Article

The International Journal of Bilingualism

Mixing things up: How blocking and mixing affect the processing of codemixed sentences

International Journal of Bilingualism
2019, Vol. 23(2) 584–611

DOI: 10.1177/1367006917752570
journals.sagepub.com/home/ijb
SAGE

Michael A Johns
The Pennsylvania State University, USA

Jorge R Valdés Kroff
The University of Florida, USA

Paola E Dussias
The Pennsylvania State University, USA

Abstract

Aims and objectives/purpose/research questions: The goal of this study is to determine if the way in which codemixed sentences are presented during experimental lab sessions affects the way they are processed, and how experimental design approximates (or not) patterns of language use in bilingual populations.

Design/methodology/approach: An eye-tracking study was conducted comparing reading times on codemixed and unilingual Spanish sentences across two modes of presentation: (a) a blocked mode, where one block contained unilingual Spanish sentences and another one contained codemixed sentences; and (b) a mixed mode, where both unilingual and codemixed sentences were mixed together in a randomized fashion.

Data and analysis: 20 heritage speakers of Spanish were tested. Four reading measures extracted from the eye-tracking data were subjected to linear mixed-effects regression, with significance determined via backwards likelihood ratio tests, to examine differences across modes of presentation.

Findings/conclusions: Codemixes took significantly longer to process in the blocked mode than in the mixed mode. This is in line with corpus data suggesting that intra-sentential codemixing does not occur for long stretches of time and is broken up by unilingual discourse.

Originality: While a few studies have hinted at the potential confounds related to the presentation of codemixed or language-switching stimuli, the direct effects of experimental manipulation coupled with insights from sociolinguistic or corpus-based studies have not been tested.

Corresponding author:
Michael A Johns, The Pennsylvania State University, 442 Burrowes Building, University Park, PA 16801, USA.
Email: mjohns@psu.edu

Significance/implications: To better understand bilingual codemixing, as well as the cost (or lack thereof) associated with it, lab-based studies of codemixing should take insights from sociolinguistic and corpus-based research. The results of this study suggest that the experience that participants bring into the lab can interact with experimental design and result in unexpected results.

Keywords
Codemixing, experimental design, psycholinguistics, corpus linguistics, eye-tracking

Introduction

Since the 1980s, research on bilingual codemixing has greatly increased: from sociolinguistic studies of spontaneous and elicited data (i.e., Broersma, 2009; Poplack, 1980; Torres Cacoullos & Travis, 2018) to brain imaging studies investigating the neural correlates of language switching (Abutalebi & Green, 2016; Hernandez, Dapretto, Mazziotta, & Bookheimer, 2001; Price, Green, & Von Studnitz, 1999), codemixing has provided a rich testing ground for cognitive and linguistic theories of multilingualism and monolingualism alike. The goal of these studies has been to discover the underlying cognitive and linguistic mechanisms that regulate codemixing. We now know that codemixing is not haphazard and requires a great deal of proficiency and cognitive control in two languages (Abutalebi & Green, 2016; Green & Wei, 2014; Poplack, 1980, p. 601; Torres Cacoullos & Travis, 2015, pp. 369–371); that codemixing costs are modulated by the linguistic experience that bilinguals have with codemixing (Guzzardo Tamargo, Valdés Kroff, & Dussias, 2016; Valdés Kroff, Guzzardo Tamargo, & Dussias, 2018; Valdés Kroff, Román, & Dussias, 2016); and that codemixed words are processed differently from within-language synonyms (Moreno, Federmeier, & Kutas, 2002). However, many of these lab-based studies have not considered the social context in which codemixing occurs, nor have they looked to corpora of codemixed speech to inform experimental stimuli and design (cf. Guzzardo Tamargo et al., 2016; Hofweber, Marinis, & Treffers-Daller, 2016; Perrotti, 2017; Valdés Kroff, Dussias, Gerfen, Perrotti, & Bajo, 2017), despite the fact that studies from usage-based perspectives demonstrate that cognitive and linguistic mechanisms are intimately linked to language experience (e.g., see Bybee & Beckner, 2010, for a review). For example, while several corpus studies of Spanish-English codemixing have shown that so-called 'mixed noun (determiner) phrases' are produced predominantly with a determiner in Spanish and a noun phrase in English (Herring, Deuchar, Couto, & Moro Quintanilla, 2010; Otheguy & Lapidus, 2003; Valdés Kroff, 2016), psycholinguistic and lab-based studies do not always find that these types of codemixes are easier to process than less frequently occurring codemixes (e.g., Fairchild & Van Hell, 2015). One reason may be the nature of the experimental design itself: because lab-based studies are normally carefully controlled to eliminate potential confounds, they may not always represent actual language use (Blanco-Elorrieta & Pylkkänen, 2017; Gullberg, Indefrey, & Muysken, 2009; Valdés Kroff & Fernández-Duque, 2017; Valdés Kroff et al., 2018). In support of this, recent research examining the interaction between experimental design and language experience has shown that the way in which stimuli are presented to participants affects how they are processed and the outcomes that participants generate. For example, Jaeger and Snider (2013) investigated two modes of presentation on syntactic priming effects. In one mode, primes and targets were blocked by construction: participants saw only one construction for the first half of the experiment and the other during the second half. In a second mode, the constructions alternated by trial. This is exemplified in Figure 1, where dark and light rectangles

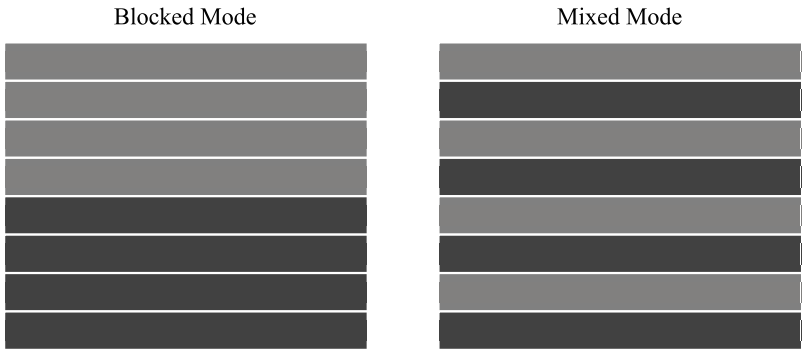

Figure 1. Blocked and mixed modes of presentation, from Jaeger and Snider (2013).

represent the two constructions. The authors found two different patterns of priming and surprisal—defined here as the likelihood of finding a particular construction in a given context—depending on the mode of presentation. In the blocked mode, surprisal spiked at the transition between the two constructions, while in the mixed mode it declined gradually over the course of the block. Since surprisal is implicated in priming (Dell & Chang, 2014, p. 3), the authors found that an increase in surprisal resulted in an increased probability of the participant being primed, with the end result being that the strength of priming depended on the mode of presentation. What this suggests is that language users employ both recent (surprisal-based) and prior (exposure-based) experience to adapt to incoming stimuli, changing their processing strategies accordingly.

Other research has found that the composition of stimuli and filler items in an experiment can also modulate language processing. Titone, Libben, Mercier, Whitford, and Pivneva (2011) examined the processing of cognates by English-French bilinguals and found that the inclusion of French filler items resulted in greater cognate facilitation in English, compared to an experiment where the filler items were in English. The authors argued that the inclusion of French filler items 'increased cross-language activation during L1 English sentence reading' (p. 1422).

Studies such as these suggest that processing is malleable and can be influenced by aspects of experimental design and properties of stimuli. In other words, the design choices that experimenters make interact with participants' own language experience; depending on what these choices are, they may lead to unexpected or puzzling effects. In the lab-based study of codemixing, one candidate is the mode of presentation: whether codemixed and unilingual stimuli are presented separately (in a blocked mode) or interleaved (in a mixed mode). Not only may these two modes differentially modulate aspects like priming and surprisal, as Jaeger and Snider (2013) show, they may also not adequately reflect the way bilinguals codemix in their daily lives.

Corpus-based studies of codemixing have shown that intra-sentential codemixing (the alternation between languages within a single sentence) occurs alongside copious stretches of unilingual discourse. For example, the Bangor Miami corpus contains over 43,000 utterances, but only 2527 (approximately 6%) contain a codemix (Guzzardo Tamargo et al., 2016, p. 142). As such, one may expect that the density of intra-sentential codemixing may not be particularly high in bilingual discourse, something that a blocked design (where codemixes would appear consecutively) does not reflect. The effects of experimental design may yield conflicting results across different studies, such as those examining switch costs. This has largely been addressed through studies of cued language-switching, where some studies report switch costs (Costa & Santesteban, 2004; Meuter & Allport, 1999, among others), while others report no switch costs (Gullifer, Kroll, & Dussias,

2013). Sometimes, switch costs surface (or not) depending on the task that the participant must complete (i.e., Von Studnitz & Green, 2002). While not all disparate results can be accounted for by the effects of blocking and mixing stimuli, it is certainly indicative of a methodological concern behind the study of codemixing: namely, sociolinguistic and corpus-based insights are often ignored in the lab-based study of codemixing, potentially leading bilinguals to process stimuli differently than when it is encountered in their daily lives.

