Cognition 242 (2024) 105639

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Does bilingual experience influence statistical language learning?

Jose A. Aguasvivas[a,b,*,1], Jesús Cespón[a,d,1], Manuel Carreiras[a,b,c]

[a] BCBL, Basque Center on Cognition, Brain and Language, San Sebastian, Spain
[b] Universidad del País Vasco / Euskal Herriko Unibersitatea (UPV/EHU), San Sebastian, Spain
[c] Ikerbasque, Basque Foundation for Science, Bilbao, Spain
[d] Achucarro Basque Center for Neuroscience, Leioa, Spain

ARTICLE INFO

Keywords:
Statistical learning
Bilingualism
Language learning

ABSTRACT

Statistical language learning (SL) tasks measure different aspects of foreign language learning. Studies have used SL tasks to investigate whether bilingual experience confers advantages in acquiring additional languages through implicit processes. However, the results have been inconsistent, which may be related to bilingualism-related features (e.g., degree of dissimilarity between the specific language pair) and other variables such as specific processes that are targeted by the SL task. In the present study, we compared the performance of one Spanish monolingual and two bilingual (Spanish-Basque and Spanish-English) groups across three well-established SL tasks. Each task targeted a different aspect of foreign language learning; specifically, word segmentation, morphological rule generalization, and word-referent learning. In Experiment 1, we manipulated sub-lexical phonotactic patterns to vary the difficulty of three SL tasks, with the results showing no differences between the groups in word segmentation. In Experiment 2, we included non-adjacent dependencies to target affixal morphology rule learning, but again no group-related differences were found. In Experiment 3, we addressed word learning using an audio-visual SL task combining exclusive and multiple word-referent mappings, and found that bilinguals outperformed monolinguals, suggesting that bilingualism may exert influences on SL at the lexical level. This advantage might have been mediated by the high working memory demands required to perform the task. Summarizing, this study shows no evidence for a general bilingual advantage in SL, although bilinguals may outperform monolinguals under specific experimental conditions such as SL tasks that place high demands on working memory processes. In addition, the similar performance of Spanish-Basque and Spanish-English bilinguals across all three SL tasks suggests that the degree of dissimilarity between pairs of spoken languages does not modulate SL skills.

The human brain is remarkably attuned to structure in the environment. Both infants and adults can quickly and implicitly learn from patterns presented repetitively, a mechanism broadly known as statistical learning (Saffran, 2003; Conway, 2020). Early studies suggested that statistical learning occurs regardless of the input type (auditory, visual, or tactile), hinting towards a domain-general mechanism to track and learn from regularities in the environment (Kirkham et al., 2002). Nevertheless, recent research has revealed low correlations between statistical learning abilities across various cognitive domains (Frost et al., 2015; Siegelman & Frost, 2015), raising doubts about the view of statistical learning as a general mechanism. Within the cognitive domain of language, statistical language learning (SL) has generated particular interest in the field of language acquisition as a potentially important mechanism through which humans implicitly learn languages (Perruchet & Pacton, 2006; Romberg & Saffran, 2010) and process the linguistic input (Nemeth et al., 2011). However, the extent to which prior linguistic experiences affect SL remains largely unexplored.

In the present study, we examined whether and to what extent one such linguistic experience—namely, bilingualism—influences performance in SL tasks. In particular, we compared one monolingual and two bilingual groups of young adults in three well-established SL tasks that targeted different aspects of foreign language learning. A main objective of this study is to investigate whether and to what extent bilingualism confers advantages in SL and whether such advantages occur regardless of the type of SL task, which would point to the existence of a general SL mechanism, or are restricted to tasks tapping specific SL processes.

* Corresponding author at: BCBL, Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 2nd floor, 20009 Donostia/San Sebastian, Spain.
E-mail address: joseaamanzano@gmail.com (J.A. Aguasvivas).

[1] These authors contributed equally.

https://doi.org/10.1016/j.cognition.2023.105639
Received 4 October 2022; Received in revised form 3 October 2023; Accepted 6 October 2023
Available online 17 October 2023

Moreover, we investigate whether the relationship between bilingualism and enhanced SL skills depends on the specific languages known by bilinguals. We expect that results from this study will shed some clues on how bilingualism can facilitate acquisition of a new language.

