Memory & Cognition (2021) 49:163–180
https://doi.org/10.3758/s13421-020-01069-7

The effect of lexical accessibility on Spanish-English intra-sentential codeswitching

Justin T. Sarkis[1] · Jessica L. Montag[2]

Published online: 10 July 2020

Abstract
Bilingual speakers sometimes codeswitch, or alternate between languages, in a single utterance. We investigated the effect of lexical accessibility of words, defined as the ease with which a speaker retrieves and produces a word, on codeswitching in Spanish-English bilinguals. We first developed a novel sentence-production paradigm to elicit naturalistic codeswitches in the lab. We then predicted items on which speakers were more or less likely to codeswitch as a consequence of the relative lexical accessibility of those items' labels across a speaker's two languages. In a Spanish sentence-production task, greater lexical accessibility in English was associated with an increased rate of codeswitching and longer speaking durations on trials on which speakers codeswitched, as well as on trials on which speakers did not codeswitch. Codeswitches were more frequent on trials where speakers likely experienced more competition from the other-language label, suggesting that codeswitching may be a tool that bilingual speakers use to alleviate difficulty associated with cross-language lexical competition. Given findings that comprehenders are able to learn lexical distributions and subtle acoustic cues to predict upcoming codeswitches, the present work suggests that demands on speakers during language production may play a role in explaining how those patterns come to exist in the language environment.

Keywords Codeswitching · Bilingualism · Lexical accessibility · Language production

Introduction

When producing an utterance, speakers have many options for conveying their intended message. During the sentence-production process, a speaker must choose a single set of words, in a particular order, out of multiple available alternatives. A major goal of the field of language production has been to understand why speakers make the choices they do, and the cognitive processes that underlie a speaker's ability to converge on a single sequence of words. For some bilingual speakers, one additional choice that is available is to codeswitch, or switch languages within a single utterance or conversation. These codeswitches can be inter-sentential, meaning that the language switch occurs between sentences or utterances or they can be intra-sentential, meaning that the switch occurs in the middle of a sentence or utterance (Gumperz, 1977), as in the Spanish sentence *El libro está en eldesk* [The book is on the desk]. The present work is primarily concerned with understanding why speakers make intra-sentential codeswitches.

Codeswitching, and what it reveals about language production in bilingual speakers, has been the topic of a great deal of prior research. Codeswitching does not occur randomly in speech, but rather is highly patterned such that it reflects high levels of proficiency in both languages (Bentahila & Davies, 1983; Otheguy & Lapidus, 2003; Poplack, 1980). For this reason, it has been a key arena in which the inhibitory and other cognitive demands associated with navigating multiple languages have been investigated. The implications for this research span from how bilingual speakers plan and process utterances, to more general cognitive consequences of bilingualism (Bialystok, Craik, Green, & Gollan, 2009; Declerck & Philip, 2015; Green & Abutalebi, 2013; Hernandez, Dapretto, Mazziotta, & Bookheimer, 2001; Meuter & Allport, 1999). The cognitive demands associated with codeswitching tasks in the lab are thought to be a window into cognitive demands associated with navigating two languages in day-to-day life.

✉ Jessica L. Montag
jmontag@illinois.edu

[1] Department of Psychology, University of California, Riverside, CA, USA

[2] Department of Psychology, University of Illinois, Urbana-Champaign, 603 E. Daniel Street, Champaign, IL 61820, USA

The approach of the present study departs from many previous codeswitching studies. Our goal was to understand when and why speakers make the decision to codeswitch in naturalistic speech. We developed a novel sentence-production paradigm that elicited naturalistic codeswitches using a picture description task in the lab. Our task reliably elicited codeswitches from speakers without explicit instruction to codeswitch. We then made and tested predictions for *items* and item properties that are more or less likely to elicit codeswitches. We made our predictions in the context of the broader sentence-production literature, specifically what is known about lexical selection processes during sentence production. Not all possible utterances are equally difficult to plan and produce, and various features of the to-be-produced utterance are known to affect speakers' word and sentence choices. A common motivation of many choices that speakers make is to alleviate difficulty associated with language planning and production (e.g., Bock, 1982; Ferreira & Dell, 2000; MacDonald, 2013; Zhan & Levy, 2018). We aimed to understand bilingual codeswitching in the context of this difficulty-reduction framework. By gaining a better understanding of when and why bilingual speakers codeswitch in natural speech, we can better understand both the cognitive processes that underlie bilingual language use and language production processes more generally.

Codeswitching is not typically cast within a difficulty reduction framework but rather is classically thought to be a cognitively demanding behavior. For example, when speakers are directed to switch languages during a picture-naming task, naming latencies are longer on individual trials in which the naming language is different than the previous trial (Bobb & Wodniecka, 2013; Costa & Santesteban, 2004; Meuter & Allport, 1999) and in mixed language versus single language blocks (Christoffels, Firk, & Schiller, 2007; Hernandez & Kohnert, 1999). These increased naming latencies are interpreted as a marker of the cognitive demands associated with codeswitching, specifically, as well as the inhibitory demands of speaking two languages, generally. Despite the difficulty associated with switching languages, codeswitching is thought to persist because it provides other advantages for speakers by allowing greater control over utterances; to be more precise, to distance themselves from negative emotions, or to signal group membership, among others (Bond & Lai, 1986; Heredia & Altarriba, 2001; Myers-Scotton, 1989). Codeswitching is often conceived of a choice that is itself associated with greater cognitive demands, but that may fulfill other important communicative functions for a speaker.

However, recent lines of work suggest that codeswitching may not be inherently cognitively demanding under many circumstances. A number of task-related factors contribute to the observed difficulty of laboratory codeswitching tasks and may not be representative of codeswitching in naturalistic contexts (Blanco-Elorrieta & Pylkkanen, 2018). For example, in many studies, speakers are instructed when to switch languages. However, during natural speech, speaker-internal factors determine if and when a speaker will codeswitch. Whether speakers choose to, or are instructed to, switch languages has consequences for the measured difficulty of codeswitching. In picture-naming tasks in which speakers can choose when to switch languages (often called "voluntary" or "bottom-up" codeswitching), overall naming latencies are faster than when speakers are told when to switch languages (deBruin, Samuel, & Dunabeitia, 2018; Gollan & Ferreira, 2009; Gollan, Kleinman, & Wierenga, 2014b; Gross & Kaushanskaya, 2015; Kleinman & Gollan, 2016) and neural signatures reflect reduced processing demands (Zhang et al., 2015). Allowing speakers to freely switch languages, more like natural codeswitches, suggests a smaller cost associated with codeswitching. The reduction of a switching cost in more naturalistic settings along with the anecdotal observation that bilingual speakers do not perceive codeswitching to be cognitively demanding (Blanco-Elorrieta & Pylkkanen, 2018) suggests that casting codeswitching within a difficulty-reduction framework, like many other choices that speakers make, may be plausible.

If codeswitching is to be cast in a difficulty-reduction framework, why might it be helpful to speakers? To answer this question, we consider frameworks that suggest that a bilingual speaker's two languages are active independent of which language a speaker is currently using (Kroll, Bobb, & Wodniecka, 2006; Thierry & Wu, 2007; cf. Costa, La Heij, & Navarrette, 2006). One possible consequence of language co-activation is that bilingual speakers face competition during lexical selection between their two languages (Kroll, Bobb, & Hoshino, 2014) such that when choosing a word to produce, speakers may require inhibitory processes to resolve the competition and select a single word in a single language (Costa & Santesteban, 2004; Finkbeiner, Gollan, & Caramazza, 2006; Green, 1998). In the present work, we propose that consequences of language co-activation, specifically lexical competition between translation equivalents, may be important for understanding intra-sentential codeswitches. Codeswitching may be a tool available to the speaker to alleviate difficulty associated with this lexical competition during language production.

The present work makes the theoretical prediction that lexical competition between translation equivalents across a bilingual's two languages is analogous to the competition that monolingual speakers experience between synonyms. For monolingual speakers, naming latencies are typically slower for pictures with multiple common labels (e.g., couch/sofa) than for words with a single dominant label (e.g., chair), likely due to the co-activation of the multiple possible labels and then inhibition of all but one alternative (Lachman, 1973; Levelt et al., 1991; Peterson & Savoy, 1998; Vitkovitch &Tyrrell, 1995). Just as lexical competition creates difficulty for monolingual speakers, lexical competition across bilingual

speakers' two languages may also create cognitively demanding contexts for language production (Ivanova & Costa, 2008; Marian & Spivey, 2003). Codeswitching may be a tool available to bilingual speakers to alleviate the planning difficulty associated with competition between labels. When faced with the difficulty associated with producing a single label and inhibiting alternative labels, codeswitching may allow the speaker to produce the label that is easiest to recall and plan, regardless of the language of that label.

Lexical accessibility, or the ease with which a word can be recalled from long-term memory and planned, is commonly implicated in many different sentence-production choices. Speakers often choose an utterance form (e.g., an active or a passive sentence) that places the more lexically accessible words or phrases earlier in an utterance, presumably to leave more time to plan more difficult words or phrases (Arnold, Losongco, Wasow, & Ginstrom, 2000; Ferreira, 2008; MacDonald, 2013). For example, speakers tend to place the easier to plan discourse-given versus discourse-new words (Bock, 1982; Bock & Irwin, 1980), animate versus inanimate words (McDonald, Bock, & Kelly, 1993; Prat-Sala & Branigan, 2000), and frequent versus infrequent words (Griffin & Bock, 1998; Jescheniak & Levelt, 1994) earlier in utterances. Not all of a speaker's multiple alternatives may be equally easy to plan and produce, and when an option exists, speakers often choose easier over more difficult alternatives.

