Phonetic variation in bilingual speech: A lens for studying the production–comprehension link

Melinda Fricke [a,*], Judith F. Kroll [a], Paola E. Dussias [b]

[a] Center for Language Science, Department of Psychology, Pennsylvania State University, United States
[b] Center for Language Science, Department of Spanish, Italian, and Portuguese, Pennsylvania State University, United States

A B S T R A C T

We exploit the unique phonetic properties of bilingual speech to ask how processes occurring during planning affect speech articulation, and whether listeners can use the phonetic modulations that occur in anticipation of a codeswitch to help restrict their lexical search to the appropriate language. An analysis of spontaneous bilingual codeswitching in the Bangor Miami Corpus (Deuchar, Davies, Herring, Parafita Couto, & Carter, 2014) reveals that in anticipation of switching languages, Spanish–English bilinguals produce slowed speech rate and cross-language phonological influence on consonant voice onset time. A study of speech comprehension using the visual world paradigm demonstrates that bilingual listeners can indeed exploit these low-level phonetic cues to anticipate that a codeswitch is coming and to suppress activation of the non-target language. We discuss the implications of these results for current theories of bilingual language regulation, and situate them in terms of recent proposals relating the coupling of the production and comprehension systems more generally.

© 2015 Elsevier Inc. All rights reserved.

Introduction

Despite the many potential pitfalls encountered during spontaneous conversation, communication between adult native speakers is generally relatively fluid and effortless. How do speakers and listeners coordinate the numerous subgoals involved in fluent language use, and to what extent does this coordination rely on sensitivity to the distributional properties of natural speech? This paper reports three studies that exploit the unique properties of bilingual language production and comprehension to investigate the resources available to interlocutors during spontaneous communication. Specifically, we examine the processing of phonetic variation related to cross-language activation during English–Spanish codeswitching. The overarching hypothesis is that pressures on the production system can give rise to regularities in the phonetic variation present in the speech stream, and that members of the speech community can ultimately come to exploit these regularities during comprehension. In essence, this proposal is a phonetic analogue of MacDonald's (2013) Production–Distribution–Comprehension account; but where MacDonald focuses on morphosyntactic aspects of linguistic form and processing – asking how processes related to memory retrieval affect word choice and syntactic formulation – we argue that the same logic can be applied to processes involved in the production and comprehension of phonetic variation. We note here that neither the production nor comprehension sides of this proposal are entirely novel. Previous studies of language production have explored processing-related sources of phonetic variation (for monolingual speakers, cf. Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Gahl, Yao, & Johnson, 2012; Goldrick &

* Corresponding author at: Center for Language Science, Psychology Department, 112 Moore Building, Pennsylvania State University, University Park, PA 16802, United States.

E-mail address: mdf18@psu.edu (M. Fricke).

Blumstein, 2006; for bilingual speakers, cf. Amengual, 2012; Goldrick, Runnqvist, & Costa, 2014; Jacobs, Fricke, & Kroll, 2015), and studies of language comprehension have repeatedly demonstrated that listeners develop acute sensitivity to low-level phonetic regularities (for monolingual listeners, cf. Beddor, McGowan, Boland, Coetzee, & Brasher, 2013; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; McMurray, Tanenhaus, & Aslin, 2002; for bilingual listeners, cf. Ju & Luce, 2004). The novel aspect of our proposal is its focus on the interplay between these processes, and on the mechanisms that support them. To better understand the processes and mechanisms involved, we take advantage of a particular type of language use that sheds light on the relationship between psycholinguistic processing and phonetic variation: codeswitching.

Why codeswitched speech?

Codeswitching offers a window into the relation between linguistic form and processing. Codeswitching is a specialized form of language use, subject to a unique set of linguistic (e.g., Myers-Scotton, 2002; Poplack, 1980; Torres Cacoullos & Travis, 2015) and psycholinguistic (e.g., Broersma & De Bot, 2006; Hartsuiker & Pickering, 2008; Kootstra, Van Hell, & Dijkstra, 2012) constraints, and engaging a dedicated mode of language control processes (e.g., Green & Abutalebi, 2013). Habitual codeswitchers are able to regulate the activation of psycholinguistic representations in a way that allows them to fluidly interleave their two languages without obvious disruptions in processing. The surface form of codeswitched speech thus ultimately reflects the end of a long chain of complex processing events, including production-internal processes (Levelt, 1989) as well as interactions between the production and comprehension systems (e.g., Kootstra, van Hell, & Dijkstra, 2010; Loebell & Bock, 2003). On the flip side, listeners' responses to codeswitched speech can provide an index of their expectations given previous experience processing a particular linguistic input (Valdes Kroff, Dussias, Gerfen, Perrotti, & Bajo, 2015). In sum, codeswitched speech presents a rich and relatively transparent opportunity for investigating the ways in which the members of a speech community come to produce and comprehend variation in linguistic form.

At this point, it is important to draw a distinction between the study of codeswitched speech and the study of language switching. In studies using the language switching paradigm (e.g., Costa & Santesteban, 2004; Meuter & Allport, 1999), participants are typically cued to switch between their two languages while naming digits or pictures, and the requirement to switch languages generally results in a "switch cost" to response times. It is therefore reasonable to hypothesize that the need to regulate two languages constitutes a source of pressure on the production planning system that could impact the surface form of codeswitched speech. However, experimentally manipulated language switching differs from codeswitching in a number of important ways. Language switching studies typically examine single word production, where the target language can vary at random and is determined by the experimenter (but see Gollan &

Ferreira, 2009). During codeswitching, by contrast, grammatical planning mechanisms are fully engaged, the language of all lexical and morphosyntactic elements is fully under the control of the speaker, and in a normal conversational setting, production must additionally be coordinated with comprehension of the interlocutor's speech (Gullberg, Indefrey, & Muysken, 2012). A critical question, then, is whether the switch costs (and cross-language phonological influence; see below) that have previously been observed in experimental settings are particular to laboratory speech, or whether there is any evidence that language regulation has appreciable consequences during spontaneous conversation; it is in no way clear that production planning in these two settings is subject to the same set of demands.

If production patterns during spontaneous codeswitching come to reflect the language regulation processes of bilingual speakers, sensitivity to these regularities would undoubtedly prove beneficial for bilingual listeners. Bilingual comprehension, like production, is widely thought to be language non-selective: a multitude of evidence indicates that even highly proficient bilinguals continue to activate representations in the non-target language, despite the fact that such non-target activation may incur a processing cost (Thomas & Allport, 2000; Von Studnitz & Green, 2002). However, a small amount of work indicates that under some circumstances, bilinguals may take advantage of exogenous cues to language identity to minimize the influence of the non-target language during comprehension (e.g., Ju & Luce, 2004; Libben & Titone, 2009; Schwartz & Kroll, 2006), and interestingly, there may be reason to believe that such cues are relatively more accessible during auditory processing. We return to this point in further depth below. Importantly, very little work (in either the production or comprehension domains) has focused on habitual codeswitchers, arguably the group of language users most likely to develop sensitivity to any statistical regularities in the speech stream that could act as cues to language regulation. The few psycholinguistic studies of codeswitching have examined proficient bilinguals who don't normally engage in codeswitching (e.g., Kootstra et al., 2010), but it is as yet largely unknown whether accumulated experience with codeswitching in particular is associated with quantitative or qualitative changes in the mechanisms involved in language regulation. A goal of the current paper is to begin to address this question: can the phonetic form of spontaneous speech provide insight into the language regulation mechanisms of habitual codeswitchers, and if so, can members of a codeswitching community capitalize on such phonetic variation during auditory comprehension?

The remainder of the paper is organized as follows. First, we present an overview of the ways in which bilingual language regulation has been hypothesized to affect the production planning process. We then describe the similarities and differences between the hypothesized pressures on the bilingual production system, and the demands placed on the comprehension system. Subsequently we present three studies: two corpus studies ask whether the phonetic form of spontaneously produced codeswitched speech reflects the processing demands

specific to bilingual language production, and an eye tracking study using the visual world paradigm asks whether similarly small fluctuations in phonetic information can be perceived and exploited by habitual codeswitchers during comprehension. We conclude with a discussion of the ways in which these studies contribute to our understanding of bilingual language regulation, in particular, and of the linkages between the production and comprehension systems more generally.

Background

Pressures on the bilingual production system

As alluded to above, the activation of psycholinguistic representations during bilingual production planning is widely thought to be language non-selective; even for highly proficient bilinguals, and even when the experimental setting would be highly conducive to "turning off" the non-target language, studies have consistently provided evidence for transient activation of the non-target language (Kaushanskaya & Marian, 2007; Wu & Thierry, 2012). In certain instances, the activation of the non-target language may be beneficial to production planning processes. For example, bilingual word and picture naming studies typically find relatively faster processing of cognates, words that share both form and meaning in a bilingual's two languages, relative to non-cognates (e.g., Colomé & Miozzo, 2010; Costa, Caramazza, & Sebastian-Galles, 2000; Hoshino & Kroll, 2008; Jacobs et al., 2015). While these findings suggest that a certain amount of cross-talk between languages may be unavoidable, they also demonstrate that cross-language activation is not necessarily detrimental to language processing.

The picture becomes somewhat more complicated when processes other than lexical selection are considered. While the literature on cognate production rather unequivocally demonstrates that cross-language representational overlap can be a boon to lexical access, recent phonetic work indicates that the activation of non-target phonological representations can interfere with accurate phonetic production. Work by Amengual (2012), Goldrick et al. (2014), and Jacobs et al. (2015) demonstrates that the voice onset time (VOT; a timing parameter involved in the production of stop consonants) of cognate words is subject to stronger phonological influence from the non-target language than that of non-cognates. Importantly, the phonetic effects in these studies appear to be the result of online processing demands, rather than (or perhaps in addition to) any qualitative differences in the long-term representation of cognate words. Both Amengual and Goldrick et al. find that (unpredictable) language switching is associated with cross-language phonological influence, suggesting that given adequate time and/or resources, bilinguals can control or resist the effects of cross-language activation on phonetic production. The findings of Jacobs et al. lend credence to this hypothesis: in that study, three groups of bilinguals differing in language proficiency and immersion context all exhibited effects of cross-language activation on naming response times, but only for the

non-immersed group of lower proficiency learners did cross-language activation spill over to influence phonetic production. Again, this suggests that the ability to regulate cross-language activation – an ability likely subject to changes in language proficiency and immersion status – determines the extent to which the surface form of bilingual speech overtly reflects the activation of the non-target language during planning.

While these studies indicate that phonetic variation can be used as a tool to investigate the regulation of cross-language activation during speech planning, it is not yet clear whether the results from laboratory studies of language learners and bilinguals who do not codeswitch regularly can be extended to the spontaneous speech of habitual codeswitchers. A number of proposals have hypothesized that bilingual language regulation may be subject to tuning over the lifespan (Blumenfeld & Marian, 2011; Green, 1986, 1998), a process that may yield quantitative and qualitative differences in the regulation mechanisms engaged by habitual versus non-habitual codeswitchers (Green & Abutalebi, 2013). If the presence of phonetic variation in bilingual speech is dependent on disruptions in processing and/or a lack of experience regulating cross-language influence, then it is possible that the highly tuned language regulation capabilities of habitual codeswitchers could eliminate or greatly reduce the amount of language regulation-related phonetic variation produced by these speakers. One recent study does speak against this hypothesis, however: Balukas and Koops (2015) examined the spontaneous speech of habitual codeswitchers from New Mexico and found that the VOT of voiceless English stops varied along with the distance from Spanish, such that English /ptk/ were produced with more Spanish-like VOTs the closer they were to Spanish words. This finding suggests that even for habitual codeswitchers, phonetic variation may provide an index of language regulation during spontaneous speech, although to our knowledge, the Balukas and Koops study is the only one to address this issue thus far. The goal of the production study presented here is to expand on these findings and dig more deeply into the psycholinguistic mechanisms involved in language regulation during spontaneous codeswitching.

Pressures on the bilingual comprehension system

In some respects, the pressures placed on the bilingual comprehension system are similar to pressures on the production system; in both cases, a large body of evidence supports the idea that the activation of psycholinguistic representations is largely language non-selective (see Kroll & Dussias, 2013, for a recent review) and that this non-selectivity may necessitate the recruitment of additional or more finely tuned cognitive resources relative to monolingual processing (Green, 1998). Specifically within the study of bilingual comprehension, a recurring research question concerns the types of cues bilinguals may exploit to help restrict their attention to the target language. Some studies in this vein have focused on contextual and/or syntactic cues to language identity. Schwartz and Kroll (2006), for example, compared word-naming times for cognate and

non-cognate words presented in either highly constraining or relatively non-constraining sentence contexts. In their study, both groups of bilinguals (one highly proficient and one of intermediate proficiency) exhibited cognate facilitation in non-constraining sentences, but no cognate facilitation in highly constraining sentences, suggesting that given an adequately constraining context, bilinguals can use language-specific syntactic cues to restrict activation to the target language. It is important to note, however, that the number of studies reporting *no* mitigating effect of syntactic context on non-target language activation is quite great, perhaps especially in the written modality (cf. Duyck, Assche, Drieghe, & Hartsuiker, 2007; Gullifer, Kroll, & Dussias, 2013; Libben & Titone, 2009; Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009). Similarly negative results have been found for other potential cues, such as orthographic script (cf. Hoshino & Kroll, 2008 for evidence of persistent cross-language activation in Japanese–English bilinguals) and semantic context (cf. Van Assche et al., 2011; but see Van Hell & De Groot, 2008).

In comparison to the body of work on the written modality, however, relatively little work has examined bilingual comprehension in the spoken modality. Importantly, spoken language offers an additional, extremely rich source of potential language cuing: the phonetic form of the speech stream. A few previous studies suggest that phonetic information can modulate cross-language activation patterns during bilingual comprehension. An early study by Grosjean (1988) used a gating paradigm to investigate whether the recognition of codeswitched words was affected by their phonetic integration into the "base language"; in one condition, French–English bilinguals were asked to transcribe interlingual homophones (words with the same phonological form but different meanings in English and French) whose phonetic realization either did or did not clearly mark their membership in the "guest" language. For example, the last word in the French sentence *Il faudrait qu'on* [kul] ("It is necessary to [kul]") could either be transcribed as the English word *cool* or the French word *coule* ("sink"). Grosjean found that whether the last word was produced with an English or French accent affected whether it was transcribed as an English or French word, although this is perhaps not surprising given that in this condition, phonetic detail was the *only* cue available to assist participants in determining the language of the target word.