The goal of this paper is to determine what effects, if any, mode of presentation has on the online processing of codemixed stimuli, and which mode may more adequately reflect the codemixing that bilingual speakers encounter. To accomplish this, the present study uses eye-tracking to compare the processing of Spanish-English codemixed stimuli in a blocked mode and a mixed mode of presentation. The present study considers findings from recent corpus-based experimental studies of codemixed language (e.g., Guzzardo Tamargo et al., 2016) to ask if any differences between the two modes might arise due to task-specific variables, sociolinguistic variables, or both. To do this, we target a construction that has been extensively researched in past codemixing literature involving Spanish and English: mixed noun phrases. Corpora of codemixed speech (e.g., the Bangor Miami corpus: Deuchar, Davies, Herring, Couto, & Carter, 2014; Otheguy & Lapidus, 2003) have shown that some Spanish-English bilingual communities exhibit an overwhelming tendency to produce determiner–noun phrase codemixes where the determiner is in Spanish and the noun phrase in English (Herring et al., 2010; Liceras, Fuertes, Perales, Pérez-Tattam, & Spradlin, 2008; Valdés Kroff, 2016; Valenzuela, Faure, Ramírez-Trujillo, & Barski, 2012). An example of this type of codemix is given below. The mixed noun phrase *un neighborhood* corresponds to the Spanish masculine noun *barrio*, and appears here with the masculine indefinite article *un*.

1. ahí está Sunset Lakes, que es la escuela donde yo estaba que es **un** neighborhood very upscale
   'there is Sunset Lakes, which is the school I was that is a very upscale neighborhood'

Given that single noun insertions are amongst the most frequent syntactic sites for a codemix (e.g., Poplack, 2018) and much is known about their production and processing (see references above), they was selected here to study the effect of blocked and mixed experimental designs on the processing of codemixed stimuli.

The experiment reported here focuses largely on reading codemixes. Like most monolingual communication, codemixing occurs frequently in the spoken domain, so there may be a question about the validity of using a reading task to study codemixing. Nevertheless, there is broad consensus that reading activates the system employed in auditory language processing (see, e.g., Perfetti, 1994). In the parsing literature, for example, Fodor (1998) proposes that syntactic processing during reading proceeds through subvocal phonological encoding of the prosody that guides auditory comprehension. Steinhauer and Friederici (2001) provided confirming evidence for this, showing that the processing of phrasing in both reading and auditory comprehension is reflected in the same closure positive event-related potential (ERP) component. We thus focus on reading because reading data linked with the speed of performance have long been used in psycholinguistics to identify the cognitive processes associated with—and to develop theories of—language comprehension. In this way, our results can be interpreted in the context of the vast general literature on reading and comprehension processes. In addition, it is important to emphasize that codemixing is not exclusively a production phenomenon: codemixed structures must also be processed by the comprehension system. Finally, for the language pair under investigation here, Spanish-English codemixing is increasingly present in text and thus in the bilingual's

reading experience as well. This is particularly true of email and chat environments, as seen in the written codemixing corpus of Montes-Alcalá (2005).[1]

Methodology

Participants

Twenty heritage Spanish speakers participated in this experiment. Participants received monetary compensation for their participation. Informed consent was obtained at the beginning of each experimental session. Four measures of language proficiency were administered: a language history questionnaire, a standardized test of proficiency in Spanish and in English, a verbal fluency task, and a picture naming task. Each is described in turn.

Language history questionnaire. The language history questionnaire collected information about participants' acquisition and use of Spanish and English, as well as their codemixing practices and attitudes. For example, the questionnaire asked about place of birth, how long participants had been speaking and writing in English and Spanish, and the use of Spanish and English in the home and with friends. The questionnaire also asked participants to self-rate their linguistic ability in speaking, listening, reading, and writing in both languages. The last section inquired about attitudes and patterns of codemixing, asking how often participants engaged in codemixing, with whom, and in which domains (spoken, email, texting, social media). All participants indicated using and/or being exposed to codemixing. On average, participants reported 'sometimes' using and encountering codemixing in the written domain ($M = 2.61$, SD = 0.81; where 1 is 'never', 2 is 'rarely', 3 is 'sometimes', 4 is 'most of the time', and 5 is 'always'). We use the questionnaire for primarily descriptive purposes.

Overall, the participants were balanced in their self-rated proficiencies in English and Spanish (see Table 1; scale is out of 10). Paired *t*-tests revealed no significant differences between participants' self-rated abilities in speaking, listening, writing, or reading English and Spanish ($p > 0.05$ for all measures), but English scores were marginally higher than Spanish scores for speaking, writing, and reading abilities—a pattern repeated in the proficiency measures below.

Table 1. Self-rated proficiency measures.

	Speaking	Listening	Writing	Reading
English	9.35	8.81	8.52	9.42
Spanish	8.97	9.45	8.39	8.97

Verbal fluency. Participants completed a verbal fluency task in Spanish and in English. In this task, participants were asked to generate in 30 seconds as many exemplars belonging to a prescribed set of semantic categories as possible. We chose this task as a measure of language proficiency for two reasons. Firstly, previous studies have shown a decline in the accessibility to first language (L1) words in a second language (L2) environment (Baus, Costa, & Carreiras, 2013). For example, Linck, Kroll, and Sunderman (2009) found that L2-immersed L2 learners produce a smaller number of exemplars in their L1 than classroom learners without immersion experience, suggesting that the L1 was attenuated in the L2 environment. Secondly, there is some evidence suggesting a strong correlation between the verbal fluency task and objective measures of language proficiency (Dussias, Halberstadt, & Carlson, in prep).

Eight different categories were named, four in each language (e.g., clothing, fruits, vegetables, furniture). The categories were counterbalanced by language, such that one participant named clothing in Spanish, and the next participant named the same category in English. Participants were given 1 point per word named, and culture-specific words (such as *elote* for corn as opposed to the more general *maíz*) were accepted. Between-group comparisons of participants' total number of exemplars revealed that they produced significantly more items in English ($M = 42.4$) than in Spanish ($M = 35.5$); $t(19) = 3.59$, $p < 0.01$.

Picture naming. In this task, participants were instructed to name pictures as quickly as possible in both English and Spanish. The order of the language was counterbalanced by participant, such that one participant would name in English before Spanish, and the next in Spanish before English. The pictures were all black-and-white line drawings and were matched in frequency across languages. Participants named 66 pictures in each language, and were scored according to whether or not they were able to correctly identify the picture. As with the verbal fluency task, culture-specific words that correctly identified the picture were counted as correct. Both accuracy and reaction times were analyzed. Paired *t*-tests revealed that participants were significantly more accurate in English ($M = 98.3\%$) than in Spanish ($M = 82.8\%$; $t(19) = 13.02$, $p < 0.01$), and significantly faster to respond in English ($M = 1028.91$ ms) than in Spanish ($M = 1093.79$ ms; $t(19) = -2.25$, $p = 0.03$).

Language proficiency tests. In addition to the two online measures of language proficiency administered above, participants also completed a modified version of the Michigan English Language Institute College Entrance Test (MELICET) and the Advanced Test of the Diplomas de Español como Lengua Extranjera (DELE, 'Diplomas of Spanish as a Foreign Language'). The MELICET is an advanced level English language test created by the University of Michigan English Language Institute (http://www.michigan-proficiency-exams.com/melicet.html) to examine ability in different English language areas. It is primarily used to test nonnative speakers of English by educational institutions as an admissions or placement test. The DELE is a standardized test of Spanish issued by the Ministry of Education, Culture, and Sport of Spain, which assesses proficiency in Spanish at seven levels (http://diplomas.cervantes.es/en). The test administered here was the Nivel Superior C2, the highest level of accreditation. Each grammar test contained 50 multiple-choice items, which evaluated grammar, vocabulary, and reading competence in isolated sentences, as well as longer stretches of discourse. Participants scored an average of 33.7 points on the DELE and 41.8 points on the MELICET; the difference was statistically significant ($t(19) = 4.2$, $p < 0.001$).

It is important to note two things with respect to the English dominance of the participants that the above measures suggest: firstly, participants are immersed in English, and primarily educated in English. Thus, it is no surprise that they are English dominant. Secondly, while the participants are all English dominant, they are still nonetheless highly proficient in Spanish, both rating themselves highly in their abilities in Spanish, and scoring highly on the Spanish proficiency measures.

Cognitive control tasks. Two measures of cognitive control were also administered. Their primary purpose was to determine if the effects of experimental design could be modulated by individual differences in executive function. The goal was to tease apart cognitive-based effects (i.e., goal maintenance and inhibition) from sociolinguistic effects (past experience with codemixing). The two tasks that were chosen were the Flanker task and the AX-CPT task. The version of the Flanker task employed here consisted of four blocks: the first block was a control block consisting of trials

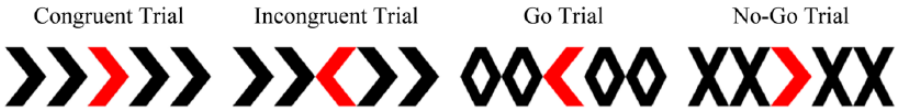

Figure 2. Flanker trial types.

where a single arrow appeared on screen without distractors; participants indicated the direction of the arrow using a Chronos button box (Psychological Software Tools). The second and third blocks were counterbalanced by participant: in one block, trials consisted of a target arrow in red surrounded by four black distractor arrows, pointing in either the same (congruent) or opposite (incongruent) direction. Participants were instructed to ignore the black arrows and indicate the direction of the red arrow. The other block was a Go/No-Go task: on trials where the red arrow appeared surrounded by diamonds, participants indicated its direction; when the red arrow was surrounded by Xs, participants waited until the end of the trial without making a response. The fourth block was a mixed block, with half of the trials consisting of the congruent/incongruent Flanker trials, and the other half of Go/No-Go trials. Accuracy and reactions times across conditions and blocks will be correlated with participants' reading times on the experimental task. An example of the various Flanker trials is given in Figure 2.

