# Effects of language mixing on bilingual children's word learning

Krista Byers-Heinlein* , Amel Jardak*, Eva Fourakis** and Casey Lew-Williams**

*Concordia University, Department of Psychology, Montreal, Canada and **Princeton University, Department of Psychology, Princeton, USA

## Abstract

Language mixing is common in bilingual children's learning environments. Here, we investigated effects of language mixing on children's learning of new words. We tested two groups of 3-year-old bilinguals: French–English (Experiment 1) and Spanish–English (Experiment 2). Children were taught two novel words, one in single-language sentences ("Look! Do you see the dog on the teelo?") and one in mixed-language sentences with a mid-sentence language switch ("Look! Do you see the *chien/perro* on the walem?"). During the learning phase, children correctly identified novel targets when hearing both single-language and mixed-language sentences. However, at test, French–English bilinguals did not successfully recognize the word encountered in mixed-language sentences. Spanish–English bilinguals failed to recognize either word, which underscores the importance of examining multiple bilingual populations. This research suggests that language mixing may sometimes hinder children's encoding of novel words that occur downstream, but leaves open several possible underlying mechanisms.

## Introduction

By definition, children growing up in bilingual environments encounter switches between their languages. Some switches occur when different people use different languages, while others occur when a bilingual speaker mixes two languages. Switches are more common in some families than others, and they can happen at both very short and very long intervals. Many switches even occur within individual sentences (see Ervin-Tripp & Reyes, 2005; Sridhar & Sridhar, 1980). In a natural-play study with Spanish–English bilingual families, all caregivers mixed two languages within sentences at least once, and some parents mixed languages in as many as one-third of their utterances (Bail, Morini & Newman, 2015). Indeed, parents sometimes produce mixed-language sentences even when they overtly try to avoid it (Bail et al., 2015; Goodz, 1989). Given that language mixing is a normal part of bilingual environments, it is important to understand its relationship to children's language learning. Here, we investigated whether and how language mixing affects 3-year-old children's learning of new words.

Language mixing has been shown to affect toddlers' comprehension of familiar words. In an eye-tracking paradigm, 20-month-old French–English bilinguals saw two familiar objects on a screen (e.g., dog, book), and one object was labeled in either a single-language sentence ("Find the dog!") or a mixed-language sentence ("Find the *chien*!"; Byers-Heinlein, Morin-Lessard & Lew-Williams, 2017). Toddlers were less accurate in looking to the target object when they heard it labeled in the mixed-language sentence than in the single-language sentence, particularly when switching occurred from their dominant to non-dominant language. This finding has been replicated with Spanish–English bilingual toddlers (Morini & Newman, 2019; Potter, Fourakis, Morin-Lessard, Byers-Heinlein & Lew-Williams, 2019). However, not all types of language mixing affected word comprehension: when the language switch occurred at a sentence boundary, toddlers' looking to the target object was not disrupted (Byers-Heinlein et al., 2017).

If language mixing can affect the processing of familiar words, then this could affect the processing of subsequent unfamiliar words. Support for this idea comes from studies of incremental, real-time language processing with monolinguals. For example, in one study, toddlers heard sentences that ended with a word that was unfamiliar to them, such as "Mommy feeds the ferret". Their looking patterns indicated that they were able to use the verb "feeds" to infer that "ferret" was an animal rather than an inanimate object (Goodman, McDonough & Brown, 1998). Children's efficiency in using familiar words to facilitate processing of subsequent words has been linked to vocabulary size and later vocabulary growth in both monolinguals (Fernald & Marchman, 2012; Fernald, Perfors & Marchman, 2006; Lany, 2018; Lew-Williams & Fernald, 2007) and bilinguals (DeAnda, Hendrickson, Zesiger, Poulin-Dubois & Friend, 2018; Hurtado, Grüter, Marchman & Fernald, 2014; Legacy, Zesiger, Friend & Poulin-Dubois, 2016; Marchman, Fernald & Hurtado, 2010). Based on this work,

we hypothesized that the momentary processing delays associated with language mixing would impact the learning of a novel word that appeared later in the sentence.

To date, few studies have investigated links between language mixing and vocabulary growth in young children, and their results are equivocal. One large study ($N = 181$) with children aged 18 and 24 months learning English and a variety of other languages found that parents who reported more language mixing often had children with smaller vocabulary sizes (Byers-Heinlein, 2013). A smaller study did not find clear links between parental language mixing in the laboratory and vocabulary size in 18- to 24-month-olds (Bail et al., 2015). Two studies that measured mixing more indirectly (i.e., as the proportion of 30-minute blocks of time where both English and Spanish were spoken) also reported no links between mixing and 2.5-year-old children's language development (Place & Hoff, 2011, 2016). The correlational nature of these studies makes it difficult to establish causation, as they cannot separate effects of mixing from other potentially confounding factors, such as parents' language proficiency, or children's absolute quantity of language exposure.

The goal of the current research was to examine whether and how mid-sentence language mixing might affect children's learning of a novel word that is spoken moments later. Given that language mixing is a common practice both within and across sentences and that children are exposed to thousands of new words throughout the first years of life, we sought to understand the intersection of these important aspects of bilingual development.

As a secondary goal, we sought to understand the generalizability of our results across different groups of bilingual learners. Testing two groups of bilingual toddlers in the same paradigm has been quite rare in the literature, likely because of the challenge of recruiting and testing such populations (Byers-Heinlein et al., 2020). Our strategy was to test the most representative populations of bilinguals in two communities: French–English bilingual children growing up in Montreal, Canada (Experiment 1), and Spanish–English bilingual children growing up in New Jersey, USA (Experiment 2). Previous work has demonstrated that language mixing affects real-time word comprehension among children from both of these communities, which sets the stage for informed interpretation of data in the current study (Byers-Heinlein et al., 2017; Morini & Newman, 2019; Potter, Fourakis, Morin-Lessard, Byers-Heinlein & Lew-Williams, 2018; Potter et al., 2019)

French–English in Montreal and Spanish–English bilinguals in New Jersey differ along multiple dimensions, and here we highlight three particularly important differences. Most obviously, the two groups are learning different language pairs. However, because French and Spanish are typologically similar languages, we would not necessarily expect different effects of language mixing on word learning. The two groups also differ in terms of family/community norms for language use. Prior studies suggest that parents of Spanish–English bilinguals from some communities may code-switch more often in interactions with their infants and toddlers than parents of French–English bilinguals in Montreal (Bail et al., 2015; Kremin, Alves, Orena, Polka & Byers-Heinlein, 2020). Thus, it is possible that Spanish–English bilinguals would have more everyday exposure to the type of code-switched stimuli that we included in our study, which could make it easier for them to process (Kroff, Dussias, Gerfen, Perrotti & Bajo, 2017). Finally, French–English bilingual participants in Montreal are generally of mid-to-high socioeconomic status (SES), while Spanish–English bilingual participants in New Jersey have a wider range of SES, including many families of lower SES. Previous research

using related experimental tasks has found that toddlers of higher SES have stronger language processing skills than those of lower SES (Fernald & Marchman, 2012; Fernald, Marchman & Weisleder, 2013), which makes it plausible that French–English bilinguals may demonstrate more successful word learning in our task compared to Spanish–English bilinguals. Although any differences observed between the two populations could be attributed to one or more of these dimensions of difference, our comparison is an important first step for evaluating the generalizability of findings across communities of bilingual learners.

Our experiments tested 3-year-old children using an eye-tracking task with a learning phase and a test phase, and was based on an existing paradigm designed to test how slow-downs in processing of one word might affect learning of a subsequent word (Fernald, Marchman & Hurtado, 2008). During the learning phase, children were taught two novel words. One novel word always appeared at the end of a single-language sentence (e.g., "Look! Do you see the dog on the teelo?"), and the second novel word always appeared at the end of a mixed-language sentence, i.e., a sentence containing one word borrowed from the other language (e.g., "Look! Do you see the *chien/perro* on the walem?"). We chose to code-mix at the point of a familiar word that preceded a novel word, rather than at the novel word itself (e.g., "Do you see the *tileau*?"), for two reasons. First, this enabled us to examine effects of language mixing on downstream learning, which would not be possible if the switch occurred at the novel word itself. Second, this allowed children to use both their existing lexical knowledge as well as phonological cues to detect that a code switch had occurred. Given that this is the first study to examine this research question, we aimed to give children as many cues as possible. In our design, the novel word always appeared in the same language in which the sentence began, and this language was randomly assigned to be children's dominant or non-dominant language. While listening to these sentences, children viewed pairs of familiar animals on top of novel objects (e.g., a bunny atop a novel purple object, and a dog atop a novel green object; see Figure 1). The visual scene required children to accurately identify the familiar referent in order to then locate the target novel referent. That is, children needed to first look to the dog in order to determine which of the two novel objects was being labeled. During the subsequent test phase, the two novel objects appeared side by side, with no familiar objects present, and children were assessed on their learning of label-object mappings.

Our main prediction was that children would show better learning of the novel word encountered in single-language sentences than in mixed-language sentences. Based on previous findings (Byers-Heinlein et al., 2017; Morini & Newman, 2019; Potter et al., 2019), we hypothesized that this would be due to a temporary disruption in language processing upon hearing the borrowed word. Specifically, upon hearing a language switch, children would be less able to identify and encode the mapping between the novel label and novel object. This would be evident in children's behavior during the learning phase. Specifically, we predicted that on mixed-language trials (relative to single-language trials), children would be less successful in identifying the correct familiar object and, in turn, would have spent less time gazing at the target novel object.