Li (1996) expanded on Grosjean's findings by asking Chinese–English bilinguals to transcribe codeswitched words that were either phonetically integrated into the base language or not, but that were presented in highly constraining or less constraining sentence contexts.[1] Li found main effects of both phonetic integration and sentence context, but no interaction between them: codeswitched words whose phonetic production matched the language of the carrier phrase were more difficult to

recognize than words that retained their language-specific phonetic cues (similarly to Grosjean, 1988), and more constraining sentence contexts were associated with improved recognition, regardless of phonetic integration. The lack of interaction between these factors indicates that, contrary to Schwartz and Kroll (2006), the constraining sentence contexts in Li's study did not wipe out cross-language effects associated with the target word's form, perhaps suggesting that a word's phonetic form provides a more robust source of language-specific cues than its orthographic form.

While language-specific phonetic information can in some cases function as a source of confusion (Grosjean, 1988; Li, 1996), under the right circumstances phonetic information can be exploited to the listener's benefit. Ju and Luce (2004), for example, demonstrated that during single word recognition, language-specific phonetic cues can mitigate the activation of lexical representations in the non-target language. Using the visual world paradigm, they found that Spanish–English bilinguals relied on VOT to restrict their attention to the lexicon of the target language; when listeners heard a Spanish word (e.g., *playa*, "beach") produced with Spanish-like VOT, they were no more likely to look at an interlingual distractor picture (e.g., *pliers*) than at an unrelated distractor (e.g., *ruler*). However, when they heard Spanish words produced with English-like VOT, they spent significantly longer gazing at the interlingual (i.e., English) distractor picture, indicating that low-level cues such as stop consonant VOT are involved in the formation of initial hypotheses concerning the language membership of lexical items. Again, this may suggest that cues to language identity are more robust or more readily accessible in the auditory than in the written modality (cf. Hoshino & Kroll, 2008), although it should be noted that not all studies of bilingual auditory word recognition converge on this conclusion (cf. Lagrou, Hartsuiker, & Duyck, 2011; Spivey & Marian, 1999; Weber & Cutler, 2004).

Of course, phonetic information does not exist in a vacuum; it must be integrated with information at other levels of linguistic processing. Recent work by Lagrou, Hartsuiker, and Duyck (2013) investigated the interaction between phonetic information and sentence processing by asking whether the global accentedness of a speaker's voice impacts cross-language activation during auditory word recognition in a sentence context. Dutch–English bilinguals performed an auditory lexical decision task on interlingual homophones embedded in low- or high-constraint English sentences that were produced by either a native Dutch or native English speaker. Both the highly constraining sentence context and the English-accented speech reduced (but did not eliminate) the degree of Dutch activation during English word recognition. These findings are again consistent with the idea that bilingual comprehension is fundamentally language non-selective, but that under some circumstances, cues to language identity may allow listeners to reduce the activation of the non-target language.

Taken together, these studies suggest that the phonetic realization of speech provides an important source of language cuing information, leading to either confusion or facilitation as a function of the particular task and perhaps

---

[1] Both Grosjean and Li examined an additional variable, the phonotactic make-up of the target words, which yielded interesting interactions across the two studies but is not immediately relevant to the present discussion.

the group of bilinguals being tested. Interestingly, the vast majority of studies of bilingual auditory comprehension have investigated how the phonetic realization of the target word itself affects the activation of lexical representations in the non-target language. To our knowledge, no study thus far has asked whether anticipatory phonetic cues to language switching modulate the degree of cross-language activation during word recognition, likely in part because the presence of such cues is not well established in the production literature. As argued above, however, if habitual codeswitchers' phonetic production reliably reflects cross-language activation patterns experienced during speech planning, then listeners may be able to key into these patterns as a means of predicting upcoming language switches.

Bilingual production, comprehension, and the bigger picture

The studies reviewed above are consistent with the claim that bilingual production and comprehension are largely language non-selective processes. In both domains, the majority of the available evidence indicates that the activation of representations in the non-target language is automatic and nearly unavoidable. The interesting questions currently lie in determining the particular contexts that promote reduction or suppression of the non-target language, and the degree to which speakers' past experience allows them to modulate this reduction or suppression. While the answers to these questions are obviously critical for understanding bilingual language regulation, they also provide important data on the relationships between different levels of linguistic processing, on the relation between language processing and more general cognitive processes (i.e., prediction, inhibition, learning), and on the extent to which these relationships hold for both language production and comprehension. The latter issue is particularly important for understanding the similarities and differences between production and comprehension processing: are both processes impacted by the same factors, and to the same extent?

Thus while the specific phenomena that we examine here are necessarily unique to bilingual speech, we contend that these effects provide a window into processes that are relevant for understanding the relationship between production and comprehension more generally. Bilingual phonetic variation allows us to probe the circumstances under which pressures on the production planning system affect the surface form of spontaneously produced speech, and consequently, bilingual comprehension serves as a natural testing ground for the question of listener sensitivity to naturally occurring distributional patterns. The logic of the present study was therefore to first identify the types of phonetic variation associated with language regulation in codeswitched speech, and then to ask whether listeners can exploit these potential cues to anticipate when a codeswitch is about to occur.

Corpus Study 1: Speech rate

The first production study examined speech rate (and speech disfluency) in the Bangor Miami Corpus of sponta-

neous codeswitching (Deuchar, Davies, Herring, Parafita Couto, & Carter, 2014). The Bangor Miami Corpus consists of 56 conversations, each around 30 minutes long, between 85 highly balanced Spanish–English bilinguals. In total, the corpus contains approximately 250,000 words and 35 hours of speech. Speakers typically knew each other quite well; many were family members or good friends. The age range represented is from 9 to 78, with a median age of 29.5. The analyses presented here rely on the CHAT transcripts prepared by Deuchar et al., which consist of word-level transcription and language tagging. The corpus also contains syntactic category information, which was automatically generated using the Bangor Autoglosser (http://bangortalk.org.uk/autoglosser.php).

Speech rate analysis

Method

The Bangor Miami Corpus is divided into a series of "utterances" consisting of one main clause each. For the analyses described here, we used an automated script to categorize each utterance as unilingual English ($n = 26{,}801$), unilingual Spanish ($n = 13{,}999$), or codeswitched ($n = 2527$). Unilingual utterances were those that contained words in only one language, while codeswitched utterances contained at least one word in English and one in Spanish. For codeswitched utterances, the script used the word-level language tags to determine the location of any language switches.

Inclusion/exclusion criteria. The speech rate analysis was restricted to codeswitches characterized by a single insertion of "other language" material; for example, English–Spanish switches and English–Spanish–English switches were included, but English–Spanish–English–Spanish switches were not. This resulted in the exclusion of 141 codeswitched utterances (5.6% of all switches). Because previous research has demonstrated that the more dominant language tends to be the most affected under language mixing conditions (e.g., Guo, Liu, Misra, & Kroll, 2011), we restricted the analysis to switches from the predominant language of the conversation to the non-predominant. For example, in a predominantly Spanish conversation, only utterances that began in Spanish and switched into English were included in the codeswitched sample, under the hypothesis that anticipation of the switch into the non-predominant language would more strongly affect the predominant language than vice versa. This resulted in the exclusion of 1085 (45.5%) of the remaining codeswitched utterances. To reduce any influence from the language of the preceding utterance (thereby isolating anticipatory changes in articulation), we included only those codeswitched utterances that were preceded by a unilingual "matching-language" utterance; for example, utterances that switched from English to Spanish were only included if they were preceded by a unilingual English utterance. This resulted in the exclusion of 373 (28.7%) of the remaining codeswitches. Next, because we wished to examine speech rate preceding codeswitches, only utterances with switches that occurred on the third word or later were included, ensuring that all

measures of preceding speech rate were based on a minimum of two words. 185 (19.9%) of the remaining codeswitches were excluded at this step. Finally, 10 additional codeswitches were excluded due to errors in language tagging, leaving a total of 733 codeswitched utterances that met the inclusion criteria.

After identifying the set of eligible codeswitched utterances, a set of "matching" unilingual utterances were identified to serve as a control comparison. Unilingual utterances were matched according to the following criteria: (1) they must begin in the same language as their matched codeswitched utterance; (2) they must occur in the same conversation, and be produced by the same speaker, as their matching counterpart; (3) they must be the same length in words; (4) they must also be preceded by a unilingual utterance in the same language, and (5) the syntactic category at the point of the language switch (plus or minus one word) must be the same. Two examples of utterance matching are given below. All codeswitched utterances lacking a match were excluded from the analysis (*n* = 219), and up to five matching unilingual utterances were included for each codeswitch (with an average of 2.8 unilingual matches per codeswitch). The final number of codeswitched utterances included in the analysis was 514, and the number of matched unilingual utterances was 1436. Table 1 provides summary information for the final data set examined in Analysis 1. The final sample represented 68 different speakers, in 48 different conversations.

(1a)	Donde siempre tenemos el [Where we always have the]	**delay**. (**N**)	(5 words)
(1b)	Yo las vi en [I saw them at]	**casa**. (**N**) [home.]	(5 words)
(2a)	Did you think like projects	**en** un día? (**PREP**) [in one day?]	(8 words)
(2b)	They gave me two tickets	**to** the movies. (**PREP**)	(8 words)

Measuring speech rate. The measure of speech rate examined here is the average syllable duration for the portion of the utterance leading up to the switch point (or matched non-switch point, for the unilingual utterances), excluding pauses. Previous work has sometimes drawn a distinction between speech rate, which includes pauses, and articulation rate, which excludes them from the calculation (Crystal & House, 1990; Verhoeven, De Pauw, & Kloots, 2004). According to this distinction, then, the present analysis concerns articulation rate. To derive this measure, it was necessary to determine how many syllables preceded the switch point or matched non-switch point. Since Spanish orthography is almost entirely transparent, with only a

Table 1
Summary information for the data set used in the analysis of speech rate and disfluency in the Bangor Miami Corpus.

	Codeswitched utterances	Matched unilingual utterances
n Total	514	1436
% Disfluent	37.6%	24.5%
n English(-to-Spanish)	285	808
n Spanish(-to-English)	228	628
Mean length in syllables (*SD*)	12.9 (6.4)	10.2 (5.3)
Mean *n* syllables before switch point (*SD*)	6.8 (4.6)	5.8 (3.9)
Most common syntactic category at switch point	Noun (*n* = 182)	Noun (*n* = 618)
Mean syllable duration (*SD*)	177.6 ms (74.3 ms)	161.4 ms (66.9 ms)

few lexical exceptions, an automated script was used to count the syllables in each Spanish word based on its orthography. For English words, syllables were counted based on the pronunciation given in the Carnegie Mellon Pronouncing Dictionary (Weide, 2007). The precise onset and offset of each utterance were hand-labeled by a research assistant using the Praat program for acoustic analysis (Boersma & Weenink, 2014), as were the onset and offset of the switch word or matched non-switch word, as well as the onset and offset of any stretches of speech that were perceived as pauses.[2] Pause durations were subtracted out from the portion of the utterance preceding the switch point, and this duration was divided by the number of syllables leading up to the switch point (or matched non-switch point).

In the course of labeling time points for the speech rate analysis, all utterances containing speech disfluencies (at any point, whether before or after the switch) were hand-tagged. All disfluent utterances were excluded from the speech rate analysis, but a subsequent analysis asked whether there was a significant difference in the rate of disfluency in codeswitched versus unilingual utterances (see below).

As a reliability check, a subset of 10% of the utterances (144 unilingual and 51 codeswitched) were randomly chosen for recoding by the same research assistant approximately one year after the initial measurements were taken. The correlation between mean syllable durations was 0.97 (*t*(164) = 52.6, *p* < .0001), with an average difference of 1 ms. An exact binomial test indicated that significantly more utterances were coded as disfluent in the reliability check (30%, vs. 22% for the full data set, *p* < .01), but this increase was roughly equivalent across utterance types: 10% more of the unilingual utterances were coded as disfluent, versus 8% more of the codeswitched utterances, perhaps indicating that generally stricter inclusion criteria were used for the reliability check.

---

[2] About 14% (*n* = 170) of the utterances in the speech rate analysis contained pauses preceding the switch point, the average duration of which was 497 ms. An analysis that excluded these utterances returned qualitatively similar results.

Statistical analysis and results

The mean syllable durations formed a skewed distribution and were log-transformed for the purpose of statistical analysis. Mixed-effects linear regression was used to obtain the best fitting model predicting average syllable duration. A baseline model including by-speaker and by-conversation random intercepts was used as the starting point, and variables were then added in a step-wise fashion to determine which predictors significantly improved model fit; this was evaluated via chi-squared tests of model log-likelihoods, with a relaxed alpha level for control variables of 0.15. By-speaker random slopes for the predictors of interest were included wherever possible, with correlations between random effects. Following construction of the control model, leave-one-out comparisons were used to verify that all predictors remained significant with all other variables in the model. The lmerTest package in R was used to estimate p values for individual beta coefficients using MCMC sampling (Kuznetsova, Brockhoff, & Christensen, 2013); these are included for reference.

The following control variables were examined before testing for an effect of codeswitching: language (English vs. Spanish) preceding the switch point or matched non-switch point, utterance length in words, utterance length in syllables, location of the switch point within the utterance, syntactic category at the switch point, position of the utterance within the conversation, mean conditional probability of bigrams preceding the switch point (cf. Myslín & Levy, in press), conditional probability of the bigram spanning the switch point, mean conditional probability of bigrams including the switch point, and mean conditional probability of all bigrams in the utterance (cf. Jurafsky, Bell, Gregory, & Raymond, 2001). The best fitting control model contained significant effects of language, position of the switch point within the utterance, syntactic category at the switch point, and the mean conditional probability of bigrams leading up to the switch point. The average syllable duration for English was 183.0 ms, versus 144.0 ms for Spanish ($\chi^2(1) = 17.6$, $p < .0001$ in leave-one-out comparisons; $\beta = -0.027$, $t = -4.7$, $p_{MCMC} < .0001$). Switch points (and matched non-switch points) that occurred later in the utterance were associated with faster speech rates ($\chi^2(1) = 17.5$, $p < .0001$, $\beta = -0.016$, $t = -4.4$, $p_{MCMC} < .0001$). Note that the Switch Point predictor entered into competition with both of the utterance length predictors, but model comparisons indicated that the position of the switch point accounted for the most unique variance (neither utterance length predictor remained significant with Switch Point in the model), so it was retained. Because the vast majority of codeswitches occurred on a noun, "noun" was taken as the reference level for the Syntactic Category predictor ($\chi^2(12) = 26.6$, $p < .01$). Switches that occurred on pronouns and conjunctions were associated with significantly longer preceding syllable durations ($\beta = 0.015$, $p_{MCMC} < .01$; $\beta = 0.014$, $p_{MCMC} < .05$, respectively), while switches on proper nouns were associated with marginally shorter syllable durations ($\beta = -0.017$, $p_{MCMC} = .08$). Finally, higher mean conditional probability of bigrams leading up to the switch point was
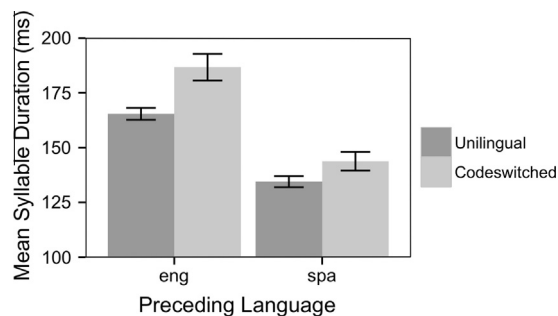

Fig. 1. Effect of codeswitching on speech rate in the Bangor Miami Corpus.

associated with marginally faster speech rate ($\chi^2(1) = 2.6$, $p = .11$, $\beta = -0.036$, $t = -1.6$, $p_{MCMC} = .10$).