Lastly, participants completed the AX-CPT (Morales, Gómez-Ariza, & Bajo, 2013; Rosvold, Mirsky, Sarason, Bransome Jr., & Beck, 1956). In this task, participants saw a string of five letters appear in the center of the screen, one at a time. The first and last letters appeared in red, and the three letters between them appeared in white. Participants were asked to press the YES button (side counterbalanced by participant) on the last red letter if and only if this last letter was an X, and the first red letter was an A. All other combinations received a NO response, as did the first four letters in the sequence. Seventy percent of the trials followed this AX pattern, biasing participants toward a YES response when seeing either an A or an X; the rest of the trials either began with A and ended with a letter besides X (AY trials, 10%), ended with X but began with a letter other than A (BX trials, 10%), or neither began nor ended with A or X (BY trials, 10%). This task measures two different mechanisms of cognitive control: proactive control, which refers to an individual's ability to maintain goal-relevant information (monitoring), and reactive control, which refers to an individual's ability to react to changing task demands (inhibition; see Braver, 2012, for an overview of these mechanisms). For the purposes of this study, monitoring will be indexed by reaction times on the AY trials, with longer reaction times indicating stronger monitoring abilities (maintaining the goal-relevant A cue), while inhibition will be indexed by reaction times on the BX trials, with shorter reaction times indicating stronger inhibitory abilities (reacting and inhibiting the 'X means yes' bias). This measure will be correlated with participants' reading times on the experimental task. An example of an AX-CPT trial is given in Figure 3.

Design and materials

Three blocks of 84 sentences each were created: a Spanish-only block, where every sentence was in Spanish; a codemixed-only block, where all stimuli contained a codemix; and a mixed block, where half of the sentences were in Spanish, and the other half contained a codemix. Of these 84 sentences, 28 were experimental sentences and 56 were fillers. Across blocks, there were a total of 84 experimental sentences and 168 fillers. A comprehension question followed each sentence. Experimental sentences contained a target determiner phrase in direct object position. The noun

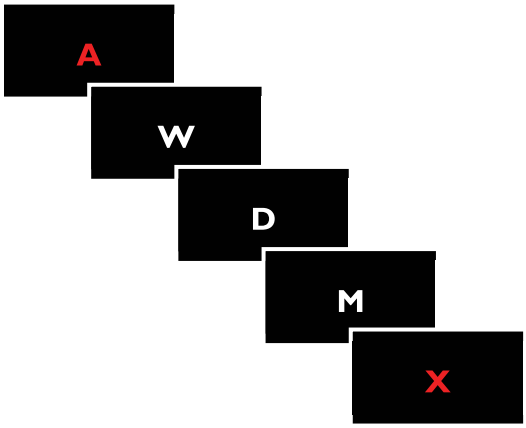

Figure 3. AX-CPT trial.

within the determiner phrase was the target word. For unilingual Spanish sentences, the sentence and target noun were all in Spanish; for codemixed sentences, the sentence was also in Spanish, but the noun was in English, resulting in an insertional codemix. The determiner was always congruent in gender with the Spanish noun, or the Spanish translational equivalent of the English noun for codemixed sentences. Appendix 3 contains all target nouns used in the stimuli.

Eighty-four target Spanish nouns were selected using the EsPal database (Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013) to control for frequency, imageability, familiarity, and concreteness. Half of the nouns were feminine, and half were masculine in gender. The nouns were selected such that when they were translated, the lexical frequency between the Spanish nouns and their English translations did not differ significantly ($p > 0.1$). Sentences could appear in one of two sentence types: unilingual, where the target noun was in Spanish; or codemixed, where the target noun was the English translation equivalent of the Spanish noun. The target noun always occurred as the direct object in a simple transitive sentence, and was followed by a prepositional modifier in almost all cases (in four stimuli, the noun was followed by a clausal modifier[2]); no other structures were used. Codemixed sentences continued in Spanish after the English noun so that this region could be adequately compared to the unilingual Spanish sentences. Previous sociolinguistic literature (e.g., Poplack, Sankoff, & Miller, 1988, p. 75; Sankoff, Poplack, & Vanniarajan, 1990, p. 94; Torres Cacoullos & Aaron, 2003) has suggested that single-word insertions such as those employed in the present study may be morphosyntactically integrated, behaving differently from multi-word stretches of codemixed speech. For now, we will continue to refer to these single-noun insertions as 'codemixes' for simplicity's sake, and remain agnostic on their status as bona fide codemixes. The term may be considered analogous to Poplack's (2018, p. 5) 'language mixing', which refers to 'various combinations of *overt* lexical material from two or more languages' (emphasis original). We will return to the question of these nouns as codemixes in the discussion.

Six lists were created so that both the unilingual and codemixed variations of each experimental sentence occurred in all three blocks. Filler items introduced codemixes at various points throughout the sentence (within the grammatical subject of the sentence, at the verb, or in an adjunct phrase following the direct object); this was done to distract participants from the determiner–noun codemixes in the experimental items. For the Spanish-only block, all 28

Table 2. Number of items per condition across blocks and modes.

Sentence type	*Gender*	**Blocked mode**		**Mixed mode**
		Spanish-only block	*Codemix-only block*	*Mixed block*
Unilingual	M	14		7
	F	14		7
Codemixed	M		14	7
	F		14	7
Total exp. sentences		28	28	28

Table 3. Experimental sentences.

Unilingual	Codemixed	Question
La criada encontró el **jabón** en el gabinete.	La criada encontró el **soap** en el gabinete.	¿El jabón estaba encima de la mesa?
El profesor tomó la **cerveza** en el bar.	El profesor tomó la **beer** en el bar.	¿Al profesor le gusta la cerveza?
'the maid found the soap in the cabinet'	'the professor drank the beer in the bar'	'is the soap on the table?' 'did the professor like the beer?'

Table 4. Filler sentences.

Unilingual	Codemixed
La científica y sus colegas descubrieron el elemento nuevo.	La científica and her colleagues discovered the new element.
El bombero miró el fuego con mucha ansiedad.	El bombero watched the fire with great anxiety.
Alejandro bebió el refresco después de correr.	Alejandro bebió el refresco after running.
Carla observó que el tren nunca llegaba tarde.	Carla observó que the train never arrived late.

'the scientist and her colleagues discovered the new element'
'the firefighter watched the fire with great anxiety'
'Alejandro drank the beverage after running'
'Carla observed that the train never arrived late'

experimental sentences were in Spanish; for the codemixed-only block, all 28 experimental items contained an insertional codemix at the target noun; for the mixed block, 14 experimental sentences were in Spanish; the remaining 14 were codemixed. The same pattern was followed for the filler sentences: all 56 fillers in the Spanish-only blocks were in Spanish, and all 56 fillers in the codemixed-only block contained a codemix; 28 fillers were in Spanish and the remaining 28 contained a codemix in the mixed block. Together, the Spanish-only and codemixed-only blocks constitute the *blocked mode*; in this mode, all sentences of one type (unilingual or codemixed) are presented first and before those in the other condition. The mixed block constitutes the *mixed mode*, with the two sentence types presented interleaved in a randomized fashion. Table 2 details the experimental design visually, while Tables 3 and 4 provide examples of experimental and filler sentences, respectively.

Predictions

Based on the above experimental design, two different predictions can be made: in considering the sociolinguistic literature, if the density with which codemixing occurs in bilingual discourse is an important factor, then it is the blocked mode of presentation that should yield overall greater processing difficulty compared to the mixed mode. Specifically, codemixed sentences should be easier to process in the mixed mode than in the blocked mode because the density of codemixing more closely approximates that found in natural bilingual discourse. However, based on the language-switching literature, the mixed mode of presentation would be predicted to yield overall greater processing difficulty than the blocked mode, due to the presence of intra-sentential codemixes as well as the alternation between unilingual and codemixed sentences.

Procedure

Participants completed an informed consent form before each session, and were paid for both sessions separately. Participants were tested individually in a quiet room. They were seated in front of an EyeLink 1000 Plus desktop-mounted eye-tracker (SR Research) and a BenQ monitor with their heads resting in a chin-rest to ensure stability during recording. The eye-tracker sampled eye movements at 1000 Hz using the corneal reflection of the participant's right eye. Button presses were recorded using a Vpixx ResponsePixx Dual Handheld button box.

Before beginning each block, participants completed a nine-point calibration procedure to ensure accurate tracking. One participant was turned away due to technical difficulties during calibration resulting in the previously mentioned total of 20 participants; no other participants' data were removed. Once calibration was successfully completed and the error was below 0.5°, the experiment began. Each of the three blocks began with a 10-item practice session, allowing the participants to familiarize themselves with the procedure. After completing the practice, they then continued to the experiment proper. For each trial, participants first focused on a fixation point placed at the same location as the start of the sentence; this was done instead of positioning the fixation point in the center of the screen to avoid overshoot on the first word as the eyes move from the center of the screen to the start of the sentence. The experimenter manually performed a drift correct at this point to ensure accurate tracking. The sentence was displayed on the screen until the participant pressed any button on the button box to continue, at which point the sentence disappeared and the comprehension question was displayed on screen. Both the sentence and the question were displayed in the center of the screen, left-aligned such that the beginnings of each corresponded to the same location on screen (the same location as the fixation point). Participants indicated a YES response with the left hand, and a NO response with the right hand. Questions were always presented in unilingual Spanish, regardless of the block, with SÍ and NO printed in the lower left- and right-hand corners of the screen (respectively).

