evaluation cannot treat *rabbit* as followed by a vowel, and tapping will not occur.

The above-cited studies support the idea that cross-word rules can depend on retrieval of the second word. Kilbourn-Ceron et al. (2020) found that the more predictable the second word (like *ears*), the more often tapping applies, presumably because more-predictable words are retrieved more quickly. Kilbourn-Ceron and Goldrick (2021) found that the more frequent the second word is, the more often tapping applies, and the frequency advantage decreased if participants were made to wait between seeing a written phrase and speaking it. Their interpretation is that if the speaker is allowed time for advance planning, then regardless of the second word's frequency, it will likely be retrieved in time to plan tapping. Tilsen's 2012 study of the English rhythm rule, which allows *Jàpanése géckos*, with a stress clash in the underlined syllables, to become *Jápanèse geckos*, also found an effect of preparation time: participants did not apply the rhythm rule when instructed to respond as soon as possible, but did when told to silently rehearse before speaking.

We did not design our materials to test for effects of frequency, predictability, or preparation time. However, by varying the language of Word2, by including participants with a range of language-dominance, and by presenting the stimuli twice, we effectively created conditions of varying Word2 retrievability: for the English-dominant participants, Spanish words are effectively lower frequency (all else being equal) than English words, which should make them slower to retrieve, especially the first time an utterance is encountered.

We are thus proposing that the English tableau for 'got' only *sometimes* includes the following word. When it does, tapping can apply, as shown in the tableau on the left in (5). When it does not, tapping does not apply (tableau on the right).

(5)

English: following word retrieved			
/**gɑt**/(+abexas)	*VTV	IDENT (place)	IDENT (cont)
a. **gɑt** abexas	*!		
b. ☞ **gɑɾ** abexas			*
c. **gɑʔ** abexas		*!	

English: following word not retrieved			
/**gɑt**/	*VTV	IDENT (place)	IDENT (cont)
d. ☞ **gɑt**			
e. **gɑɾ**			*
f. **gɑʔ**		*!	

Olson's results for Spanish *s*-voicing are fairly consistent with our delayed-retrieval account. Olson found that mixed utterances like *escucha**s** noises* had less *s*-voicing than all-Spanish utterances like *escucha**s** niños*. English-dominant participant showed little difference, consistent with reliably fast retrieval of *noises*. Balanced bilinguals showed a greater difference, and Spanish-dominant bilinguals the greatest, consistent with a tendency towards delayed retrieval of *noises* that could prevent /s/ from being voiced.[7]

Olson's results for Spanish spirantization of /b, d, g/ are not predicted by the delayed-retrieval account, but do not contradict it. Overall, Olson's participants produced less spirantization in mixed sequences like *flee **g**uerras* than in all-Spanish sequences like *exige **g**uerras*. This is surprising, because the target sound occurs *after* the word boundary: by the time the second word is being planned, it should already be known that the first word ends in a vowel. Across the spectrum of language dominance, Olson's participants showed a similar level of spirantization in the mixed versus all-Spanish sequences. Olson speculates that the reason is hyper-articulation: although results have been mixed, the literature suggests that immediately following a code-switch there may be hyper-articulation, including greater pitch range and longer duration (see Olson 2016, but also Aly 2017), perhaps as a correlate of unpredictability (see Myslín & Levy 2015). Hyper-articulation could block a word-initial lenition rule, like Spanish spirantization, but should not affect a word-final rule, such as Spanish *s*-voicing or English tapping. Hyper-articulation could also explain why we saw tokens like *nee[d ʔabejas]*, with a glottal stop inserted at the beginning of the Spanish word, more often in mixed sequences than in all-English sequences (even when excluding tapped tokens): the glottal stop could be part of hyper-articulating the word-initial vowel after the language switch.

3.3 */t/ vs. /d/* We found higher rates of tapping for /d/ than for /t/. Previous studies of American English tapping have found the same result (Herd, Jongman & Sereno 2010), the opposite (Kilbourn-Ceron, Clayards & Wagner 2020), or no difference between /t/ and /d/ (Raymond, Dautricourt & Hume 2006). One reason for the lack of consistency in the literature is that the outcome is not just binary (tapped vs. untapped), but includes a third major category of glottalization/glottal stop. In our results, /t/ is usually pronounced as creakiness or a glottal stop, and /d/ rarely is.