Once the best fitting control model was obtained, the effect of Utterance Type (unilingual vs. codeswitched) was examined. The addition of Utterance Type to the control model resulted in a significant gain in model log-likelihood ($\chi^2(1) = 18.2$, $p < .0001$): mean syllable durations preceding a codeswitch were longer than in matched unilingual utterances, on average by about 16 ms ($\beta = 0.016$, $t = 3.9$, $p_{MCMC} = .001$). This effect is depicted in Fig. 1. There was no interaction between Language and Utterance Type, and the main effect of Utterance Type remained significant when by-conversation random slopes were included in the model.[3] The final model is included as an appendix.

Disfluency analysis and results

The overall rates of disfluency are given along with the other descriptive statistics for the data set in Table 1: 24.5% of unilingual utterances were disfluent, versus 37.6% of codeswitched utterances. To determine whether this difference was significant, a mixed-effects logistic regression was constructed in the same manner, and examining most of the same predictors, as in the speech rate analysis. (The exception being that the conditional probability of disfluent sequences was not considered; the only measure of bigram frequency examined in this analysis was conditional probability spanning the switch point.) The best-fitting control model contained only two predictors: the position of the switch point within the utterance and the utterance length in words. A greater probability of disfluency was observed for earlier switch points and matched non-switch points ($\chi^2(1) = 3.1$, $p = .08$ in leave-one-out comparisons; $\beta = -0.321$, $t = -1.7$, $p_{MCMC} = .09$), and also for longer utterances ($\chi^2(1) = 116.7$, $p < .0001$, $\beta = 2.079$, $t = 10.0$, $p_{MCMC} < .0001$). When added to this control model, Utterance Type was a significant predictor: codeswitching was reliably associated with a greater probability of speech disfluency ($\chi^2(1) = 10.2$, $p = .001$, $\beta = 0.444$, $t = 3.2$, $p_{MCMC} = .001$), and this effect was significant when

---

[3] When by-speaker random slopes for the effect of Utterance Type were entered, the model would not converge. By-conversation slopes were therefore included instead.

by-speaker random slopes for the effect of Utterance Type were included in the model. The final model is included as an appendix.

Discussion for Corpus Study 1

To our knowledge, these data provide the first evidence for a processing cost associated with switching languages in spontaneous speech. While speech rate can vary for many reasons (e.g., emphasis, clarification), some of which could also be associated with codeswitching, the co-occurrence of the speech rate and disfluency effects in this data set strongly suggest that processing-related factors are involved. Interestingly, the effect of the Syntactic Category predictor may also support this conclusion. We found that speech rate leading up to proper nouns was marginally faster than speech rate leading up to common nouns. In fact, the estimated coefficient associated with proper nouns completely canceled out the coefficient for the Utterance Type predictor ($\beta = -0.017$ vs. $\beta = 0.016$, respectively), indicating no apparent cost associated with switching into a proper noun. This is potentially consistent with the Triggering Hypothesis (Broersma & De Bot, 2006; Clyne, 2003), the idea that spontaneous codeswitching can be triggered as a result of cross-language phonological overlap. Broersma and De Bot provided the first quantitative evidence that codeswitches were significantly more likely to occur within clauses containing phonological trigger words – words with the same phonological representation across a bilingual's two language, such as proper nouns. Our findings suggest a possible extension of this hypothesis: in a new, much larger corpus, representing a different pair of languages, we find that speech rate leading up to trigger words (proper nouns) is marginally faster than speech rate leading up to common nouns; in fact, speech rate leading into codeswitched trigger words was no slower than speech rate leading into non-codeswitched common nouns, perhaps suggesting that the same cross-language phonological overlap that promotes triggered codeswitching also facilitates the production of trigger words more generally.

Of course, the missing piece of information is whether speech rate is consistently faster leading up to proper nouns than common nouns in the absence of cross-language phonological overlap, and this is not currently known. To determine whether the apparent "triggering effect" on speech rate was significantly different in unilingual versus codeswitched utterances, we added an interaction between Syntactic Category and Utterance Type to the fitted speech rate model. While this interaction term marginally improved the model ($\chi^2(12) = 20.8$, $p = .05$), the estimated coefficient for speech rate preceding proper nouns in codeswitched utterances was not significantly different from that of proper nouns in unilingual utterances ($\beta = -0.002$, $t = -0.12$, $p_{MCMC} = .9$). The only syntactic categories reliably associated with different coefficients in a codeswitching versus unilingual context were pronouns ($\beta = 0.019$, $t = 1.86$, $p_{MCMC} = .06$) and verbs ($\beta = 0.039$, $t = 3.44$, $p_{MCMC} < .001$), which were both associated with relatively slowed speech rate leading into the switch point. These findings are clearly not conclusive, but certainly sug-

gestive of differential processing costs associated with structurally different types of codeswitches. Future studies should explicitly address the intersection of triggered codeswitching, bilingual syntactic planning, and changes in speech rate and disfluency.

The finding that codeswitching may be associated with a processing cost even in the spontaneous speech of habitual codeswitchers is highly pertinent to the recent debate surrounding the question of how language experience could give rise to the cognitive and neural changes associated with bilingualism (Green & Abutalebi, 2013; Green & Wei, 2014). As discussed above, previous evidence suggesting that bilingual language regulation may require the recruitment of additional cognitive resources has largely been based on the language-switching paradigm. While the language-switching paradigm provides a high degree of experimental control, it is quite far-removed from typical language use in a number of ways, and its corresponding lack of ecological validity raises questions as to whether the bilingual's everyday linguistic experience engages the type of control mechanisms that could eventually foster an advantage in executive function. The finding that codeswitching appears to be associated with a processing cost in spontaneous conversation thus adds a crucial piece of information to the puzzle: even when highly balanced, proficient codeswitchers retain full control over the choice to switch languages, this switch does not come without a cost.

While this finding fits nicely into a large body of experimental literature, it is also worth considering two alternative – although not mutually exclusive – possibilities. First, in some cases, codeswitching may serve as a repair strategy for already disrupted speech planning. That is, while we think it likely that the intention to switch languages can give rise to a processing cost, it is also possible that when faced with difficulties in lexical retrieval and/or syntactic formulation, bilinguals may choose to switch languages rather than persist in the current, temporarily problematic language. The finding that most of the switches in the corpus occurred on common nouns tends to support the idea that lexical retrieval difficulty may tend to promote codeswitching, at least in some instances. Second, it is possible that the slowed speech rate preceding codeswitches reflects some difference in prosodic organization in codeswitched versus unilingual utterances[4]; for example, it is possible that codeswitching tends to take place at prosodic boundaries (cf. Shenk, 2006). If this is the case in the present data set, then the slowed speech rate could in part reflect phrase-final lengthening (Klatt, 1976). Both of these issues should be the subject of future research. For now we simply note that these explanations are not at odds with one another; it is certainly possible that all three factors play a role.

Corpus Study 2: Voice onset time

The starting point for the voice onset time (VOT) analysis was the companion comprehension study that

---

[4] We thank an anonymous reviewer for this suggestion.

we present below, where we ask whether proficient codeswitchers can use phonetic modulations in an English carrier phrase to anticipate codeswitches into Spanish. We accordingly restricted the VOT analysis to variation in the production of the English voiceless stops /ptk/. As a first pass, the set of utterances identified for inclusion in the speech rate analysis was searched for instances of /ptk/ that preceded a stressed vowel (and that did not follow the sound /s/). A total of 562 /ptk/ words were initially identified for analysis: 315 in unilingual English utterances, and 247 preceding codeswitches into Spanish. However, a very large number of tokens were found to be unanalyzable for various reasons: approximately 60% were excluded due to poor sound quality in the recordings, overlapping speech, speech disfluency, and /t/ flapping, the phenomenon in American English whereby /t/ sounds preceding unstressed syllables are produced similarly to /d/, as in the word *butter*.[5] This first pass analysis comprised 227 tokens in total: 133 from unilingual English utterances, and 94 from codeswitched utterances. A mixed-effects regression predicting log-transformed VOT was constructed following the same procedure as in previous analyses, but no effect of Utterance Type was found (mean VOT in both utterance types was 40 ms; $\chi^2(1) = 0.10$, $p = .75$).

To increase statistical power, we considerably relaxed the search criteria for tokens of /ptk/ and conducted a second analysis adopting some aspects of the approach taken in Balukas and Koops (2015). In a second pass through the corpus, we searched *all* codeswitched utterances (not just those from the speech rate analysis) for tokens of English /ptk/ that occurred before a stressed vowel (but not after /s/), with no restrictions on the type of codeswitch or the language of the preceding utterance. This resulted in the identification of 1169 possible English VOT tokens, in 828 unique codeswitched utterances. To serve as a control comparison, we then identified up to five unilingual English utterances that matched each codeswitched utterance according to the speaker, conversation, and utterance length in words, where the preceding utterance was also unilingual English. These matching unilingual English utterances were then searched for tokens of English /ptk/, resulting in the identification of 1992 possible VOT tokens in 1277 unique unilingual utterances.

A research assistant examined all 3161 possible VOT tokens and hand-labeled the utterances for the following time points: the onset and offset of the utterance, the onset and offset of any switches into Spanish, the onset and offset of the word containing the VOT, and the VOT itself, measured from the stop release burst to the onset of periodicity associated with the following vowel. In cases where the burst and/or vowel onset were not clearly visible on the waveform display, the token was marked as unusable: VOT could not be reliably determined for a total of 1781 (56.3%) of the possible tokens. Of the remaining 1380 tokens, 591 (42.8%) contained a speech disfluency that did not affect the word containing the VOT of interest;

because the statistical analysis indicated that the inclusion versus exclusion of utterances containing speech disfluencies did not qualitatively alter the results, we report the results including the disfluent utterances. The final data set contained usable VOT tokens from 345 unique words, produced by 67 speakers in 49 conversations; 498 of these were in codeswitched utterances, versus 882 in unilingual utterances.

As a reliability check, the first author (blinded to the utterance type of each token) labeled the VOT of 33% of the tokens that were marked as usable. The correlation between these measurements and those of the research assistant was 0.94 ($t(454) = 57.3$, $p < .0001$), with an average measurement difference of 1 ms.

Two analyses were then conducted: the first analysis compared (log-transformed) English VOT in codeswitched versus control (unilingual) utterances, and the second examined the effect of proximity to Spanish in only the subset of codeswitched utterances. The following control predictors were examined in both analyses: consonant (/p/ vs. /t/ vs. /k/; cf. Lisker & Abramson, 1964), vowel height (high vs. low; cf. Klatt, 1975), stress of the target syllable (primary vs. secondary; cf. Klatt, 1975), speech rate (i.e., average syllable duration calculated over the entire utterance; cf. Boucher, 2002), length of the target word in syllables, utterance length in words and in syllables, position of the target word within the utterance, and language of the final word of the preceding utterance (English vs. Spanish). All models included random intercepts by speaker, conversation, and target word, and by-speaker and by-word random slopes were also included for the predictors of interest.

In the analysis comparing codeswitched to unilingual utterances, the best fitting control model included significant effects of consonant ($\chi^2(1) = 60.6$, $p < .0001$ in leave-one-out comparisons; see appendix for beta coefficients), number of syllables in the target word ($\chi^2(1) = 5.4$, $p = .02$, $\beta = -0.040$, $t = -2.3$, $p_{MCMC} = .02$), and speech rate ($\chi^2(1) = 58.7$, $p < .0001$, $\beta = 0.308$, $t = 7.1$, $p_{MCMC} < .0001$). When added to this control model, the fixed effect of Utterance Type was significant ($\chi^2(1) = 4.7$, $p < .05$), and this was true when by-speaker and by-word random slopes for the effect of Utterance Type were included. English VOT was overall significantly shorter in codeswitched utterances than in matched unilingual English utterances, though only very slightly (38.7 vs. 40.7 ms; $\beta = -0.05$, $p_{MCMC} = .03$). The final model is included as an appendix.

While this analysis indicates that cross-language phonological activation affects the realization of English VOT in codeswitched utterances, it does not address the timing of the effect: is it the anticipation of *upcoming* Spanish that affects English VOT, or residual activation of Spanish following a switch *out of* Spanish (or both)? To answer this question, we conducted a second analysis of only the English VOT tokens that occurred in codeswitched utterances, splitting the data set into the 348 tokens that were preceded by some amount of Spanish within the same utterance, and the 191 tokens that were followed by Spanish within the same utterance. (There were 41 tokens that were both preceded and followed by Spanish; these tokens were included in both subsets of the data.)

---

[5] Even though we restricted the analysis to consonants occurring in the onset of syllables whose citation forms receive primary or secondary stress, prosodic reorganization in connected speech can give rise to /t/ flapping; think of the variable realization of the word *to*.

For the second analysis, the control model obtained from the analysis of all codeswitched tokens was used for each subset of the data, and the amount of time between the target VOT (measured from the stop burst) and either the offset or onset of the nearest Spanish word (in the Preceding versus Following Spanish analyses, respectively) was log-transformed and entered as a predictor. Tokens of /ptk/ ranged from 0.091 to 6.351 s preceding switches into Spanish (providing an index of anticipatory changes to articulation), and tokens ranged from 0.003 to 5.066 s following switches (corresponding to perseverative changes). In the analysis of Preceding Spanish, the amount of time between the offset of the nearest Spanish word and the English VOT that followed it did not improve the control model ($\chi^2(1) = 0.3$, $p = .58$). In the analysis of Following Spanish, however, the time between the target English VOT and the onset of the nearest following Spanish word did significantly improve the control model ($\chi^2(1) = 5.8$, $p = .01$, $\beta = 0.215$, $t = 2.5$, $p_{MCMC} = .01$), which was true when random by-speaker and by-word slopes for the effect of Following Spanish were included. These analyses indicate that the effect of cross-language phonological activation on English VOT is anticipatory in nature; this effect is depicted in Fig. 2, and the final model is included as an appendix.