The present study tests the prediction that bilingual speakers may be more likely to codeswitch when they experience competition between translation equivalents, and the other-language label is more accessible. In other words, the relative lexical accessibility of labels should predict codeswitches in a naturalistic language-production task. Sentential contexts in which a bilingual speaker experiences the most lexical competition from an other-language label – which contributes to those contexts being inherently harder – should be when speakers are most likely to codeswitch. The idea that codeswitching is associated with inherently more difficult contexts is consistent with previous work showing that speakers are more likely to switch from their non-dominant into their dominant language when producing lower-frequency words (Gollan & Ferreira, 2009; Gross & Kaushanskaya, 2015) and that words are named more slowly in speakers' non-dominant language (de Bruin, Samuel & Dunabeitia, 2018). Our item-based approach aims to assess the relative accessibility of word-translation equivalent pairs and then use that relative accessibility to predict the presence or absence of a codeswitch and the time course of sentence production.

We developed a laboratory paradigm to elicit naturalistic, intra-sentential codeswitches from Spanish-English bilinguals without a confederate. Our goal was to predict codeswitches as a consequence of the relative lexical accessibility of the competing Spanish and English labels for target items. To assess lexical accessibility, speakers named pictures freely in either Spanish or English. The relative lexical accessibility was defined as the proportion of English or Spanish labels produced across participants. Then, a different group of Spanish-English bilingual speakers produced sentences in our picture description paradigm in one of three conditions: Codeswitch-permitted (participants were instructed to perform the task in Spanish but codeswitches into English were permitted), Spanish-only (participants were instructed to only use Spanish), or English-only (participants were instructed to only use English). This design allowed us to compare production choices – specifically, whether utterances contained a codeswitch – and the initiation latencies and speaking duration of the produced utterances, as a consequence of the lexical accessibility of the target items across multiple experimental conditions. If codeswitching is motivated in part by the difficulty associated with competition across translation-equivalent labels during lexical selection, then in the Spanish sentence-production tasks, words that are more accessible in English should be more likely to be codeswitched and should take longer to produce even when not codeswitched.

The two single-language conditions (Spanish-only and English-only) allow us to investigate consequences of lexical accessibility for the time course of sentence production, independent of a codeswitching choice. In the Spanish-only condition, speakers should be *slower* to produce words that are more accessible in English, indexing the difficulty of these items, even when they do not codeswitch. In the English-only condition we predict an opposite effect, such that speakers should be *faster* to produce items whose label was more accessible in English. However, given that our participants are largely English-dominant, the effect size may be reduced in the English language task.

Method

Item selection

In order to assess the effect of lexical accessibility of competing labels across English and Spanish, we needed a set of items with high name agreement in both languages so that lexical competition during language production would arise from competition across languages, rather than from multiple available labels within a single language. To this end, we selected a set of pictures that participants consistently named with a single label in each language.

Participants Nineteen Spanish-English bilinguals from the psychology participant pool at the University of California, Riverside participated for course credit. An additional six participants were run and excluded due to: equipment failure (three), the participant failing to follow task instructions (two), and experimenter error (one). Table 1 shows the

Table 1 Demographic characteristics of participants. Standard deviations are shown in parentheses

	N	Age	Born in USA	English dominant	English proficiency	Spanish proficiency
Picture naming	19	19.8 (1.46)	18 / 94.7%	16 / 84.2%	9.42 (0.77)	7.89 (1.48)
Lexical accessibility	24	18.9 (0.90)	18 / 75.0%	18 / 75.0%	9.17 (0.82)	8.38 (1.17)
Sentence-production task						
Codeswitch permitted	35	19.30 (2.31)	31 / 88.6%	27 / 77.1%	9.20 (0.96)	7.51 (1.22)
Spanish only	35	19.51 (2.03)	29 / 82.9%	29 / 82.9%	8.74 (1.12)	8.09 (1.15)
English only	58	19.53 (1.58)	51 / 87.9%	53 / 91.4%	9.26 (0.98)	7.69 (1.39)

demographic characteristics of these and all other participants included in this study. English dominance was self-reported, and English and Spanish proficiency refer to the self-reported speaking proficiency on a 1–10 scale. All participants began learning Spanish from birth and reported that they grew up in a Spanish-speaking household. Consistent with this self-report, all but five participants across all tasks reported that their primary caregiver in childhood spoke Spanish as their first language. For all but one participant whose primary caregiver's first language was not Spanish, Spanish was the second language of their primary caregiver *and* the first language of their secondary caregiver. The participants could be described as typical Southern California heritage bilinguals (e.g., Gollan, Starr & Ferreira, 2015). Table 2 shows the self-reported codeswitching behavior of participants included in this study.

Picture-naming procedure Participants sat in front of a computer screen and named 525 pictures from the International Picture Naming Project (Bates et al., 2000) in random order. Participants completed the task in either English or Spanish. After the task, participants completed a language history questionnaire.

Picture-naming results This norming procedure yielded a set of 40 pictures with high name agreement in both English and Spanish. All selected items were inanimate (no humans or animals were considered as possible items) due to the well-established tendency for animacy to affect lexical accessibility, specifically (Branigan, Pickering, & Tanaka, 2008), and cognitive processes, more broadly (Caramazza & Shelton, 1998; Narine, VanArsdall, & Cogdill, 2017). We wanted to ensure that participants could indeed generate both Spanish and English labels for items, so we took a conservative approach to name agreement such that failures to generate any label for an item were considered label disagreement. This procedure ensured that we only selected items that our population of participants could consistently name. Further, providing an additional adjective to describe an item (e.g., "empty glass" instead of "glass") was considered label disagreement. The 40 items included in the study (1) had high name agreement in each language and (2) were items that participants in this naming task could consistently name. Average name agreement across items was 90.6% for English labels and 91.9% for Spanish labels, and no individual item had agreement less than 75% in either language. For reference, average name agreement of our items according to the International Picture Naming Project norms was 97.7% for English and 95.1% for Spanish, with no single item below 77% agreement (English) and 67% agreement (Spanish).

Assessment of lexical accessibility

Following the item selection procedure described above, where we selected 40 experimental items that participants could name in both English and Spanish and consistently named with a single label in both languages, we assessed the tendency across speakers to label the image in English or

Table 2 Participants' self-reported codeswitching frequency

	N	Never	Very infrequently	Occasionally	Frequently	Always
Picture naming	19	1	4	6	4	4
Lexical accessibility	24	1	5	9	6	3
Sentence-production task						
Codeswitch permitted	35	0	7	13	11	4
Spanish only	35	2	6	9	12	6
English only	58	4	8	21	16	9

Spanish when allowed to use either language freely. We used this measure – the proportion of speakers who produced the English versus Spanish label – as a proxy for the relative lexical accessibility of the two labels. Given competition between two available labels, speakers may be more likely to produce the label that is more accessible, so this paradigm would allow us to capture systematicities in lexical accessibility in a population (akin to using a corpus to assess word frequency) that may predict behavior in other tasks.

We assessed lexical accessibility because we predicted that the relative accessibility of translation equivalent labels would contribute to production choices. The degree to which an object has a single label or multiple labels is often referred to as "name agreement" or "codability" and has previously been used as an independent measure predicting speaking times and other behavioral measures in sentence-production tasks. Items with multiple possible labels often take longer to name and are prone to more errors than items with a single high-agreement label (Griffin, 2001; Hartsuiker & Notebaert, 2010; Vitkovitch & Tyrell, 1995). Following this literature, we predicted that competition of translation equivalents across languages, similar to competition between multiple labels in a single language, may similarly make lexical selection more difficult and contribute to production choices, specifically codeswitching, and the time course of sentence production.

Participants Twenty-four Spanish-English bilinguals from the introductory psychology participant pool at the University of California, Riverside participated for course credit. An additional ten participants were run but excluded from analyses for several reasons: completing the task entirely in Spanish (four), completing the task entirely in English (four), equipment failure (one), and experimenter error (one). Demographic characteristics and self-reported codeswitching behavior are shown in Tables 1 and 2.

Procedure Participants sat in front of a computer screen. On each trial, participants were presented with the items in random order and asked to name the picture in their choice of either English or Spanish. They were instructed to name the pictures as quickly as they could without making errors. Participants were told that they could use either English or Spanish and switch languages as often as they liked. After the task, participants completed a language history questionnaire.