We also considered several other potential results. First, it is possible that even when language mixing disrupts real-time processing, young children's uptake of new words is unaffected. Second, it is possible that we would not replicate previously reported effects of language mixing on language comprehension. In this case, we would either find that language mixing does not
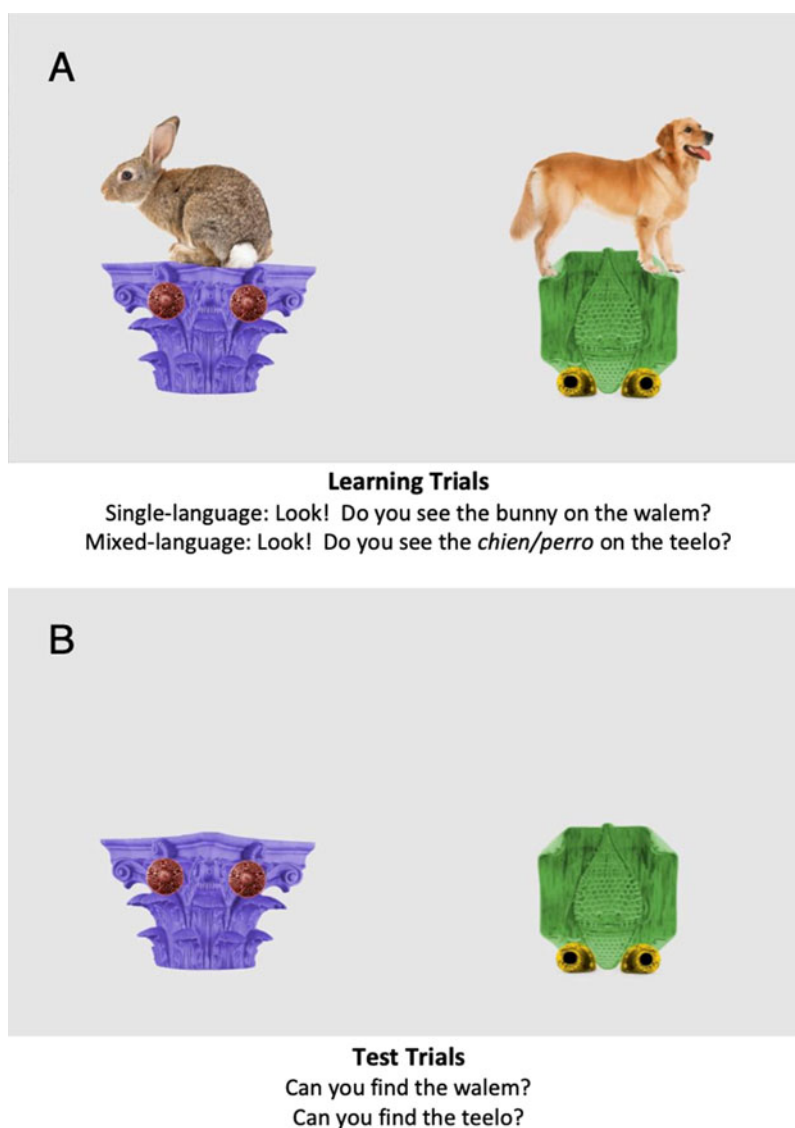
**Figure 1.** Auditory and visual stimuli for learning trials (A) and test trials (B) in Experiment 1 (French–English bilinguals) and Experiment 2 (Spanish–English bilinguals).

affect word learning, or possibly that it does affect word learning but not due to in-the-moment processing difficulties during the learning phase. Finally, considering the population differences noted above, French–English and Spanish–English bilingual children might show different patterns of looking, listening, and learning. We thus not only compared the two groups, but also explored whether individual differences in performance during the learning phase, exposure to language mixing at home, vocabulary size, or socioeconomic status predicted successful word learning in both the single-language and the mixed-language conditions.

## Experiment 1: French–English bilinguals in Montreal

### Method

The methods of Experiment 1 were approved by the Concordia University Human Research Ethics Committee ("Monolingual and Bilingual Language Development"; approval #10000493). Data were collected between November 2016 and April 2017, and parents provided informed consent prior to their child's

participation. As is common in laboratories that test hard-to-recruit populations such as bilingual children, children participated in a second, separate study, either immediately prior to or following participation in this experiment (the order of the two studies was counterbalanced). The results of that study are not reported here. Stimuli, data, and analysis scripts are available via the Open Science Framework at https://osf.io/q2nzr/.

The general concept for the study design was inspired by Fernald et al. (2008), which examined how children's processing of adjective-noun sequences can sometimes derail children's learning of novel words that appear later in the sentence (e.g., "There's a blue cup on the deebo"). Here, children heard a sentence either with or without a mid-sentence language switch, which was then followed by a novel noun.

### Participants

Children were recruited from a database of families interested in participating in research, principally identified via government

birth lists. A total of 19 healthy French–English bilingual children were included in the final sample. A sensitivity analysis indicated that this sample size has 80% power to detect an effect size of $dz$ = .68 or greater in a two-tailed repeated measures $t$-test, which is the average effect size observed in word recognition paradigms in children older than 16 months according to MetaLab (https://metalab.stanford.edu; Bergmann et al., 2018). All children were growing up in Montreal. Both French and English are widely spoken in the community, and the community generally holds positive attitudes toward both languages (Kircher, 2014). Childcare and elementary education are widely available in both languages.

Children had an average age of 3.50 years (range: 3.10–4.10). There were 12 males and 7 females. A modified version of the Language Experience and Proficiency Questionnaire (Marian, Blumenfeld & Kaushanskaya, 2007) was used to assess the children's language background and proficiency. As part of the questionnaire, parents were asked to rate their child's proficiency in French and in English. A predetermined inclusion criterion was that children were rated at least 7/10 in comprehension in both languages. Language dominance was established for each child as the language that was rated the highest for comprehension. When ratings were equal in both languages, the child's dominant language was considered to be the mother's dominant language. There were 10 English-dominant children and 9 French-dominant children. Age of acquisition of the languages varied across the sample: there were 8 children who were regularly exposed to both English and French from birth (i.e., were simultaneous bilinguals), and 11 who began regular exposure to their second language later in life (in all cases prior to age 18 months), usually upon starting daycare. Another 10 children were tested but not included in the final sample due to fussiness ($n = 3$), reported health or developmental issues ($n = 3$), not meeting pre-established language criteria ($n = 3$), or because of technical difficulties ($n = 1$).

Children's exposure to intra-sentential language mixing was measured via the Language Mixing Scale, a parent-report questionnaire that assesses the frequency of language mixing by the primary bilingual caregiver (Byers-Heinlein, 2014). The scale ranges from 0 (no language mixing) to 30 (frequent language mixing). The average score in this sample was 13.50 (range: 4–28).

Socioeconomic status was assessed via mothers' highest educational attainment. All mothers had completed high school. Specifically, 4 mothers had completed a graduate degree, 9 had completed a bachelor's degree, 5 had completed other post-secondary training (e.g., college diploma, some university, trade school), and 1 had completed high school but had not received further formal education. For subsequent analysis, these categories were then converted to approximate years of education (e.g., high school = 12 years, bachelor's degree = 16 years). Mothers' average educational attainment was 15.80 years ($SD = 2.50$).

Children's productive vocabulary was measured in each of their languages. We assessed vocabulary via a parent report measure, rather than using a direct measure such as the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007). This choice was made in order to reduce the length of the experimental session for children, as they were also participating in a second eye-tracking study. Moreover, some of our research assistants did not have sufficient proficiency to administer the test in both languages. We chose a vocabulary measure appropriate for this age group, the Developmental Vocabulary Assessment for Parents (DVAP; Libertus, Odic, Feigenson & Halberda, 2015), which consisted of a checklist of words known by children ages

2–18 years based on words used in the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007). The DVAP shows a high level of concurrent and predictive validity (Libertus et al., 2015). To assess children's vocabulary size in French, we adapted a checklist similar to the DVAP, based on words used in the French adaptation of the PPVT (Échelle de vocabulaire en images Peabody; Dunn, Dunn & Thériault-Whalen, 1993). The words are ordered from easy (e.g., boy, chair) to hard (e.g., honing, angler), and parents were asked to indicate which words their child could say. A parent who was familiar with the child's vocabulary in a particular language filled out the form in that language. In some cases, each parent filled out a form in the language they usually used to interact with their child, while in other cases the same parent filled out both forms. One parent, who was not familiar with the language her child used at daycare, had the form filled out by a daycare worker. Data were missing for one language for one participant. Of the words on the DVAP, toddlers produced an average of 77 words in their dominant language (range: 37–177) and 48 words in their non-dominant language (range: 2–131), [$t(17) = 6.08$, $p < .001$, $d = 1.43$]. The average number of words produced across both languages was 125 (range: 39–308).

### Visual stimuli

Visual stimuli consisted of three animals (bunny, dog, fish) and two novel objects created for this study. Familiar animals were selected among words known by at least 50% of 18-month-olds in English, and at least 50% of 16-month-olds in English (see Fenson et al., 2007) and Québec French (Boudreault, Cabirol, Trudeau, Poulin-Dubois & Sutton, 2007; Frank, Braginsky, Yurovsky & Marchman, 2017). None of the animals or their French translations had phonological overlap at onset, nor were they cognates. Images of the animals were chosen from free online libraries. Images of platform-like novel objects were found online and digitally modified, such that they did not closely resemble any known object.

On learning trials, the objects were combined in a scene with each familiar object standing or sitting on top of a different novel object, with one novel–familiar object pair located on each side of the screen (Figure 1). The placement of familiar and novel objects was counterbalanced within and across children such that children saw two different familiar objects on each trial, but these varied across trials. On test trials, only the two novel objects were presented, without animals, and their left–right location was counterbalanced both within and across children.

### Auditory stimuli

Auditory stimuli were recorded in a child-directed manner by a female native bilingual French–English speaker with no perceptible accent in either language. Two nonsense words, "walem" ("walème" in French) and "teelo" ("tileau" in French), were chosen to label the novel objects. They were produced with French phonology when embedded in a French carrier phrase, and English phonology when embedded in an English carrier phrase, as judged by native speakers of each language. The two words were carefully chosen to each have two syllables, different phonological onsets, different rhymes, and sound masculine in French grammatical gender based on the judgment of native speakers. Further, their onsets were different from the familiar objects' onsets. The speaker recorded several versions of each stimulus sentence, and the final stimuli were chosen so that the single-

language and mixed-language sentences were as similar as possible in length, pitch, prosody, and naturalness.