Discussion for Corpus Study 2

Consistent with the only other known study of VOT in spontaneous codeswitching (Balukas & Koops, 2015), we found that English VOT varied systematically as a function of proximity to Spanish words, such that English words produced closer to a codeswitch boundary were realized with significantly more Spanish-like VOT. However, our analyses build on Balukas and Koops' work in a number of ways. For one, Balukas and Koops did not report whether VOT in their data set was conditioned by speech rate. Given the results of our Corpus Study 1, it was crucial to demonstrate that any differences in consonant VOT in unilingual versus codeswitched speech were not a byproduct of systematic changes in speech rate associated with switching languages. Because we find an effect of cross-language

activation when speech rate is taken into account, our results corroborate and strengthen the findings of Balukas and Koops.

We also build on these authors' findings by investigating the question of whether cross-language phonological influence in spontaneous speech can be considered primarily anticipatory or perseverative. When we examined English words that occurred before versus after switches into Spanish, we found that the influence of Spanish on English was significant *only* in the anticipatory direction. This finding provides crucial data regarding the time course of language regulation during spontaneous bilingual speech. When speaking English, habitual codeswitchers in the Bangor Miami Corpus exhibit increased Spanish activation in anticipation of switching into Spanish, but once the language switch has occurred, they show no evidence of residual Spanish activation. This observation raises numerous questions, i.e., Does a similar effect occur when these speakers speak Spanish? Does either the language that is affected or the direction of the effect (anticipatory versus perseverative) vary as a function of language dominance? Are these results specific to habitual codeswitchers, or would similar findings be observed in other populations? These questions must be left up to future research, but two main points bear discussion here: the direction of causality leading up to the codeswitch, and the lack of perseveration following it.

As in the speech rate analysis, it is not possible to determine the direction of causality from the present data. On the one hand, it is possible that speakers evince increased cross-language phonological influence leading up to a codeswitch because they intend to switch languages, and the anticipation of the language switch leads them to preemptively begin activating the language not currently in use. Under this interpretation, it is the intention to switch languages that causes Spanish influence to creep into the phonetic realization of English words (and perhaps by the same token, that necessitates the engagement of a suppression mechanism associated with the slowed speech rate in Corpus Analysis 1). On the other hand, it is also possible that the *unintended* activation of Spanish representations while speaking English (due to some unknown factor)
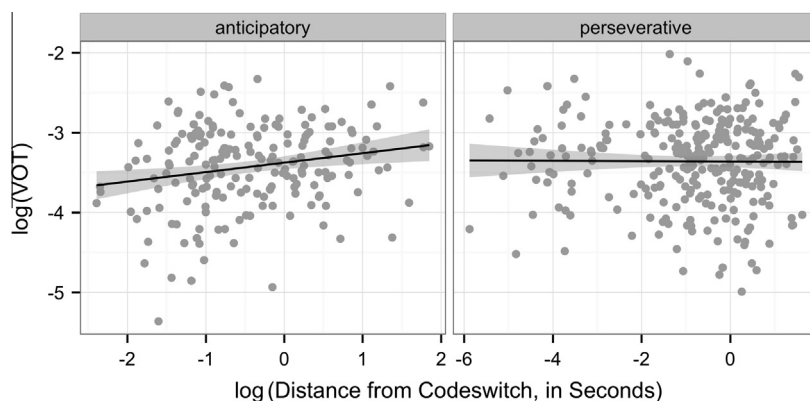

Fig. 2. Effect of codeswitching on voice onset times for English /ptk/ in the Bangor Miami Corpus.

in effect triggers the codeswitch. Under this interpretation, it is the change in the relative activation levels of the two languages that causes the codeswitch, and that is reflected in the low-level phonetic detail of words prior to the moment when the decision to switch languages is made.

Regardless of the causal relationship between the act of codeswitching and the changes in language activation that temporally lead up to it, what is clear from our data is that once the decision to switch languages occurs, the activation of the now-non-target language is immediately greatly reduced. Taken together, a possible interpretation of our findings is that of active anticipation and suppression on the part of the speaker; a speaker who intends to switch languages may begin preemptively increasing the activation of the language not in use (giving rise to anticipatory cross-language phonological influence) while simultaneously suppressing the activation of the language currently in use (yielding slightly slowed speech rate), allowing them to seamlessly switch into the other language (and show no signs of perseverative activation of the previously active language). It will be important to test this hypothesis in other populations of bilinguals, and to pinpoint the specific instances where this general pattern holds (or not) within the population of habitual codeswitchers. Moreover, we readily submit that this is not the only possible interpretation of our corpus findings, and it does not address the idea of "codeswitching as a recovery device" that was proposed in the discussion of the disfluency results. We maintain that the picture is likely far more complicated, and will require a more nuanced consideration of the factors involved, than we have proposed here. What is clear, however, is that phonetic variation in the speech of habitual codeswitchers reveals multiple ways in which bilingual language regulation impacts the surface form of spoken language.

Comprehension study

Given the findings of our production studies, the goal of the comprehension study was to determine whether the types of fine-grained phonetic changes associated with cross-language activation could be used by bilingual listeners to predict when a codeswitch is about to happen, and whether such a prediction could aid in the suppression of the non-target language. Also, as mentioned in the background section on bilingual auditory comprehension, the present study was aimed not only at making the link between bilingual production and comprehension more explicit, but also at extending previous findings in several ways. First, we ask whether *anticipatory* changes in articulation, rather than changes in the realization of the target word itself, affect listeners' processing of the target. Second, we present the target stimuli within a mixed language experimental context, with the goal of making the experimental setting somewhat more similar to that in which bilingual codeswitchers normally find themselves than a blocked or single-language design allows. And finally, the phonetic cues present in our stimuli are explicitly modeled on (and derived from) naturally occurring low-level differences found in natural speech.

Method

Participants

A total of 36 Spanish–English bilinguals were tested. Four participants (all heritage Spanish speakers) were excluded from the analysis because they scored lower than one standard deviation below the group mean on the Spanish portion of the Boston Naming Test. Two additional participants were also excluded: one highly proficient L2 Spanish speaker who reported learning Spanish in school, and one participant who reported a childhood language disorder that resulted in involuntary language switching. Table 2 summarizes the language background information for the remaining 30 participants. The average age was 22.9 (*SD* = 7.1). All participants reported exposure to both Spanish and English from an early age, as well as continued use of both languages on a daily basis, including language mixing with friends and family. Twenty-three participants reported that Spanish was their more dominant language (vs. seven for English), and eight reported some knowledge of a language other than Spanish or English. Thirteen participants were Puerto Rican, with the remainder of the group composed of participants born in the U.S. (*n* = 8), Central America (*n* = 7), or elsewhere in the Caribbean (*n* = 2).

Materials

Picture stimuli. All picture stimuli were drawn from two sources: half were colorized versions of the Snodgrass and Vanderwart set of line drawings (Rossion & Pourtois, 2004), and half were from the Peabody Picture Vocabulary Test (Dunn & Dunn, 2007).

Auditory stimuli. Auditory stimuli were recorded in a sound-attenuated booth using an Audix HT-5 head-mounted microphone. The speaker, a female highly balanced Spanish–English bilingual from Puerto Rico, was asked to simply read sentences as they appeared on a computer screen, in a clear but natural speaking style. All target words were recorded in the English carrier phrase, "Click on the picture of the [target]." The speaker first recorded the full set of English targets, followed by the full set of Spanish targets, followed by all of the English targets again. Each target word was presented twice within each recording

Table 2
Language background information for participants in the comprehension study.

	English Mean (*SD*)	Spanish Mean (*SD*)
n Correct Items, Boston Naming Test (max. 30)	18.8 (4.9)	15.0 (5.6)
% Exposure at Home	30.3 (21.7)	68.8 (21.2)
% Exposure at Penn State	68.8 (18.1)	29.0 (16.2)
Age Began Acquiring	4.8 (3.0)	1.0 (1.0)
Age Became Fluent	9.1 (4.5)	5.0 (2.4)
Self-Rated Speaking	9.1 (0.8)	9.5 (1.1)
Self-Rated Understanding	9.4 (0.8)	9.8 (0.8)
Self-Rated Reading	9.3 (0.8)	9.3 (1.1)
Self-Rated Foreign Accent	4.7 (3.1)	2.4 (2.9)

Table 3
Summary of the experiment design for the comprehension study.

Condition	*n* Unilingual English Trials	*n* Codeswitched Trials	"Click" VOT	"Picture" VOT	"Click on the" Duration	"Picture of the" Duration
Uncued	192	168	65–75 ms	20–30 ms	410–430 ms	612–654 ms
Cued	–	24	55–60 ms	12–17 ms	410–430 ms	660–680 ms

block, in randomized order. The speaker's productions were subsequently examined and manipulated using the Praat software for phonetic analysis (Boersma & Weenink, 2014). When the upcoming target word was English, the speaker produced average VOTs of 68 and 22 ms in "click" and "picture", respectively. When the upcoming word was Spanish, these shortened slightly to 64 and 20 ms, respectively. The average duration of the phrase "click on the" was 427 ms in anticipation of an English target word, and 429 ms in anticipation of switching to Spanish. The latter portion of the carrier phrase, "picture of the", averaged 647 ms in unilingual English sentences, versus 661 ms in codeswitched sentences, a lengthening of approximately 14 ms. This "case study" corroborates previous work on cross-language effects on phonetic production (e.g., Amengual, 2012; Balukas & Koops, 2015; Goldrick et al., 2014) as well as the results of our corpus analyses. We note that the durational differences that the speaker naturally produced while recording the stimuli were quite small, on the order of 5–10% for VOT and around 2% for speech rate. While previous work has demonstrated that monolingual listeners are sensitive to VOT differences as small as 5 ms (McMurray et al., 2002), work on the perception of speech rate has typically examined differences on the order of 40% (Baese-Berk et al., 2014).

All stimuli were spliced and manipulated in the same way using Praat, ensuring that any effects could not be due to differences in acoustic manipulation across conditions. For each target word, a single acoustic token was selected from among the speaker's productions and spliced on to a different carrier phrase from the one in which it originally appeared. Two different carrier phrase productions were used for each target, such that participants heard 192 different carrier phrases during the experiment (each presented twice). Table 3 summarizes the distribution of trials in the experiment, along with the durational aspects of the stimuli that were explicitly controlled. A total of 24 trials (12 carrier phrases, each heard twice) contained anticipatory phonetic cues signaling an upcoming codeswitch: these carrier phrases were originally produced by the speaker during the codeswitching block, and had their VOTs and word durations scaled to fall consistently within a slightly shorter (VOT) or longer (word duration) range than the carrier phrases in the "uncued" condition. Note that other phonetic cues may well have been present in the cued carrier phrases (for example, the speaker seemed to produce a "clearer", more Spanish-like /l/ in some codeswitched utterances), but only the durational cues were explicitly controlled. For the rest of the 168 codeswitched trials, the target Spanish target word was spliced onto a carrier phrase that originally preceded an

English word; for example, the word "pot" in "Click on the picture of the [pot]" was replaced with the Spanish word "pato". These uncued codeswitch trials also had their VOTs and speech rate adjusted to fall within the restricted range given in Table 3. The result was that all stimuli sounded extremely natural, with no noticeable coarticulatory mismatches, and that VOT and speech rate were reliable, albeit extremely subtle and only occasional, cues to codeswitching within the context of the experiment.

Design

On each trial of the experiment, participants were instructed to "click on the picture of the [target]", where the target could be produced in either English or Spanish. Twelve target-interlingual distractor pairs were constructed for the purpose of gauging cross-language activation during word recognition, taking the Ju and Luce (2004) stimuli as a point of departure. These critical pairs are given in Table 4. We introduced several additional constraints to the stimulus set: all critical Spanish targets were two syllables, with stress on the first syllable, and all English interlingual distractors either followed this same pattern or were one syllable words. Additionally, none of the targets or distractors in the experiment were cognates, and only half of the critical targets began with voiceless stop consonants; because it was unclear whether or how the phonetics of the target word would interact with the phonetics of the carrier phrase, we selected six critical targets beginning with /ptk/ and six additional critical targets beginning with other sounds (liquids, nasals, and the voiceless fricative /s/).[6] The experiment used a within-participant design, with each listener responding to the twelve critical target words in each of four different contexts: participants heard all critical Spanish targets twice with the English interlingual distractor (ILD) picture present on the screen (once with anticipatory phonetic cues signaling a codeswitch, and once without such cues), and twice without an ILD present (once with phonetic cuing, and once without). This design allows us to ask whether phonetic cuing has an effect on target recognition absent any strong competition from a non-target lexical item (cf. Dahan et al., 2001) versus when a non-target language competitor is immediately present in the display (e.g., Ju & Luce, 2004; Spivey & Marian, 1999).

A critical aspect of the experiment was that the probability of a codeswitched trial be the same as the probability of a non-codeswitched (unilingual English) trial. We

---

[6] Unfortunately, this design does not provide enough statistical power to ask whether the production of the carrier phrase interacts with any language-specific phonetic cues within the target word itself. Future work should examine this question systematically.

Table 4
Critical target word-interlingual distractor pairs in the comprehension study.

Critical Target (English Translation)	Interlingual Distractor
pato (duck)	pot
perro (dog)	parrot
taza (cup)	tie
queso (cheese)	carrot
cama (bed)	comb
cola (tail)	corn
sapo (toad)	sock
libro (book)	leaf
lápiz (pencil)	lock
mano (hand)	money
niña (girl)	knee
huesos (bones)	whistle

additionally aimed to reduce the possibility that participants would become aware of the target-ILD relationships. Given these constraints, participants identified the twelve critical items four times in Spanish with the cuing/distractor manipulations described above, but also four times in English with no such manipulations. Likewise, the interlingual distractors were identified four times in English and four times in Spanish. We also included an additional 24 filler items, each identified four times in each language, for a total of 384 trials (48 total items × 2 languages × 4 presentations in each language). Because the critical target items were presented twice in the context of their ILDs, we additionally added the constraint that all targets must occur twice with each of their potential distractors. We therefore paired each target word with 12 possible distractor pictures, ensuring that if participants were able to learn the co-occurrence restrictions on the target-distractor pairs, this would be uniform across trial types and would not allow the critical target items to be more easily predicted than any other items. The result was that each picture was presented a total of 32 times (8 times as a target, 24 times as a distractor).

The presentation of the trials was pseudo-randomized as follows. All participants responded to the same four practice trials at the beginning of the experiment, and were then asked whether they had any questions about the procedure. Participants then saw the same twelve items, in the same fixed order, none of which contained a critical target. The remainder of the experiment was organized into 31 blocks of twelve items each such that each block contained no repetitions of the same target (in either language), as well as no more than three critical trials (those containing a critical target item, whether cued or uncued, with or without an ILD) per block. The order of these 31 blocks was randomized by the experimental program used to present the stimuli (Experiment Builder; SR Research, Ottawa, Canada). Within each block, the trial order was fixed, with approximately half of the trials in unilingual English and half codeswitched, although this ratio varied up to 8:4 (or 4:8). Critical trials never occurred as the first or last item of the block, to ensure that a minimum of two non-critical trials always intervened between critical items. Additionally, critical trials always occurred immediately following a unilingual English trial. Since it was not known

whether switching from unilingual English to a codeswitch would affect the results, we opted to keep this factor constant.