The idea of SL stems from observing the naturalistic language learning. For example, with sufficient exposure to the words in the phrase *baby monkey*, infants and adults learn that the syllables within the words (i.e., *ba-by*) predict each other more reliably than the syllables at the boundary between the words (i.e., *by mo*) (Erickson & Thiessen, 2015). Typical SL experimental tasks consist of exposing participants to a continuous stream of artificial linguistic input and testing them on congruent items —those that follow the statistics of the input—versus incongruent items from the stream (Siegelman et al., 2017). Ideally, the participants should implicitly discover the boundaries between words solely from the statistics of the artificial language and mentally represent them as separate units. Researchers have employed this paradigm to examine how individuals implicitly learn different aspects of a novel language, including segmenting words from speech (Saffran, 2003; Saffran et al., 1996), acquiring sub-lexical units (Maye et al., 2002; McMurray et al., 2009), discovering morphological structures or patterns (Frost & Monaghan, 2016; Peña et al., 2002), and learning word-referent pairs (Smith & Yu, 2008; Yu & Smith, 2007).

Individual linguistic experiences may influence performance in SL tasks. For instance, a recent study showed that 16 months old bilingual infants are more skilled than their monolingual peers at segmenting words from dual artificial speech streams (Antovich & Graf Estes, 2020). This finding aligns with the difficulty experienced by monolingual infants in tracking statistical regularities in two speech streams presented sequentially (Benitez et al., 2020). However, some studies have not found differences between monolingual and bilingual infants in word segmentation in a simulated language mixing environment (Tsui et al., 2021). Also, studies have shown that the stress pattern of the native language can interfere with SL performance in 9-month-olds, but not in younger infants with less exposure to the native language (Jusczyk et al., 1999; Thiessen & Saffran, 2003). Adults also seem to struggle in SL tasks that contain syllables that are not plausible in their native language (Finn & Hudson Kam, 2008). Furthermore, the SL performance of both infants and adults seems to be biased towards the word order of their native language (Onnis et al., 2016; Onnis & Thiessen, 2013). Overall, it seems that the specific properties of a known language (e.g., stress patterns, word order) may facilitate or interfere with SL performance. This is in line with other research that has shown that prior linguistic experiences modulate the learning of novel linguistic material, as evidenced by cross-linguistic transfer studies (Alonso, 2016).

In bilingual adults who regularly use both their languages, experience of conflicting statistics may influence performance in SL tasks (Weiss et al., 2020). To test this idea, Wang and Saffran (2014) compared two monolingual (English and Mandarin) and two bilingual (Mandarin-English, and English-Spanish) groups in a challenging SL task that combined syllables and tones to emulate a foreign tonal language. They found that the monolingual groups performed at chance level in this task, whereas both bilingual groups performed similarly and significantly better than monolinguals. They concluded that this "bilingual advantage" was irrespective of experience of tonal languages and may have emerged from enhanced phonological working memory in bilinguals. This interpretation aligns with studies showing relationships between bilingualism and enhanced performance in some working memory tasks (Calvo et al., 2016), which suggests that high working memory demands in a SL task may give rise to stronger differences between monolinguals and bilinguals.Additional research has supported the idea of enhanced phonological working memory in bilinguals using other challenging SL tasks, such as learning a Morse Code language with interfering statistics (Bartolotti et al., 2011), simultaneous learning of two grammars (Onnis et al., 2018), and learning multiple word-referent pairs (Poepsel & Weiss, 2016).

Despite the emerging evidence supporting enhanced SL skills in bilinguals compared to monolinguals, some studies have failed to find differences between monolinguals and bilinguals in SL tasks (Bulgarelli & Weiss, 2016; Yim & Rudoy, 2013). This may be due to the high variety of cognitive (linguistic and nonlinguistic) factors and linguistic variables (i.e., context of learning, switching habits, language similarity, etc.) related to novel language learning (Hirosh & Degani, 2018) and it is also consistent with the low correlations between statistical learning skills across various cognitive domains (Frost et al., 2015; Siegelman & Frost, 2015). Thus, taking into account that SL tasks measure different aspects of foreign language learning, a main aim of the present study is to examine whether bilinguals show an overall advantage across SL tasks or any such advantages emerge at specific levels of implicit foreign language learning. In order to address this question, in the present study we carried out three different SL experiments in which we focused on three central aspects of foreign language learning (namely, word segmentation, morphological rule generalization, and word-referent learning) by means of three well-established SL paradigms.