During the learning phase, each auditory stimulus consisted of a carrier phrase ("Look! Do you see the…" or *"Regarde! Vois-tu le…"*) that was followed by a familiar target word (bunny, dog, or fish) and a novel target word. Each sentence then labeled one of the animals and its corresponding novel object, such that children had to process the familiar word in order to determine the referent of the novel word. Stimuli were produced naturalistically, in either a single-language context where the carrier phrase, familiar target word, and novel target word were in the same language (e.g., "Look! Do you see the dog on the teelo?" or *"Regarde! Vois-tu le chien sur le tileau?"*), or in a mixed-language context where the familiar word was in a different language from the carrier phrase and the target novel word (e.g., "Look! Do you see the *chien* on the teelo?" or *"Regarde! Vois-tu le* dog *sur le tileau?"*). The familiar word was always pronounced with the phonology of the language it appeared in; "dog" was pronounced with an English accent, and *"chien"* was pronounced with a French accent. The target novel word was always pronounced with the phonology of the carrier phrase language. The choice to present the code-switch following the determiner was consistent with previous experiments (Byers-Heinlein et al., 2017; Potter et al., 2019). Moreover, direct observations of French–English bilingual parents' code switching suggests that they are equally likely to switch within a syntactic phrase as they are to switch between syntactic phrases (Kremin et al., 2020).

During the test phase, auditory stimuli consisted of simple sentences that directed children's attention to a target novel object. Sentences during this phase were all in either English (e.g., "Can you find the teelo/walem?") or French (e.g., *"Peux-tu trouver le tileau/walème?"*), depending on which language corresponded to the carrier phrase that the child encountered during the learning phase.

## Procedure

Children were seated on their parents' laps or directly on a chair in front of a 24-inch Tobii T60XL corneal reflection eye-tracking system which collected gaze data at a rate of 60 Hz (Tobii Group, Stockholm, Sweden) located in a sound-attenuated, dimly-lit room, which was used to present the stimuli and record children's eye gaze. Parents were asked to wear darkened sunglasses and headphones so they would be unaware of the stimuli. They were also asked not to talk or interact with the child during the experiment. Calibration was completed using a built-in five-point calibration routine. The sound was set at an average of 70 dB. Children viewed the stimuli in one of four randomly assigned experimental orders consisting of 12 learning trials (6 single-language sentences, 6 mixed-language sentences) immediately followed by 4 test trials. A colourful attention-getter was presented prior to each trial to reorient the child to the screen. Children were randomly assigned to hear the carrier phrases of both the learning and test phases either always in English or always in French. This resulted in 9 children being tested in their dominant language, and 10 children being tested in their non-dominant language. A sample trial order for a single child is presented in the Appendix.

During each of 12 learning trials, the two novel objects appeared side by side, with a different familiar object on top of each of them (e.g., a dog on the teelo, and a bunny on the walem). After 2000 ms, the auditory stimulus directed the child to look at the familiar target atop the novel target, either in the single-language or mixed-language sentence. The familiar word onset occurred exactly at 4500 ms, and the audio for each trial ended at around 6500 ms, leaving approximately 2000 ms of looking time after the target novel offset; the total length of each learning trial was 8500 ms. Each novel object was consistently labeled in a single-language or mixed-language sentence, and the assignment of label to sentence type was counterbalanced across children. Each familiar target in the learning phase appeared eight times, counterbalanced within children for whether it served as a target or a distractor, and whether it was encountered in a single-language or mixed-language sentence. The two novel objects appeared on every trial, half as target and half as distractor for each child. Assignment of label–object pairs was counterbalanced by sentence type across participants (i.e., half of participants consistently encountered "teelo" in single-language sentences and "walem" in mixed-language sentences, while the other half consistently encountered "walem" in single-language sentences and "teelo" in mixed-language sentences). Trial order during the learning phase was semi-randomized.

During each of four test trials, only the two novel objects appeared on the screen, presented side by side without their corresponding familiar objects. The novel objects appeared in silence for approximately 2000 ms, and then the auditory stimulus directed children's attention to a target object such that the target novel word onset occurred at exactly 3000 ms. Each test trial lasted 6000 ms. Each novel object appeared twice as target and twice as distractor, and the side of presentation was counterbalanced both within and across children.

The experimenter monitored the status of the experiment via the eye tracker's built-in camera and controlled the experiment from a computer in an adjacent room using Tobii Studio software. The total duration of the experiment was approximately 4 minutes. Parents completed questionnaires either before or after the eye-tracking experiment.

## Results and Discussion

Analyses were conducted in R (R Core Team, 2018), primarily using the eyetrackingR package (Dink & Ferguson, 2015).

### Test phase

To determine if children successfully learned each of the words presented during the learning phase, we examined the proportion of time that they looked to the labeled object, averaged across the two test trials of each type. Children's looking was measured within areas of interest defined around each object. Following previous approaches, data for each trial were collapsed across a time window that began 360 ms after the onset of the target word label and ended 2000 ms after, such that the duration of the window of analysis was 1640 ms (Swingley, 2012). Trials where the child was inattentive were excluded from analyses (i.e., when they looked at the two objects together for less than 750 ms). Children completed an average of 1.79 out of 2 trials testing the novel word that was encountered during single-language sentences in the learning phase (range: 1–2) and 1.53 out of 2 trials testing the novel word that was encountered during mixed-language sentences (range: 1–2).

Our main analysis directly compared children's performance at test across the two conditions. A two-tailed paired-samples *t*-test indicated that children looked significantly longer at the target object when its label had been encountered in single-language sentences [$M = 0.64$, $SD = 0.05$] than when its label had been
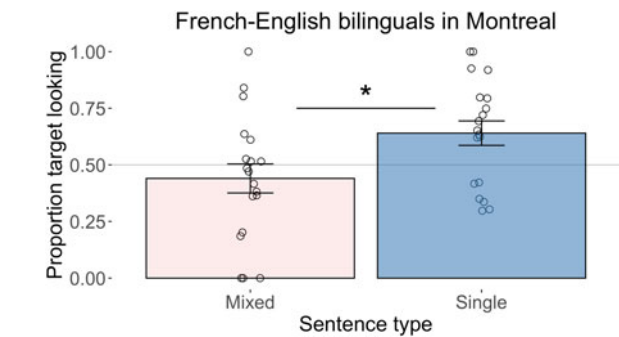
Figure 2. Average proportion looking to target object for French–English bilinguals (Experiment 1). Circles show individual data points. Bars represent standard error of the mean.

**Table 1.** French–English bilinguals (Experiment 1): Descriptive statistics for looking times during learning phase.

| Target type | Sentence type | Familiarity | Mean | SD |
|---|---|---|---|---|
| target | single | familiar | 0.31 | 0.15 |
| target | single | novel | 0.44 | 0.18 |
| target | mixed | familiar | 0.32 | 0.12 |
| target | mixed | novel | 0.47 | 0.18 |
| distractor | single | familiar | 0.08 | 0.07 |
| distractor | single | novel | 0.09 | 0.09 |
| distractor | mixed | familiar | 0.10 | 0.08 |
| distractor | mixed | novel | 0.07 | 0.05 |

encountered in mixed-language sentences [$M = 0.44$, $SD = 0.06$, $t(18) = 3.01$, $p = 0.01$, $d = 0.69$]. Children's looking to the target novel object was significantly above chance (.5) when its label had been heard in single-language sentences, [$t(18) = 2.69$, $p = 0.01$, $d = 0.62$], indicating successful recognition. However, children's looking to the target novel object was not significantly above (or below) chance when its label had been heard in mixed-language sentences, [$t(18) = -0.97$, $p = 0.34$, $d = -0.22$], indicating lack of successful recognition. Together, these results suggest that mid-sentence mixing of an informative familiar word can block three-year-olds' learning of a novel word that follows moments later. Results are illustrated in Figure 2.

### Learning phase

To better understand children's overall unsuccessful learning of the word presented in mixed-language sentences, we investigated their looking patterns during the learning phase. We expected that language mixing would affect learning in this study by slowing children's processing of the familiar word, which, in turn, would block children from having sufficient time to encode the novel word–object pair. If correct, this would manifest as less looking to the labeled objects (both familiar and novel) during mixed-language sentences than during single-language sentences.

We computed the proportion of time that children looked toward each of the four objects onscreen: the familiar target, the novel target, the familiar distractor, and the novel distractor. As during the test phase, data for each trial were collapsed across a time window 360–2000 ms after the familiar word onset. We note that while it is possible that children would need additional time to process the novel word that followed the familiar word, this time window afforded maximum sensitivity for detecting potential differences in the single-language versus mixed-language contexts (Swingley, 2012). Trials where the child was inattentive were excluded from these analyses (i.e., when they looked at the four objects on the screen for less than 750 ms total). Children contributed data for an average of 5.53 single-language trials (range: 3–6) and 5.63 mixed-language trials (range: 4–6) out of 6 possible trials per type. Means and standard deviations of proportion looking to each object type (target familiar, target novel, distractor familiar, distractor novel) during single-language and mixed-language sentences are displayed in Table 1.

Proportion of looking to the objects was analyzed with a 2 (target type: target, distractor) x 2 (familiarity: familiar, novel) x 2 (sentence type: single-language, mixed-language) repeated measures ANOVA. Statistics are displayed in Table 2; note that chance looking is not 0.5, as there were two groups of two objects on the screen.