Procedure

Data collection took place in a sound-attenuated booth using an EyeLink 1000 eye tracker (SR Research, Ottawa, Canada) with a chin rest. Participants wore a set of professional quality headphones and were told that on each trial, they would see four pictures displayed on the computer screen, and their task was to simply click on the correct picture as quickly and accurately as possible. They were also told that they would be listening to a native speaker of Puerto Rican Spanish who would name some of the pictures in English, and some in Spanish. Calibration of the eye tracker took place at the beginning of the experimental procedure, and then throughout the experiment as needed. Participants were given the opportunity to take a short break approximately 25%, 50%, and 75% of the way through the experiment. The eye tracking task lasted approximately one hour, after which participants completed an abridged version of the Boston Naming Test (Kaplan, Goodglass, Weintraub, Segal, & Loon-Vervoorn, 2001) containing 30 English and 30 Spanish items (15 high and low frequency words in each language), as well as a detailed language history questionnaire (the LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007), which was supplemented with a set of questions concerning participants' experience with language mixing. The entire experimental procedure lasted approximately 90 minutes, and participants were paid $15 for their time.

Results

In all analyses presented here, we focus on only the critical target items. Recall that these twelve items were presented four times in Spanish (twice each in the cued vs. uncued conditions, and twice each with an ILD present vs. no ILD) and four times in English (with no cuing or ILD manipulation). In each analysis, we first ask whether participants' response to the codeswitched trials (all conditions collapsed together) differ significantly from the English trials. We then examine the effects of the ILD and cuing manipulations within the codeswitched trials only. Following Barr, Levy, Scheepers, and Tily (2013), each model uses the maximal random effects structure that would converge and statistical significance is evaluated using chi-squared comparisons of model log-likelihood. Estimated p values for individual beta coefficients were obtained using MCMC sampling via the lmerTest package in R (Kuznetsova et al., 2013) and are included for reference. The full model specifications are given in the appendix.

Mouse click data

Accuracy for the mouse click data was at ceiling; of a total of 2880 critical trials, 7 were responded to incorrectly or timed out, yielding an overall accuracy rate of 99.8%. For the mouse click response time (RT) analysis, RT was calculated as the interval between the onset of the target word

Table 5

Mean response times by trial type in the comprehension study. Response time was computed as the time of the mouse click minus the auditory onset of the target word. See text for details.

Trial type	Mean RT (*SD*), ms
Unilingual English	1080.5 (252.3)
Uncued – No ILD Present	1115.8 (239.9)
Cued – No ILD Present	1084.0 (241.4)
Uncued – ILD Present	1143.9 (244.4)
Cued – ILD Present	1126.2 (264.3)

and the press of the mouse button (for accurate trials only). Table 5 gives the mean RTs for the critical items in English and the four codeswitching conditions. RTs were log-transformed for the purposes of statistical analysis. In the analysis comparing English to codeswitched trials, a mixed effects regression returned significant main effects of trial number and of language: RTs sped up over the course of the experiment ($\chi^2(1) = 15.8$, $p < .0001$, $\beta = -0.0001$, $t = -4.0$, $p_{MCMC} < .0001$), and RTs were overall significantly slower to codeswitched than to English trials ($\chi^2(1) = 22.1$, $p < .0001$, $\beta = 0.035$, $t = 4.2$, $p_{MCMC} < .0001$). These predictors did not interact.

In the analysis of the subset of codeswitched trials, there was still a main effect of trial number ($\chi^2(1) = 8.1$, $p < .01$, $\beta = -0.0001$, $t = -2.8$, $p_{MCMC} < .01$). There was also a main effect of Distractor, such that mouse clicks were slower on trials where the ILD was present in the display ($\chi^2(1) = 8.1$, $p < .01$, $\beta = 0.029$, $t = 2.8$, $p_{MCMC} < .01$), and a main effect of Cuing, such that mouse clicks were faster on trials where anticipatory phonetic cues were present ($\chi^2(1) = 6.0$, $p = .01$, $\beta = -0.025$, $t = -2.2$, $p_{MCMC} < .05$). These predictors did not interact.

Eye fixation data

The proportion of looks to the target, interlingual distractor (ILD), and unrelated distractors (averaged together) is plotted for codeswitched versus English trials in Fig. 3 (but recall that there was never an ILD present on English trials). The analysis of the eye fixation data focuses on three time windows: an early target activation window (from 0 to 800 ms following auditory onset of the target word) corresponding to the initial activation of the target, a late target activation window (from 800 to 1800 ms) corresponding to the decay in target activation, and the competitor activation window (from 0 to 1000 ms) corresponding to the activation and decay of the non-target competitors. These windows were chosen based on visual inspection of the fixation time course plots.

The eye fixation data were modeled using growth curve analysis (Mirman, 2014), a statistical method that allows the use of mixed effects regression modeling, and that does not violate the independence assumption.[7] Following Mirman (2014), orthogonal polynomials were used to model changes in fixation proportions over time. In the analysis of looks to the target picture, the proportion of fixations within

each time window was modeled using third-order orthogonal polynomial time terms, which were permitted to interact with each of the experimental variables of interest. The analysis of looks to non-target competitors was similar but used fourth-order orthogonal polynomial time terms to capture the more complex shape of the fixation functions. Under this approach, effects of the experimental variables on the model intercept reflect differences in the overall number of fixations throughout the entire time window, while interactions between experimental variables and the time terms reflect differences in the rate of change in fixation proportions.

In the analysis of English versus codeswitched trials in the early target activation window (0–800 ms following target word onset), there was a main effect of Language on the intercept ($\chi^2(1) = 116.3$, $p < .0001$) reflecting the fact that overall, participants spent more time looking at the target on English trials than on codeswitched trials ($\beta = -0.035$, $t = -4.5$, $p_{MCMC} < .0001$). There was also a significant interaction between Language and the quadratic time term ($\chi^2(1) = 71.2$, $p < .0001$), reflecting the different time course of target activation in the two language conditions. On codeswitched trials, the increase in target activation was initially delayed relative to English trials (likely due in part to the presence of interlingual distractors on some trials), but by the end of the window, the proportion of fixations to the target in the two language conditions was approximately equal ($\beta = 0.114$, $t = 4.9$, $p_{MCMC} < .0001$). These effects are shown in Fig. 4, which plots the empirical data along with smoothed lines depicting the model fit.

In the analysis of English versus codeswitched trials in the late window, there was a main effect of Language on the intercept ($\chi^2(1) = 95.6$, $p < .0001$) reflecting the fact that there were overall more looks to the target on codeswitched trials than on English trials in this window ($\beta = 0.033$, $t = 3.5$, $p_{MCMC} < .01$). This effect is depicted in Fig. 5. There were no other effects of Language on the time course of target fixations in the late window.

The next analysis examined the effects of the ILD and phonetic cuing on target activation for the subset of codeswitched trials in the early window (0–800 ms). There were main effects of both Distractor ($\chi^2(1) = 18.9$, $p < .0001$) and Cuing ($\chi^2(1) = 65.5$, $p < .0001$) on the intercept: participants spent less time overall looking at the target when an ILD was present in the display ($\beta = -0.025$, $t = -1.9$, $p_{MCMC} = .06$), but more time when anticipatory phonetic cues indicated that a codeswitch was about to happen ($\beta = 0.032$, $t = 2.5$, $p_{MCMC} < .05$). There was an interaction between Distractor and the quadratic time term ($\chi^2(1) = 7.7$, $p < .01$) indicating that the target activation function was significantly more curved when an ILD was present in the display ($\beta = 0.081$, $t = 2.1$, $p_{MCMC} < .05$). Finally, there was a three-way interaction between Distractor, Cuing, and the linear time term ($\chi^2(1) = 12.7$, $p < .001$) such that on trials where the ILD was present, the rate of increase in target activation was actually slower in the cued than in the uncued condition ($\beta = -0.135$, $t = -3.6$, $p_{MCMC} < .001$). These results are depicted in Fig. 6, and we return to them in the discussion.

The last analysis of target fixations examined the effects of the ILD and cuing manipulations on target activation in

---

[7] Similar approaches have been suggested by Barr (2008) and by Magnuson, Dixon, Tanenhaus, and Aslin (2007), among others.
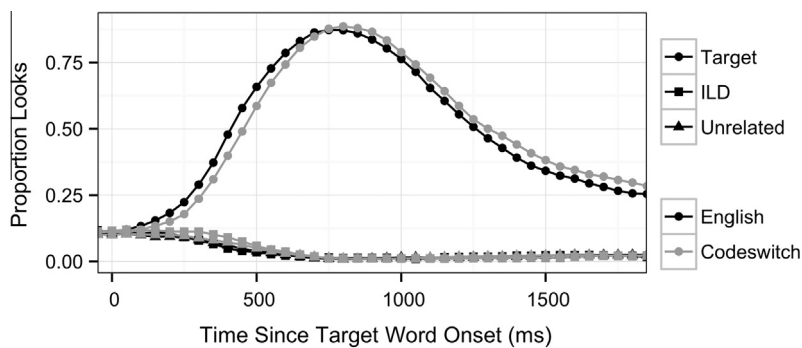
Fig. 3. Empirical plot showing the overall time course of looks to the target, interlingual distractor, and unrelated distractors in the comprehension study.

Fig. 4. Empirical data (points) and model fit (lines) for looks to the target picture during the early window, with all codeswitched trials collapsed together.

Fig. 5. Empirical data (points) and model fit (lines) for looks to the target picture during the late window, with all codeswitched trials collapsed together.

Fig. 6. Empirical data (points) and model fit (lines) for looks to the target picture during the early window, for codeswitched trials only.

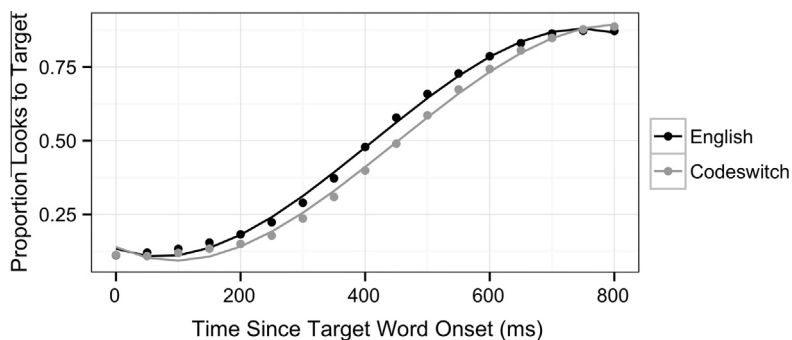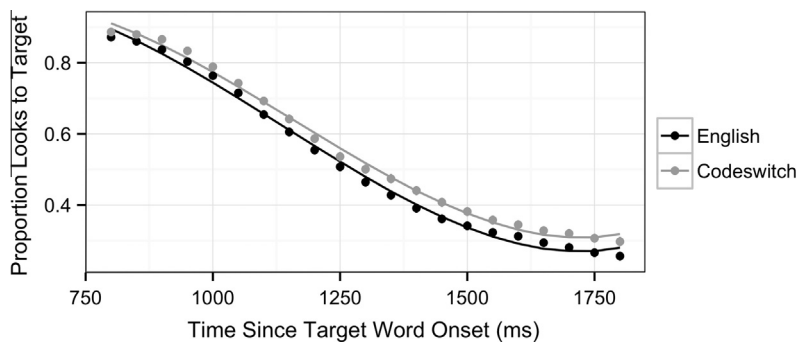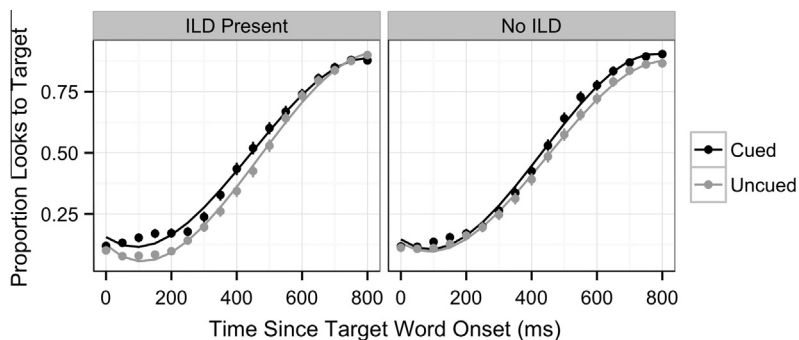the late window (800–1800 ms) for the subset of codeswitched trials. There was a main effect of Distractor on the intercept ($\chi^2(1)$ = 13.8, $p < .001$), reflecting the fact that overall, participants spent marginally more time gazing at the target when an ILD was present ($\beta = 0.020$, $t = 1.4$, $p_{MCMC} = .16$). There were also two three-way interactions: a significant interaction between Distractor, Cuing, and the linear time term ($\chi^2(1)$ = 7.8, $p < .01$, $\beta = 0.128$, $t = 2.9$, $p_{MCMC} < .01$) and a marginally significant interaction between Distractor, Cuing, and the quadratic time term ($\chi^2(1)$ = 3.2, $p = .07$, $\beta = 0.078$, $t = 1.8$, $p_{MCMC} = .07$). These interactions, depicted in Fig. 7, indicate that when an ILD was present in the display, the rate of decay in target activation was significantly slower in the cued than uncued condition.

Eye fixations to distractor pictures

The final analysis examined looks to the non-target distractor pictures only on codeswitched trials where the ILD was present in the display, from the auditory onset of the target word until 1000 ms later, when looks to the competitors were at floor. There was a main effect of Object ($\chi^2(1)$ = 79.7, $p < .0001$): participants spent more time overall looking at the interlingual distractors than at the unrelated distractors ($\beta = 0.022$, $t = 4.5$, $p_{MCMC} < .0001$).