The experiment took place over two sessions, each on a separate day. The time between sessions varied by participant; as no effects of priming nor habituation are expected with respect to reading strategies, the time between sessions was not fixed. After informed consent was obtained, the first session began with the language history questionnaire. Afterwards, the participants completed both the Spanish-only and codemixed-only blocks, with the order counterbalanced by participant. At the end of the first session, the participants completed the DELE and MELICET tests. In the second session, participants returned and completed the mixed block, after which they completed the category fluency, picture naming, Flanker, and AX-CPT tasks (in that order). The mixed block was placed on a separate day to mitigate any potential effects that could influence the Spanish-only or codemixed-only blocks. Likewise, this adequately reflected two separate experiments: one using a blocked mode of presentation and the other a mixed mode.

Analysis

Data cleaning. Initial data cleaning took place in DataViewer (SR Research). Each trial was manually inspected for errors. Trials where technical glitches occurred and fixations were not recorded were removed (8/1680 trials). Any fixations that fell outside of interest areas were associated with the closest interest area into which they fell (Rayner, Chace, Slattery, & Ashby, 2006).

Four eye-tracking measures were extracted from the data: (1) gaze duration, which is the sum of all fixations that occurred in an interest area when it was first entered until it was first exited to the left or right; (2) right bounded duration, which is the sum of all fixations in an interest area that occur before the interest area is exited to the right for the first time (re-reads from the left are included); (3) regression path duration, which is the sum of right bounded duration plus the duration of saccades made before exiting the interest area to the right; and (4) total duration, which is the sum of all fixations made on an interest area in a given trial. Gaze duration has been suggested to reflect earlier measures of processing, and right bounded and regression path duration reflect processes in initial re-reading. Total duration reflects later stages of processing. Including these measures provides more information about the time course of processing (see Clifton, Staub, & Rayner, 2007, for an overview of these reading measures). These four measures were calculated for three regions: the determiner preceding the target noun, the target noun, and a spillover region consisting of the following three words. The spillover region contained multiple words because the following two words were almost always function words (prepositions and determiners), which are not fixated on as frequently as content words; the third word in the region was almost always a content word. This ensured that there were sufficient informative data to look for a spillover effect.

Outliers for each of the four measures at each region were removed using the median absolute deviation method (MAD; see Leys, Ley, Klein, Bernard, & Licata, 2013, for the specific procedure). Firstly, reading measures were log-normalized to adjust for skewness in the reading times (a common procedure for reading and reaction times: see Ratcliff, 1993). Then, a *Z*-score based on the MAD was calculated for each data point using the normalize function in the Rling package (v. 1.0; Levshina, 2015, p. 60). This *Z*-score indicated the normalized distance of each point from the median, and a cutoff of 2.5 deviations away from the median was chosen to remove outliers: values whose *Z*-score was greater than 2.5 or less than −2.5 were removed (8.2% of the data). The remaining log-normalized values served as the dependent measure for the analysis below.

Modeling. Linear mixed-effects models were created using the lmer function in the lme4 package (v. 1.1-12; Bates, Maechler, Bolker, & Walker, 2015) in the statistical program R (v. 3.3.2; R Core Team, 2016). Maximal models were first created: mode of presentation, sentence type, and their interaction were included in the fixed-effects and (participant and item) random-effects structures. The random-effects structures for each model were subjected to a principle components analysis using the rePCA function in the RePsychLing package (v. 0.0.4; Baayen, Bates, Kliegl, & Vasishth, 2015). This function performs a principle components analysis on the random-effects structure to determine which random effects and interactions the data are minimally capable of supporting (Bates, Kliegl, Vasishth, & Baayen, 2015). This ensures that the model is not overly complex nor underspecified in its random-effects structure. Backward elimination model comparisons were conducted via likelihood ratio tests using the analysis of variance (anova) function in the R base package. This method compares a maximal model with all fixed effects and interactions to a smaller model with one of the fixed effects or their interaction removed. For example, a model containing the main effects of mode of presentation and sentence type, and their interaction, would be compared to a model with only the main effects of mode of presentation and sentence type. The anova function then determines if the smaller

Table 5. Summary of significant main effects. (All times in ms; standard deviation in parentheses.).

Target noun			**Spillover region**		
Sentence type	*Right bounded*	*Total*	Mode	*Regression path*	*Total*
Unilingual	327 (163)	430 (258)	*Blocked*	942 (605)	951 (539)
Codemixed	304 (137)	403 (244)	*Mixed*	894 (548)	875 (498)

model is significantly worse at fitting or explaining the data compared to the maximal model by reporting a chi-squared test statistic and its associated p-value. A significant value for this test indicates that, as in the above example, the interaction between sentence type and mode significantly adds to the explanatory power (or fit) of the model above and beyond the two main effects alone. A backwards elimination model comparison was conducted for the main effect of sentence type, the main effect of mode of presentation, and the interaction between the two at each of the three regions for all four reading measures.

Results

Linear mixed-effects models

Appendix 1 contains the results of all model comparisons conducted as well as the R code for the relevant models; we summarize the findings here. No significant main effects or interactions were found for the four fixation duration measures at the determiner, indicating that there were no baseline effects before encountering the noun. At the target noun, the main effect of sentence type significantly improved the fit of the model for both right bounded ($\chi^2 = 4.55$, $p = 0.03$) and total duration ($\chi^2 = 4.01$, $p = 0.05$) measures. For both measures, durations were longer in unilingual sentences than codemixed sentences (i.e., longer in Spanish target nouns than English target nouns). No other main effects or interactions were significant.

In the spillover region, the main effect of mode of presentation significantly improved the fit of the model for both regression path ($\chi^2 = 5.62$, $p = 0.02$) and total duration ($\chi^2 = 6.45$, $p = 0.02$) measures. For both measures, durations were longer in the blocked mode than in the mixed mode. The main effect of mode was approaching significance for right bounded duration as well ($\chi^2 = 3.08$, $p = 0.08$), with the direction of the effect identical to regression path and total duration. Means and standard deviations (in parentheses, both in ms) for significant effects are summarized in Table 5, and visualized in Figures 4 and 5.

Switch costs

Although assessing switch costs was not the primary goal of the work presented here, we nevertheless investigated the presence of switch costs as this is of wide interest in the literature on codemixing and language switching (i.e., Gollan & Ferreira, 2009; Gullifer et al., 2013). These previous studies (see also Wylie & Allport, 2000) have calculated switch costs by subtracting the value (i.e., reaction time, reading time, amplitude, etc.) for unilingual stimuli from codemixed stimuli (or same-language trials from different-language trials). Positive values indicate a greater processing cost associated with codemixed stimuli. The same method of analysis was used for the reading measures extracted from the eye-tracking data in the present study. In this case, switch costs were calculated for each participant for each reading measure. Within the blocked mode, participants'

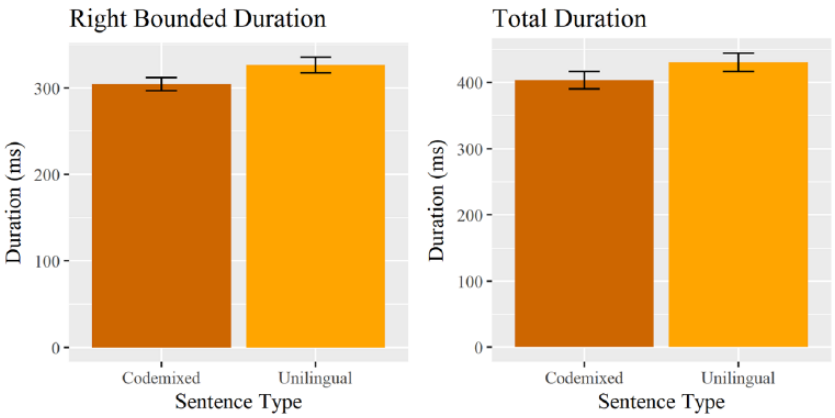

Figure 4. Significant main effects on the target noun.

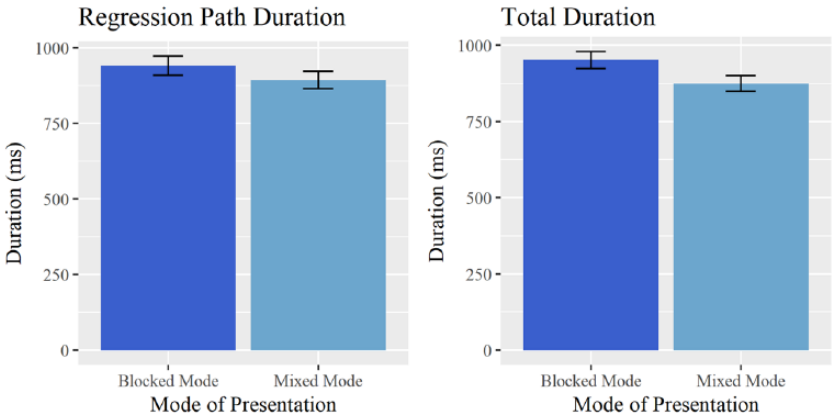

Figure 5. Significant main effects on the spillover region.

average reading durations for unilingual sentences were subtracted from those durations for codemixed sentences; the same procedure was applied within the mixed mode. This was done for the determiner, target noun, and following word.

For all four reading measures at all three regions, two-tailed paired *t*-tests revealed no significant differences between switch costs across the two modes of presentation (all *p*-values greater than 0.05). The difference between switch costs in the two modes at the spillover region was marginally significant for right bounded duration ($t(19) = 1.83$, $p = 0.08$), such that switch costs in the blocked mode were marginally greater than in the mixed mode (blocked difference: 13.3 ms; mixed difference: –0.3 ms).