We posit that a larger range of language knowledge in a given linguistic domain (e.g., phonology, grammar) should be related to enhanced SL skills to learn a new language and that this could be a major reason for enhanced SL in bilinguals compared to monolinguals. In this context. The structural sensitivity hypothesis claims that exposure to two or more languages can improve bilinguals' skills to detect the parameters that vary when learning a new language because they have experienced a larger range of linguistic input (Weiss et al., 2020). Specifically, the structural sensitivity hypothesis states that proficiency in two languages may increase the salience of similarities and differences between structural properties of two languages, strengthening the ability of bilinguals to extract or detect the features of a new language. Consequently, bilinguals should be more skilled than monolinguals at detecting new patterns in linguistic input (Bialystok, 1986; Cromdal, 1999; Foursha-Stevenson & Nicoladis, 2011; Kuo & Anderson, 2010, Kuo & Anderson, 2012; Kuo & Kim, 2014). In line with this reasoning, a higher degree of dissimilarity between two languages in a given process should be associated with enhanced skills at detecting regularities in a new language due to the higher diversity of linguistic imput. Although the structural sensitivity hypothesis is mainly related to learning of the linguistic structure, other theoretical models at other linguistic levels (e.g., phonology) may involve similar predictions. For example, the perceptual assimilation model (Best & Tyler, 2007) states that L2 phonemes are classified into different categories depending on the similarities and differences with respect to native sounds. Thus, it it is possible that the larger the range of phonological sounds spoken by a given individual, the more precise their ability is to classify a given sound when learning a new language. For instance, bilinguals could be more skilled at segmenting words in a new language if they speak two languages that are phonologically very different from each other due to their expanded range of phonological familiarity. In fact, another question addressed by the present research is whether the degree of dissimilarity between particular pairs of spoken languages can modulate SL skills.

In the present study, we recruited groups of Spanish monolingual young adults (functional monolinguals, whose L2 knowledge, if any, was insufficient to maintain a basic conversation), Spanish-Basque bilinguals, and Spanish-English bilinguals. It is important to highlight that SL skills seem to be enhanced in individuals who are relatively skilled in an L2 language but not in those who have only a very basic grasp of an L2 (Ellis, 2015). All participants were from Spain, although Spanish-Basque bilinguals came from a different region (Basque Country) than the Spanish monolingual and Spanish-English groups (Murcia). Prior studies have discussed the similarities and differences between Spanish and English (Rivera, 2019). Here, we briefly comment on the commonalities and differences between the Spanish and Basque languages. Spanish and Basque do not share any common root, but Basque possesses many Spanish loan words due to its shared geographic location within Spain. Despite some language-specific bigrams differentiating Basque from Spanish (e.g., "tx", "tz"), the two languages are phonologically similar. However, the Basque language possesses a predominantly subject-object-verb (SOV) word order, coupled with a postpositional and agglutinative morphology—i.e., morphemes and determiners are appended to the end of word roots (e.g., *eskolan – school the in*). These grammatical properties differentiate Basque from Spanish, English, and many other Indo-European languages (Bengtson, 2011). Therefore, according to the structural sensitivity hypothesis, phonological and morphological generalization should elicit performance differences between Spanish-Basque and Spanish-English bilinguals.

In short, we investigate three SL processes that could be improved by bilingual experience (i.e., word segmentation, morphological rule generalization, and word-referent learning) to study whether there is a general bilingual advantage in SL or it is restricted to specific SL tasks. Also, we study the role of the dissimilarity between known spoken language pairs to test whether a more varied range of linguistic input further increases SL improvement during foreign language acquisition. At the phonological level, English is more complex than Spanish, whereas Spanish and Basque are much closer. Consequently, we predict that Spanish-English bilinguals will outperform Spanish-Basque bilinguals for word segmentation in Experiment 1. In contrast, at the morphological and grammatical levels, Basque is both highly complex and different from Spanish and English. This means Spanish-Basque bilinguals internalize a greater variety of grammatical rules than Spanish-English bilinguals. Accordingly, we hypothesize that Spanish-Basque bilinguals should perform better than Spanish-English bilinguals at morphological rule generalization (specifically, a non-adjacent dependency learning task) in Experiment 2. In Experiment 3, we do not expect to find any difference between both bilingual groups in word-referent learning. However, the higher working memory requirements of the task in Experiment 3 compared to the tasks in Experiments 1 and 2 may lead to more pronounced differences between both bilingual groups and the monolingual group. The rationale of the hypotheses for each task and the specific predictions are schematically represented in Fig. 1.