Object also interacted with two of the time terms ($\chi^2(1)$ = 13.2, $p < .001$, $\beta = -0.039$, $t = -1.5$, $p_{MCMC} = .14$ for the interaction with the linear term; and $\chi^2(1)$ = 3.1, $p = .08$, $\beta = 0.047$, $t = 2.3$, $p_{MCMC} < .05$ for the interaction with the cubic term). These interactions reflect the overall different shape of the fixation function for the ILD as compared to the unrelated distractors, with looks to the ILD generally reaching a higher peak and then exhibiting a sharper drop-off (see Fig. 8). These differences were additionally qualified by a two-way interaction between Cuing and the linear time term ($\chi^2(1)$ = 12.9, $p < .001$) indicating that the decay function was marginally steeper in the cued condition ($\beta = 0.028$, $t = 1.4$, $p_{MCMC} = .15$), and by a three-way interaction between Object, Cuing, and the cubic time term ($\chi^2(1)$ = 7.3, $p < .01$) indicating that the decay was especially steep for the ILD in the cued condition ($\beta = -0.056$, $t = -2.7$, $p_{MCMC} < .01$). These effects are clearly seen in Fig. 8: there are initially more looks overall to the ILD than to the unrelated distractors, but looks to all distractors are reduced to nearly zero by about 750 ms following the onset of the target word. Strikingly, however, the activation function for the ILD in the uncued condition is significantly more peaked than in the cued condition; the presence of anticipatory phonetic cues affected the proportion of time participants spent looking at the ILD especially in the early competitor activation window.
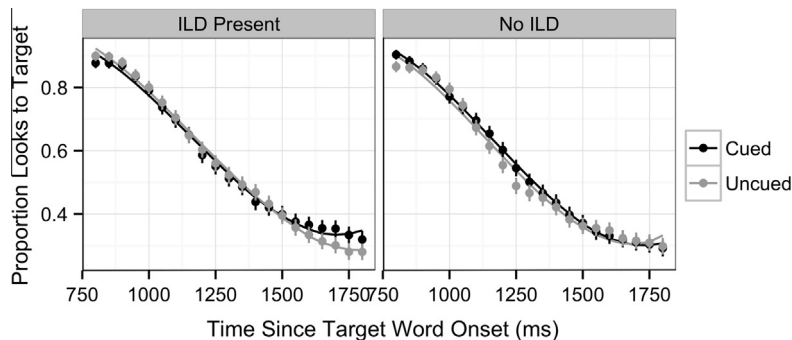

Fig. 7. Empirical data (points) and model fit (lines) for looks to the target picture during the late window, for codeswitched trials only.
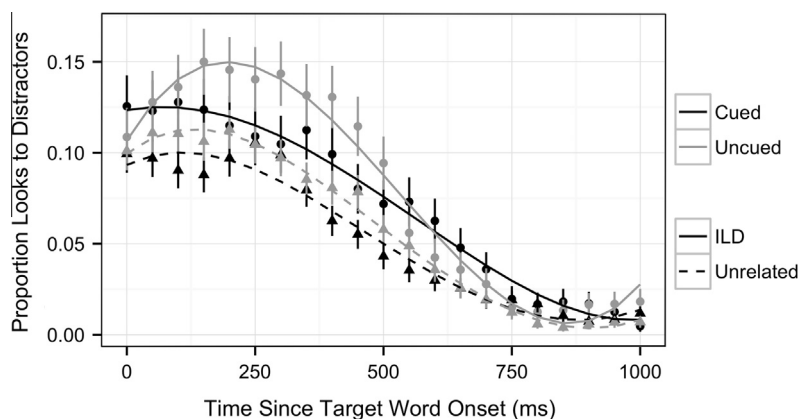

Fig. 8. Empirical data (points) and model fit (lines) for looks to the distractor pictures, for codeswitched trials where an interlingual distractor was present in the display.

Comprehension study discussion

Consistent with previous studies of bilingual auditory word recognition (e.g., Ju & Luce, 2004; Marian & Spivey, 2003; Spivey & Marian, 1999), we found that bilingual listeners generally encountered difficulty in suppressing the activation of phonologically similar competitor items from the non-target language. In contrast to earlier studies, however, we investigated the time course of this competitive process in a mixed language context, using habitual codeswitchers as participants. When an interlingual distractor was present in the display, we found that recognition of the target was initially hindered, but this competition was quickly resolved; looks to the target picture were roughly equivalent in all conditions by 800 ms following the auditory onset of the target word.

The most unique aspect of our study was its goal of examining the effects of anticipatory phonetic cuing on target recognition. It is important to note that in studies of cross-language influence in bilingual speech production, the magnitude of the phonetic effects uncovered is typically extremely small. In our Corpus Study 1, for example, we found that articulatory slowing in anticipation of a codeswitch amounted to approximately 16 ms per syllable. While Ju and Luce (2004) demonstrated that bilinguals can exploit phonetic cues to language identity that occur *within* the target word itself, the present results indicate for the first time that bilinguals can also perceive and exploit fine-grained, naturalistic cues to an *upcoming* language switch. Because our stimuli were only minimally altered from natural productions, our results suggest that bilingual interlocutors may be able to take advantage of naturally occurring phonetic variation that signals an upcoming codeswitch during spontaneous conversation.

The results provide evidence for at least two distinct mechanisms relating phonetic cuing to improved recognition of a codeswitched target word. In the analysis of looks to the target picture in the early window, we found that the time course of target activation was affected by both the ILD and cuing manipulations. As expected, listeners spent less time overall looking at the target when an ILD was visible, since their visual attention was split between the target and ILD. They also spent overall more time looking at the target when anticipatory phonetic cues were present, indicating that they perceived the cues and used them to their advantage. It may be surprising, then, that we found a slower rate of increase in target activation on phonetically cued trials where the ILD was present. To understand this result, it is necessary to examine the window from 0 to 200 ms following the target onset in codeswitched trials (Fig. 6). Because it takes approximately 200 ms to program an eye movement (Hallett, 1986), looks to the target within this region cannot be driven by auditory recognition of the target word itself. Nonetheless, when an ILD is present, we find more looks to the target in the cued than uncued condition. Our interpretation of this result is that participants were able to use the co-occurrence of the ILD and phonetic cues to anticipate specific target items within the experiment. To test this hypothesis, we plotted looks to the target on trials where an ILD was present, for each third of the experiment.

Fig. 9 indicates that in the first third of the experiment, participants did not begin anticipating the target item before its auditory onset. By the second third of the experiment, however, they had begun making strong predictions based on the presence of both the ILD and the slowing in speech rate that occurred just prior to the onset of the target word.

The finding that listeners responded to the slightly slowed speech rate preceding cued codeswitches in some respects resembles findings that monolingual listeners are sensitive to naturalistic hesitation phenomena (Bosker, Quené, Sanders, & de Jong, 2014; Corley, MacGregor, & Donaldson, 2007).[8] We think it quite reasonable that habitual codeswitchers could develop associations between hesitations and codeswitching much in the same way that monolinguals learn to associate hesitations with new information, for example (Arnold, Tanenhaus, Altmann, & Fagnano, 2004).

In the context of our experiment, the ability to associate subtly slowed speech rate with specific items appears to have affected the time course of both competitor activation and target decay. With respect to competitor activation, the effects of cuing on looks to the ILD indicate that the ability to predict a specific lexical item allowed participants to quickly suppress the activation of non-target items. The fact that participants still looked longer overall at the ILD than at the unrelated distractors, however, indicates that despite their predictions, listeners were unable to keep from initially activating the non-target language distractors. This finding resembles the results of Libben and Titone (2009), who found that language-specific cues in the written modality allowed readers to quickly deactivate non-target representations, but not to avoid activating them in the first place.

With respect to target decay, we found multiple effects of cuing on the time course of target deactivation. When the ILD was present in the display, and especially when phonetic cues were available, participants spent relatively longer gazing at the target even after having clicked the mouse button. This suggests that the item-specific predictions discussed above resulted not only in faster deactivation of the competitors, but also in longer-lived activation of the target itself. It is possible that this persistent target activation is related in some way to participants receiving confirmation of their earlier predictions, but for now this issue must be left up to future research.

While some of the benefit to target recognition appears to have come from item-specific predictions, our results also provide evidence that participants used phonetic cues to generate more global predictions regarding the language membership of upcoming lexical items. In the analysis of looks to the target picture in the early window when no ILD was present (Fig. 6, right panel), we found that target activation increased more rapidly in the cued than in the uncued condition. To confirm that this was statistically reliable, we ran a follow-up analysis on only the trials where no ILD was present, and found effects of Cuing on both the intercept ($\chi^2(1) = 24.4$, $p < .0001$, $\beta = 0.032$,

---

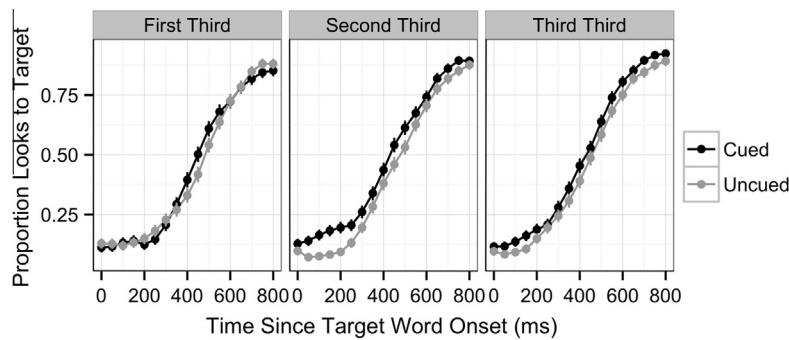[8] We thank an anonymous reviewer for this suggestion.

Fig. 9. Empirical plots showing looks to the target picture during the early window, for codeswitched trials only. The first, second, and third third of the experiment are each plotted separately.

$t = 4.9$, $p_{MCMC} < .0001$) and the linear time term ($\chi^2(1)$ = 3.7, $p = .05$, $\beta = 0.051$, $t = 1.9$, $p_{MCMC} = .05$), demonstrating that even when participants were unable to predict specific lexical items (cf. the window between 0 and 200 ms), phonetic cuing resulted in more looks to the target overall and a faster rate of increase in target activation. This is corroborated by the leftmost panel of Fig. 9, which likewise shows no item-specific predictions prior to the onset of the target word, but nonetheless a faster rise in target activation on cued trials. These findings strongly suggest that participants came into the experiment able to use phonetic cues to anticipate language switches, and that the ability to anticipate language switches is associated with different activation functions for codeswitched targets. This difference in the rate of target activation could be modeled with an anticipatory, global activation boost for the upcoming language, and/or by some manner of biasing mechanism.

In sum, the results of our comprehension study advance the understanding of bilingual auditory word recognition in several ways. We find that when placed in a mixed language context, habitual codeswitchers are sensitive to low-level, realistic, anticipatory phonetic cues signaling an upcoming codeswitch. Moreover, these cues promote more robust recognition of the target word both by allowing listeners to make item-specific predictions (in the context of this experiment), but also by boosting the rate of increase in target activation in the absence of specific predictions. The latter finding suggests a more global component of language regulation targeting the entire target versus non-target languages.

General discussion

In a set of studies examining the production and comprehension of phonetic variation in bilingual speech, we found that habitual codeswitchers produced low-level acoustic cues reflecting the demands placed on the production system during planning, and that listeners were able to exploit these cues during the comprehension of mixed language sentences to improve the recognition of codeswitched target items. In this final section, we summarize the importance of these results for current accounts of bilingual language regulation, and we conclude by relating our findings to the relationship between production and comprehension more broadly.

Using bilingual phonetic variation to study language processing

In recent discussions of the cognitive and neural consequences of bilingualism (e.g., Green & Abutalebi, 2013), codeswitching has often been invoked as an example of how the neural networks that support both language and cognitive control may be differentially tuned by language experience. Bilinguals who are equally proficient and balanced in their two languages may differ in whether they codeswitch or not, and also in whether their most frequent interlocutors are similarly bilingual or not. These different contexts of bilingual language use impose distinct demands on the cognitive system to negotiate language selection. In addition to allowing comparisons between habitual and non-habitual codeswitchers as a means of testing alternative accounts of the production–comprehension link, the type of research undertaken here enables new tests of the ways that language processing may create consequences for executive function. Few of the recent reports on consequences of bilingualism have identified specific aspects of language use that may be associated with the consequences that have been claimed for cognitive control mechanisms and for the neural circuitry that supports them. The few that have been suggested are focused primarily on lexical processes (e.g., Blumenfeld & Marian, 2011; Wu & Thierry, 2013). The significance of the present results is thus in identifying a specific instance of the tuning of the language comprehension system to a set of naturally occurring features in the output from the language production system, and in suggesting that such tuning may be a critical aspect of enabling fluent performance. On this account, examining the joint relationship between production and comprehension can help to develop specific hypotheses about which aspects of language performance may alter cognition more generally.

The research we report in this paper also demonstrates the utility of bilingualism as a lens for investigating subtle aspects of language processing that are otherwise difficult or impossible to observe in monolingual speakers (Kroll,

Bobb, & Hoshino, 2014; Kroll, Dussias, Bice, & Perotti, 2015). Bilingualism is of course interesting in its own right, but the present research exemplifies an instance in which it may also help to reveal interactions that characterize the relationship between comprehension and production more broadly. As described in the Introduction, work with monolingual speakers has similarly suggested that processing-related mechanisms can impact the phonetic form of monolingual speech (e.g., Gahl et al., 2012; Goldrick & Blumstein, 2006), and there is ample evidence that the comprehension systems of listeners of diverse language backgrounds are exquisitely tuned to regularities in the input (e.g., Dahan et al., 2001; Ju & Luce, 2004). The present work demonstrates that bilingual phonetic variation in particular constitutes an additional tool that can be fruitfully employed to study the interface between the production and comprehension systems, as well as the relationships between language experience, representations, and processing.

Linking production and comprehension

As noted in the introduction to the paper, the present findings can broadly be viewed as a phonetic analogue of MacDonald's (2013) Production–Distribution–Comprehension account. Under this account, the linguistic variation produced by a community of speakers is in part composed of distributional regularities introduced by limitations on the production system, and the comprehension systems of listeners who belong to that community consequently become adapted to these regularities. In our view, this makes an interesting prediction vis à vis the phenomenon of codeswitching. While all proficient bilinguals generally have access to codeswitching as a potential conversational resource, bilingual speakers differ widely in their deployment of this resource; some speakers belong to communities that regularly engage in codeswitching, following established conventions and community norms, while some speakers maintain a high degree of separation between the languages at their disposal. The PDC account makes the prediction that in general, proficient bilinguals who habitually codeswitch between the same two languages should tend to experience the same types of pressure on the production system, and members of the speech community should be attuned to the resulting norms. Our results are broadly consistent with this prediction, but it should be noted that the habitual codeswitchers in our production and comprehension studies were not members of the *same* codeswitching community (i.e., the corpus data concern bilinguals from Miami, while the listeners in our experiment hailed largely from Puerto Rico). A strong usage-based account predicts that particular communities may develop their own norms, which could, of course, diverge (see Torres Cacoullos & Travis, 2015, for an overview of relevant corpus research). An important question for future research will be to determine how the properties of specific linguistic systems (or combinations of linguistic systems) and the language history of a partic-

ular community impact the types of variation produced, as well as listener sensitivity to this variation.

A related prediction of the PDC is that *non*-habitual codeswitchers will *not* be attuned to the distributional regularities specific to proficient codeswitching. Even if the bilingual production system is subject to similar demands across speakers, thereby conditioning similar types of phonetic variation across populations, the PDC posits that it is membership in a speech community that fosters sensitivity to these potentially useful regularities. This stands in contrast to another recent proposal relating the production and comprehension systems, Dell and Chang's (2014) P-chain model, which hypothesizes that predictions made during comprehension processing are generated by the listener's own production system. The P-chain model therefore predicts that both habitual and non-habitual codeswitchers should be able to take advantage of phonetic cues associated with cross-language activation, to the extent that their own production systems would generate such cues. Our data cannot adjudicate between these models, but future work should directly address this issue by comparing the production and perception of phonetic variation in these different populations.