Accuracy to comprehension questions

Accuracy to the comprehension questions was analyzed using general linear mixed-effects regression. The results indicated no differences in accuracy across the two modes of presentation, nor across sentence type (unilingual, codemixed). Overall accuracy across participants was 95% (SD = 21%).

Correlations with proficiency measures

Reading times at the target noun and in the spillover region were correlated with the various language proficiency measures and cognitive tasks administered. Only results for gaze duration will be reported, as this reflects early stages of processing and will be the most informative of the four reading measures collected. Because correlations between gaze durations and the measures extracted from the Flanker and AX-CPT tasks were spurious and did not form a coherent pattern, they will not be discussed. R-squared and *p*-values are given in Appendix 2.

At the target noun, Spanish proficiency as measured by performance in the Spanish picture naming task and the DELE anticorrelated significantly with reading times on both English nouns (in codemixed sentences) and Spanish nouns (in unilingual sentences). This was true for items in both the blocked and mixed modes of presentation. These anticorrelations indicated that as proficiency in Spanish increased, gaze durations on the target noun decreased.

A similar pattern emerged in the spillover region: proficiency in Spanish anticorrelated significantly with gaze durations for both unilingual and codemixed sentences in the blocked and mixed modes of presentation. It is worth noting that proficiency in English as measured by the English picture naming, category fluency, and MELICET did not significantly correlate with gaze durations at any region.

Discussion

The goal of the present study was to determine if the experimental design—namely, the use of a blocked versus a mixed mode of presentation—affected how bilinguals processed codemixed sentences, and if this might result in differing or even conflicting results across the two modes. To test this, we compared the two modes of presentation in a within-subjects design. To summarize the results, right bounded and total durations were significantly longer at the target noun for unilingual sentences (i.e., *La cocinera usó el aceite que compró en Italia*, 'the cook used the oil that she bought in Italy') compared to codemixed sentences (*La cocinera usó el oil que compró en Italia*). In the spillover region following the target noun (*...aceite que compró en Italia*), regression path and total durations were significantly longer in the blocked mode than in the mixed mode. Spanish proficiency was significantly anticorrelated with reading times in both unilingual and codemixed sentences: higher Spanish proficiency led to faster reading times for both Spanish *and* English nouns, as well as in the Spanish spillover region across conditions.

The effects found at the target noun do not align with previous studies that have found switch costs in both production and processing (see the Introduction). In this experiment, switching into English seems to have been easier than staying in Spanish. This effect is potentially due to the participants' dominance in English. This account, however, is less likely when considering an unexpected set of findings that were found in the correlation tests; specifically, only Spanish proficiency measures—but not English proficiency—resulted in significant correlations with gaze durations on English nouns. That is, while a higher Spanish proficiency led to faster reading times on English nouns, English proficiency did not correlate with reading times on the English nouns. One possible explanation for the lack of correlation with the English proficiency measures may have been that participants, immersed and educated in English, were simply not variable enough in their scores to merit correlation.

Turning now to the effect of mode of presentation found in the spillover region, the primary goal of the study, the results suggest that when codemixed and unilingual stimuli are presented mixed together, reading times in Spanish decrease as compared to when the two sentence types are presented separately. To determine the locus of the differences across modes, the regression path and

Table 6. Block differences in the spillover region.

Reading measure	Spanish-only	Codemixed-only	Mixed
Regression path	890 (548)	994 (655)	894 (548)
Total	946 (525)	957 (553)	875 (498)

total durations were compared across the three individual blocks in a post-hoc analysis. Linear mixed-effects models were created to predict each reading measure by block (Spanish-only, codemixed-only, or mixed). The results for regression path duration showed the following: block type (Spanish-only, codemixed-only, or mixed) was found to significantly improve the model (χ^2 = 7.95, p = 0.02), indicating that the difference in modes was being driven by significantly longer reading times in the codemixed-only block compared to both the Spanish-only and mixed blocks. For total duration, block type also significantly improved the model (χ^2 = 6.89, p = 0.03), but instead reading times in the mixed block were found to be significantly *faster* compared to both the Spanish-only and codemixed-only blocks. Means and standard deviations are given in Table 6 (in ms, standard deviations in parentheses).

Why would more codemixing lead to longer durations compared to no codemixing at all or codemixing half as often? We propose here an explanation based on the distribution of codemixes in bilingual production: as previously mentioned, intra-sentential codemixing is not particularly dense, occurring in only 6% of the utterances of the Bangor Miami corpus (Guzzardo-Tamargo et al., 2016). In addition, the use of single-word insertions is also relatively sparse. In Poplack et al. (1988, p. 57), this type of mixing accounts for 'under 1% (0.83%) of the total verbal output' and only '3.3% of the total vocabulary.' Such low rates were also found in a sample from the New Mexico Spanish-English Bilingual corpus (cf. Torres Cacoullos & Travis, 2018) by Aaron (2015, p. 461), who reports that English-origin nouns in otherwise Spanish speech occur with normalized frequencies ranging from 6 to 215 tokens per 10,000 words.[3] Thus, prolonged codemixing—as in the codemixed-only block, where every sentence contained a codemix of some type, be it single- or multi-word—does not align with bilingual experience, and thus could yield longer regressions as participants re-read sentences. Regressions in the mixed block are shorter than in the codemixed-only block, however, potentially because the decreased rate of codemixing may better approximate what these bilinguals encounter in their daily lives.

The results for total duration tell a slightly different story: here, the mixed block shows the shortest total durations, while the Spanish-only and codemixed-only blocks do not differ from one another. The reduced total durations in the mixed block could be the result of two factors working together: language dominance and codemixing experience. Because intra-sentential codemixing is not likely to occur for long stretches of discourse, breaking up codemixes with unilingual sentences aligns more closely with production data and in turn results in reduced processing difficulty compared to the codemixed-only block. Compared to the Spanish-only block, total durations in the mixed block may be shorter due to the English dominance of the participants in this study. Thus, even though participants still read unilingual sentences in this block (and reading times did not differ across sentence types in the mixed block), the presence of English alongside Spanish in the mixed block could have acted to 'release' some of the processing difficulty associated with reading in the less dominant language—an explanation also in line with Green and Wei's (2014, p. 504) cooperative or coupled control mode. In summary, the presence of English combined with the reduced density of codemixing combined to yield faster total durations for the mixed block compared to the other two blocks.

Returning to the main question of this study, no evidence was found that mixing is costlier than blocking, nor that switch costs differ between the two modes (as no switch costs were found). Instead, the results suggest that a mixed mode of presentation may align more closely with bilinguals' everyday experience with codemixing. This agrees with psycholinguistic models of production and processing that posit a tight link between production and comprehension, such as the Production-Distribution-Comprehension (PDC) model by MacDonald (2013). In this model, production is constrained by general cognitive mechanisms, which in turn shape the distribution of the input that feeds into the comprehension system. As MacDonald (and others: see Dell & Chang, 2014) argues, the comprehension system tunes to patterns of production because the two are inextricably linked, and this tuning can even occur within a single experiment. As discussed by Kroll, Dussias, Bice, and Perrotti (2015, p. 7.8), 'if the parser's configuration is related to language exposure … and language contact, then bilinguals' processing routines are expected to change as a function of the frequency with which the relevant structure appears in an experimental session.'

Insights from sociolinguistic studies

One question that recurs in lab-based studies of codemixing, particularly with respect to the cost of codemixing, is why codemixing should prove cognitively difficult in a lab setting if bilinguals seem to readily engage in codemixing in their day-to-day lives. In other words: if codemixing is costly, why do it at all?

Previous lab-based studies of codemixing have often used cued language-switching to assess switch costs. For example, Meuter and Allport (1999) report asymmetric switch costs in their seminal study on language switching: switching from the weaker into the dominant language results in greater switch costs than switching from the dominant into the weaker language. In another influential study, Gollan and Ferreira (2009) report that switch costs can be mitigated if the bilingual can choose when to switch; cued language switching resulted in the same asymmetric pattern found in previous studies.

Outside of the lab, however, the cost of codemixing is not obviously present. Analyses of production data have revealed that cues are often used and exploited by bilingual speakers to aid in the production and processing of codemixes: for example, Fricke, Kroll, and Dussias (2016) analyzed the Bangor Miami corpus and found that speech rate decreased around codemix sites, and that voice onset time (VOT) for English stop consonants was significantly more Spanish-like when close to the codemix site. In a follow-up visual-word study using eye-tracking, they found that bilinguals exploit these as cues to an upcoming codemix that 'promote[s] more robust recognition of the target word both by allowing listeners to make item-specific predictions…, but also by boosting the rate of increase in target activation in the absence of specific predictions' (Fricke et al., 2016, p. 127). A similar corpus analysis conducted by Piccinini and Arvaniti (2015) found similar effects on VOT for Mexican-American heritage speakers of Spanish in California. Likewise, Balukas and Koops (2015) report a similar finding on VOT in Spanish-English bilinguals in New Mexico. Lastly, recent neurolinguistic evidence suggests that when producing spontaneous as opposed to elicited language switching, or processing naturalistic as opposed to artificial codemixing, bilingual brains show less activation in regions associated with executive control, and participants exhibit no behavioral switch costs (Blanco-Elorrieta & Pylkkänen, 2017).