1. Experiment 1

Experiment 1 tested the word segmentation performance of monolingual and bilingual adults in three SL streams that differed in their sublexical phonotactic patterns. Phonotactics define the allowed combinations of phonemes in a given language. Studies have shown that violating the phonotactic patterns of the native language can hinder the word learning performance of infants and adults (Estes et al., 2016; Finn & Hudson Kam, 2008). For instance, exposing English-speaking adults to a SL stream containing syllables that do not exist in English (e.g., */tfobu/*) leads to lower performance than stimuli that only contain legal English syllables (Finn & Hudson Kam, 2008). Sublexical phonotactics—defined here as the plausibility of constituent syllables—provide an excellent way of adjusting the difficulty of SL tasks. This is an important issue because the degree of difficulty of a task may modulate the existence of group-related differences. Specifically, low demanding SL tasks may not be challenging enough to be sensitive to differences between monolinguals and bilinguals in SL, given that differences between monolinguals and bilinguals in other cognitive domains tend to emerge only at relatively high levels of demand (Bialystok et al., 2014; Kuipers & Westphal, 2021).

Bilinguals experience distinct phonotactics in both languages, potentially affecting their word-learning performance (Kaushanskaya & Marian, 2009). Consequently, we expected both bilingual groups to show an advantage compared to monolinguals in this task and that this advantage could be even greater in Spanish-English compared to Spanish-Basque bilinguals, as Spanish-English bilinguals are familiar with a larger range of phonological sounds compared to Spanish-Basque bilinguals. Additionally, a specific question in this experiment was whether task difficulty could modulate this advantage, thus, participants listened to three SL streams. The first one was constructed using simple consonant-vowel syllables, whereas the other two contained consonant clusters that were either legal or illegal in any of the three languages that could potentially be spoken by the participants. We also predicted that the more complex syllables in these last two streams would impose higher difficulty compared to simple consonant-vowel syllables, as indicated in a pilot study.

1.1. Methods

Anonymized participant information, experimental data, materials, and annotated analyses scripts for all experiments are available at https://osf.io/86sr4/.