Results Trials on which participants failed to label the target object, mis-identified the object (e.g., "cantaloupe" instead of "orange"), or produced a label other than the label identified as the high name-agreement label during item selection were excluded. Overall, 5.9% of trials contained one of these errors and were excluded. Participants failed to generate any label on 1.9% of trials, generated an English label other than the high name-agreement label on 2.3% of trials, and a Spanish label other than the high name-agreement label on 1.8% of trials. The remaining 94.1% of trials on which participants produced either the high name agreement English or Spanish label were analyzed. Given the low rate of error responses (5.9%), we do not believe that these errors substantially contribute to our measure of lexical accessibility. Participants produced on average 64.7% (range = 39.1.0–95.8%; SD = 15.5) of their labels in English. In order to generate predictions for our sentence-production task, we computed by-item averages of the proportion of participants who named that item in English. We observed a range of proportions, with some items more frequently produced in English, and others more frequently produced in Spanish. Figure 1 illustrates these proportions in each of the 40 items.

The observed variability in speakers' language preferences across items allowed us to make predictions for the sentence-production task. If lexical competition contributes to codeswitching choices, speakers should more often codeswitch when the item label is relatively more accessible in their other language. When performing the sentence-production task in Spanish, words on the right side of Fig. 1 (words that in the lexical accessibility assessment were more likely to be produced in English) should be more likely to be codeswitched. For these items, the English label should intrude more, and be harder to inhibit, than items that were more often named in Spanish on the lexical accessibility assessment. Further, because of the increased competition from words that are more accessible in English, even when speakers choose not to codeswitch, we expected to see longer speaking durations on these same items, which are more accessible in English, relative to those that are less accessible in English (and thus more accessible in Spanish). When performing the sentence-production task in English, we predicted the opposite pattern of results, such that words on the left side of Table 1 should yield longer initiation latencies, though this effect may be attenuated or absent given the overall higher self-reported English proficiency of the participants in our sample.

Sentence-production task

After selecting 40 items that one group of participants could consistently name in both languages, using a consistent label (item selection task), and then assessing the relative lexical accessibility of the English versus Spanish labels of these items in a second group of participants (assessment of lexical accessibility), a final group of participants completed a picture-description sentence-production task using these 40 items. This task was designed to allow us to investigate the consequences of lexical accessibility on the language production choices that speakers make and the time course of the sentence-production process. We investigated language production choices in three conditions: (1) a Spanish condition in which participants were permitted to codeswitch into English,

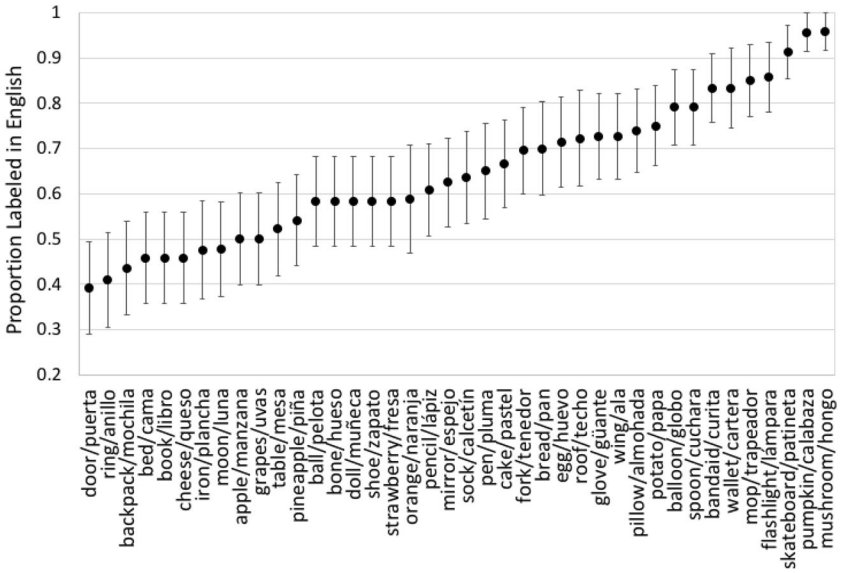

Fig. 1 Distribution of lexical accessibility norms of the 40 target items. Error bars are standard deviations calculated for a binomial distribution

(2) a Spanish-only condition in which participants were not permitted to codeswitch into English, and (3) an English-only condition. These conditions allowed us to investigate not only the trials on which participants were most likely to codeswitch, but also initiation latencies and speaking durations as a consequence of whether or not speakers had an option to codeswitch.

Participants A total of 128 Spanish-English bilinguals from the introductory psychology subject pool at the University of California, Riverside participated for course credit. Thirty-five participants each were included in the Codeswitch-permitted and Spanish-only conditions and 58 participants were included in the English-only condition. Other participants were run but excluded from analyses: One participant each were removed from the Codeswitch-permitted and Spanish-only conditions due to equipment failure, and two participants were removed from the English-only condition, one for failing to complete the task in the allotted amount of time and the other because responses on the language history questionnaire were not recorded. Additional participants who were run but excluded on the basis of task performance are discussed in the data-cleaning section below. Demographic characteristics and self-reported codeswitching behavior are shown in Tables 1 and 2.

Sample size choices were made prior to data collection. We decided to include 35 participants in each of the two Spanish conditions and a larger number, 60 participants, in the English condition, because we expected effect sizes to be smaller in the English condition due to the greater English language proficiency of our sample. Only 58 participants were included in the English condition due to two unexpected participant exclusions that were not evident until after data collection had concluded.

Procedure Participants sat in front of a computer screen. Participants were assigned to the Codeswitch-permitted, Spanish-only, or English-only condition. In all three conditions, participants watched a short video in which a bilingual research assistant described the task. In the Codeswitch-permitted condition, instructions were presented mostly in Spanish, but the speaker occasionally and naturalistically codeswitched into English. In addition to the implicit codeswitching cue provided by the video, the instructions explicitly told participants to complete the task in Spanish but that they could use English if they wanted to. Participants were never instructed on which trials to codeswitch, or where in the sentence to codeswitch, if they chose to do so. In the Spanish-only condition, the instructions were given fully in Spanish, and participants were explicitly told to use only Spanish in their responses. In the English-only condition, the instructions were given fully in English, and participants had no reason to believe the task had anything to do with Spanish. All three videos featured the same speaker. All materials and data are available online at https://osf.io/uchpv/.

Participants were told that they would be answering questions about the pictures on the screen. In the critical test trials participants were asked to describe the location of one image relative to another. In the experiment instructions, participants were given a matrix frame sentence with which to respond. On these test trials, participants heard the prompt (translated into English) "Tell me about the X." Participants were instructed to give their responses in the form "The X is above/below the Y." This procedure is illustrated in Fig. 2. Participants were encouraged to answer prompts in this frame for two reasons. First, uniformity in the sentence structure of responses would allow us to better assess the contribution of our key manipulation, the lexical accessibility of target items, to codeswitching and the time course of sentence production.

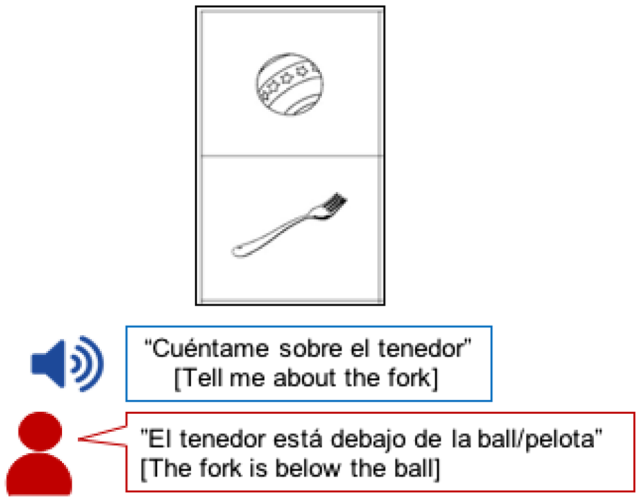

Fig. 2 Illustration of an experimental trial. Participants saw the two pictures of the ball and the fork, heard an auditory prompt, and provided a verbal response

Second, in the sentence frame, the item that was named in the prompt would be the first item that the participant named in their response. By having participants begin their utterances with a discourse-given(as opposed to discourse-new) noun, participants should be able to begin speaking quickly, before they have planned the second noun, increasing our ability to detect consequences of our critical manipulation during online sentence production on the second noun in the utterance (Griffin, 2001).

Participants completed 40 test trials. In addition, participants completed 60 filler trials in which they described either the number of items or presence of specific items in the images (e.g., *Hay dos manzanas*/There are two apples). The filler trials were included to ensure that participants did not habituate to a single syntactic form. The trials were presented such that there were always either one or two filler trials between test trials. In the two Spanish conditions (Codeswitch-permitted and Spanish-only), all task prompts were presented in Spanish. The English-only condition was identical, except that all materials were recorded in English. The same bilingual speaker recorded the audio prompts in all conditions. After participating in the task, participants filled out a language history questionnaire.

We coded each utterance for the presence or absence of a codeswitch in the second noun phrase, most often between the determiner (*el* or *la*) and the noun (e.g., between the determiner *la* and the noun *pelota* in the example trial in Fig. 2 to yield *debajo la ball*) but occasionally before the determiner (e.g., *debajo the ball)*. Again, speakers were never instructed where to codeswitch, but codeswitches virtually never occurred in any other location in the utterance, which reflects both how naturalistic our elicited codeswitches were and that naturalistic codeswitching occurs in specific and predictable locations in an utterance. In addition to codeswitching, initiation latencies (time from the sentence prompt offset to the beginning of the speaker's utterance) and speaking durations, defined as the duration from the speaker's utterance onset through the determiner before the second noun and until (but not including) the target noun, were measured to provide finer-grained detail about the time course of sentence production. Figure 3 illustrates how we defined initiation latencies and speaking durations.