Finally, an assumption underpinning the framing of our study has been that costs to the production system – instantiated here by the cost that is hypothesized to accompany language regulation – can and do affect the surface form of spoken language. Importantly, however, it has often been argued that speakers adjust their speech to accommodate potential comprehension difficulties on the part of their listeners (Aylett & Turk, 2004; Clark & Fox Tree, 2002; see also Jaeger, 2013). Instead of taking a speaker-centric view as a point of departure, the story told here could have been that because cross-language activation imposes a cost during bilingual comprehension processing, proficient codeswitchers accommodate their listeners by providing cues to upcoming codeswitches. While we do not wish to suggest that bilingual speakers never accommodate their listeners, or that information-theoretic accounts (e.g., Jaeger, 2010) do not provide insight into the relation between linguistic form and processing, the fact that language regulation has repeatedly and independently been demonstrated to impose a cost during production planning leads us to prefer the speaker-oriented account of our production data, which in turn leads us to frame our results in terms of the PDC model.

Conclusion

Much psycholinguistic research has traditionally reflected a somewhat artificial divide between production and comprehension. While some accounts (perhaps especially within phonetics; cf. Liberman & Mattingly, 1985) have explicitly argued that the two types of processing are inextricably linked (see also Scott, McGettigan, & Eisner, 2009), the general theoretical decoupling of production and comprehension seems to have resulted more from a tendency toward narrowing empirical focus rather

than from principled argumentation. In a recent commentary, however, Pickering and Garrod (2014) point out that novel methodological approaches are beginning to challenge the default assumption that production and comprehension are modular processes. In our view, an exciting aspect of recent work (e.g., Dell & Chang, 2014; MacDonald, 2013; see also Silbert, Honey, Simony, Poeppel, & Hasson, 2014) and of this special issue is that we are beginning to see a reorientation toward the ways in which production and comprehension must be integrated to enable fluent communication. Our own investigation contributes to this reorientation by combining careful examination of the properties of spontaneous speech with controlled laboratory experimentation examining listener sensitivity to these properties. By exploiting the resources available to bilingual speakers, and by integrating corpus analyses and laboratory experimentation, this approach enriches our understanding of how this complex coordination process is ultimately achieved.

Acknowledgments

Appendix A. Details of the statistical model examining speech rate in the Bangor Miami Corpus

Formula: $\log(\text{SpeechRate}) \sim (1|\text{Speaker}) + (\text{UttType}|\text{Conversation}) + \text{Lang} + \log(\text{SwitchPos}) + \text{SynCat} + \log(\text{AvgPrecCondProb}) + \text{UttType}$

Scaled residuals				
Min	1Q	Median	3Q	Max
−2.6839	−0.6032	−0.1186	0.4409	5.6239

Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Speaker	(Intercept)	0.0001045	0.01022	
Conversation	(Intercept)	0.0001493	0.01222	
UttTypeCS		0.0002044	0.01430	−0.26
Residual		0.0020750	0.04555	

Number of obs: 1216, groups: Speaker, 68; Conversation, 48

Fixed effects					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	1.725e−01	6.937e−03	4.122e+02	24.870	<2e−16***
LangSpa	−2.667e−02	5.637e−03	3.390e+01	−4.731	3.85e−05***
log(SwitchPos)	−1.600e−02	3.597e−03	1.163e+03	−4.449	9.45e−06***
SynCatADJ	3.059e−03	4.968e−03	1.134e+03	0.616	0.53821
SynCatADV	1.825e−03	7.004e−03	1.156e+03	0.261	0.79448
SynCatCONJ	1.400e−02	6.784e−03	1.057e+03	2.063	0.03936*
SynCatDET	−6.061e−03	6.028e−03	1.131e+03	−1.006	0.31483
SynCatIM	3.459e−02	2.709e−02	1.146e+03	1.277	0.20189
SynCatname	−1.711e−02	9.836e−03	1.112e+03	−1.739	0.08232
SynCatNUM	9.643e−04	1.962e−02	1.121e+03	0.049	0.96080
SynCatPREP	8.303e−03	7.064e−03	1.162e+03	1.175	0.24009
SynCatPRON	1.497e−02	4.868e−03	9.276e+02	3.075	0.00217**
SynCatSV	−5.470e−03	3.370e−02	1.074e+03	−0.162	0.87110
SynCatunknown	−1.292e−02	1.162e−02	1.160e+03	−1.112	0.26646
SynCatV	−1.925e−03	5.104e−03	1.004e+03	−0.377	0.70614
log(AvgPrecCP)	−3.599e−02	2.218e−02	1.173e+03	−1.623	0.10496
UttTypeCS	1.599e−02	4.091e−03	1.620e+01	3.909	0.00123**

Appendix B. Details of the statistical model examining the rate of disfluency in the Bangor Miami Corpus

Formula: Disfluent ∼ (UttType|Speaker) + (1|Conversation) + log(SwitchPos) + log(UttLengthWords) + UttType

Scaled residuals				
Min	1Q	Median	3Q	Max
−2.3964	−0.5451	−0.3553	0.5826	5.0562

Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Speaker	(Intercept)	0.198480	0.44551	
	UttTypeCS	0.008618	0.09283	−1.00
Conversation	(Intercept)	0.477901	0.69130	

Number of obs: 1644, groups: Speaker, 72; Conversation, 48

Fixed effects				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−4.8085	0.3611	−13.318	<2e−16***
log(SwitchPos)	−0.3206	0.1878	−1.707	0.08787
log(UttLengthWords)	2.0785	0.2069	10.046	<2e−16***
UttTypeCS	0.4441	0.1369	3.243	0.00118**

Appendix C. Details of the statistical model comparing voice onset time in codeswitched versus unilingual utterances in the Bangor Miami Corpus

Formula: log(VOT) ∼ (UttType|Speaker) + (1|Conversation) + (UttType|Word) + Consonant + NSyll + SpeechRate + UttType

Scaled residuals				
Min	1Q	Median	3Q	Max
−4.3190	−0.5819	0.0553	0.6116	3.3266

Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Word	(Intercept)	2.859e−02	0.16909	
	UttTypeCS	5.713e−06	0.00239	1.00
Speaker	(Intercept)	3.746e−02	0.19355	
	UttTypeCS	1.030e−04	0.01015	1.00
Conversation	(Intercept)	1.296e−02	0.11385	
Residual		1.677e−01	0.40951	

Number of obs: 1376, groups: Word, 345; Speaker, 67; Conversation, 49

Fixed effects				
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	−2.64756	0.08958	614.00000	−29.554	<2e−16***
ConsonantP	−0.32902	0.04942	280.80000	−6.657	1.46e−10***
ConsonantK	0.03918	0.04120	218.30000	0.951	0.3426
NSyll	−0.04035	0.01749	456.10000	−2.306	0.0215*
SpeechRate	0.30826	0.04369	1331.40000	7.055	2.76e−12***
UttTypeCS	−0.05287	0.02497	1124.10000	−2.117	0.0345*

Appendix D. Details of the statistical model examining the effect of Following Spanish on voice onset time in codeswitched utterances in the Bangor Miami Corpus

Formula: log(VOT) ∼ (1|Speaker) + (0 + log(SpAfter + 1)|Speaker) + (1|Conversation) + (0 + log(SpAfter)|Word) + (1|Word) + Consonant + NSyll + SpeechRate + log(SpAfter)

Scaled residuals

Min	1Q	Median	3Q	Max
−2.99396	−0.46030	0.07131	0.51819	2.23527

Random effects

Groups	Name	Variance	Std.Dev.
Word	(Intercept)	0.07058	0.2657
Word.1	log(SpAfter)	0.00000	0.0000
Speaker	log(SpAfter)	0.00000	0.0000
Speaker.1	(Intercept)	0.04985	0.2233
Conversation	(Intercept)	0.01984	0.1409
Residual		0.13329	0.3651

Number of obs: 191, groups: Word, 95; Speaker, 47; Conversation, 40

Fixed effects

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	−3.20635	0.27400	186.15000	−11.702	<2e−16***
ConsonantP	−0.34650	0.12785	98.57000	−2.710	0.00793**
ConsonantK	−0.05826	0.09998	78.11000	−0.583	0.56174
NSyll	−0.03521	0.05355	94.99000	−0.657	0.51252
SpeechRate	0.08142	0.13278	178.80000	0.613	0.54053
log(SpAfter)	0.21473	0.08725	166.82000	2.461	0.01487*

Appendix E. Details of the statistical model examining looks to the target picture in the early window (0–800 ms following auditory onset of the target word), contrasting unilingual English and codeswitched trials

Formula: PropLooks ∼ (ot1 + ot2 + ot3|Subject) + (ot1 + ot2 + ot3|Subject:Lang) + (ot1 + ot2 + ot3) ∗ Lang

Scaled residuals

Min	1Q	Median	3Q	Max
−2.86476	−0.59475	−0.01748	0.54541	2.92828

Random effects

Groups	Name	Variance	Std.Dev.	Corr		
Subject:Lang	(Intercept)	0.0007596	0.02756			
	ot1	0.0072219	0.08498	0.13		
	ot2	0.0053335	0.07303	−0.69	−0.06	
	ot3	0.0021756	0.04664	−0.20	−0.80	0.43
Subject	(Intercept)	0.0146299	0.12095			
	ot1	0.0285911	0.16909	0.06		
	ot2	0.0336542	0.18345	−0.86	0.26	
	ot3	0.0049103	0.07007	−0.59	−0.83	0.31
Residual		0.1308870	0.36178			

Number of obs: 48,719, groups: Subject:Lang, 60; Subject, 30

Fixed effects					
	Estimate	Std. Error	*df*	*t* value	Pr(>|*t*|)
(Intercept)	0.486067	0.022767	31.900000	21.349	<2e−16***
ot1	1.187297	0.035849	38.760000	33.120	<2e−16***
ot2	0.031338	0.037295	36.460000	0.840	0.406
ot3	−0.230951	0.018098	48.050000	−12.761	<2e−16***
LangCS	−0.035421	0.007835	30.030000	−4.521	8.96e−05***
ot1:LangCS	0.002573	0.025773	29.950000	0.100	0.921
ot2:LangCS	0.114408	0.023202	29.960000	4.931	2.84e−05***
ot3:LangCS	0.022031	0.018105	40.710000	1.217	0.231

Appendix F. Details of the statistical model examining looks to the target picture in the late window (800–1800 ms following auditory onset of the target word), contrasting unilingual English and codeswitched trials

Formula: PropLooks ∼ (ot1 + ot2 + ot3|Subject) + (ot1 + ot2 + ot3|Subject:Lang) + (ot1 + ot2 + ot3) ∗ Lang

Scaled residuals				
Min	1Q	Median	3Q	Max
−2.5167	−0.7572	0.1215	0.7298	2.4529

Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Subject:Lang	(Intercept)	0.001215	0.03485	
	ot1	0.012877	0.11348	−0.02
	ot2	0.014270	0.11946	−0.74 −0.09
	ot3	0.001356	0.03682	0.27 −0.96 −0.19
Subject	(Intercept)	0.023912	0.15464	
	ot1	0.086491	0.29409	0.68
	ot2	0.027194	0.16491	−0.64 −0.33
	ot3	0.004415	0.06644	0.16 −0.45 −0.66
Residual		0.164507	0.40560	

Number of obs: 57,600, groups: Subject:Lang, 60; Subject, 30

Fixed effects					
	Estimate	Std. Error	*df*	*t* value	Pr(>|*t*|)
(Intercept)	0.521893	0.029041	31.800000	17.971	<2e−16***
ot1	−0.943429	0.058625	35.440000	−16.092	<2e−16***
ot2	0.166632	0.038800	44.160000	4.295	9.45e−05***
ot3	0.085002	0.017727	51.560000	4.795	1.43e−05***
LangCS	0.033241	0.009619	30.150000	3.456	0.00165**
ot1:LangCS	0.023232	0.033262	30.450000	0.698	0.49020
ot2:LangCS	−0.015553	0.034603	30.280000	−0.449	0.65629
ot3:LangCS	0.006231	0.018280	59.980000	0.341	0.73438

Appendix G. Details of the statistical model examining looks to the target picture in the early window (0–800 ms following auditory onset of the target word), examining effects of the interlingual distractor and phonetic cuing on codeswitched trials

Formula: PropLooks $\sim$ (ot1 + ot2 + ot3|Subject) + (ot1 + ot2 + ot3) + Dist $*$ Cued + Dist:ot1 + Dist:ot2 + Dist:ot3 + Cued:ot1 + Cued:ot2 + Cued:ot3 + Dist:Cued:ot1 + Dist:Cued:ot2 + Dist:Cued:ot3 + (ot1 + ot2 + ot3|Subject:Dist) + (ot1 + ot2 + ot3|Subject:Cued)

Scaled residuals				
Min	1Q	Median	3Q	Max
−2.93056	−0.56701	−0.03619	0.51878	2.89490

Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Subject:Cued	(Intercept)	0.001886	0.04343	
	ot1	0.012077	0.10989	0.25
	ot2	0.012316	0.11098	−0.64 −0.04
	ot3	0.002511	0.05011	−0.15 −0.99 0.02
Subject:Dist	(Intercept)	0.002037	0.04513	
	ot1	0.012587	0.11219	0.36
	ot2	0.011126	0.10548	−0.80 −0.25
	ot3	0.003627	0.06022	−0.49 −0.98 0.42
Subject	(Intercept)	0.013137	0.11462	
	ot1	0.023036	0.15177	0.24
	ot2	0.029586	0.17201	−0.82 0.36
	ot3	0.003945	0.06281	−0.94 −0.57 0.57
Residual		0.125264	0.35393	