There were significant main effects of familiarity (greater looking at novel than familiar objects) and target type (greater looking at target than distractor objects). There was also a significant interaction between familiarity and target type: looking times to target objects were higher overall, with significantly greater looking to novel than to familiar targets, and looking times to distractor objects were lower overall, with no significant difference in looking to novel vs. familiar distractors. Critically, there were no effects or interactions with sentence type, indicating similar patterns of looking across single-language and mixed-language sentences. Across both single-language and mixed-language sentences, children looked significantly longer at the target objects than the distractor objects after hearing the label ($p$s < .05). Visual inspection of time course plots (Figure 3) confirms that on both single-language and mixed-language sentences, children showed similar increases in looking to the labeled objects after naming. These analyses suggest that language mixing did not affect children's accuracy in looking to the labeled targets during the learning phase.

## Experiment 2: Spanish–English bilinguals in New Jersey

In Experiment 1, we found that mid-sentence language mixing reduced French–English bilingual children's abilities to encode a sentence-final novel word. Instead of using these findings to draw conclusions about the nature of bilingual development, we examined their generalizability to a different bilingual population. The purpose of Experiment 2 was to study how language mixing affects word learning among Spanish–English bilingual 3-year-olds in New Jersey.

### Method

The method of Experiment 2 was approved by the Princeton University Institutional Review Board ("Language learning and communication"; approval #7117). Data were collected between March 2017 and January 2018, and parents provided informed consent prior to their child's participation. As in Experiment 1, participants in Experiment 2 participated in an additional experiment either prior to or after participating in this experiment. Stimuli, data, and analysis scripts are available via the Open Science Framework at https://osf.io/q2nzr/.

### Participants

A total of 21 healthy Spanish–English bilingual children were included in the final sample. All children were growing up in

**Table 2.** French–English bilinguals (Experiment 1): ANOVA for learning trials.

| | Effect | DFn | DFd | *F* | *p* | *p* < .05 | ges |
|---|---|---|---|---|---|---|---|
| 2 | sentence type | 1.00 | 18.00 | 0.76 | 0.39 | | 0.00 |
| 3 | familiarity | 1.00 | 18.00 | 5.05 | 0.04 | * | 0.07 |
| 4 | target type | 1.00 | 18.00 | 141.90 | 0.00 | * | 0.62 |
| 5 | sentence type * familiarity | 1.00 | 18.00 | 0.06 | 0.81 | | 0.00 |
| 6 | sentence type * target type | 1.00 | 18.00 | 0.49 | 0.49 | | 0.00 |
| 7 | familiarity * target type | 1.00 | 18.00 | 9.94 | 0.01 | * | 0.09 |
| 8 | sentence type * familiarity * target type | 1.00 | 18.00 | 0.56 | 0.47 | | 0.00 |
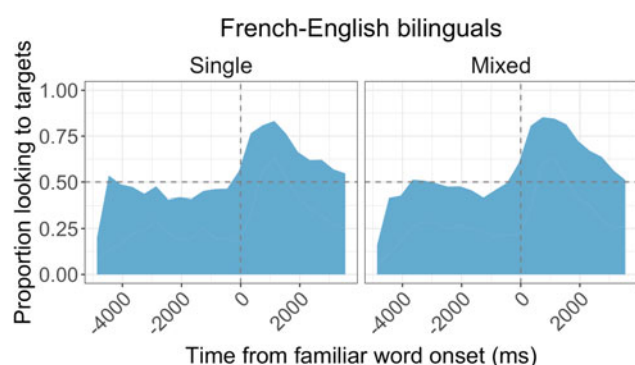


**Figure 3.** Timecourse of looking to the labeled target for French–English bilingual children (Experiment 1).

New Jersey. English is the majority language and Spanish is the minority language in this region. Participants were primarily recruited via various nonprofit organizations in the region.

Children had an average age of 3.70 years (range: 3.20–4.10). There were 9 males and 12 females. As in Experiment 1, a modified version of the Language Experience and Proficiency Questionnaire (Marian et al., 2007) was used to assess the children's language background and proficiency. Inclusion criteria and definitions of language dominance were the same as in Experiment 1. There were 8 English-dominant children and 13 Spanish-dominant children. For Spanish, all children were exposed regularly from birth. For English, 8 children were exposed regularly from birth (i.e., were simultaneous bilinguals), 9 were exposed after birth but within the first year of life, and 4 were exposed between 12 and 36 months. Another 5 children were tested but not included in the final sample due to not meeting pre-established inclusion criteria for bilingual exposure (*n* = 4) or reported language delay (*n* = 1).

As in Experiment 1, children's exposure to parental language mixing was measured via the Language Mixing Scale Score (Byers-Heinlein, 2014). The average score in this sample was 16.50 (range: 0–30). This was not statistically different from the amount of mixing reported by parents in Experiment 1 who had a mean of 13.50 [*t*(38) = 1.18, *p* = .244, *d* = 0.37].

Socioeconomic status was assessed via mothers' highest educational attainment: 7 had completed a graduate degree, 4 had completed some university, 5 had completed high school, 3 had not completed high school, and 2 did not report. That is, among those reporting, 56% of mothers of Spanish–English bilingual children had completed education beyond high school; in

contrast, 95% of mothers of French–English bilingual children had done so. As in Experiment 1, these categories were then converted to approximate years of education, and the average educational attainment among mothers of the Spanish–English sample was 14.40 years (*SD* = 4.10). Compared to the Montreal sample, the overall mean educational level of the two samples was not dissimilar (Montreal = 15.80, New Jersey = 14.40), and these were not different in a Welch's *t*-test, which does not assume equal variances [*t*(29.50) = −1.35, *p* = .188, *d* = -0.44]. However, the level of education was more uniform in the Montreal sample (*SD* = 2.50) and more variable in the New Jersey sample (*SD* = 4.10), a difference which was significant by Levene's test [*F*(1), 36 = 6.36, *p* = 0.02]. Thus, we infer that the sample of Spanish–English bilingual children had a different profile of socioeconomic status, including more mothers with lower educational attainment, compared to the sample of French–English bilingual children.

As in Experiment 1, children's productive vocabulary size in English was assessed using the Developmental Vocabulary Assessment for Parents (DVAP; Libertus et al., 2015) and a Spanish adaptation we created for the purposes of this study based on the Spanish version of the PPVT (Test de Vocabulario en Imagenes Peabody; Dunn, Padilla, Lugo & Dunn, 1986). Data were missing for one language for 1 participant. Of the words on the DVAP, toddlers produced an average of 54 words in their dominant language (range: 3–105) and 41 words in their non-dominant language (range: 1–107), a difference which was not statistically reliable, [*t*(19) = 0.90, *p* = .380, *d* = 0.20]. The total number of words produced across both languages was 93 words on average (range: 31–155). This was significantly fewer than French–English bilinguals, who produced an average of 125 words, [*t*(36) = 2.05, *p* = .047, *d* = 0.67]. This difference in vocabulary size may be related to differences in socioeconomic status between the two samples (e.g., Hart & Risley, 1995; Gilkerson et al., 2017).

### Stimuli

Auditory stimuli for Experiment 2 were equivalently structured to those used in Experiment 1, and were created with the same constraints (i.e., different phonological onsets, different rhymes, and sounded masculine in Spanish grammatical gender). Sentences were produced by a female Spanish–English bilingual speaker (e.g., "Look! Do you see the dog/*perro* on the *teelo*?" / "*¡Mira! ¿Puedes ver el perro/dog encima del teelo?*"). All other aspects of the stimuli were identical to Experiment 1. As in Experiment 1, the language of the carrier phrase was consistent for each child (i.e., always in Spanish or always in English). Children were
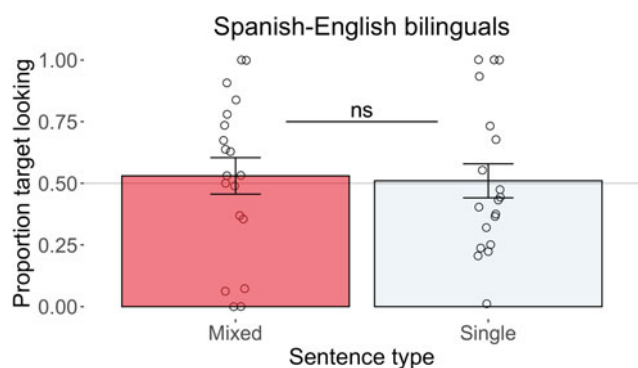
**Figure 4.** Average proportion looking to target object for Spanish–English bilinguals (Experiment 2). Circles show individual data points. Bars represent standard error of the mean.

randomly assigned to the language of testing: 10 children were tested in their dominant language, and 11 were tested in their non-dominant language.

### Procedure and reliability coding

The procedure was similar to Experiment 1. Children sat on a parent's lap, and parents wore occluded glasses and headphones with music so that they were unaware of the stimuli. Children viewed stimuli approximately 70 cm from a large, flat-screen television monitor (68 x 122 cm). As no eye tracker was available at the satellite lab where data were collected, the experimenter recorded children's eye movements via a video camera below the screen, which had a frame-rate of 29.97 Hz. Research comparing automatic eye tracking to manual coding suggests that the latter results in less data loss, but overall patterns of looking are similar across the two approaches (Venker et al., 2020). After testing, a researcher who was unaware of the experimental condition coded eye movements, indicating the direction of looks as: left, right, shifting, or away. Another researcher also coded 25% of the videos. Inter-coder reliability was 98.7% across all video frames, and 97.5% for video frames surrounding only shift events.

### Results

#### Test phase

Analyses for Experiment 2 were similar to those for Experiment 1, focusing on children's looking patterns during the 360-2000ms time window after the object was labeled during the test phase, again excluding trials with less than 750 ms of overall looking during this time window. Children's proportion of looking to target was averaged across the two test trials of each type. Children completed an average of 1.68 out of 2 trials testing the novel word that was encountered during single-language sentences in the learning phase (range: 1– 2) and 1.58 out of 2 trials testing the novel word that was encountered during mixed-language sentences (range: 1– 2).