Results

Data-cleaning procedure

In all conditions, participants were excluded if they had 15 or more error trials out of 40 critical trials. These error trials included utterances that did not follow the instructed sentence frame, that is, participants did not produce a sentence such as, *El tenedor está debajo de la ball/pelota* [The fork is below the ball]. Error trials also included trials on which participants failed to name the target item, used a label other than the normed label, or if the utterance contained a revision or restart. Nine participants were excluded in the codeswitch-permitted condition, 12 from the Spanish-only condition, and ten from the English-only condition for producing too many error trials.

Next, despite being instructed not to use any English, some participants in the Spanish-only condition codeswitched. Though this behavior reflects how natural codeswitching is in this population and how naturalistic our elicited codeswitches were, it provided a challenge for data analysis. Eight participants were excluded for codeswitching more than twice during the experiment, and all remaining codeswitched trials in the Spanish-only condition (11 trials) were removed. Rather than eliminate all participants who ever codeswitched, we chose to only exclude individuals who codeswitched more than twice, to minimize any systematic population differences between those participants included and excluded across conditions.

For participants included in the final analyses, all error trials were removed. The percentage of valid trials included in the final analyses was 85.2% for the Codeswitch-permitted condition, 85.0% in the Spanish-only condition, and 90.3% in the English-only condition.

Data analysis

All inferential statistics reported were performed in R (version 3.5.1) using mixed-effects logistic or linear regression (glmer or lmer) analysis (Baayen, Davidson, & Bates, 2008) with the lme4 package (version 1.1-18-1) (Bates, Maechler, Bolker, & Walker, 2015). We used the maximal crossed random effects structure for participants (p) and words (w) in which the model converged (Barr, Levy, Scheepers, and Tily, 2013). When the full model did not converge, we report the most complete model that did converge. When two different models of equal complexity were the most complete model that converged, we report the model with the better model fit. All models

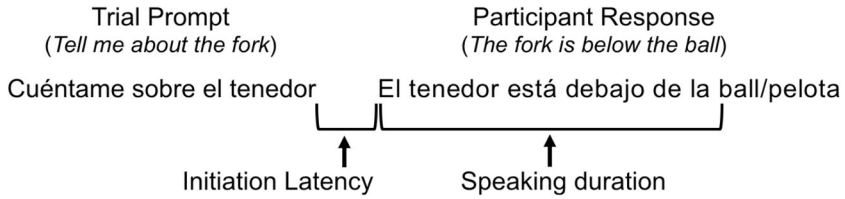

Fig. 3 Definitions of initiation latencies and speaking durations in the elicited utterances

contained random intercepts, and all random slopes that were included in the model are indicated in each table in which results are reported. For the linear mixed effect regression models, p-values were generated using the lmerTest package, version 3.1-1 (Kuznetsova, Brockhoff, and Christensen, 2017) using Satterthwaite's method.

Codeswitching results

The average codeswitching rate in the Codeswitch-permitted condition was 6.9% (standard deviation over participants = 7.4%). To determine the effect of English lexical accessibility on codeswitching, we performed a mixed-effects logistic regression analysis predicting codeswitching behavior (1 or 0) with lexical accessibility and Spanish proficiency. The lexical accessibility measure (proportion of labels produced in English) and Spanish proficiency measure were z-scored. The model predicted either no codeswitching (reference group) or codeswitching for target words. The model is summarized in Table 3. Lexical accessibility significantly predicted codeswitching on target words such that participants were more likely to codeswitch on items with high English lexical accessibility. Neither participants' self-reported Spanish proficiency nor the interaction between Spanish proficiency and lexical accessibility significantly predicted codeswitching behavior. The removal of high-influence items did not change the effect.[1] To allow for visualization of the relationship between English lexical accessibility and codeswitching, the codeswitching proportion by items is shown in Fig. 4, though the inferential statistics were computed with logistic mixed effect models over the trial data.

As predicted, participants were more likely to codeswitch from Spanish into English on

items that were a priori determined to be more lexically accessible in English. This finding is consistent with the idea that codeswitching behavior may in part be motivated by aspects of a word's lexical accessibility. When item labels compete, and when speakers are permitted to codeswitch, they are more likely to do so when the other-language label is more lexically accessible.

As an exploratory analysis, we investigated the role of lexical gender in codeswitching choices, and the locus of codeswitching in utterances relative to the determiner (e.g., the, a/an) in speaker utterances, which are gender-marked in Spanish. Spanish nouns are assigned either a masculine or feminine lexical gender, and the determiner that precedes a noun must agree with the gender of the noun. For example, the determiner "the" in Spanish is "el" when it precedes a masculine noun and "la" when it precedes a feminine noun. Given the challenges associated with selecting items that participants could name, and name with a single label in both languages, we did not consider lexical gender when selecting our items. However, 23 of our target items were feminine and 17 were masculine. A total of 85 trials in the Codeswitch-Permitted condition yielded codeswitches. Of these, 56 were feminine and 29 were masculine, but a Fisher's exact test shows that this ratio is not significantly different from the overall ratio of feminine to masculine items (p = 0.42), so speakers did not seem to codeswitch more often on trials with masculine or feminine target nouns.

Next, we found that on 84.7% of codeswitched trials (72 trials), speakers codeswitched between the determiner and the noun. This means that participants produced a determiner in Spanish followed by a noun in English. This locus of codeswitching is consistent with previous findings that speakers often codeswitch within a noun phrase, between the determiner and noun (Valdés Kroff, 2016). On 10.6% of codeswitched trials (nine trials), participants produced both the determiner and noun in English. On the remaining 4.7% of trials, participants either produced both a determiner in Spanish followed by a determiner in English (three trials) or no determiner (one trial). In our laboratory-based sample, speakers overwhelmingly codeswitched between the determiner and target noun, consistent with codeswitching patterns reported in analyses of naturalistic corpora.

We then investigated gendered determiner use on codeswitched trials. We investigated whether the present pattern of codeswitching is consistent with previous work. Specifically, prior work has found an asymmetry in determiner usage, such that participants often use the masculine determiner before nouns that have either masculine or feminine translation equivalents in Spanish but use the feminine determiner only with nouns whose translation equivalent in Spanish is feminine (Valdés Kroff, 2016). When eliminating the three utterances that contained determiners in both English

[1] To ensure that a few items were not disproportionally driving the observed effect, we checked for high-influence items using the influence.ME (version 0.9-9) R package (Nieuwenhuis, Manfred, & Pelzer, 2012). Three items were identified as high-influence items (iron, mushroom, and bandaid), and the pattern of results did not change when these items were excluded.

Table 3 Results of mixed-effects logistic model predicting codeswitches in the Codeswitch-permitted condition from lexical accessibility (z-scored)

	Coefficient	SE	z	*p*	Random slope
Intercept	-4.21	0.41	-10.21	< .001	
Lexical accessibility	1.62	0.26	6.24	< .001	
Spanish proficiency	-0.013	0.36	-0.104	0.97	w*
Spanish proficiency × lexical accessibility	-0.10	0.21	-0.48	0.6	

*In this model and in all other models we report, we indicate the inclusion of random slopes for participants (p) and words (w). All models contained random intercepts for participants and words

and Spanish, we found that when the target noun was masculine (28 trials) speakers produced the noun with the masculine determiner on 20 trials and the feminine determiner on eight trials. However, when the target noun was feminine (54 trials), speakers produced the masculine determiner on 27 trials, the feminine determiner on 17 trials, an English determiner on nine trials, and no determiner on one trial. The observed data is consistent with previous results that found that feminine codeswitched nouns are often used with masculine determiners but not vice versa; 28% of masculine nouns were used with feminine determiners, but 50% of feminine nouns (61% when excluding English or missing determiners) were used with masculine determiners. A Fisher's exact test confirms that the pattern of gender-matched determiner usage use varied between masculine and feminine codeswitched target nouns ($p < 0.01$) such that speakers were more likely to use a masculine determiner with a feminine target item than feminine determiner with a masculine target noun. Though the sample size is small, and the present study was not designed to investigate the effect of lexical gender on codeswitching, we found patterns in our laboratory-based picture description study that are consistent with the patterns of determiner use in other studies, including corpus analyses of codeswitching behavior.