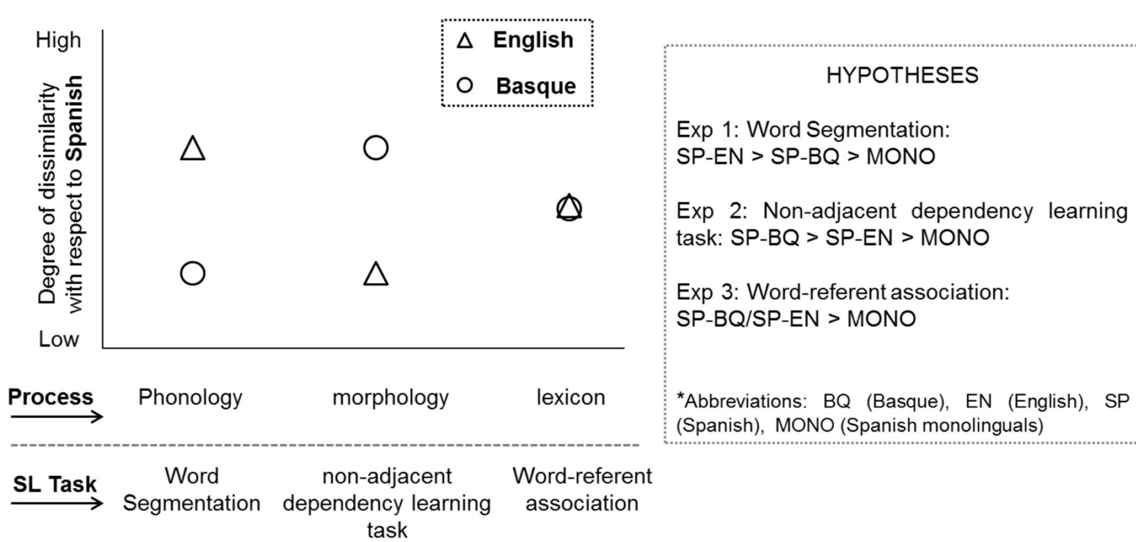

Fig. 1. Degree of dissimilarity between pairs of spoken languages (left panel) and the corresponding hypotheses for each Experiment (right panel).
Note: The left panel shows a visual representation of the degree of dissimilarity between pairs of spoken languages (i.e., Spanish-Basque and Spanish-English). The hypotheses for the three experiments are specified in the right panel. The first two experiments assess whether there is a relationship between increased repertory of bilingual skills (due to greater dissimilarity between the spoken languages) and enhanced SL task performance. Experiment 3 explores whether increased working memory demands of the task relate to increased differences in SL performance between monolinguals and bilinguals.

1.1.1. Participants

Forty Spanish monolinguals ($M_{age} = 21.8$, SD = 2.6; 37 females), forty Spanish-Basque bilinguals ($M_{age} = 21.2$, SD = 1.9; 35 females), and thirty-seven Spanish-English bilinguals ($M_{age} = 20.9$, SD = 2.3; 33 females) participated in this experiment. All participants reported normal or corrected-to-normal vision and no history of hearing or other neurological disorders. The study protocol was approved by the Ethics Committee of the Basque Center for Cognition, Brain and Language (BCBL) and carried out following the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. Before their inclusion in the study, all participants s provided written informed consent. They received monetary compensation for their participation.

Prior to the study, participants completed a language background questionnaire, a picture-naming task, and a matrix reasoning task from the Kaufman Brief Intelligence Test (Kaufman & Kaufman, 2014) to measure non-verbal IQ. The linguistic questionnaire included information about their age of acquisition, total exposure to each language, self-rated proficiency scales for each known language, and demographic questions such as age and gender. The picture naming task consisted of sixty-five images from the Basque, English, and Spanish Test (BEST) (de Bruin et al., 2017). Participants in the monolingual group named the images in Spanish and English (as an objective check of knowledge of English, although the participants had reported it was null or minimal), whereas the bilingual groups named them in Spanish and their L2 (English or Basque). Additionally, participants in the bilingual groups completed the LexTALE test (Lemhöfer & Broersma, 2012) in Spanish and their respective L2. The combination of these tests provided information on productive and receptive L2 vocabulary in the bilingual groups.

For the three experiments, we recruited monolinguals by asking for participants who could not maintain conversations (even at a very basic level) in a language other than Spanish. To recruit bilinguals, we asked for participants who could easily maintain complex conversations in Spanish and Basque (Spanish-Basque bilinguals) or in Spanish and English (Spanish-English bilinguals). After recruitment, the procedure and criteria to classify a given participant as monolingual or bilingual was the following: participants self-rated their proficiency in the L2 on a scale of 0–10. Thus, the monolingual group self-assessed their skills in the L2 as 6 or lower, whereas the bilingual groups self-assessed their skills in the L2 at 7 or higher. In addition, we adminstered the BEST test as an objective measure of the level of L2 proficiency of the participants. In the BEST test (range of scores: 0–65), all the monolingual participants scored <35 and all the bilinguals above 45, confirming the great differences in their L2 skills. Of note, the reported age of acquisition of L2 by participants belonging to the monolingual group was later than that of participants belonging to the Spanish-English bilingual group (5 years old vs. 3 years old). This suggests that Spanish-English bilinguals started L2 acquisition before schooling and achieved high proficiency due to exposure outside the public school system.

We compared the three groups using ANOVAs with Helmert contrasts, contrasting the Spanish monolingual group against the two bilingual groups in the first level and both bilingual groups against each other in the second level. A complete table with demographic information and statistical comparisons can be found in Appendix A. The results suggested that the three groups did not differ in age ($p = 0.204$, $BF_{10} = 0.203$), non-verbal IQ ($p = 0.619$, $BF_{10} = 0.103$), self-reported Spanish proficiency ($p = 0.419$, $BF_{10} = 0.166$) or in the Spanish picture naming task ($p = 0.245$, $BF_{10} = 0.260$). As expected, the monolingual and bilingual groups differed in their exposure to Spanish (that is, bilinguals reported less exposure to Spanish compared with monolinguals due to their exposure to their L2), L2 age of acquisition, L2 exposure, L2 self-rated proficiency, and BEST picture naming performance in L2 (all $p < 0.001$). Even if monolinguals could have some knowledge of L2, they were functionally monolinguals, as they could not maintain a very basic conversation in their L2. Crucially, the two bilingual groups did not differ in their L1 and L2 exposure, age of acquisition, self-rated proficiency, BEST picture naming and LexTALE scores (all $p > 0.05$).