Overall, children in Experiment 2 appeared unsuccessful at learning either word, as displayed in Figure 4. Two-tailed single-samples *t*-tests showed that children's looking to the target object was not significantly different from chance (.5) for the novel object heard in single-language sentences [$M = 0.51$, $SD = 0.07$, $t(18) = 0.43$, $p = 0.67$, $d = 0.10$] or in mixed-language sentences [$M = 0.53$, $SD = 0.07$, $t(18) = 0.10$, $p = 0.92$, $d = 0.02$]. A two-tailed paired-samples *t*-test confirmed that children showed similar

**Table 3.** Spanish–English bilinguals (Experiment 2): Descriptive statistics for looking times during learning phase.

| Target type | Sentence type | Familiarity | Mean | SD |
|---|---|---|---|---|
| target | single | familiar + novel | 0.72 | 0.23 |
| target | mixed | familiar + novel | 0.67 | 0.25 |
| distractor | single | familiar + novel | 0.28 | 0.23 |
| distractor | mixed | familiar + novel | 0.33 | 0.25 |

looking to the target novel objects heard in the single-language and mixed-language sentences, [$t(18) = 1.45$, $p = 0.16$, $d = 0.33$].

#### Learning phase

In these analyses, we investigated whether children's apparently unsuccessful learning of the novel words was due to difficulty locating the labeled referent during the learning phase. Because data in Experiment 2 were coded by humans rather than an eye tracker, children's looks were coded in terms of whether looking was at the target familiar-novel combination of objects or at the distractor familiar-novel combination of objects; finer resolution to the area of interest surrounding each familiar and novel object was not possible, which is a limitation of hand coding relative to eye-tracking. Children contributed an average of 5.53 single-language trials (range: 2–6) and 5.05 mixed-language trials (range: 3–6) out of 6 possible trials per type. Means and standard deviations of proportion looking to each object type (target familiar, target novel, distractor familiar, distractor novel) during single-language and mixed-language sentences are displayed in Table 3.

A two-tailed, paired-samples *t*-test showed that children's attention to the target was similar for the single-language and mixed-language sentences, [$t(18) = -0.97$, $p = .344$, $d = -0.22$]. Children's looking to the target side was significantly above chance during both single-language sentences [$M = 0.72$, $SD = 0.23$, $t(18) = 4.10$, $p = .001$, $d = 0.94$] and mixed-language sentences [$M = 0.67$, $SD = 0.25$, $t(18) = 2.86$, $p = .010$, $d = 0.66$]. Time course plots shown in Figure 5 confirm that children showed similar increases in looking to the labeled objects after naming for both sentence types. This indicated that language mixing did not affect children's success in orienting to the labeled object during the learning phase.

#### Comparison of French–English (Experiment 1) and Spanish–English (Experiment 2) bilinguals

Next, we directly compared data for French–English (Experiment 1) and Spanish–English (Experiment 2) bilingual children. First, we examined performance during the test phase in a 2 (population: French–English, Spanish–English) x 2 (sentence type: single-language, mixed-language) mixed ANOVA (Table 4). The largest observed effect was an interaction of population and sentence type, which was expected as French–English bilinguals looked longer at the novel target that had been encountered in single-language trials than in mixed-language trials, but Spanish–English bilinguals showed similar looking to the targets regardless of the presence or absence of mixing. However, this interaction did not reach statistical significance ( $p = 0.06$).

Second, we examined performance during the learning phase. A 2 (population: French–English, Spanish–English) x 2 (sentence type: single-language, mixed-language) mixed ANOVA revealed a
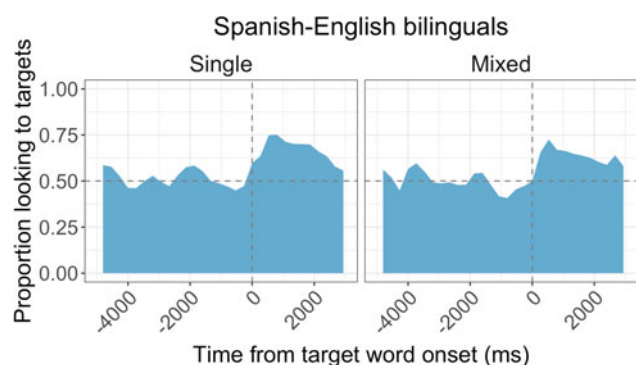
**Figure 5.** Timecourse of looking to target during learning trials for Spanish–English bilinguals (Experiment 2).

statistically significant main effect of population, whereby the French–English bilinguals looked to the labeled target with greater accuracy than did the Spanish–English bilinguals (Table 5). The main effect of sentence type and the interaction between sentence type and population were not statistically significant. Thus, although both groups did look more to the target objects than the distractor objects during the learning phase, the French–English bilinguals did so more reliably.

### Relationship between learning phase and test performance
In the next two sections, we combined datasets from the two populations to explore potential individual differences in performance. First, we examined whether performance at test was predicted by performance during the learning phase, exploring whether this relationship was moderated by sentence type (single-language, mixed-language) and population (French–English, Spanish–English). Scatterplots visualizing these relationships are shown in Figure 6. Because one of our predictors was continuous rather than categorical (performance during learning), we moved from the ANOVA framework used in the previous analyses to a linear mixed-effects model framework. This approach had the additional benefit of modeling the data at an individual trial level, which could potentially increase statistical power. Based on our previous results, our model included fixed effects of performance during the learning phase, sentence type, and population, the interaction between sentence type and population (since French–English but not Spanish–English toddlers showed different performance by sentence type), as well as random intercepts by participants. Results are shown in Table 6. The reference level for population was French–English bilinguals, and the reference level for sentence type was single-language. The model equation was prop.test ~ prop.learning + sentence.type*population + (1|id).

The analysis revealed a significant main effect of performance during the learning phase, revealing that toddlers who performed better during the learning phase performed better at test. There was also a significant interaction between sentence type and population, replicating the finding from our previous analysis that French–English bilinguals did better on single-language than mixed-language trials, but that Spanish–English toddlers performed similarly on both sentence types. This is also reflected in the significant main effect of sentence type, which is interpreted at the reference level of French–English bilinguals. We compared this model to a larger model that included interactions of all predictors, but found no evidence that the more complicated model was a significantly better fit to the data, based on a chi-squared test comparing the models.

### Individual differences
Next, we were interested in whether performance was related to vocabulary knowledge (dominant language vocabulary, non-dominant language vocabulary, and total vocabulary), exposure to language mixing, or socioeconomic status. As a first step, we computed reliability estimates for infants' performance in the learning and test phases. The reliability of proportion target looking across the 16 learning trials was estimated with an intraclass correlation coefficient (ICC), based on a mean-rating (k = 16), consistency, 2-way random-effects model (ICC3k) using the psych package in R (Revelle, 2021). The estimated ICC for training trials was 0.79, 95% CI [0.70, 0.87]. Using a similar approach for proportion target looking during the 4 test trials (k = 4), the estimated ICC for test trials was 0.81, 95% CI [0.71, 0.88]. These values reflect good reliability (Koo & Li, 2016), indicating that children showed stable individual variability in our task,

**Table 4.** Comparison of French–English (Experiment 1) and Spanish–English (Experiment 2) bilinguals' performance during the test phase.

| Effect | DFn | DFd | SSn | SSd | F | p | p < .05 | ges |
|---|---|---|---|---|---|---|---|---|
| (Intercept) | 1.00 | 36.00 | 21.39 | 3.49 | 220.47 | <.01 | * | 0.78 |
| population | 1.00 | 36.00 | 0.01 | 3.49 | 0.09 | 0.76 | | 0.00 |
| sentence type | 1.00 | 36.00 | 0.16 | 2.43 | 2.34 | 0.13 | | 0.03 |
| population * sentence type | 1.00 | 36.00 | 0.26 | 2.43 | 3.80 | 0.06 | | 0.04 |

**Table 5.** Comparison of French–English (Experiment 1) and Spanish–English (Experiment 2) bilinguals' performance during the learning phase.

| Effect | DFn | DFd | SSn | SSd | F | p | p < .05 | ges |
|---|---|---|---|---|---|---|---|---|
| (Intercept) | 1.00 | 36.00 | 16.33 | 1.67 | 352.05 | <.01 | * | 0.88 |
| population | 1.00 | 36.00 | 3.97 | 1.67 | 85.61 | <.01 | * | 0.65 |
| sentence type | 1.00 | 36.00 | 0.01 | 0.48 | 0.71 | 0.40 | | 0.00 |
| population * sentence type | 1.00 | 36.00 | 0.02 | 0.48 | 1.17 | 0.29 | | 0.01 |

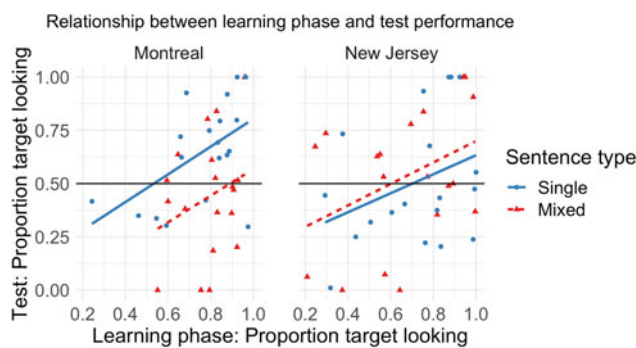Relationship between learning phase and test performance



**Figure 6.** Scatterplot showing performance during the learning phase and performance at test by population and sentence type. Lines represent best linear fit.

**Table 6.** Linear mixed effects model testing effects of population, sentence type, and performance during the learning phase on performance at test.

| Variable | Estimate | *SE* | *t* | *p* |
|---|---|---|---|---|
| intercept | 0.192 | 0.135 | 1.420 | 0.160 |
| proportion target learning | 0.599 | 0.159 | 3.760 | <.001 |
| sentence type (mixed) | −0.228 | 0.074 | −3.090 | 0.004 |
| population (Spanish–English) | −0.115 | 0.087 | −1.320 | 0.192 |
| sentence type * population | 0.283 | 0.105 | 2.700 | 0.011 |

which could potentially be accounted for by our predictors of interest (for a discussion of the role of reliability in infant research, see Byers-Heinlein, Bergmann & Savalei, 2021).