It is also worth mentioning that even in monolingual production and processing, some structures are more 'costly' than others, despite their use in naturalistic speech. For example, object relatives are more difficult to process than subject relatives (Traxler, Morris, & Seely, 2002), but experience with object relatives can make them easier to parse (Wells, Christiansen, Race, Acheson, & MacDonald, 2009), and their difficulty can be modulated by factors like animacy of the modified

noun (Reali & Christiansen, 2007). Likewise, differences in linguistic experience, in particular those derived from education, have also been found to modulate implicit grammatical knowledge even in the native language (Dąbrowska, 2012). Thus, even though one structure or language mode may be more 'difficult' than another, speakers have a variety of tools at their disposal to ease this burden (see MacDonald, 2013) and these tools are refined according to experience and input (Christiansen & Chater, 2016).

The connection between production and processing has also been tested in recent work by Guzzardo Tamargo and colleagues (2016). Again, a corpus analysis formed the basis of a processing study, where the asymmetry of production of two potential codemixes predicted their asymmetric processing in an eye-tracking study. Similarly, Stammers and Deuchar (2012) show that word frequency derived from corpus data is an excellent predictor of soft mutation in Welsh-English codemixing, such that highly frequent English words are more likely to undergo the mutation than less frequent English words when inserted into a Welsh frame.

The link between language use and language processing is not necessarily surprising, especially from a usage-based perspective, where tokens of use come to influence subsequent encounters of that same structure (e.g., Bybee, 2010, p. 19; Tamminga, 2016). Given the results of the present study, and the findings summarized above, it is crucial that lab-based studies of codemixing take this into consideration, especially when addressing questions surrounding the cost associated with codemixing: when production data are ignored and bilinguals encounter structures that do not conform to (or that violate) their prior experience, processing may change and not adequately reflect how bilinguals process speech in normal conversation. This poses problems not only for the study of codemixing itself, but also for the generalizability of these studies to the broader bilingual population.

Lastly, the very nature of switch costs and the fact that they are found in both processing (Litcofsky & Van Hell, 2017) and production (Bultena, Dijkstra, & Van Hell, 2015) may not be particularly informative or reflective of bilingual practices in actual language use, but instead may be more similar to gradience in processing difficulty, as found in monolingual parsing. Considering models such as MacDonald's (2013) PDC model, producing and comprehending speech is in and of itself a difficult task, but what matters is not its difficulty but rather the ways in which the processing and production mechanisms deal with this difficulty. As such, the question of why switch costs occur may not be of particular theoretical interest; rather, the nature of how those switch costs are resolved may prove more fruitful in the study of codemixed processing and production (Adler, Valdés Kroff, & Novick, 2017).

The results of the present study suggest that language dominance and codemixing experience combine to affect how both unilingual and codemixed sentences are processed. A high density of codemixes clashes with patterns found in bilingual speech, resulting in processing difficulties; 'thinning out' codemixes with intervening unilingual material seems to lessen this burden. Thus, it seems that the processing strategies used by participants are indeed sensitive to the way in which stimuli are presented; these effects appear to be strongest at later stages of processing, as suggested by the effects of mode of presentation on the right bounded and total duration measures.

Single-noun insertions: Codemixing or borrowing?

We return now to a question raised earlier concerning whether the single-noun insertions employed in this study can, in fact, be considered codemixes. The construction under study has been considered by some researchers to be an instance of insertional codemixing (Muysken, 2000, 2013; see also Myers-Scotton, 1993, pp. 180–183). Since the late 1980s, however, there has been extensive debate on the status of these single-item insertions (sometimes referred to as *lone other-language items*, or LOLIs) in spontaneously occurring bilingual codemixing.

Poplack and her colleagues (Poplack et al., 1988; Poplack & Meechan, 1998; Sankoff et al., 1990) have proposed that LOLIs overwhelmingly behave like established borrowings (e.g., like *troca* for 'truck', a borrowing from English into Spanish attested in the New Mexican Spanish-English bilingual corpus (NMSEB)) and overwhelmingly *unlike* multi-word stretches of code-mixing. Poplack (2012, p. 644) states that the 'accepted twin assumptions of the era' are that established loanwords are fully integrated into the recipient language, while codemixes are not. Thus, LOLIs that are morphosyntactically integrated into the recipient language despite their low frequency of occurrence or diffusion throughout the speech community (see Poplack et al., 1988, p. 72) cannot be codemixes. These have instead been termed *nonce borrowings* because they occur infrequently enough (in the original definition, only once in a corpus) to not be considered established borrowings, but nonetheless behave like them. Poplack (2012, p. 645) states that 'speakers not only code-switch spontaneously, but may also BORROW spontaneously' (emphasis original). This has been deemed the Nonce Borrowing Hypothesis (NBH).

It is possible that because the English nouns in the current study were presented as LOLIs—that is, immediately preceded and followed by Spanish—that participants may have treated them like nonce borrowings instead of codemixes. Although this is certainly a possibility, testing this hypothesis within the current study is not possible. Firstly, the NBH was originally formulated to explain a phenomenon based upon the analysis of large corpora of spontaneous speech production. Here, we test comprehension; without access to the complete linguistic history of our participants, it is not possible to determine whether these items are nonce *to the participants*. Secondly, the status of nonce borrowings as being morphosyntactically integrated or not into the recipient language (in this case Spanish) was not manipulated, but was instead held constant: all English nouns were preceded by a gender-congruent Spanish determiner, and since no adjectives were used, word order (another common criterion for integration; see Poplack et al., 1988, pp. 69–70) was congruent across languages. Likewise, all nouns appeared in the singular, so no morphological inflection from Spanish was present, further obscuring their status as integrated or not. Therefore, because we have no comparison between integrated and unintegrated English nouns *nor* between these single-word insertions and multi-word stretches of codemixing, it is difficult to determine the status of these items as codemixes or borrowings in the present study under the NBH framework. Nonetheless, the finding that these single-noun insertions resulted in distinct processing patterns compared to their unilingual counterparts suggests that there is more at play. Given the limitations of the experiment, however, we remain agnostic on the status of these items as codemixes or nonce borrowings, but leave the question of their processing to future research.

To mix, or not to mix?

The results of the present study suggest that the way bilinguals process codemixed and unilingual stimuli can differ depending on how the stimuli are presented and that sometimes this can result in unexpected findings. In particular, while many studies typically report switch costs associated with language-switching (see above), the present study found an *advantage* for codemixed sentences compared to their unilingual counterparts. The findings of this study then serve to illustrate how changing task demands come to affect online language processing, and that these effects should be taken into consideration with respect to both previous work on switch costs and future studies of codemixing.

In comparing these two modes of presentation, which are commonly used in lab-based studies of codemixing, another goal of the present study was to determine how each compared to naturalistic bilingual speech. One prediction was that the mixed mode of presentation would lead to overall lesser processing difficulty. When looking to corpus data, such as in the Bangor Miami corpus (Deuchar et al., 2014) or NMSEB (Torres Cacoullos & Travis, 2018), we see why: a blocked mode

of presentation is not representative of bilingual speech, where codemixing is locally introduced and otherwise surrounded by unilingual discourse. In fact, it is the mixed block that begins to more closely approximate these observations: codemixes occurred in 50% of the stimuli (as opposed to 100% in the codemixed-only portion of the blocked mode) and were broken up by intervening unilingual material. As such, it may be a mixed mode of presentation that should be pursued, as opposed to a blocked mode, although this may also vary given the goals of the experiment at hand. As Bialystok (2016) argues, one must not only understand the population one is surveying, but also select the right measure (or constellation of measures) that allows the researcher to paint a fuller picture of the phenomenon at hand. Indeed, this line of thinking is not novel: Labov (1972, 2012) comments on the primacy of the speech community in both the study of the variable linguistic phenomenon and in the understanding of language use and acquisition. This perspective, however, has only more recently appeared in psycholinguistic studies.

What is important to note, however, is that precisely due to the usage-based assumptions that we advocate here, the pattern of results found here need not apply to all bilingual communities or language combinations. For example, Broersma (2009) mentions that English-Dutch codemixing can be particularly dense, given the high similarity between the two languages. Likewise, Muysken (2013) discusses that different codemixing strategies (insertion, alteration, congruent lexicalization, and backflagging) are adopted by bilinguals under different linguistic and social circumstances. Therefore, differences between a mixed versus a blocked mode of presentation may be exacerbated in some bilinguals, but negligible in others.

The importance of the speech community and language use in context has formed the foundation of sociolinguistic research. Given the social nature of codemixing, lab-based approaches serve to benefit greatly from the incorporation of these sociolinguistic insights in both the design and implementation of experiments. Future studies should take into consideration the way bilinguals codemix in their day-to-day lives, as well as the possibility that not all bilingual communities approach codemixing in the same way. By integrating these two approaches—the sociolinguistic and the psycholinguistic—it becomes possible to correlate findings from online language processing with patterns of language use, further illustrating the intimate connection between the two.

Acknowledgements

We would like to thank Aliza Winneg for her help in data collection, as well as Grant M. Berry and Matthew Carlson for their assistance with statistical analyses. We also thank the reviewers for their close reading of the paper and for insightful comments, and we have benefitted greatly from discussions with Rena Torres Cacoullos. Preliminary versions of this work were presented at Interdisciplinary Perspectives on Code-Switching, Bilingualism in the Hispanic and Lusophone World, and the 11th International Symposium on Bilingualism. All errors remain our own.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE1255832 to Michael A. Johns. This research was supported in part by NSF Grant BCS-1535124, NSF Grant OISE-1545900, NIH Grant HD082796, and NIH Grant HDO71758 to Paola Dussias, and a University of Florida Rothman Summer Fellowship to Jorge Valdés Kroff.