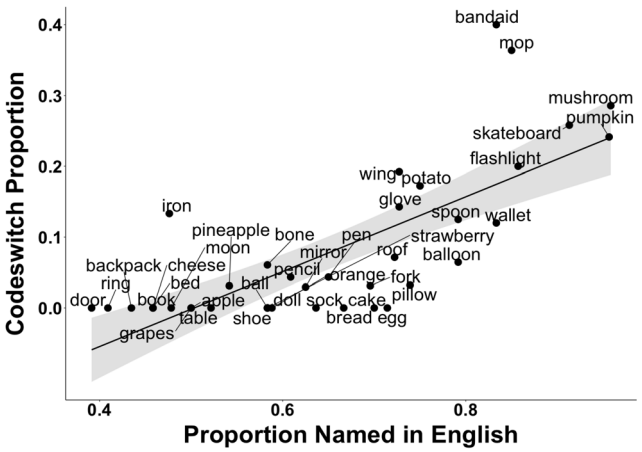

Fig. 4 Scatterplot of codeswitching proportion and lexical accessibility for the 40 target words (r (38) = .75, $p < .001$). Transforming the proportion of items named in English to log odds to be more consistent with the mixed effects logistic regression inferential statistics (and due to the non-linear relationship between the variables) yields a similar correlation (r (38) = .77, $p < .001$)

Next, we turn to analyses of the time course of sentence production. We found that codeswitching from Spanish to English is associated with trials on which the English label for a target object is more lexically accessible, presumably leading to greater lexical competition and more intrusions from the English label. If codeswitching is indeed associated with trials with increased lexical competition (in other words, a priori more difficult trials) then we should see additional markers of this competition in the time course of sentence production even when speakers chose not to codeswitch, or when codeswitching was not permitted.

Initiation latencies

Initiation latency was defined as the time between the end of the audio prompt and the beginning of the participant's utterance. The initiation latencies across conditions is shown in Table 4. Table 5 shows the results of a linear mixed effect regression model predicting initiation latencies by condition, Spanish proficiency and English lexical accessibility. Initiation latencies were faster in the English condition than in either of the two Spanish conditions, which is not surprising given the self-reported English dominance and higher English proficiency of most of the participants. Higher Spanish proficiency was also associated with shorter initiation latencies, but as the condition-by-proficiency interaction suggests, only in the two Spanish conditions, not in the English-only condition. There was no difference in initiation latencies between either of the two Spanish conditions. Finally, lexical accessibility did not contribute to initiation latencies in any condition. The lack of a difference in initiation latencies by condition or by codeswitching is not surprising. The task was designed such that participants would begin their utterances with the noun given in the prompt, to encourage speakers to begin their utterances quickly without fully planning the rest of the sentence. The logic for this design was that participants could

[0] From the model in Table 9 it may appear that English lexical accessibility did not interact with codeswitching, such that speaking durations for codeswitched utterances was also associated with English lexical accessibility. However, there were very few codeswitched utterances relative to unswitched utterances, and a follow-up model of only the codeswitched utterances showed no effect of English lexical accessibility or Spanish Proficiency. This model is included in the Supplemental Online Materials.

Table 4 Initiation latencies (in seconds) for codeswitched and non-codeswitched trials in both the Spanish-only and Codeswitch-permitted conditions

	Codeswitch-permitted	Spanish-only	English-only
Codeswitch	1.15 (0.72)	-	-
No codeswitch	1.04 (0.36)	1.08 (0.50)	0.92 (0.33)

begin speaking with minimal planning in order to maximize the possible influence of lexical properties of the target noun later in the utterance (e.g., Griffin, 2001). In essence, the task was designed to keep initiation latencies constant across conditions to maximize our ability to detect effects of our manipulation later in the utterance, on speaking durations.

Next, Table 6 shows the results of a linear mixed effect regression model predicting initiation latencies by Spanish proficiency, English lexical accessibility and whether or not the utterance contained a codeswitch in the Codeswitch-permitted condition. There was no difference in initiation latencies between utterances that were or were not codeswitched when codeswitching was permitted. Consistent with the analysis in Table 5, there was a marginal effect of Spanish proficiency on initiation latencies, and English lexical accessibility did not contribute to initiation latencies.

Speaking durations

Speaking duration was defined as the time from the beginning of an utterance to before (but not including) the utterance of the target word for each trial. The speaking durations associated with each condition are shown in Table 7 and Fig. 5. Visually, codeswitching had a clear effect on speaking durations, which is confirmed below with inferential statistics. Speaking durations in the Codeswitch-permitted condition on trials in which participants chose to codeswitch were longer than when speakers chose not to codeswitch. The challenge in interpreting the increase in speaking duration is understanding the locus of the increase. Codeswitched trials were disproportionately selected from the most difficult trials (more competition from the English label in a Spanish sentence-production task). Therefore, even in the absence of a cost associated with the codeswitch itself, we might expect codeswitched utterances to have longer speaking durations. Alternately, slowing may indeed be due to costs associated with the codeswitch itself. These two alternatives are not mutually exclusive. We return to the interpretation of the very long speaking durations associated with codeswitched utterances in the *General discussion*.

Table 8 shows the results of a linear mixed-effect regression model predicting speaking durations of utterances that were not codeswitched with multiple predictors: Condition, Spanish Proficiency and English lexical accessibility. As with initiation latencies, speaking durations were overall faster in English than in both Spanish conditions. Again, given the self-reported English dominance and higher self-reported English proficiency of our participants, this result is not surprising. There was no difference in speaking duration across the two Spanish conditions. Finally, greater Spanish proficiency was also associated with shorter speaking durations, but as with initiation latencies, only in the two Spanish conditions.

Table 8 also shows the key comparison of interest, the effect of lexical accessibility on speaking durations. If lexical competition contributed to codeswitching choices, then we should see evidence of this lexical competition even on trials that were not codeswitched in the form of longer speaking durations. Consistent with this prediction, English lexical accessibility contributed to speaking durations such that

Table 5 Results of linear mixed-effects linear regression model predicting initiation latencies (seconds) with experiment condition, self-reported Spanish proficiency (z-scored), and English lexical accessibility. Experiment conditions were Helmert contrast coded, such that the first contrast compares the English-only condition to the two Spanish conditions (Codeswitch-permitted and Spanish-only) and the second contrast compares the two Spanish conditions to each other

	Coefficient	SE	*t*	*p*	Random slope
Intercept	0.031	0.048	0.64	0.5	
Contrast 1 (English-Only vs. Both Spanish)	0.053	0.027	1.97	< 0.06	w
Contrast 2 (CS-Permitted vs. Spanish-only)	0.043	0.047	0.92	0.4	w
Spanish proficiency	-0.085	0.037	-2.27	< 0.05	w
Lexical accessibility	0.039	0.030	1.31	.2	p
Contrast 1 × Spanish proficiency	-0.054	0.022	-2.42	< 0.05	w
Contrast 2 × Spanish proficiency	-0.033	0.045	-0.73	0.5	w
Contrast 1 × lexical accessibility	0.010	0.014	0.69	0.5	
Contrast 2 × lexical accessibility	-0.0004	0.019	-0.02	0.98	
Spanish proficiency × lexical accessibility	-0.012	0.016	-0.80	0.43	

Table 6 Results of linear mixed-effects linear regression model predicting initiation latencies (seconds) with self-reported Spanish proficiency (z-scored), English lexical accessibility, and whether or not the utterance contained a codeswitch in the Codeswitch-permitted condition

	Coefficient	SE	*t*	*p*	Random slope
Intercept	0.016	0.078	0.21	0.8	
Spanish proficiency	-0.12	0.061	-1.93	< 0.07	w
Lexical accessibility	0.090	0.058	1.55	0.1	p
Codeswitch	-0.14	0.17	-0.84	0.4	w, p
Spanish proficiency × lexical accessibility	-0.027	0.033	-0.82	0.4	
Spanish proficiency × codeswitch	0.045	0.22	0.20	0.8	w
Lexical accessibility × codeswitch	0.12	0.18	0.68	0.5	p
Spanish proficiency × lexical accessibility × codeswitch	0.18	0.21	0.87	0.4	

speaking durations were longer when the target item was more lexically accessible in English. This effect persisted in both Spanish conditions but not in the English-only condition. This effect means that not only was English lexical accessibility associated with greater incidence of codeswitching but was *also* associated with longer speaking durations when speakers did not codeswitch. This result is consistent with the co-activation of a bilingual speaker's two languages during sentence production (Kroll, Bobb, & Hoshino, 2014; Marian & Spivey, 2003). We find evidence of production difficulty proportional to the lexical competition that we would expect across item labels if both languages were active during sentence production. The effect of English lexical accessibility on speaking durations did not interact with Spanish proficiency, suggesting that these longer speaking durations were not being driven by speakers with lower Spanish proficiency. As in previous analyses, this pattern of results did not change when high-influence items were identified and excluded.

The results in Table 8 illustrate a key prediction of the present study, that speakers took longer to produce utterances in which the target noun label was more lexically accessible in English, even when the utterance was not codeswitched. Tables 9 and 10 further illustrate this key finding by showing the effect of lexical accessibility separately in the Codeswitch-permitted and Spanish-only conditions. Table 9 shows the results of linear mixed effect regression models predicting speaking duration from the target noun's English lexical accessibility, the participant's Spanish proficiency, and whether or not the utterance contained a codeswitch, for the trials in the Codeswitch-permitted condition. This model shows that, as is evident in Fig. 5 and the means in Table 7, speaking durations were indeed longer when the utterance contained a codeswitch. Table 10 shows the results of a linear mixed effect regression models predicting speaking duration from the target noun's English lexical accessibility and the participant's Spanish proficiency in the Spanish-only condition. Crucially, Tables 9 and 10 show similar effects of English lexical accessibility and Spanish proficiency. In Spanish utterances that did not contain a codeswitch when permitted, and in utterances in which codeswitching was not permitted, speaking durations were longer when producing items that were more lexically accessible in English.[2] Again, though Spanish proficiency contributed to speaking durations, Spanish proficiency did not interact with lexical accessibility, suggesting that these longer speaking durations were not being driven by speakers with lower Spanish proficiency. The simple by-item correlations between speaking duration and English lexical accessibility are shown in Fig. 6. These figures are for illustrative purposes; the inferential statistics were computed using mixed effect linear regression models.