Next, we performed a sensitivity power analysis to determine the magnitude of correlations we would be able to detect given our sample size with 80% power, using the pwr package in R (Champely, 2020). The analysis indicated that we could detect an observed correlation of $r = 0.60$ for participants within each community ($n = 19$/community) and $r = 0.44$ for the two groups combined ($N = 38$). Thus, we had sufficient power to detect medium to large correlations, with greater power for the combined sample than for each community separately.

Finally, we computed correlations with the predictors of interest separately for the learning and test phases, both for each community separately as well as for the two communities combined. Thus, a total of 30 different correlations were computed (see Table 7). Two of the correlations approached but did not reach statistical significance, although not when corrected for multiple comparisons. Specifically, there were positive correlations in the combined sample between performance on the learning phase and total vocabulary size ($p = 0.07$), as well as performance at test and maternal education ($p = 0.08$). None of the other correlations approached statistical significance, even without controlling for multiple comparisons ($ps \geqslant .10$).

## General discussion

We tested whether and how language mixing affects novel word learning in 3-year-old bilingual children from two language communities: French–English bilinguals in Montreal, and Spanish–English bilinguals in New Jersey. Children viewed pairs of familiar animals, each perched atop different novel objects, and had the opportunity to learn two novel words. One novel word was consistently encountered within single-language sentences (e.g., "Look! Do you see the dog on the teelo?"), and one was consistently encountered within mixed-language sentences (e.g., "Look! Do you see the *chien/perro* on the teelo?"). This design required children to first identify the familiar referent in order to then locate the referent of the target novel noun. At test, children saw the two novel objects without the familiar objects, and a sentence directed their attention to the target.

Children's accuracy in learning novel words differed across the two language communities. French–English bilingual children successfully recognized the word that had been encountered in single-language sentences, but not the word encountered in mixed-language sentences. However, Spanish–English bilingual children did not recognize the word encountered in either context. Results from French–English bilinguals support the hypothesis that language mixing can (at least sometimes) disrupt initial encoding of novel words.

Our investigation was motivated by the idea that language mixing can sometimes derail children's recognition of a familiar word, which in turn could derail learning of a novel word that occurs moments later. Indeed, substantial previous work has demonstrated that both children and adults experience processing

**Table 7.** Correlations with proportion target looking.

| Phase | Variable | All | French–English | Spanish–English |
|---|---|---|---|---|
| Test | language mixing | −0.10 | −0.25 | 0.02 |
| Test | dominant language vocabulary | −0.02 | −0.08 | 0.01 |
| Test | non-dominant language vocabulary | −0.10 | −0.19 | −0.03 |
| Test | total vocabulary | −0.07 | −0.14 | −0.02 |
| Test | maternal education | 0.21 | 0.16 | 0.24 |
| Learning | language mixing | −0.06 | 0.14 | −0.11 |
| Learning | dominant language vocabulary | 0.16 | 0.25 | −0.01 |
| Learning | non-dominant language vocabulary | 0.13 | 0.18 | 0.05 |
| Learning | total vocabulary | 0.22 | 0.23 | 0.11 |
| Learning | maternal education | 0.12 | −0.16 | 0.17 |

costs for language mixing in a range of paradigms including eye tracking and speech production (Altarriba, Kroll, Sholl & Rayner, 1996; Byers-Heinlein et al., 2017; Chan, Chau & Hoosain, 1983; Grainger & Beauvillain, 1988; Macnamara & Kushnir, 1971; Meuter & Allport, 1999; Moreno, Federmeier & Kutas, 2002; Morini & Newman, 2019; Potter et al., 2019; Price, Green & von Studnitz, 1999; Proverbio, Leoni & Zani, 2004; Soares & Grosjean, 1984; Thomas & Allport, 2000), and our stimuli were closely modeled after those used in previous studies. However, our analyses of children's looking behavior during the learning phase did not necessarily converge with these findings. We found that children in both groups did successfully identify the familiar object during the learning phase, regardless of whether they encountered its label in a single-language or a mixed-language sentence. Eye-tracking data, which was only available for French–English bilinguals, additionally revealed that French–English bilinguals successfully directed their gaze to the target novel object itself. That is, children across both communities generally shifted their eyes to the 'correct' location independent of sentence type, community, or their previous experience with language mixing. This presents a key problem: despite consistently successful looking toward the correct familiar/novel referent pair, Spanish–English bilinguals did not appear to successfully learn either of the novel word-object mappings (perhaps because they were less efficient overall in orienting to the target novel referent), yet French–English bilinguals showed asymmetry in doing so across the two sentence types.

Samuelson and colleagues (Horst & Samuelson, 2008; Samuelson & McMurray, 2017) make an important distinction between two phases in word learning. The first phase is referent selection: determining the likely referent of a novel word. The second phase is retention: linking the sound of the novel word to its meaning. A number of recent studies have shown that these are dissociable processes in word learning, as children sometimes look toward the correct referent without recognizing the word later (Bion, Borovsky & Fernald, 2013; Horst & Samuelson, 2008; O'Connell, Poulin-Dubois, Demke & Guay, 2009). Our findings – using the quite distinct domain of language mixing – confirm the dissociable nature of these two components of word learning. That is, during the learning phase, children looked toward the labeled object in all contexts, but only French–English bilinguals showed successful word learning, and only in the single-language context.

The cognitive demands of our word learning task might have contributed to children's overall difficulty with word learning. First, our learning phase presented children with complex visual scenes involving two familiar and two novel objects, which we had modeled on previous research with monolingual English-learners of this age (Fernald et al., 2008). However, some studies have suggested that the presence of multiple objects can make word learning challenging. For example, Horst and Samuelson (2008) presented 24-month-old monolingual children with three objects: two familiar objects (e.g., car, cow) and one novel object (e.g., 'splicket'). They instructed children to "Look at the splicket! Which one is the splicket?" Children's knowledge of the familiar objects helped them identify the novel object as the referent of the novel word (i.e., referent selection), yet when tested they did not appear to retain this link (i.e., retention). Only when provided with additional cues that heightened the novel object's salience, such as ostensive naming, did children retain the label. The authors suggested that although the familiar objects helped children identify the novel object, they also competed as possible

referents for the novel word. Such complexity was also present in our design. On each learning trial, children not only encountered the target referent (e.g., teelo), but also three other competitors: the familiar target (e.g., dog), the familiar distractor (e.g., bunny), and the novel distractor (e.g., walem). Moreover, when children heard "Look! Do you see the dog on the teelo?", they had to infer that the teelo was the object underneath the dog – an additional processing step that might have made the task especially challenging for some children (see Bion et al., 2013; Goodman et al., 1998; Horst & Samuelson, 2008; and Carey & Bartlett, 1978 for studies that required children to make comparable inferences). Our results converge with this theoretical framework, given that French–English and Spanish–English bilinguals who spent more time looking at the labeled object during the learning phase performed better at test.

In light of these design-related complexities, the data for Spanish–English versus French–English bilingual children merit special attention. Previous work has demonstrated that language mixing affects real-time word comprehension among children from both of these communities (Byers-Heinlein et al., 2017; Potter et al., 2018, 2019). Such comparisons are essential for evaluating the generalizability of findings across communities of bilingual learners. Spanish–English bilinguals did not learn words in single- or mixed-language sentences, yet French–English bilingual children were able to learn in the single-language condition. While both groups showed above-chance looking toward the target side during the learning phase, French–English bilinguals were more efficient in doing so. This likely gave French–English bilinguals more time to encode the word-referent link. However, we do not currently know why the two groups differed in their orientation behaviors during the learning phase. We cannot rule out the possibility that small differences in experimental setups, stimuli, and coding approaches between the two labs contributed to our divergent findings. We took as many steps as possible to minimize these differences, and both labs have a history of detecting subtle processing effects in related paradigms with both bilinguals and monolinguals of this approximate age. That said, there are several ways that future research could evaluate the robustness and generalizability of these findings. While we were able to test a total of 38 children across the two communities, larger samples are needed to verify the replicability of the patterns we observed. Moreover, while existing paradigms for studying novel word learning in young children are often constrained to a small number of items, it will be important to develop new approaches for examining children's learning of larger stimulus sets.

We believe that the differences we observed are likely to be attributable to differences between the populations themselves. Indeed, there are wide-ranging differences between our two groups of learners, which collectively highlight the importance of examining multiple populations in bilingual research. In our dataset, the two samples were similar in terms of how much language mixing they encountered (according to parental report), although this still leaves open the possibility that they might encounter different types of language mixing, which could be either similar to or distinct from our stimuli. We did observe that the families of French–English bilingual participants had less variable socioeconomic status than the Spanish–English bilinguals (which included more mothers who had not completed high school), and that French–English bilingual children had larger vocabularies. Previous research points to socioeconomic status as an important influence on early language processing and vocabulary growth (e.g., Fernald

et al., 2013), and theories of language acquisition may need to more explicitly consider interactions between bilingualism and socio-economic status. Indeed, we found weak evidence that, across both communities, children with larger vocabularies were more accurate during the learning phase, and those from higher socio-economic backgrounds were more accurate in looking to the target object at test, although these results should be interpreted cautiously as correlations were marginally significant even prior to correction for multiple comparisons.

However, it is safe to assume that other dimensions of language experience which we were not able to measure directly also play a role in shaping bilingual learning. For example, we do not know how typical our chosen novel words were of other words in their vocabularies (even though the words were always pronounced with language-specific phonology). Moreover, these bilinguals' learning takes place in different cultural milieus, with potentially different family-level and/or society-level differences in language attitudes. Research is needed to better understand the microstructure of language use in different bilingual households and different bilingual communities (including the communities studied here), and how these in turn affect children's language processing and vocabulary growth over time. Efforts to understand variation in dual-language use across languages and cultures are currently underway in several countries (e.g., Carbajal & Peperkamp, 2020; Orena, Byers-Heinlein & Polka, 2020), which will be critical for furthering knowledge about early bilingual development.