Notes

1. An anonymous reviewer noted that differences in processing may nonetheless emerge as a result of using written as opposed to aural stimuli. We agree, and posit that the difference between mixing and blocking would be *enhanced* when using aural stimuli, as greater exposure to spoken codemixing would yield stronger biases and predictions that, when violated, would result in more profound processing difficulty. We leave this question, however, to planned future studies.
2. Given that prepositional phrases form a new constituent while relative clauses are part of the noun phrase, an anonymous reviewer noted that this may affect the processing of the nouns, and result in notable differences in the spillover region. Only four items were followed by a relative clause. To address the reviewer's comment, we removed these items from the data and re-ran the analyses. We found no difference in the results; see Appendix 4.
3. The first author also conducted an analysis on a sample of approximately 10,000 intonation units extracted from the NMSEB and found that only 7% (739) contained a single-word insertion (either of English into Spanish or Spanish into English). Likewise, less than 3% (296) contained a multi-word codemix.

ORCID iD

Michael A Johns https://orcid.org/0000-0003-1868-8141

References

Aaron, J. E. (2015). Lone English-origin nouns in Spanish: The precedence of community norms. *International Journal of Bilingualism*, *19*(4), 459–480.

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition*, *19*(4), 689–698.

Adler, R. M., Valdés Kroff, J. R., & Novick, J. M. (2017). Does code-switching engage cognitive control? Paper presented at the *International Symposium on Bilingualism 11*, Limerick, Ireland, 11–15 June.

Baayen, H., Bates, D., Kliegl, R., & Vasishth, S. (2015). *RePsychLing: Data sets from psychology and linguistics experiments*. R package version 0.0.4. Retrieved from https://github.com/dmbates/RePsychLing

Balukas, C., & Koops, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*, *19*(4), 423–443.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*. Retrieved from https://arxiv.org/pdf/1506.04967.pdf

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

Baus, C., Costa, A., & Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychologica*, *142*(3), 402–409.

Bialystok, E. (2016). The signal and the noise: Finding the pattern in human behavior. *Linguistic Approaches to Bilingualism*, *6*(5), 517–534.

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual language switching in the lab vs. in the wild: The spatio-temporal dynamics of adaptive language control. *Journal of Neuroscience*, *37*(37), 9022–9036.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113.

Broersma, M. (2009). Triggered codeswitching between cognate languages. *Bilingualism: Language and Cognition*, *12*(4), 447–462.

Bultena, S., Dijkstra, T., & Van Hell, J. G. (2015). Language switch costs in sentence comprehension depend on language dominance: Evidence from self-paced reading. *Bilingualism: Language and Cognition*, *18*(3), 453–469.

Bybee, J. L. (2010). *Language, usage and cognition*. Cambridge, UK: Cambridge University Press.

Bybee, J. L., & Beckner, C. (2010). Usage-based theory. In B. Heine, & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 827–856). Oxford, UK: Oxford University Press.

Christiansen, M., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62. doi:10.1017/S0140525X1500031X.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (Vol. 27, pp. 341–372). Washington, DC: American Psychological Association.

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*(4), 491–511.

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*(3), 219–253.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 1–9.

Deuchar, M., Davies, P., Herring, J., Couto, M. P., & Carter, D. (2014). Building bilingual corpora. In E. Thomas, & I. Mennen (Eds.), *Advances in the study of bilingualism* (pp. 93–111). Bristol: Multilingual Matters.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, *45*(4), 1246–1258.

Dussias, P. E., Halberstadt, L., & Carlson, M. (in prep). *Short-term exposure can trigger changes in L1 parsing strategies*. Manuscript is in preparation.

Fairchild, S., & Van Hell, J. G. (2015). Determiner-noun code-switching in Spanish heritage speakers. *Bilingualism: Language and Cognition*, *20*(1), 150–161.

Fodor, J. D. (1998). Learning to parse. *Journal of Psycholinguistic Research*, *27*(2), 285–319.

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*(3), 640–665.

Green, D. W., & Wei, L. (2014). A control process model of code-switching. *Language, Cognition, and Neuroscience*, *29*(4), 499–511.

Gullberg, M., Indefrey, P., & Muysken, P. (2009). Research techniques for the study of code-switching. In B. E. Bullock, & J. A. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 21–39). Cambridge, UK: Cambridge University Press.

Gullifer, J. W., Kroll, J. F., & Dussias, P. E. (2013). When language switching has no apparent cost: Lexical access in sentence context. *Frontiers in Psychology*, *4*, 1–13.

Guzzardo Tamargo, R. E., Valdés Kroff, J. R., & Dussias, P. E. (2016). Examining the relationship between comprehension and production processes in code-switched language. *Journal of Memory and Language*, *89*, 138–161.

Hernandez, A. E., Dapretto, M., Mazziotta, J., & Bookheimer, S. (2001). Language switching and language representation in Spanish–English bilinguals: An fMRI study. *NeuroImage*, *14*(2), 510–520.

Herring, J. R., Deuchar, M., Couto, M. C. P., & Quintanilla, M. M. (2010). 'I saw the madre': Evaluating predictions about codeswitched determiner-noun sequences using Spanish–English and Welsh–English data. *International Journal of Bilingual Education and Bilingualism*, *13*(5), 553–573.

Hofweber, J., Marinis, T., & Treffers-Daller, J. (2016). Effects of dense code-switching on executive control. *Linguistic Approaches to Bilingualism*, *6*(5), 648–668.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*(1), 57–83.

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

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, *1*(1), 97–120.

Labov, W. (2012). What is to be learned: The community as the focus of social cognition. *Review of Cognitive Linguistics*, *10*(2), 265–293.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.

Liceras, J. M., Fuertes, R. F., Perales, S., Pérez-Tattam, R., & Spradlin, K. T. (2008). Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional–lexical mixings. *Lingua*, *118*(6), 827–851.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*(12), 1507–1515.

Litcofsky, K. A., & Van Hell, J. G. (2017). Switching direction affects switching costs: Behavioral, ERP, and time-frequency analyses of intra-sentential codeswitching. *Neuropsychologia*, *97*, 112–139.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 1–6.

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*(1), 25–40.

Montes-Alcalá, C. (2005). "Dear amigo": Exploring code-switching in personal letters. In L. Sayahi, & M. Westmoreland (Eds.), *Selected proceedings of the Second Workshop on Spanish Sociolinguistics* (pp. 102–108). Somerville, MA: Cascadilla Proceedings Project.

Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology*, *25*(5), 531–546.

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language*, *80*(2), 188–207.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge, UK: Cambridge University Press.

Muysken, P. (2013). Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition*, *16*(4), 709–730.

Myers-Scotton, C. (1993). *Duelling languages*. New York, NY: Oxford University Press.

Otheguy, R., & Lapidus, N. (2003). An adaptive approach to noun gender in New York contact Spanish. In R. Núñez-Cedeño, L. López, & R. Cameron (Eds.), *A Romance perspective on language knowledge and use: Selected papers from the 31st linguistic symposium on Romance languages (LSRL)* (pp. 209–229). Philadelphia, PA: John Benjamins Publishing.

Perfetti, C. A. (1994). Psycholinguistics and reading ability. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849–894). San Diego, CA: Academic Press.

Perrotti, L. (2017). *Investigating community norms and linguistic mechanisms in codeswitching: Bridging linguistic theory and psycholinguistic experimentation*. Unpublished doctoral dissertation, The Pennsylvania State University, PA, USA.

Piccinini, P., & Arvaniti, A. (2015). Voice onset time in Spanish–English spontaneous code-switching. *Journal of Phonetics*, *52*, 121–137.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en Español: Toward a typology of code-switching. *Linguistics*, *18*(7–8), 581–618.

Poplack, S. (2012). What does the Nonce Borrowing Hypothesis hypothesize? *Bilingualism: Language and Cognition*, *15*(3), 644–648.

Poplack, S. (2018). *Borrowing: Loanwords in the speech community and in the grammar*. Oxford, UK: Oxford University Press.

Poplack, S., & Meechan, M. (1998). Introduction: How languages fit together in codemixing. *International Journal of Bilingualism*, *2*(2), 127–138.

Poplack, S., Sankoff, M., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, *26*(1), 4–104.

Price, C. J., Green, D. W., & Von Studnitz, R. (1999). A functional imaging study of translation and language switching. *Brain*, *122*(12), 2221–2235.

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241–255.

Reali, F., & Christiansen, M. H. (2007). Word chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology*, *60*(2), 161–170.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D. Jr., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, *20*(5), 343–350.

Sankoff, D., Poplack, S., & Vanniarajan, S. (1990). The case of the nonce loan in Tamil. *Language Variation and Change*, *2*(1), 71–101.

Stammers, J. R., & Deuchar, M. (2012). Testing the nonce borrowing hypothesis: Counter-evidence from English-origin verbs in Welsh. *Bilingualism: Language and Cognition*, *15*(3), 630–643.

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*(3), 267–295.

Tamminga, M. (2016). Persistence in phonological and morphological variation. *Language Variation and Change*, *28*(3), 335–356.

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1-L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1412–1431.

Torres Cacoullos, R., & Aaron, J. E. (2003). Bare English-origin nouns in Spanish: Rates, constraints, and discourse functions. *Language Variation and Change*, *15*(3), 289–328.

Torres Cacoullos, R., & Travis, C. E. (2015). Gauging convergence on the ground: Code-switching in the community. *International Journal of Bilingualism*, *19*(4), 365–386.

Torres Cacoullos, R., & Travis, C. E. (2018). *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge, UK: Cambridge University Press.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing Subject and Object Relative Clauses: Evidence from Eye Movements. *Journal of Memory and Language*, *47*(1), 69–90.