Eddington and Taylor (2009) and Eddington and Channer (2010) argued that glottal realizations of /t/ in just this context—word final, intervocalic—were on the rise in American English, especially among young women and Californians, and especially in read-aloud speech. The participants in our study performed a reading task and were mainly young, female Californians, so it is not surprising that they had such very high rates of glottalization, pushing down their tapping rate for /t/. In Section 3.4 below, we model this with a constraint against word-final [t].

3.4 *Full grammar model* We first present our model using the all-English utterance *got asparagus*. As illustrated in Figure 2, probabilistic outcomes arise from two points in the model. The first source of variation is speech planning (the top box). If second word is already available when the first word is being planned, a tableau will evaluate both words at once; in the model as fitted to our experimental data, this happens 90% of the time. The other 10% of the time, the second word is not available, and each word is evaluated in turn. The second source of variation is the grammar, which, as explained in more detail below, assigns a probability to each output candidate (boxes beginning "tableau"). In the left branch of the figure, where *got asparagus* is evaluated together, the most probable outputs are *go[ʔ]asparagus* (79%) and *go[ɾ]asparagus* (20%). In the right branch of the figure), the grammar first evaluates *got* alone. If the output is [gɑt] (3%), then the tableau for *asparagus* determines whether a glottal stop is inserted (99%) or not (1%). If the output is [gɑʔ] (97%) or [gɑɾ] (0%), the tableau for *asparagus* nearly always outputs faithful [əspɛɹəgəs], so for simplicity the figure does not show other possibilities. The overall probability of [gɑɾəspɛɹəgəs] is then predicted to be 90% * 20% + 10% * 0% * 100% = 18% (after rounding), close to the 21% that we observed.

[7] Olson's results are different from ours in that balanced bilinguals in his study (dominance score between -40 and 40 for him, between -55.11 and 35.79 for us) were sensitive to the language of the second word. Specifics of the task could affect where on the language-dominance scale participants' responses become noticeably sensitive to the language of the second word. In our study, stimuli included the same words recurring several times, facilitating their retrieval.