Regarding the findings for French–English bilingual children, it is important to consider why the single-language sentences (relative to the mixed-language sentences) enabled encoding of the link between the novel word and its referent. In addition to children's equal gaze to target referents during the learning phase, many factors were identical across these sentence contexts: the same number of objects was visible, the same competition between objects was at play, the same inferences were required in both cases, and the same word–object association demands were required. What differed was whether the label for the mid-sentence familiar animal did or did not match the language used in the rest of the sentence, and possibly whether that word appeared in the child's dominant or non-dominant language. Why were French–English bilinguals unable to encode the word-referent link in the mixed-language condition?

One possibility is that language mixing could present ambiguity or confusion about what language the to-be-learned word is in. Some theories of bilingual language acquisition posit that even young bilingual learners do encode the language of a word (Byers-Heinlein, 2014; Curtin, Byers-Heinlein & Werker, 2011). But in contexts with language mixing, the language of a word is incrementally less clear, as the preceding sentence suddenly includes sounds, syllables, words, and/or phrases from both languages. Although our target novel words had clear phonological characteristics of one language or the other (i.e., words were pronounced either in an English-like or French/Spanish-like accent), these cues might have been insufficient or too fleeting.

A second possibility is that bilingual children learn over time that language mixing does not offer an optimal moment for learning, and thus they either implicitly or explicitly disengage and/or do not encode novel words that appear in mixed-language sentences. Adult bilinguals sometimes mix their languages when they have a lexical gap (Grosjean, 1982) or have difficulty retrieving a word (Heredia & Altarriba, 2001). Given evidence that toddlers prefer learning word labels that come from certain over

uncertain speakers (Birch, Vauthier & Bloom, 2008; although see also Buac et al., 2019), it may be that children interpret some types of language mixing as a signal of a speaker's lexical uncertainty in the moment, which may hinder real-time encoding of a novel label. This possibility would need to be tested empirically in a future study that broadly examines children's uptake of new information in communicative contexts with frequent vs. infrequent language mixing.

Another approach to evaluating these contrasting – but complementary – ideas would be to teach children words in language-mixing contexts with simpler visuals than those used in the current design. For example, in an eye-tracking task, the child could view just one familiar–novel object pair at a time (e.g., a dog on a teelo, with no distractor), or only one novel object at a time, in isolation (e.g., only the teelo). Any heightened success in learning words under these simpler task demands would indicate that the cognitive load of the current design was too high. However, if language mixing is implicitly or explicitly perceived as a sub-optimal learning opportunity, children would avoid learning and/or show failure to learn a novel word that is presented in a mixed-language sentence, regardless of task complexity. Indeed, there is some evidence to support this possibility. In a study of Mandarin–English bilingual 18-month-olds using the Switch task (which presents one novel object at a time accompanied by an isolated label), children readily encoded words with full phonetic detail in a single-language context, but not in a switched-language context (Singh, Fu, Tay & Golinkoff, 2018).

It will also be important for future research to test the same question but in a range of language processing contexts, in order to better understand when code mixing does and does not impact word learning. In our study, children heard a code-mixed noun prior to encountering a novel word. However, code-mixing comes in many other forms, at different parts of the phrase, and for different parts of speech. Moreover, parents sometimes code-mix by providing immediate translations of words and phrases, which is likely to reflect a conscious effort to teach their children new words (Byers-Heinlein, 2013; Kremin et al., 2020). In the end, it is likely that some types of code-mixing would actually support rather than derail word learning (e.g., Read, Contreras, Rodriguez & Jara, 2020).

To conclude, these experiments provide two main contributions to existing research. First, the data suggest that language mixing can sometimes hinder bilingual children's learning of new words that occur downstream, but leave open important questions as to why. Second, the two experiments show that different populations of bilingual children sometimes perform differently on equivalent language tasks. This highlights the need for caution in generalizing findings to bilinguals in general, and calls for an increased practice of comparing multiple populations of bilingual learners.

## References

Altarriba, J, Kroll, JF, Sholl, A, & Rayner, K (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition*, **24**(4), 477–492. https://doi.org/10.3758/BF03200936

Bail, A, Morini, G, & Newman, RS (2015). Look at the gato! Code-switching in speech to toddlers. *Journal of Child Language*, **42**(5), 1073–1101. https://doi.org/10.1017/S0305000914000695

Bergmann, C, Tsuji, S, Piccinini, PE, Lewis, ML, Braginsky, M, Frank, MC, & Cristia, A (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, **89**(6), 1996–2009. https://doi.org/10.1111/cdev.13079

Bion, RAH, Borovsky, A, & Fernald, A (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, **126**(1), 39–53. https://doi.org/10.1016/j.cognition.2012.08.008

Birch, SA, Vauthier, SA, & Bloom, P (2008). Three-and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, **107**(3), 1018–1034. https://doi.org/https://doi.org/10.1016/j.cognition.2007.12.008

Boudreault, M-C, Cabirol, A-A, Trudeau, N, Poulin-Dubois, D, & Sutton, A (2007). Les inventaires MacArthur du développement de la communication: Validité et données normatives préliminaires. / MacArthur communicative development inventories: Validity and preliminary normative data. *Canadian Journal of Speech-Language Pathology and Audiology*, **31**(1), 27–37.

Buac, M, Tauzin-Larché, A, Weisberg, E, & Kaushanskaya, M (2019). Effect of speaker certainty on novel word learning in monolingual and bilingual children. *Bilingualism*, **22**(4), 883–895. https://doi.org/https://doi.org/10.1017/S1366728918000536

Byers-Heinlein, K (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, **16**(1), 32–48. https://doi.org/10.1017/S1366728912000120

Byers-Heinlein, K (2014). Languages as categories: Reframing the "one language or two" question in early bilingual development. *Language Learning*, **64**(s2), 184–201. https://doi.org/10.1111/lang.12055

Byers-Heinlein, K, Bergmann, C, & Savalei, V (2021). Six solutions for more reliable infant research. PsyArXiv. https://doi.org/10.31234/osf.io/u37fy

Byers-Heinlein, K, Morin-Lessard, E, & Lew-Williams, C (2017). Bilingual infants control their languages as they listen. *Proceedings of the National Academy of Sciences*, **114**(34), 9032–9037. https://doi.org/10.1073/pnas.1703220114

Byers-Heinlein, K, Tsui, A, Bergmann, C, Black, AK, Brown, A, Carbajal, MJ, … Wermelinger, S (2020). A multi-lab study of bilingual infants: Exploring the preference for infant-directed speech. *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.31234/osf.io/sqh9d

Carbajal, MJ, & Peperkamp, S (2020). Dual language input and the impact of language separation on early lexical development. *Infancy*, **25**(1), 22–45. https://doi.org/10.1111/infa.12315

Carey, S, & Bartlett, E (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, **15**, 17–29.

Champely, S (2020). *Pwr: Basic functions for power analysis.* Retrieved from https://CRAN.R-project.org/package=pwr

Chan, M-C, Chau, HLH, & Hoosain, R (1983). Input/output switch in bilingual code switching. *Journal of Psycholinguistic Research*, **12**(4), 407–416. https://doi.org/10.1007/BF01067622

Curtin, S, Byers-Heinlein, K, & Werker, JF (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, **39**(4), 492–504. https://doi.org/10.1016/j.wocn.2010.12.002

DeAnda, S, Hendrickson, K, Zesiger, P, Poulin-dubois, D, & Friend, M (2018). Lexical access in the second year: A study of monolingual and bilingual vocabulary development. *Bilingualism: Language and Cognition*, **21**(2), 314–327. https://doi.org/10.1017/S1366728917000220

Dink, J, & Ferguson, B (2015). eyetrackingR: An R Library for Eye-tracking Data Analysis. Retrieved from http://www.eyetrackingr.com

Dunn, LM, & Dunn, DM (2007). *Peabody Picture Vocabulary Test-IV*. Circle Pines, MN: American Guidance Service.

Dunn, LM, Dunn, LM, & Thériault-Whalen, CM (1993). *Échelle de vocabulaire en images Peabody (EVIP)*. Toronto: PSYCAN.

Dunn, LM, Padilla, ER, Lugo, DE, & Dunn, LM (1986). *Test de vocabulario en imagenes Peabody: TVIP: Adaptacion hispanoamericana (Peabody Picture Vocabulary Test: PPVT: Hispanic-American Adaptation)*. Circle Pines, MN: American Guidance Service (AGS), Inc.

Ervin-Tripp, S, & Reyes, I (2005). Child codeswitching and adult content contrasts. *International Journal of Bilingualism*, **9**(1), 85–102. https://doi.org/10.1177/13670069050090010601

Fenson, L, Marchman, VA, Thal, DJ, Dale, PS, Steven Reznick, J, & Bates, E (2007). *MacArthur-Bates Communicative Development Inventories* (2nd ed.). Baltimore, MD: Brookes.