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). Amsterdamn, Netherlands: John Benjamins Publishing Company.

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*(2), 163–198.

Valdés Kroff, J. R., Román, P., & Dussias, P. E. (2016). Not all codeswitches are costly: Examining semantic vs. language unexpectancy. Paper presented at *The Workshop on Interdisciplinary Perspectives on Codeswitching*, Cambridge, UK: Cambridge University.

Valdés Kroff, J. R., & Fernández-Duque, M. (2017). Experimentally inducing Spanish-English code-switching: A new conversation paradigm. In K. Bellamy, M. Child, P. González, A. Muntendam, & M. C. Parafita Couto (Eds.), *Multidisciplinary approaches to bilingualism in the Hispanic and Lusophone world* (pp. 209–231). Amsterdam, The Netherlands: John Benjamins.

Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Dussias, P. E. (2018). Experimental contributions of eye-tracking to the understanding of comprehension processes while hearing and reading code-switches. *Linguistic Approaches to Bilingualism, 8*(1). doi:10.1075/lab.16011.val

Valenzuela, E., Faure, A., Ramírez-Trujillo, A. P., & Barski, E. (2012). Gender and heritage Spanish bilingual grammars: A study of code-mixed determiner phrases and copula constructions. *Hispania*, *95*(3), 481–494.

Von Studnitz, R. E., & Green, D. W. (2002). The cost of switching language in a semantic categorization task. *Bilingualism*, *5*(3), 241–251.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250–271.
Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs". *Psychological Research*, *63*(3), 212–233.

Author biographies

Michael A Johns is a National Science Foundation Graduate Research Fellow in the Department of Spanish, Italian and Portuguese at The Pennsylvania State University. Using psycholinguistic methodologies such as eye-tracking and pupillometry, his research integrates sociolinguistic and usage-based perspectives to investigate the online processing and production of bilingual language mixing. Through this interdisciplinary approach, his work seeks to understand the strategies bilinguals employ to both control and integrate their two languages.

Jorge R Valdés Kroff is an Assistant Professor of Spanish and Linguistics at the University of Florida. He is the director of the Bilingual Sentence Processing Lab which uses eye-tracking and other behavioral measures to investigate how bilinguals adapt their parsing preferences to better anticipate upcoming code-switches. His recent work focuses on testing whether bilinguals tap into production asymmetries to guide comprehension and recruit greater engagement of cognitive control to rapidly integrate upcoming code-switches.

Paola E Dussias is a Professor of Spanish, Linguistics and Psychology at the Pennsylvania State University (U.S.A). She and her students use behavioral (eye tracking) and neuroscience (ERP) methods to examine the conditions under which syntactic information from one of the bilingual's languages influences processing in the other language. They also study the processing of code-switched language by conducting lab-based experimental studies that are informed by usage-based theories of language. Her work has been supported by grants from the National Science Foundation and the National Institutes of Health. Currently, she is the Principal Investigator on a multi-million dollar grant from the National Science Foundation to develop an international research program of training that enables Penn State undergraduate students, graduate students and post-doctoral fellows to pursue research abroad on bilingualism.

Appendix 1. Linear models.

Summary of main effects and interactions from model comparisons

Effect	Measure	Determiner		Target noun		Spillover	
		χ^2	p	χ^2	p	χ^2	p
Mode	Gaze	0.073	0.788	1.144	0.285	0.85	0.357
	Right	0.027	0.871	0.15	0.699	3.084	0.079
	Regression	0.371	0.543	0.009	0.925	5.617	0.018*
	Total	0.746	0.388	0.038	0.846	6.45	0.011*
Sentence type	Gaze	0.845	0.357	0.841	0.359	0	0.995
	Right	0.129	0.72	4.55	0.033*	0.007	0.932
	Regression	0.161	0.688	0.377	0.539	2.533	0.112
	Total	1.437	0.231	4.01	0.045*	0.185	0.667
Mode by sentence type	Gaze	0.02	0.887	0.973	0.324	0.073	0.787
	Right	0.05	0.824	0.174	0.701	1.136	0.287
	Regression	0.207	0.649	0.05	0.823	1.106	0.293
	Total	0.437	0.509	0.591	0.442	0.045	0.832

Target noun: right bounded duration

```
right5_max <- lmer(log(right) ~ cmode + cswitch + cmode:cswitch +
                   (1 + (cmode + cswitch + cmode:cswitch)|participant) +
                   (1 + (cmode + cswitch + cmode:cswitch)|group),
                   data = subset(right, word == 5), REML = F,
                   control = lmerControl(optCtrl = list(maxfun = 1e5)))
```

Target noun: total duration

```
total5_max <- lmer(log(total) ~ cmode + cswitch + cmode:cswitch +
                   (1 + (cmode + cswitch + cmode:cswitch)|participant) +
                   (1 + (cmode + cswitch)|group),
                   data = subset(total, word == 5), REML = F,
                   control = lmerControl(optCtrl = list(maxfun = 1e5)))
```

Spillover region: regression path duration

```
regression6_max <-lmer(log(regression) ~ cmode + cswitch + cmode:cswitch +
                   (1 + (cmode + cswitch)|participant) +
                   (1 + (cmode)|group),
                   data = subset(regression, word == 6), REML = F,
                   control = lmerControl(optCtrl = list(maxfun = 1e5)))
```

Spillover region: total duration

```
total6_max <- lmer(log(total) ~ cmode + cswitch + cmode:cswitch +
                   (1 + (cmode + cswitch + cmode:cswitch)|participant) +
                   (1 + (cmode + cswitch + cmode:cswitch)|group),
                   data = subset(total, word == 6), REML = F,
                   control = lmerControl(optCtrl = list(maxfun = 1e5)))
```

Appendix 2. Correlation summaries.

Correlations between Spanish proficiency and gaze durations at target noun

Mode	Sentence Type	Accuracy in Spanish PNT	DELE
Blocked	Codemixed	$p < 0.01$; $r = -0.81$	$p < 0.01$; $r = -0.62$
	Unilingual	$p < 0.01$; $r = -0.80$	$p < 0.01$; $r = -0.65$
Mixed	Codemixed	NS	$p < 0.01$; $r = -0.67$
	Unilingual	$p = 0.01$; $r = -0.61$	$p < 0.01$; $r = -0.64$

Correlations between Spanish proficiency and gaze durations in the spillover region.

Mode	Sentence Type	Accuracy in Spanish PNT	DELE
Blocked	Codemixed	$p = 0.03$; $r = -0.52$	$p < 0.01$; $r = -0.88$
	Unilingual	$p = 0.02$; $r = -0.56$	$p < 0.01$; $r = -0.88$
Mixed	Codemixed	$p = 0.01$; $r = -0.58$	$p < 0.01$; $r = -0.82$
	Unilingual	NS	$p < 0.01$; $r = -0.82$

Appendix 3. Target nouns and their translation equivalents.

Masculine nouns		Feminine nouns	
aceite	oil	bolsa	bag
algodón	cotton	cadena	chain
anillo	ring	caja	box
árbol	tree	calle	street
avión	plane	canción	song
baile	dance	cárcel	prison
baño	bathroom	carne	meat
barco	boat	carta	letter
barrio	neighborhood	casa	house
bosque	forest	cerveza	beer
calor	heat	ciudad	city
camión	truck	edad	age
cielo	sky	escuela	school
cinturón	belt	felicidad	happiness
concurso	competition	fiesta	party
corazón	heart	fuente	source
deseo	wish	guerra	war
ensayo	essay	habitación	room
espejo	mirror	jaula	cage
estadio	stadium	leche	milk
fuego	fire	llave	key
grito	scream	lluvia	rain
guante	glove	luna	moon
hielo	ice	madera	wood
hierro	iron	mente	mind
hueso	bone	miel	honey
jabón	soap	mirada	gaze
jardín	garden	mosca	fly
lago	lake	nieve	snow
lápiz	pencil	noche	night
lugar	place	nube	cloud
maíz	corn	pantalla	screen
mundo	world	pared	wall
peine	comb	pelea	fight
pincel	brush	piedra	stone
premio	prize	rama	branch
pueblo	village	red	net
puente	bridge	regla	rule
reloj	clock	revista	magazine
río	river	sangre	blood
ruido	noise	torre	tower
traje	suit	voz	voice

Appendix 4. Linear models excluding items with a relative clause.

Summary of main effects and interactions from model comparisons where items with target nouns followed by clausal modifiers have been removed

Effect	**Measure**	**Determiner**		**Target noun**		**Spillover**	
		χ^2	*p*	χ^2	*p*	χ^2	*p*
Mode	Gaze	0.072	0.788	0.976	0.323	0.947	0.330
	Right	0.003	0.959	0.085	0.771	3.311	0.069
	Regression	0.109	0.742	0.054	0.817	5.037	0.025
	Total	0.449	0.503	0.080	0.777	7.380	0.007
Sentence type	Gaze	0.821	0.365	0.992	0.319	0.019	0.889
	Right	0.004	0.949	5.277	0.022	0.019	0.891
	Regression	0.005	0.949	0.511	0.475	1.645	0.199
	Total	1.121	0.289	4.460	0.035	0.216	0.642
Mode by sentence type	Gaze	0.043	0.835	0.808	0.369	0.051	0.822
	Right	0.059	0.807	0.143	0.705	1.059	0.303
	Regression	0.443	0.506	0.029	0.866	1.123	0.289
	Total	0.339	0.560	0.419	0.517	0.036	0.849