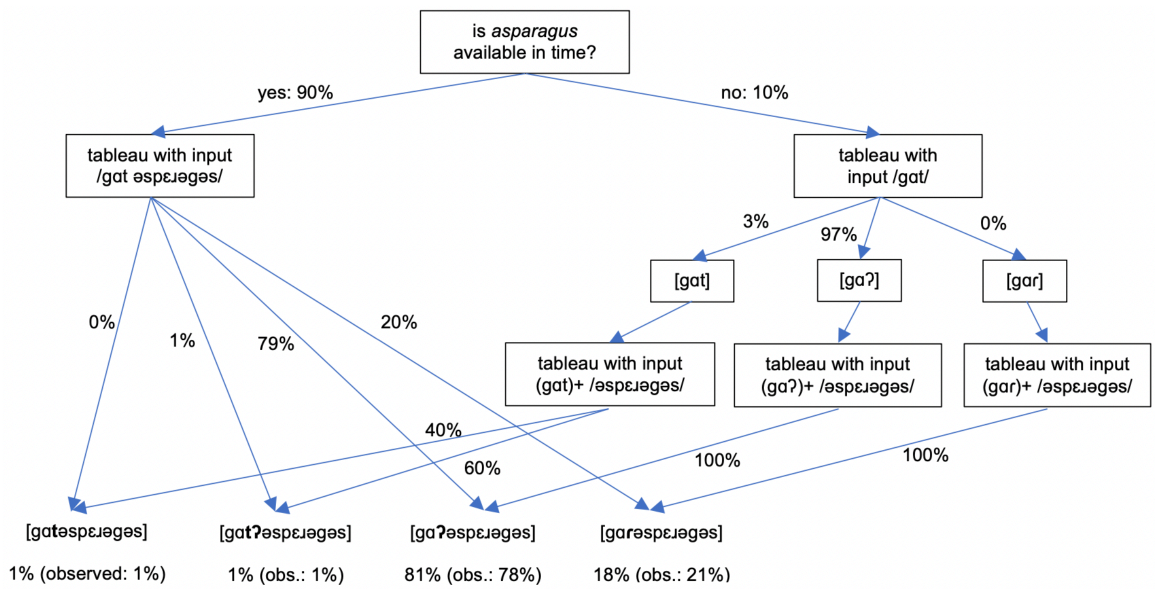

Figure 2: Full model for all-English /t/ utterances

We adopt a probabilistic grammar, specifically a Maximum Entropy constraint grammar (Goldwater & Johnson 2003), because if the only source of variation were speech planning with a binary choice between whether the second word is available or not, the model would produce only two pronunciation outcomes. As illustrated in (5), each constraint has a numerical weight, and each candidate receives a harmony score that is the weighted sum of its constraint violations. We exponentiate the negative harmony score (eHarmony) and normalize it to obtain the predicted probability that each candidate will be chosen. (Full set of tableaux is available on the project OSF page.)

(5)

English /gɑt əspɛɹəgəs/	*V{t,d}#V	IDENT(place)	IDENT(cont)	IDENT(voice)	*CODAT	*ɾ##	ONSET	NOCODA	ALIGN	DEP-C	harmony[8]	eHarmony	predicted probability
weights	1.20	1.24	0.00	2.61	4.78	18.01	0.00	0.34	13.78	0.79			
a. gɑt.əspɛɹəgəs	*				*		*	*			1.20+4.78+ 0.00+0.34 = 6.32	$e^{-6.32}$ = 0.002	0.002/0.367 = **0.00**
b. gɑ.təspɛɹəgəs	*								*		1.20+13.78 = 14.98	$e^{-14.98}$ = 0.000	0.000/0.367 = **0.00**
c. gɑt.ʔəspɛɹəgəs					*			*		*	4.78+0.34+ 0.79 = 5.91	$e^{-5.91}$ = 0.003	0.003/0.367 = **0.01**
d. gɑʔəspɛɹəgəs		*									1.24	$e^{-1.24}$ = 0.289	0.289/0.367 = **0.79**
e. gɑɾəspɛɹəgəs			*	*							0.00+2.61 = 2.61	$e^{-2.61}$ = 0.074	0.074/0.367 = **0.20**

[8] Weights are shown to two decimal places. Harmony is calculated on the full weights; discrepancies reflect rounding.

- *V{t,d}#V: penalizes any intervocalic [t] or [d], followed by a word boundary
- IDENT(place): penalizes changing coronal /t/ or /d/ to placeless [ʔ]
- IDENT(cont): penalizes changing [-continuant] /t/ or /d/ to [+continuant] [ɾ]
- IDENT(voice): penalizes changing [-voice] /t/ to [+voice] [ɾ], or changing [+voice] /d/ to [-voice] [ʔ]
- *CODAT: penalizes syllable-final [t]
- *ɾ##: penalizes utterance-final [ɾ]
- ONSET: penalizes every syllable with no onset consonant (Prince & Smolensky 1993/2004)
- ALIGN: penalizes syllabifying the final consonant of a word as an onset (McCarthy & Prince 1995)
- DEP-C: penalizes inserting [ʔ] (McCarthy & Prince 1995)
- NOCODA: penalizes every syllable that ends in a consonant (Prince & Smolensky 1993/2004)

If the second word is not present, as in (6), *V{t,d}#V cannot be violated, but the constraint *ɾ## can be. As a result, a slightly higher rate of [gɑt] outputs is predicted, and a much lower rate of [gɑɾ].

(6)

English /gɑt/	*V{t,d}#V	IDENT(place)	IDENT(cont)	IDENT(voice)	*CODAT	*ɾ##	ONSET	NOCODA	ALIGN	DEP-C	predicted probability
weights	1.20	1.24	0.00	2.61	4.78	18.01	0.00	0.34	13.78	0.79	
a. gɑt					*			*			0.03
b. gɑʔ		*		*				*			0.97
c. gɑɾ			*	*		*		*			0.00

The word *asparagus* is then evaluated in a separate tableau, with the realization of *got* already fixed.