Fernald, A, Perfors, A, & Marchman, VA (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, **42**(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Fernald, A, Marchman, V, & Hurtado, N (2008). *Input affects uptake: How early language experience influences processing efficiency and vocabulary learning*. https://doi.org/10.1109/DEVLRN.2008.4640802

Fernald, A, & Marchman, VA (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, **83**(1), 203–222. https://doi.org/10.1111/j.1467-8624.2011.01692.x

Fernald, A, Marchman, VA, & Weisleder, A (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, **16**(2), 234–248. https://doi.org/10.1111/desc.12019

Frank, MC, Braginsky, M, Yurovsky, D, & Marchman, VA (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, **44**(3), 677–694. https://doi.org/10.1017/S0305000916000209

Gilkerson, J, Richards, JA, Warren, SF, Montgomery, JK, Greenwood, CR, Kimbrough Oller, D, … Paul, TD (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology*, **26**(2), 248–265. https://doi.org/https://doi.org/10.1044/2016_AJSLP-15-0169

Goodman, JC, McDonough, L, & Brown, NB (1998). The role of semantic context and memory in the acquisition of novel nouns. *Child Development*, **69**(5), 1330–1344. https://doi.org/10.1111/j.1467-8624.1998.tb06215.x

Goodz, NS (1989). Parental language mixing in bilingual families. *Infant Mental Health Journal*, **10**(1), 25–44. https://doi.org/10.1002/1097-0355(198921)10:1<25::AID-IMHJ2280100104>3.0.CO;2-R

Grainger, J, & Beauvillain, C (1988). Associative priming in bilinguals: Some limits of interlingual facilitation effects. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, **42**(3), 261–273. https://doi.org/10.1037/h0084193

Grosjean, F (1982). *Life with two languages: An introduction to bilingualism.* Harvard University Press.

Hart, B, & Risley, T (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: Paul H. Brookes Publishing Company.

Heredia, RR, & Altarriba, J (2001). Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science*, **10**(5), 164–168. https://doi.org/https://doi.org/10.1111/1467-8721.00140

Horst, JS, & Samuelson, LK (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, **13**(2), 128–157. https://doi.org/10.1080/15250000701795598

Hurtado, N, Grüter, T, Marchman, VA, & Fernald, A (2014). Relative language exposure, processing efficiency and vocabulary in Spanish–English bilingual toddlers. *Bilingualism: Language and Cognition*, **17**(1), 189–202. https://doi.org/10.1017/S136672891300014X

Kircher, R (2014). Thirty years after Bill 101: A contemporary perspective on attitudes towards English and French in Montreal. *Canadian Journal of Applied Linguistics*, **17**(1), 20–50.

Koo, TK, & Li, MY (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, **15**(2), 155–163. https://doi.org/https://doi.org10.1016/j.jcm.2016.02.012

Kremin, LV, Alves, J, Orena, AJ, Polka, L, & Byers-Heinlein, K (2020). *Code-switching in parents' everyday speech to bilingual infants.* PsyArXiv. https://doi.org/10.31234/osf.io/rk2ah

Kroff, JRV, Dussias, PE, Gerfen, C, Perrotti, L, & Bajo, MT (2017). Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguistic Approaches to Bilingualism*, 7(2), 163–198. https://doi.org/10.1075/lab.15010.val

Lany, J (2018). Lexical-processing efficiency leverages novel word learning in infants and toddlers. *Developmental Science*, 21(3), e12569. https://doi.org/10.1111/desc.12569

Legacy, J, Zesiger, P, Friend, M, & Poulin-Dubois, D (2016). Vocabulary size and speed of word recognition in very young French–English bilinguals: A longitudinal study. *Bilingualism: Language and Cognition*, 21(1), 1–13. https://doi.org/10.1017/S1366728916000833

Lew-Williams, C, & Fernald, A (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3), 193–198. https://doi.org/10.1111/j.1467-9280.2007.01871.x

Libertus, ME, Odic, D, Feigenson, L, & Halberda, J (2015). A developmental vocabulary assessment for parents (DVAP): Validating parental report of vocabulary size in 2- to 7-year-old children. *Journal of Cognition and Development*, 16(3), 442–454. https://doi.org/10.1080/15248372.2013.835312

Macnamara, J, & Kushnir, SL (1971). Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior*, 10(5), 480–487. https://doi.org/10.1016/S0022-5371(71)80018-X

Marchman, VA, Fernald, A, & Hurtado, N (2010). How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish–English bilinguals. *Journal of Child Language*, 37(4), 817–840. https://doi.org/10.1017/S0305000909990055

Marian, V, Blumenfeld, HK, & Kaushanskaya, M (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Meuter, RFI, & Allport, A (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, 40(1), 25–40. https://doi.org/10.1006/jmla.1998.2602

Moreno, EM, Federmeier, KD, & Kutas, M (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language*, 80(2), 188–207. https://doi.org/10.1006/brln.2001.2588

Morini, G, & Newman, RS (2019). Dónde está la ball? Examining the effect of code switching on bilingual children's word recognition. *Journal of Child Language*, 46(6), 1238–1248. https://doi.org/10.1017/S0305000919000400

O'Connell, L, Poulin-Dubois, D, Demke, T, & Guay, A (2009). Can infants use a nonhuman agent's gaze direction to establish word–object relations? *Infancy*, 14(4), 414–438. https://doi.org/10.1080/15250000902994073

Orena, AJ, Byers-Heinlein, K, & Polka, L (2020). What do bilingual infants actually hear? Evaluating measures of language input to bilingual-learning 10-month-olds. *Developmental Science*, e12901. https://doi.org/10.1111/desc.12901

Place, S, & Hoff, E (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development*, 82(6), 1834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x

Place, S, & Hoff, E (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition*, 19(5), 1023–1041. https://doi.org/10.1017/S1366728915000322

Potter, CE, Fourakis, E, Morin-Lessard, E, Byers-Heinlein, K, & Lew-Williams, C (2018). Bilingual infants process mixed sentences differently in their two languages. In *Proceedings of the 40th annual meeting of the cognitive science society.*

Potter, CE, Fourakis, E, Morin-Lessard, E, Byers-Heinlein, K, & Lew-Williams, C (2019). Bilingual toddlers' comprehension of mixed sentences is asymmetrical across their two languages. *Developmental Science*, 22(4), e12794. https://doi.org/10.1111/desc.12794

Price, CJ, Green, DW, & von Studnitz, R (1999). A functional imaging study of translation and language switching. *Brain*, 122(12), 2221–2235. https://doi.org/10.1093/brain/122.12.2221

Proverbio, AM, Leoni, G, & Zani, A (2004). Language switching mechanisms in simultaneous interpreters: An ERP study. *Neuropsychologia*, 42(12), 1636–1656. https://doi.org/10.1016/j.neuropsychologia.2004.04.013

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Read, K, Contreras, PD, Rodriguez, B, & Jara, J (2020). ¿Read Conmigo?: The Effect of Code-switching Storybooks on Dual-Language Learners' Retention of New Vocabulary. *Early Education and Development*, 0(0), 1–18. https://doi.org/10.1080/10409289.2020.1780090

Revelle, W (2021). *Psych: Procedures for psychological, psychometric, and personality research.* Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Samuelson, LK, & McMurray, B (2017). What does it take to learn a word? *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(1–2), e1421. https://doi.org/10.1002/wcs.1421

Singh, L, Fu, C, Tay, Z, & Golinkoff, R (2018). Novel word learning in bilingual and monolingual infants: Evidence for a bilingual advantage. *Child Development*, 89, e183–e198. https://doi.org/10.1111/cdev.12747

Soares, C, & Grosjean, F (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory & Cognition*, 12(4), 380–386. https://doi.org/10.3758/BF03198298

Sridhar, SN, & Sridhar, KK (1980). The syntax and psycholinguistics of bilingual code mixing. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 34(4), 407. https://doi.org/10.1037/h0081105

Swingley, D (2012). The looking-while-listening procedure. *Research Methods in Child Language: A Practical Guide*, 29–42. https://doi.org/10.1002/9781444344035.ch3

Thomas, MSC, & Allport, A (2000). Language switching costs in bilingual visual word recognition. *Journal of Memory and Language*, 43(1), 44–66. https://doi.org/10.1006/jmla.1999.2700

Venker, CE, Pomper, R, Mahr, T, Edwards, J, Saffran, J, & Weismer, SE (2020). Comparing Automatic Eye Tracking and Manual Gaze Coding Methods in Young Children with Autism Spectrum Disorder. *Autism Research*, 13(2), 271–283. https://doi.org/10.1002/aur.2225

## Appendix

Sample trial order for French–English bilinguals (Experiment 1).

| study phase | trial # | auditory stimulus | sentence type | left familiar object | left novel object | right familiar object | right novel object | study phase |
|---|---|---|---|---|---|---|---|---|
| Learning | 1 | Look! Do you see the *chien* on the walem? | mixed | bunny | teelo | dog | walem | learning |
| Learning | 2 | Look! Do you see the bunny on the teelo? | single | bunny | teelo | fish | walem | learning |
| Learning | 3 | Look! Do you see the fish on the teelo? | single | dog | walem | fish | teelo | learning |
| Learning | 4 | Look! Do you see the *lapin* on the walem? | mixed | bunny | walem | dog | teelo | learning |
| Learning | 5 | Look! Do you see the *poisson* on the walem? | mixed | bunny | teelo | fish | walem | learning |
| Learning | 6 | Look! Do you see the dog on the teelo? | single | dog | teelo | bunny | walem | learning |
| Learning | 7 | Look! Do you see the *poisson* on the walem? | mixed | fish | walem | dog | teelo | learning |
| Learning | 8 | Look! Do you see the dog on the teelo? | single | fish | walem | dog | teelo | learning |
| Learning | 9 | Look! Do you see the *lapin* on the walem? | mixed | fish | teelo | bunny | walem | learning |
| Learning | 10 | Look! Do you see the fish on the teelo? | single | fish | teelo | bunny | walem | learning |
| Learning | 11 | Look! Do you see the *chien* on the walem? | mixed | dog | walem | fish | teelo | learning |
| Learning | 12 | Look! Do you see the bunny on the teelo? | single | dog | walem | bunny | teelo | learning |
| Test | 13 | Can you find the walem? | encountered in mixed | NA | teelo | NA | walem | test |
| Test | 14 | Can you find the teelo? | encountered in single | NA | teelo | NA | walem | test |
| Test | 15 | Can you find the walem? | encountered in mixed | NA | walem | NA | teelo | test |
| Test | 16 | Can you find the teelo? | encountered in single | NA | walem | NA | teelo | test |