To gain one final insight into the locus of our lexical accessibility, we investigated the effect of lexical accessibility on speaking durations in English. If the Spanish labels interfere with the English labels, we would expect the opposite direction of the effect as in the two Spanish conditions, that items which are more lexically accessible in English should have faster speaking durations. However, as shown in Table 11 and the graph in Fig. 6, there is no reliable effect of English lexical accessibility on speaking durations. Given the overall higher English proficiency of our

Table 7 Speaking durations (s) for codeswitched and non-codeswitched trials in both the Spanish-only and Codeswitch-permitted conditions

	Codeswitch-permitted	Spanish-only	English-only
Codeswitch	7.99 (2.65)	-	-
No codeswitch	3.57 (0.58)	3.45 (0.99)	1.56 (0.26)

[2] From the model in Table 9 it may appear that English lexical accessibility did not interact with codeswitching, such that speaking durations for codeswitched utterances was also associated with English lexical accessibility. However, there were very few codeswitched utterances relative to unswitched utterances, and a follow-up model of only the codeswitched utterances showed no effect of English lexical accessibility or Spanish Proficiency. This model is included in the Supplemental Online Materials.

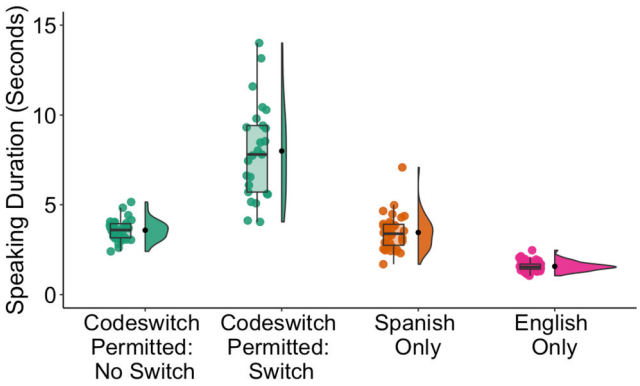

Fig. 5 Speaking durations by condition. In the Codeswitch-permitted condition, both codeswitched and not codeswitched utterances are shown. Each point is a participant mean. Raincloud plot code adapted from Allen et al. (2019)

sample and the fact that speakers produced English utterances very quickly, floor effects may have impeded our ability to detect any reliable effects. However, this null result provides some limited evidence that words that were more lexically accessible in English were not simply overall harder to produce, possibly for reasons unrelated to their relative accessibility across languages. Further, the absence of an observed effect of Spanish lexical accessibility on English speaking duration may suggest a limit to interference associated with cross-language activation, with the less dominant language more susceptible to lexical intrusions from the more dominant language than vice versa.

General discussion

In a novel sentence-production paradigm, we investigated whether codeswitching behavior could be explained by the relative lexical accessibility of translation equivalent words across Spanish-English bilingual speakers' two languages. We found that lexical accessibility predicted codeswitches in the sentence-production task, such that the more lexically accessible a word was in English, the higher the likelihood of codeswitching from Spanish into English. We also found that lexical accessibility predicted speaking durations. The more accessible a word was in English, the longer the speaking durations in a Spanish utterance, even when the speaker did not codeswitch. Put simply, trials on which speakers should experience more lexical competition or intrusions from the other-language label were more likely to contain codeswitches and took longer to produce even when they did not contain codeswitches. These findings have a number of implications for our understanding of why speakers codeswitch, and the cognitive processes that underlie codeswitching and lexical selection more generally, in both mono- and multi-lingual speakers.

An important finding from the present study is that we were able to predict codeswitches at the item level in the sentence-production task. Specifically, we found evidence consistent with lexical competition between translation equivalents. Codeswitches occurred on trials that were a priori predicted to be "harder" as a consequence of the greater lexical competition from the English language label. That a priori harder trials were trials on which speakers were more likely to codeswitch suggests how codeswitching can be situated in a difficulty-reduction framework. A word may be harder to recall and plan when it is more accessible in the language a bilingual speaker is not currently using, due to the increased lexical competition from the other-language label, analogous to how a near-synonym prime interferes with picture naming (Jescheniak & Schriefers, 1998). A speaker may alleviate this difficulty by simply producing rather than inhibiting the more accessible label, thus codeswitching. We then interpret codeswitching as a tool available to the bilingual speaker to

Table 8 Results of linear mixed-effects model predicting speaking duration (seconds) with experiment condition, self-reported Spanish proficiency (z-scored), and lexical accessibility. Experiment conditions were Helmert contrast coded, such that the first contrast compares the English-only condition to the two Spanish conditions (Codeswitch-permitted and Spanish-only) and the second contrast compares the two Spanish conditions to each other

	Coefficient	SE	*T*	*p*	Random slope
Intercept	0.26	0.082	3.13	< 0.01	
Contrast 1 (English-Only vs. Both Spanish)	0.67	0.044	15.33	< 0.001	w
Contrast 2 (CS-Permitted vs. Spanish-only)	-0.039	0.080	-0.45	0.7	w
Spanish proficiency	-0.15	0.058	-2.53	<0.05	w
Lexical accessibility	0.19	0.057	3.47	< 0.001	p
Contrast 1 × Spanish proficiency	-0.092	0.035	-2.64	<0.01	w
Contrast 2 × Spanish proficiency	-0.074	0.068	-1.08	0.3	w
Contrast 1 × Lexical accessibility	0.10	0.027	3.74	< 0.001	
Contrast 2 × Lexical accessibility	-0.020	0.046	-0.43	0.7	
Spanish proficiency × lexical accessibility	0.0014	0.027	0.052	0.96	

Table 10 Results of linear mixed-effects model predicting speaking duration (seconds) from lexical accessibility (z-scored and centered) and Spanish proficiency (z-scored) in the Spanish-only condition

	Coefficient	SE	*t*	*p*	Random slope
Intercept	0.040	0.18	0.22	.82	
Lexical accessibility	0.36	0.11	3.35	< .005	p
Spanish proficiency	-0.34	0.16	-2.07	< .05	w
Lexical accessibility × Spanish proficiency	-0.11	0.081	-1.30	.2	

alleviate the increased lexical competition associated with language co-activation, when the conversational context permits such a choice. Rather than viewing codeswitching as a choice that is itself associated with increased language production difficulty, we view codeswitching as a behavior that arises from the pressures that are inherently part of the language production process. Additional evidence for the view that codeswitching may be a tool available to the speaker to ease production difficulty comes from our analysis of speaking durations. We observed longer speaking durations for items that were highly lexically accessible in English but were not codeswitched, in both the Codeswitch-Permitted and the Spanish-Only conditions. We find that that same predictor, English lexical accessibility, contributed to both increased codeswitching and a second conventional marker of planning difficulty, increased speaking durations, even when the utterance did not contain a codeswitch. These two consequences of lexical accessibility suggest that both outcomes, codeswitching and slowed speaking rate, may be markers of language production difficulty associated with lexical competition. Of course, codeswitching need not have a single cause. However, by understanding codeswitching in the context of language production and lexical competition processes we can better understand when and why speakers codeswitch.

The question that then arises is, if a codeswitch is a choice that speakers make in response to harder-to-plan utterances in order to make planning easier, why should codeswitched utterances exhibit such a dramatic increase in speaking duration? While the present work cannot definitively answer that question, we have a number of hypotheses. One possibility is that speakers misgauge the difficulty of a codeswitch, perhaps such that recalling a lexical item in a codeswitch and the articulatory and motoric demands of executing the codeswitch are misaligned. For example, phonological encoding processes that occur after lexical selection may underlie the dramatic increase in speaking duration (e.g., Bock & Levelt, 1994). Another possibility is the potential contribution of task demands specific to our lab-based task, especially because speaking durations of the length we observe (3.5 s longer than unswitched utterances) are not characteristic of natural speech, though codeswitches are indeed associated with a slowed speaking rate in naturalistic corpora (Fricke, Kroll, & Dussias, 2016). Perhaps speakers strategically approached the task in manner unlike typical speech and interpreted the task as prioritizing producing utterances entirely in Spanish over maintaining fluency. Anecdotally, speakers occasionally paused, sometimes for seconds, between the determiner (*el* or *la*) and the target noun prior to a codeswitch, as if they were trying to recall a word in Spanish but could not, and eventually codeswitched. Interestingly, we do not see these very long speaking durations in the Spanish-only condition. Perhaps aspects about the codeswitching task itself may underlie these long durations, in the form of meta-task strategies or task demands. For example, task demands associated with the

Table 9 Results of linear mixed-effects linear regression model predicting speaking durations (seconds) with self-reported Spanish proficiency (z-scored), English lexical accessibility, and whether or not the utterance contained a codeswitch in the Codeswitch-permitted condition