In /d/ cases, the picture is very similar, except that with *CODAT and IDENT(voice) irrelevant, the tapping candidate in a two-word tableau like *need asparagus* receives a probability of 0.64. The predicted rate of tapping for *d* in all-English utterances is 90% * 64% = 58%, the same as what we observed.

We used Excel's Solver (Fylstra et al. 1998) to simultaneously fit the constraint weights and the probability of each type of tableau occurring, for the all-English case above, and for two additional cases that we will now illustrate: a mixed English-Spanish utterance such as *got abejas* under lowest difficulty (balanced bilinguals in Block 2 of the experiment) and highest difficulty (English-dominant bilinguals in Block 1). We did not allow the constraint weights to change across all three cases, but did allow the probability of retrieving the second word to change, reflecting varying processing difficulty.

The structure of the model for a mixed English-Spanish utterance such as *got abejas* is illustrated in Figure 3. The top branch of the model, "is *abejas* available in time?", still determines whether *abejas* can affect the pronunciation of *got*. In the left branch of the figure, where *abejas* is available in time, the candidate probabilities are the same as they were for *got asparagus*. And on the right branch, where *got* is evaluated on its own, the candidate probabilities are the same as they were above. Because we assume that each word is subject only to its own language's grammar, there will always have to be a second, Spanish tableau for *abejas*, which determines whether a glottal stop is inserted after [gɑt], using the Spanish weightings for ONSET and DEP-C. In the case where accessing *abejas* is easiest, the predicted probability of tapping is 70%*20% + 30%*0% = 14% (observed: 18%). In the case where accessing *abejas* is hardest, the predicted probability of tapping is 44%*20% + 56%*0% = 9% (observed: 3%).