Number of obs: 24,349, groups: Subject:Cued, 60; Subject:Dist, 60; Subject, 30

Fixed effects					
	Estimate	Std. Error	*df*	*t* value	Pr(>|*t*|)
(Intercept)	4.446e−01	2.427e−02	4.000e+01	18.317	<2e−16***
ot1	1.160e+00	4.404e−02	6.400e+01	26.342	<2e−16***
ot2	1.321e−01	4.601e−02	5.800e+01	2.870	0.005705**
ot3	−1.931e−01	2.618e−02	1.100e+02	−7.375	3.24e−11***
DistDist	−2.521e−02	1.330e−02	3.900e+01	−1.895	0.065623
CuedCued	3.189e−02	1.292e−02	4.000e+01	2.469	0.017884*
DistDist:CuedCued	1.083e−02	9.074e−03	2.407e+04	1.193	0.232734
ot1:DistDist	7.605e−02	3.925e−02	5.500e+01	1.938	0.057748
ot2:DistDist	8.125e−02	3.798e−02	5.800e+01	2.139	0.036684*
ot3:DistDist	−2.296e−02	3.070e−02	1.510e+02	−0.748	0.455661
ot1:CuedCued	5.138e−02	3.879e−02	5.400e+01	1.325	0.190799
ot2:CuedCued	−2.506e−02	3.899e−02	5.900e+01	−0.643	0.522901
ot3:CuedCued	−3.215e−02	2.943e−02	1.880e+02	−1.092	0.276070
ot1:DistDist:CuedCued	−1.349e−01	3.742e−02	2.407e+04	−3.606	0.000312***
ot2:DistDist:CuedCued	−5.737e−02	3.742e−02	2.407e+04	−1.533	0.125255
ot3:DistDist:CuedCued	4.689e−02	3.742e−02	2.407e+04	1.253	0.210116

Appendix H. Details of the statistical model examining looks to the target picture in the late window (800–1800 ms following auditory onset of the target word), examining effects of the interlingual distractor and phonetic cuing on codeswitched trials

Formula: PropLooks ~ (ot1 + ot2 + ot3|Subject) + (ot1 + ot2 + ot3) + Dist ∗ Cued + Dist:ot1 + Dist:ot2 + Dist:ot3 + Cued:ot1 + Cued:ot2 + Cued:ot3 + Dist:Cued:ot1 + Dist:Cued:ot2 + Dist:Cued:ot3 + (ot1 + ot2 + ot3|Subject:Dist) + (ot1 + ot2 + ot3|Subject:Cued)

Scaled residuals				
Min	1Q	Median	3Q	Max
−2.5970	−0.7470	0.0962	0.7108	2.5506

Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Subject:Cued	(Intercept)	0.004487	0.06699	
	ot1	0.038898	0.19722	0.47
	ot2	0.011519	0.10733	−0.58 −0.59
	ot3	0.005527	0.07435	−0.24 −0.96 0.37
Subject:Dist	(Intercept)	0.002225	0.04718	
	ot1	0.022428	0.14976	0.36
	ot2	0.008930	0.09450	−0.52 −0.02
	ot3	0.002173	0.04662	0.06 −0.88 −0.43
Subject	(Intercept)	0.023170	0.15222	
	ot1	0.078027	0.27933	0.65
	ot2	0.026773	0.16362	−0.59 −0.37
	ot3	0.001954	0.04420	0.14 −0.47 −0.59
Residual		0.157176	0.39645	

Number of obs: 28,785, groups: Subject:Cued, 60; Subject:Dist, 60; Subject, 30

Fixed effects					
	Estimate	Std. Error	*df*	*t* value	Pr(>|t|)
(Intercept)	5.443e−01	3.191e−02	3.900e+01	17.057	<2e−16***
ot1	−8.906e−01	7.154e−02	5.300e+01	−12.450	<2e−16***
ot2	1.874e−01	4.519e−02	6.300e+01	4.148	0.000103***
ot3	9.776e−02	2.804e−02	1.290e+02	3.486	0.000670***
DistDist	1.990e−02	1.388e−02	3.800e+01	1.434	0.159627
CuedCued	5.997e−03	1.853e−02	3.400e+01	0.324	0.748162
DistDist:CuedCued	−7.990e−03	9.369e−03	2.851e+04	−0.853	0.393756
ot1:DistDist	−6.701e−02	4.941e−02	4.600e+01	−1.356	0.181718
ot2:DistDist	−7.677e−02	3.915e−02	6.300e+01	−1.961	0.054308
ot3:DistDist	−2.372e−02	3.281e−02	1.900e+02	−0.723	0.470511
ot1:CuedCued	−5.540e−02	5.947e−02	4.000e+01	−0.932	0.357166
ot2:CuedCued	−3.523e−02	4.129e−02	5.800e+01	−0.853	0.397110
ot3:CuedCued	1.425e−05	3.607e−02	9.700e+01	0.000	0.999686
ot1:DistDist:CuedCued	1.276e−01	4.335e−02	2.853e+04	2.943	0.003256**
ot2:DistDist:CuedCued	7.840e−02	4.326e−02	2.853e+04	1.812	0.069931
ot3:DistDist:CuedCued	1.929e−02	4.316e−02	2.852e+04	0.447	0.654827

Appendix I. Details of the statistical model examining looks to the competitor pictures (0–1000 ms following auditory onset of the target word)

Formula: PropLooks ∼ (ot1 + ot2 + ot3 + ot4|Subject) + (ot1 + ot2 + ot3 + ot4) + Object ∗ Cued + Object:ot1 + Object:ot2 + Object:ot3 + Object:ot4 + Cued:ot1 + Cued:ot2 + Cued:ot3 + Cued:ot4 + Object:Cued:ot1 + Object:Cued:ot2 + Object:Cued: ot3 + (ot1 + ot2 + ot3|Subject:Cued) + (ot1 + ot2 + ot3|Subject:Object)

Scaled residuals				
Min	1Q	Median	3Q	Max
−1.8021	−0.3848	−0.1348	−0.0244	5.0975

Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Subject:Object	(Intercept)	0.0002025	0.01423	
	ot1	0.0067250	0.08201	−0.91
	ot2	0.0033695	0.05805	0.09 −0.38
	ot3	0.0029319	0.05415	0.61 −0.52 −0.55
Subject:Cued	(Intercept)	0.0001889	0.01375	
	ot1	0.0024317	0.04931	−0.85
	ot2	0.0011385	0.03374	−0.11 −0.34
	ot3	0.0004634	0.02153	0.67 −0.62 −0.38
Subject	(Intercept)	0.0007844	0.02801	
	ot1	0.0199573	0.14127	−0.92
	ot2	0.0044817	0.06695	0.72 −0.93
	ot3	0.0015343	0.03917	0.83 −0.98 0.98
	ot4	0.0020389	0.04515	−0.66 0.90 −1.00 −0.96
Residual		0.0387050	0.19674	

Number of obs: 29,988, groups: Subject:Object, 60; Subject:Cued, 60; Subject, 30

Fixed effects					
	Estimate	Std. Error	*df*	*t* value	Pr(>|*t*|)
(Intercept)	5.795e−02	6.662e−03	5.000e+01	8.699	1.50e−11***
ot1	−1.849e−01	3.286e−02	4.700e+01	−5.626	9.58e−07***
ot2	−5.435e−03	2.022e−02	6.900e+01	−0.269	0.78884
ot3	4.851e−02	1.653e−02	8.800e+01	2.934	0.00427**
ot4	−4.595e−03	1.223e−02	7.800e+01	−0.376	0.70817
ObjectILD	2.180e−02	4.886e−03	5.400e+01	4.463	4.14e−05***
CuedCued	−5.596e−03	4.788e−03	5.200e+01	−1.169	0.24786
ObjectILD:CuedCued	−2.752e−03	4.545e−03	2.970e+04	−0.605	0.54486
ot1:ObjectILD	−3.863e−02	2.582e−02	5.900e+01	−1.496	0.13993
ot2:ObjectILD	−2.594e−02	2.104e−02	9.500e+01	−1.233	0.22062
ot3:ObjectILD	4.665e−02	2.033e−02	1.090e+02	2.294	0.02368*
ot4:ObjectILD	5.735e−03	1.041e−02	2.969e+04	0.551	0.58189
ot1:CuedCued	2.791e−02	1.947e−02	7.300e+01	1.433	0.15607
ot2:CuedCued	1.149e−02	1.711e−02	1.250e+02	0.671	0.50330
ot3:CuedCued	−9.960e−03	1.574e−02	1.980e+02	−0.633	0.52768
ot4:CuedCued	1.126e−03	1.042e−02	2.970e+04	0.108	0.91391
ot1:ObjectILD:CuedCued	5.888e−04	2.083e−02	2.969e+04	0.028	0.97745
ot2:ObjectILD:CuedCued	4.487e−03	2.083e−02	2.969e+04	0.215	0.82946
ot3:ObjectILD:CuedCued	−5.600e−02	2.083e−02	2.970e+04	−2.688	0.00718**

References

Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition, 15*, 517–530.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new disfluency and reference resolution. *Psychological Science, 15*, 578–582.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy,

prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*, 31–56.

Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science, 25*, 1546–1553.

Balukas, C., & Koops, C. (2015). Spanish–English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism, 19*, 423–443.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457–474.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America, 133*, 2350–2366.

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*, 92–111.

Blumenfeld, H. K., & Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition, 118*, 245–257.

Boersma, P., & Weenink, D. (2014). *Praat: Doing phonetics by computer [computer program]. Version 5.4.* <http://www.praat.org> Retrieved 04.10.14.

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language, 75*, 104–116.

Boucher, V. J. (2002). Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception & Psychophysics, 64*, 121–130.

Broersma, M., & De Bot, K. (2006). Triggered codeswitching: A corpus-based evaluation of the original triggering hypothesis and a new alternative. *Bilingualism: Language and Cognition, 9*, 1–13.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73–111.

Clyne, M. G. (2003). *Dynamics of language contact: English and immigrant languages.* Cambridge University Press.

Colomé, À., & Miozzo, M. (2010). Which words are activated during bilingual word production? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 96.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition, 105*, 658–668.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1283.

Costa, A., & Santesteban, M. (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.

Crystal, T. H., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America, 88*, 101–112.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*, 507–534.

Dell, G., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*, 20120394.

Deuchar, M., Davies, P., Herring, J., Parafita Couto, M., & Carter, D. (2014). Building bilingual corpora. In E. M. Thomas & I. Mennen (Eds.), *Advances in the study of bilingualism* (pp. 93–110). Bristol, UK: Multilingual Matters.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, (PPVT-4).* Bloomington, MN: Pearson Education Inc.

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 663.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*, 789–806.

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes, 21*, 649–683.

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

Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost–benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 640.

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language, 27*, 210–223.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67–81.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*, 515–530.

Green, D. W., & Wei, L. (2014). A control process model of code-switching. *Language, Cognition, and Neuroscience, 29*, 499–511.

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes, 3*, 233–274.

Gullberg, M., Indefrey, P., & Muysken, P. (2012). Research techniques for the study of code switching. In B. Bullock & A. J. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 21–39). Cambridge: Cambridge University Press.

Gullifer, J. W., Kroll, J. F., & Dussias, P. E. (2013). When language switching has no apparent cost: Lexical access in sentence context. *Frontiers in Psychology, 4*.

Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese–English bilinguals. *NeuroImage, 56*, 2300–2309.

Hallett, P. E. (1986). Eye movements. In K. Boff, L. Kaufman, & J. Thomas (Eds.). *Handbook of perception and human performance* (Vol. 1, pp. 10–112). New York, NY: Wiley-Interscience.

Hartsuiker, R. J., & Pickering, M. J. (2008). Language integration in bilingual sentence production. *Acta Psychologica, 128*, 479–489.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition, 106*, 501–511.

Jacobs, A., Fricke, M., & Kroll, J. F. (2015). Cross-language activation begins during planning and extends into second language speech. *Language Learning, 66*.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61*, 23–62.

Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology, 4*.

Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science, 15*, 314–318.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological Studies in Language, 45*, 229–254.

Kaplan, E., Goodglass, H., Weintraub, S., Segal, O., & Loon-Vervoorn, A. (2001). *The Boston Naming Test.* Pro-Ed.

Kaushanskaya, M., & Marian, V. (2007). Bilingual language processing and interference in bilinguals: Evidence from eye tracking and picture naming. *Language Learning, 57*, 119–163.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research, 18*, 686–706.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59*, 1208–1221.

Kootstra, G. J., van Hell, J. G., & Dijkstra, T. (2010). Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language, 63*, 210–231.

Kootstra, G. J., Van Hell, J. G., & Dijkstra, T. (2012). Priming of code-switches in sentences: The role of lexical repetition, cognates, and language proficiency. *Bilingualism: Language and Cognition, 15*, 797–819.

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science, 23*, 159–163.

Kroll, J. F., & Dussias, P. E. (2013). The comprehension of words and sentences in two languages. In T. Bhatia & W. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (2nd ed., pp. 216–243). Malden, MA: Wiley-Blackwell Publishers.

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). LmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R Package Version, 2*.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 952.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language and Cognition, 16*, 508–517.

Levelt, W. J. (1989). *Speaking: From intention to articulation*. MIT Press.

Li, P. (1996). Spoken word recognition of code-switched words by Chinese–English bilinguals. *Journal of Memory and Language, 35*, 757–774.

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 381.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1–36.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384–422.

Loebell, H., & Bock, K. (2003). Structural priming across languages. *Linguistics, 41*, 791–824.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*, 1–16.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*, 133–156.

Marian, V., Blumenfeld, H., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*, 940–967.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition, 6*, 97–115.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*, B33–B42.

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*, 25–40.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. CRC Press.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford University Press.

Myslín, M., & Levy, R. (in press). Codeswitching and predictability of meaning in discourse. *Language*.

Pickering, M. J., & Garrod, S. (2014). Neural integration of language production and comprehension. *Proceedings of the National Academy of Sciences, 111*, 15291–15292.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics, 18*, 581–618.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217–236.

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language, 55*, 197–212.

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—Candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience, 10*, 295–302.

Shenk, P. S. (2006). The interactional and syntactic importance of prosody in Spanish–English bilingual discourse. *International Journal of Bilingualism, 10*, 179–205.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences, 111*, E4687–E4696.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*, 281–284.

Thomas, M. S., & Allport, A. (2000). Language switching costs in bilingual visual word recognition. *Journal of Memory and Language, 43*, 44–66.

Torres Cacoullos, R. T., & Travis, C. E. (2015). Gauging convergence on the ground: Code-switching in the community. *International Journal of Bilingualism, 19*, 365–386.

Valdes Kroff, J. R., Dussias, P. E., Gerfen, C., Perrotti, L., & Bajo, M. T. (2015). Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguistic Approaches to Bilingualism*.

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language, 64*, 88–107.

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science, 20*, 923–927.

Van Hell, J. G., & De Groot, A. M. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica, 128*, 431–451.

Verhoeven, J., De Pauw, G., & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech, 47*, 297–308.

Von Studnitz, R. E., & Green, D. W. (2002). Interlingual homograph interference in German–English bilinguals: Its modulation and locus of control. *Bilingualism: Language and Cognition, 5*, 1–23.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*, 1–25.

Weide, R. (2007). *The Carnegie Mellon pronouncing dictionary [cmudict 0.07a]*.

Wu, Y. J., & Thierry, G. (2012). Unconscious translation during incidental foreign language processing. *NeuroImage, 59*, 3468–3473.

Wu, Y. J., & Thierry, G. (2013). Fast modulation of executive function by language context in bilinguals. *The Journal of Neuroscience, 33*, 13533–13537.