	Coefficient	SE	*t*	*p*	Random slope
Intercept	-0.19	0.13	-1.49	0.1	
Spanish proficiency	-0.17	0.09	-1.87	< 0.08	w
Lexical accessibility	0.38	0.10	3.89	< 0.001	p
Codeswitch	4.80	0.65	7.34	< 0.001	w, p
Spanish proficiency × lexical accessibility	-0.013	0.054	-0.25	0.8	
Spanish proficiency × codeswitch	0.30	0.60	0.49	0.6	W
Lexical accessibility × codeswitch	-0.57	0.40	-1.41	0.2	P
Spanish proficiency × lexical accessibility × codeswitch	0.007	0.32	0.022	0.98	

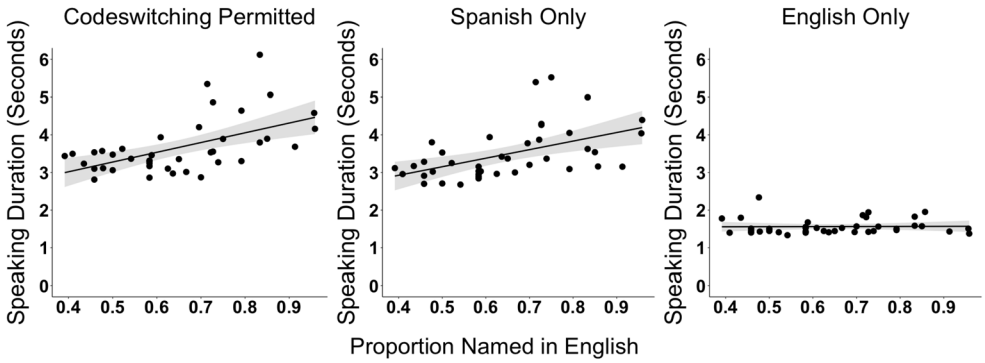

Fig. 6 Scatterplot of speaking duration and English lexical accessibility in the Codeswitch-permitted condition, when participants did not codeswitch ($r(38) = .55$, $p < .001$), the Spanish-only condition ($r(38) = .51$, $p < .001$), and the English-only condition ($r(38) = .02$, $p = .9$)

opportunity to codeswitch, or trial-by-trial alternations in codeswitching choices, may have contributed to task difficulty and long speaking times independently of lexical retrieval, (e.g., Ong et al., 2019).

A final speculation is that the increase in speaking durations for codeswitched items may relate to the tendency for bilingual speakers to experience more tip-of-the-tongue(TOT) states, where speakers experience difficulty retrieving a word that they indeed know (Gollan & Silverberg, 2001; Gollan, Ferreira, Cera, & Flett, 2014a). Some theories of TOT states in bilingual speakers explicitly relate the increase in TOT states to language co-activation, especially interference from a translation equivalent during lexical selection (Gollan et al., 2014a). The fact that speakers had an option to, or not to, codeswitch in our Codeswitch-Permitted condition, and that items were selected to observe competition between translation equivalents across languages, the present task may have put speakers in a context in which they were particularly prone to TOT states. We selected experimental items that specifically did not have highly accessible synonyms, so speakers could not substitute a different label or "talk around" a word they were having difficulty accessing a label, unlike many natural speech contexts. Again, we did not see these long speaking durations in the Spanish-only condition, which could suggest that language co-activation or the processes that are associated with lexical selection vary when speakers know that codeswitching is an option, either more generally in conversational contexts, or more narrowly specific to the demands of our task. There are multiple possible explanations for the very long speaking durations that we observed in the codeswitched utterances and the examples presented here are certainly not exhaustive, nor are they mutually exclusive. Future work is necessary to better understand how online planning and articulatory processes may interact with specific task demands our participants encountered in this task.

Table 11 Results of linear mixed-effects model predicting speaking duration from lexical accessibility (z-scored) in the English-only condition

Speaking duration	Coefficient	SE	*t*	*P*	Random slope
Intercept	0.012	0.046	0.26	.8	
Lexical accessibility	0.004	0.033	0.12	.9	p

An important note that may aid in the interpretation of the long speaking durations to codeswitched items is that experimental items were such that our population of speakers could indeed generate labels in both languages. An obvious flaw with the experimental design would be selecting items that speakers could not name in both languages. For this reason, in our item selection task, we selected study items that participants drawn from the same population could name, and that different speakers consistently named with the same label. In both English and Spanish, average name agreement of our items was over 90% (a failure to name an item contributed to label disagreement). Though our experimental design required that separate groups of participants were included in the item selection and sentence-production tasks, it was through our item selection procedure we aimed to reduce the likelihood that speakers not knowing the Spanish label of an item contributed to codeswitching choices or the time course of codeswitched utterances.

We aim to interpret our results within an existing difficulty-reduction theoretical framework for understanding language production processes. We want to emphasize that it is *not* the case that casting codeswitching in a difficulty-reduction framework would mean that codeswitched utterances should have shorter speaking durations than unswitched utterances. This is because codeswitching was associated with more difficult *contexts*. Slowed speech preceding codeswitches is consistent with corpus analyses of codeswitches in spontaneous speech (Fricke et al., 2016). The key question that our work raises is the extent to which the slowing associated with codeswitched trials indexes the cognitive demands of an upcoming switch, or the fact that words and phrases that are inherently more difficult to plan and articulate are more likely to be codeswitched. The distinction between these two different loci of difficulty is theoretically important. While either or both may act in a single context, the distinction is crucial for

interpreting codeswitching costs in past and future research. We still have a great deal to learn about the processes that underlie codeswitches and the processes that underlie lexical recall, planning, and articulation during sentence production. Future work that aims to understand the online sentence-production processes may shed light on the etiology of the long codeswitch speaking durations we observe.

Many of the findings in the present work are broadly consistent with existing findings in the literature, such that this project provides converging evidence for many key phenomena regarding codeswitching and bilingual language production. First, and in the most big-picture sense, the present results provide additional evidence for the co-activation of a speaker's two languages during language production (Kroll, Bobb, & Wodniecka, 2006; Thierry & Wu, 2007). In the present Spanish sentence-production task, participants' speaking rate slowed and speakers' likelihood of codeswitching increased when target words were more accessible in English, that is, when cross-language interference would have been stronger assuming the co-activation of both languages. Our results are also consistent with studies that find evidence of interference across a speaker's languages, possibly as a consequence of cross-language activation and subsequent inhibition of one of the competing labels (Bijeljac-Babic, Biardeau, & Grainger, 1997; Ivanova & Costa, 2008). Further, previous work also explicitly highlights similarities between lexical competition both within and across a bilingual speaker's languages, as we do in this work. Runnqvist, Strijkers, Alario, and Costa (2012) found evidence of cumulative semantic interference (increased naming latencies to members of the same category when speaker name series of pictures) whether the objects were named in the same or different languages. This interference is interpreted as evidence that during language production, item labels compete both within and across languages, which is consistent with our finding that translation equivalents compete across a speaker's two languages. Finally, the present work is certainly not the first to implicate lexical accessibility in codeswitching choices, but rather ease of lexical access is frequently identified as a possible contributor to codeswitching. Previous work has found that speakers are more likely to switch languages when the other-language label is more lexically accessible, and that these codeswitches incur a smaller cognitive cost (de Bruin, et al., 2018; Gross & Kaushanskaya, 2015; Kleinman & Gollan, 2016). Given the consistencies between our findings and other findings in the literature, we conclude that cross-language competition is likely not an artifact of object-namingtasks that are often employed in the literature. Cross-language competition may indeed be a feature of naturalistic multiword sentence-production contexts. However, an important caveat is that our failure to detect consequences of Spanish lexical accessibility in the English sentence-production task suggests that cross-language activation or interference may be reduced with greater language proficiency or in some language contexts.

Despite many consistencies between the existing literature and the present work, important inconsistencies exist as well. The asymmetry we found in the present study – greater intrusions from the more dominant language (English) onto the less dominant language (Spanish)– suggests a different underlying mechanism than studies that find the opposite result, that language switching from the more dominant to the less dominant language is *easier* than switching from the less to more dominant language (Costa & Santesteban, 2004; Meuter & Allport, 1999). We ascribe this inconsistency to differences in the task. Rather than a task that explicitly required participants to alternately inhibit languages, the present task investigated incidental intrusions from one language to the other. Thus, fewer intrusions from the weaker to the stronger language is consistent with the expertise literature, where proficiency in a task or behavior is associated with greater task automaticity and fewer intrusions from outside stimuli (Anderson, 1982; Beilock, Wierenga, & Carr, 2002; Kimble & Perlmuter, 1970). Some previous findings have also found an effect inconsistent with our results such that speakers may be more likely to switch languages when the *same*-language label is more lexically accessible (Gollan & Ferreira, 2009; Gollan, Kleinman, and Wierenga, 2014b) rather than the *other*-language label, as we find in the present work. Again, we interpret this inconsistency as stemming from experimental design choices. Gollan and colleagues defined item difficulty at the item level, such that pictures that are hard to name in one language are also hard to name in speakers' other language, rather than quantifying item difficulty as label competition across languages as we did in the present work. Both of these sources of difficulty likely contribute to language production processes. Different item-selection procedures and different experimental manipulations have allowed researchers to investigate related but distinct phenomena than what we investigate here. We believe these task-based discrepancies across studies further illustrate how different experimental methodologies can complement and inform each other to increase our understanding of important cognitive processes. Finally, though we found evidence consistent with cross-language competition here, these results are not inconsistent with other potentially mutually compatible accounts of bilingual language production such as the frequency-lag account (Gollan et al., 2008). Future work will likely reconcile multiple factors that contribute to sentence-production choices and the time course of language production in bilingual speakers.