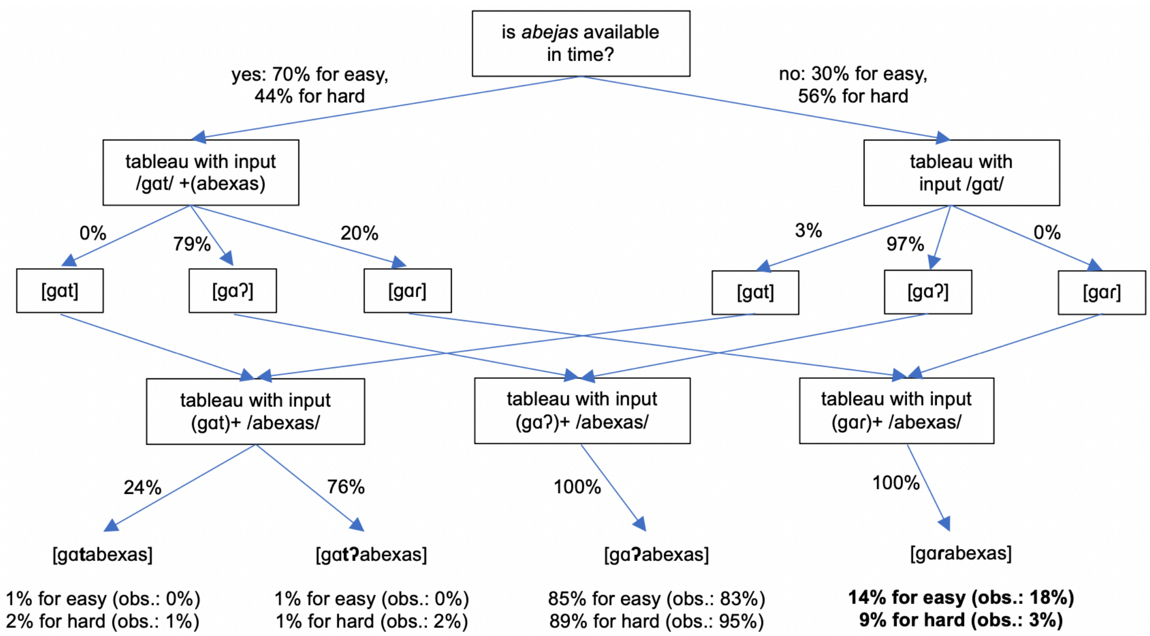

Figure 3: Full model for code-mixed /t/ utterances

4 Conclusion

We have demonstrated that English cross-word tapping applies even when part of its environment belongs to a Spanish word. However, its application rate was reduced in those cases, particularly for English-dominant bilinguals and during the first experimental block, where participants were reading each stimulus for the first time. Based on these observations, we argued that the reduction is not due to the phonological grammar itself being sensitive to language, but rather to processing factors. For English-dominant bilinguals, Spanish words are effectively lower frequency (all else being equal) than English words, potentially causing a delay in speech planning, which might prevent the vowel-initial-ness of the following word (e.g., *abejas*) from being accessed in time for it to be present in the tableau. We developed a model that incorporates these processing factors, accounting for the varied application rates across conditions and speaker groups.

5 Supplementary Files

The list of stimuli, data, analysis scripts, and the MaxEnt analysis spreadsheet are available in the OSF repository at osf.io/dexr8/?view_only=ec4b6d177fd44629b3b0dbf446487f8b.

References

Aly, Ann M. 2017. Prosodic effects of code-switching in Spanish-Basque bilinguals. UCLA PhD dissertation.

Birdsong, David, Libby Gertken & Mark Amengual. 2012. Bilingual language profile: An easy-to-use instrument to assess bilingualism. COERLL, University of Texas at Austin.

Bultena, Sybrine, Ton Dijkstra & Janet van Hell. 2015. Language switch costs in sentence comprehension depend on language dominance: Evidence from self-paced reading. *Bilingualism: Language and Cognition* 18: 453–469.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80: 1–28.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. Cambridge, MA: MIT Press.

Costa, Albert & Mikel Santesteban. 2004. Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language* 50: 491–511.

Daland, Robert & Ingrid Norrmann-Vigil. 2015. Toward a generative theory of language transfer: experiment and modeling of sC prothesis in L2 Spanish. *Open Linguistics* 1: 650-683.

de Groot, Annette M. B., Susanne Borgwaldt, Mieke Bos & Ellen van den Eijnden. 2002. Lexical decision and word naming in bilinguals: language effects and task effects. *Journal of Memory and Language* 47: 91–124.