The present study has implications beyond codeswitching by bilingual speakers and sheds light on language-production processes shared by all speakers. We observed consequences of lexical competition in online sentence production, which provides additional support for the idea that the multiple labels for a single entity indeed compete during language production (Dell, 1986; Levelt et al., 1991), and that this competition may be associated with increased production difficulty for the speaker. Speakers often have multiple possible labels for a single entity even in a single language, as with synonyms such as

"couch" and "sofa" or other alternatives such as "it" or "the blue thing." Lexical competition between labels is an inherent part of language production. Bilingualism may be one arena in which we can investigate this lexical competition, with broader implications for the processes that underlie language production regardless of the language or languages being spoken.

To conclude, if language production is defined as a series of choices made by the speaker, the present work illustrates why investigating language production is important, and how gaining a better understanding of the processes that underlie language production can provide insight into broader language phenomena. Fricke, Kroll, and Dussias (2016) found that bilingual persons' speaking durations increase before a codeswitch and that comprehenders exploit these subtle cues to facilitate comprehension. The current work offers insight into how patterns such as these speaking durations arise in the environment as a consequence of constraints on speakers. Difficulty associated with the sentence-production process may cause speakers to slow down or codeswitch, which creates reliable patterns in the language environment with respect to utterance acoustics, or the contexts in which codeswitches are more or less probable. Then, comprehenders learn and use those patterns to aid language comprehension (Beatty-Martinez & Dussias, 2017; Valdés Kroff, Dussias, Gerfen, Perrotti, & Bajo, 2017). This relationship between language production, language distributions, and language comprehension can explain other language comprehension phenomena in which comprehenders exploit diagnostic patterns in the language environment brought about by constraints on speakers (MacDonald, 2013). The idea that constraints on the language producer may play a significant role in explaining why certain language patterns come to exist may have far-reaching consequences for not only the study of language production but of language comprehension and language development as well.

Acknowledgements We thank Ms. Paloma Iñiguez for recording the instruction videos. We also thank Dr. Jakub Szewczyk for helpful comments and discussion.

Open practices statement All data and materials are available at https://osf.io/uchpv/. No studies were preregistered.

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369-406.

Arnold, J.E., Losongco, A., Wasow, T., Ryan, G. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, *76*, 28-55.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 1–43.

Bates, E., Federmeier, K., Herron, D., Iyer, G., Jacobsen T., Pechmann, T., D'Amico, S., Devescovi, A., Wicha, N., Orozco-Figueroa, A., Kohnert, K., Gutierrez, G., Lu, C., Hung, D., Hsu, J., Tzeng, O., Andonova, E., Gerdjikova, I., Mehotcheva, T., Szekely, A., & Pleh, C. (2000). Introducing the CRL international picture-naming project (CRL-IPNP). *Center for Research in Language Newsletter*, *12*, 12-1.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1-48.

Beatty-Martinez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, *95*, 173-189.

Beilock, S. L., Wierenga, S. A., & Carr, T. H. (2002). Expertise, attention, and memory in sensorimotor skill execution: Impact of novel task constraints on dual-task performance and episodic memory. *The Quarterly Journal of Experimental Psychology, 55A*, 1211-1240.

Bentahila, A., & Davies, E. E. (1983). The syntax of Arabic-French code-switching. *Lingua, 59*, 301-330.

Bialystok, E., Craik, F. I., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, *10*, 89-129.

Bijeljac-Babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition, 25*, 447-457.

Blanco-Elorrieta, E., & Pylkkänen, L. (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences, 22*, 1117-1126.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*, 1-47.

Bock, K. J., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, *19*, 467–484.

Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gerasbacher (Ed.), *Handbook of psycholinguistics* (pp. 945-984). San Diego, CA: Academic Press.

Bobb, S. C., & Wodniecka, Z. (2013). Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology*, *25*, 568-585.

Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, *118*, 172-189.

Bond, M. H., & Lai, T. M. (1986). Embarrassment and code-switching into a second language. *Journal of Social Psychology*, *126*, 179-186.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*, 1-34.

Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, *1147*, 192-208.

Costa, A., La Heij, W., & Navarrete, E. (2006). The dynamics of bilingual lexical access. *Bilingualism: Language and Cognition, 9*, 137-151.

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.

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages?. *Journal of Memory and Language*, *103*, 28-43.

Declerck, M. & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review*, *22*, 1630-1645.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.

Ferreira, V. S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation*, *49*, 209-246.

Ferreira, V. S. & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*, 296-340.

Finkbeiner, M., Gollan, T. H., & Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem?. *Bilingualism: Language and Cognition*, *9*, 153-166.

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

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.

Gollan, T. H., Ferreira, V. S., Cera, C., & Flett, S. (2014a). Translation-priming effects on tip-of-the-tongue states. *Language, Cognition and Neuroscience*, *29*, 274-288.

Gollan, T. H., Kleinman, D., & Wierenga, C. E. (2014b). What's easier: Doing what you want, or being told what to do? Cued versus voluntary language and task switching. *Journal of Experimental Psychology: General*, *143*, 2167-2195.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of memory and language, 58*, 787-814.

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition*, *4*, 63-83.

Gollan, T. H., Starr, J., & Ferreira, V. S. (2015). More than use it or lose it: The number-of-speakers effect on heritage language proficiency. *Psychonomic Bulletin & Review, 22*, 147-155.

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.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, B1-B14.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*, 313-338.

Gross, M., & Kaushanskaya, M. (2015). Voluntary language switching in English–Spanish bilingual children. *Journal of Cognitive Psychology*, *27*, 992-1013.

Gumperz, J. J. (1977). The sociolinguistic significance of conversational codeswitching. *RELC Journal*, *8*, 1-34.

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, *57*, 169-177.

Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code switch? *Current Directions in Psychological Science*, *10*(5), 164–168.

Hernandez, A. E., Dapretto, M., Mazziotta, J., & Bookheimer, S. (2001). Language switching and language representation in Spanish–English bilinguals: An fMRI study. *NeuroImage, 14*, 510-520.

Hernandez, A. E., & Kohnert, K. J. (1999). Aging and language switching in bilinguals. *Aging, Neuropsychology, and Cognition*, *6*, 69-83.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*, 277-288.

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824-843.

Jescheniak, J. D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1256-1274.

Kimble, G. A., & Perlmuter, L. C. (1970). The problem of volition. *Psychological Review, 77*, 361-384.

Kleinman, D., & Gollan, T. H. (2016). Speaking two languages for the price of one: Bypassing language control mechanisms via accessibility-driven switches. *Psychological Science*, *27*, 700–714.

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., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, *9*, 119–135.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). LmerTest: Tests for random and fixed effects for linear mixed effect models. R package, version 3.1-1.

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, *99*, 199.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98*, 122-142.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226.

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

McDonald, J. L., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, *25*, 188-230.

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

Myers-Scotton, C. (1989). Codeswitching with English: types of switching, types of communities. *World Englishes*, *8*, 333-346.

Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science*, *26*, 22-27.

Nieuwenhuis, R., Pelzer, B., & Te Grotenhuis, M. (2012). Influence. ME: Tools for recognizing influential data in mixed models. *R package version 0.9-9*.

Otheguy R, Lapidus N. (2003). An adaptive approach to noun gender in New York contact Spanish. In *A Romance Perspective on Language Knowledge and Language Use: Selected Papers from the 31st Linguistic Symposium on the Romance Languages*, ed. R Nuñez-Cedeño, L López, R Cameron, pp. 209–29. Amsterdam: Benjamins

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 539–557.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of codeswitching. *Linguistics*, *18*, 581-618.

Prat-Sala, M., & Branigan, H.P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Language and Memory*, *42*, 168-182.

Runnqvist, E., Strijkers, K., Alario, F. X., & Costa, A. (2012). Cumulative semantic interference is blind to language:

Implications for models of bilingual speech production. *Journal of Memory and Language, 66*, 850-869.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*, 12530-12535.

Valdés Kroff, J. R. (2016). Mixed NPs in Spanish-English bilingual speech: Using a corpus-based approach to inform models of sentence processing. In R. E. Guzzardo Tamargo, C. M. Mazak, & M. C. Parafita Couto (Eds.), Spanish-English codeswitching in the Caribbean and the US (pp. 281–300). Amsterdam, The Netherlands: John Benjamins.

Valdés Kroff, J. R., Dussias, P. E., Gerfen, C., Perrotti, L., & Bajo, M. T. (2017). Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguistic Approaches to Bilingualism*, *7*, 163–198.

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *The Quarterly Journal of Experimental Psychology, 48A*, 822-848.

Zhan, M. & Levy, R. (2018). Comparing Theories of Speaker Choice Using a Model of Classifier Production in Mandarin Chinese. *Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1997–2005.

Zhang, Y., Wang, T., Huang, P., Li, D., Qiu, J., Shen, T., & Xie, P. (2015). Free language selection in the bilingual brain: an event-related fMRI study. *Scientific Reports*, *5*, 11704.