de Jong, Kenneth. 1998. Stress-related variation in the articulation of coda alveolar stops: flapping revisited. *Journal of Phonetics* 26: 283–310.
Derrick, Donald & Bryan Gick. 2015. Accommodation of end-state comfort reveals subphonemic planning in speech. *Phonetica* 71: 183–200.
Derrick, Donald, Ian Stavness & Bryan Gick. 2015. Three speech sounds, one motor action: Evidence for speech-motor disparity from English flap production. *Journal of the Acoustical Society of America* 137: 1493–1502.
Eddington, David & Caitlin Channer. 2010. American English has goʔ a loʔ of glottal stops: social diffusion and linguistic motivation. *American Speech* 85: 338–351.
Eddington, David & Michael Taylor. 2009. T-glottalization in American English. *American Speech* 84: 298–314.
Friesen, Deanna, Debra Jared & Corinne Haigh. 2014. Phonological processing dynamics in bilingual word naming. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale*. 68: 179–193.
Fukaya, Teruhiko & Dani Byrd. 2005. An articulatory examination of word-final flapping at phrase edges and interiors. *Journal of the International Phonetic Association* 35: 45–58.
Fylstra, Daniel, Leon Lasdon, John Watson & Allan Waren. 1998. Design and use of the Microsoft Excel Solver. *INFORMS* 28: 29–55.
Garcia, Christina & Rebeka Campos-Astorkiza. 2015. Allophonic perception in Spanish: The case of intervocalic /s/ voicing. Paper presented at Phonetics and Phonology in Europe.
Goldwater, Sharon & Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In *Proceedings of the Workshop on Variation within Optimality Theory*, 111–120.
Henriksen, Nicholas, Andries Coetzee, Lorenzo García-Amaya & Micha Fischer. 2021. Exploring language dominance through code-switching: intervocalic voiced stop lenition in Afrikaans–Spanish bilinguals. *Phonetica* 78: 201–240.
Herd, Wendy, Allard Jongman & Joan Sereno. 2010. An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics* 38: 504–516.
Hualde, José Ignacio, Miquel Simonet & Marianna Nadeu. 2011. Consonant lenition and phonological recategorization. *Laboratory Phonology* 2: 301–329.
Kilbourn-Ceron, Oriana. 2017a. Speech production planning affects phonological variability: a case study in French liaison. In Karen Jesney & al. (eds.), *Proceedings of the 2016 Annual Meeting on Phonology*.
Kilbourn-Ceron, Oriana. 2017b. Speech production planning affects variation in external sandhi. McGill PhD dissertation.
Kilbourn-Ceron, Oriana, Meghan Clayards & Michael Wagner. 2020. Predictability modulates pronunciation variants through speech planning effects: a case study on coronal stop realizations. *Laboratory Phonology* 11: 1-28.
Kilbourn-Ceron, Oriana & Matthew Goldrick. 2021. Variable pronunciations reveal dynamic intra-speaker variation in speech planning. *Psychonomic Bulletin & Review* 28: 1365–1380.
Kilbourn-Ceron, Oriana & Morgan Sonderegger. 2018. Boundary phenomena and variability in Japanese high vowel devoicing. *Natural Language and Linguistic Theory* 36: 175-217.
MacSwan, Jeff & Sonia Colina. 2014. Some consequences of language design: codeswitching and the PF interface. In Jeff MacSwan (ed.), *Grammatical theory and bilingual codeswitching*. MIT Press: 185-210.
Makowski, Dominique, Mattan Ben-Shachar & Daniel Lüdecke. 2019. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software* 4: 1541.
McCarthy, John & Alan Prince. 1995. Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers in Linguistics* 18: Papers in Optimality Theory 10.
Meuter, Renata & Alan Allport. 1999. Bilingual language switching in naming: asymmetrical costs of language selection. *Journal of Memory and Language* 40: 25–40.
Myslín, Mark & Roger Levy. 2015. Code-switching and predictability of meaning in discourse. *Language* 91: 871–905.
Olson, Daniel. 2016. The impact of code-switching, language context, and language dominance on suprasegmental phonetics: evidence for the role of predictability. *International Journal of Bilingualism* 20: 453–472.
Olson, Daniel. 2019. Phonological processes across word and language boundaries: Evidence from code-switching. *Journal of Phonetics* 77: 1-16.
Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: toward a typology of code-switching. *Linguistics* 18: 581–618.
Prince, Alan & Paul Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Wiley.
Raymond, William, Robin Dautricourt & Elizabeth Hume. 2006. Word-internal /t,d/ deletion in spontaneous speech: modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation & Change* 18: 55–97.
Tamminga, Meredith. 2018. Modulation of the following segment effect on English coronal stop deletion by syntactic boundaries. *Glossa* 3: 1-27.
Tanner, James, Morgan Sonderegger & Michael Wagner. 2017. Production planning and coronal stop deletion in spontaneous speech. *Laboratory Phonology* 8: 1-39.
Tilsen, Sam. 2012. Utterance preparation and stress clash: Planning prosodic alternations. In Susanne Fuchs & al. (eds.), *Speech production and perception: Planning and dynamics*. Peter Lang: 119-156.
Toribio, Almeida Jacqueline, Barbara Bullock, Christopher Botero & Kristopher Allen Davis. 2005. Perseverative phonetic effects in bilingual code-switching. In Randall Gees & Edward Rubin (eds.), *Theoretical and Experimental Approaches to Romance Linguistics*. John Benjamins: 291–306.
Wagner, Michael. 2012. Locality in phonology and production planning. In *Proceedings of phonology in the 21st Century: In honour of Glyne Piggott / McGill Working Papers in Linguistics* 22.