1.1.2. Materials

We constructed three auditory SL streams with eight trisyllabic words per stream. The syllables did not repeat within the streams, maintaining the transitional probabilities within the words at a constant value of 1. The eight words for each condition were randomly concatenated into continuous speech streams, with the constraint that there was no immediate word repetition and that each word appeared 80 times per stream.

The first of these streams, the *simple condition*, was composed entirely of consonant-vowel (CV) syllables and represented a version of the earliest statistical learning experiment (Saffran et al., 1996). An example of a word in this stream is */motufi/*. In the second and third conditions—or complex conditions—each word contained one consonant cluster (CCV) syllable, either at the beginning or the end of the word. The difference between these conditions was the sub-lexical phonotactic patterns that defined the syllables. In the *complex legal* condition, the consonant cluster syllable (i.e., */fre/, /bla/, /fle/, /gli/, /gra/, /pre/, /pli/, /tre/*) was plausible at the beginning or end of the word without interfering with its segmentation (e.g., */betafre/*). However, in the *complex illegal* condition, participants heard words with consonant clusters that generally mark the boundary between syllables (i.e., */rnu/, /gma/, /lgi/, /nfu/, /rbu/, /rfo/, /sfe/, /bso/*) and were thus not plausible at the beginning or end of words (e.g., */tenobso/*). Taking "tenobso" as an example, we may consider it illegal because the last syllable (i.e., "bso") would be very weird in any of the three languages spoken by the samples of participants (for instance, in the English word "absorb", "b" and "s" belong to different syllables). So, illegal clusters are those which do not happen or they happen with a low frequency in the languages spoken by the participants. These consonant clusters were selected because they provided interfering cues—i.e., syllable boundaries—when inserted into an uninterrupted speech stream.

We synthesized the streams and each separate word using the MBROLA software (Dutoit et al., 1996), using a constant pitch of 82.63 Hz, and a duration of 200 ms per syllable. No pauses were inserted between the words in any of the auditory streams. These constraints ensured minimal interference of segmentation cues due to varying pitch, amplitudes, or co-articulation of the syllables. Finally, following previous research (Onnis et al., 2016), we introduced an amplitude ramp for the first and final 5 s of the stream. This produced a fade-in and fade-out effect, giving the impression of an unbounded speech stream. The duration of each SL stream was around 7 min.

1.1.3. Procedure

Participants completed a familiarization phase followed by a 2-alternatives forced-choice (2AFC) test. They were asked to wear headphones and sit in a quiet room for the entire experiment. During the familiarization task, they listened to the artificial SL streams. As in prior studies, they were instructed to pay close attention to the artificial languages because later, they would answer questions about them (Saffran et al., 1999; Wang & Saffran, 2014). The order of presentation of each condition was counterbalanced across the participants. The 2AFC test immediately followed the familiarization phase in each condition. During each 2AFC test, participants heard two words (one target and one foil) separated by 500 ms of silence and decided which of the words belonged to the previously heard language. They responded using the keyboard (f and j keys) to indicate which word they thought was more similar to the previously heard language and their responses were automatically recorded. The trials and the order of presentation of target and foil words were fully randomized.

We included two types of foils during the test phase, synthesized in the same manner as the original words. The first type was non-words, created by inverting the syllables of the words in the stream. The second type of foils was part-words. We created these foils by combining the ending of one word and the beginning of another, maintaining the correspondence between syllables. For instance, at test, a word in the simple condition (*/bukoni/*) would be paired both with a non-word (*/nikobu/*) and with a part-word (*/konito/*) in separate trials. The inclusion of two types of foils allowed us to double the number of trials per condition to sixteen. With three familiarization and three test phases, the entire experiment lasted around 30 min. Participants were encouraged to take a small break between each condition. The tasks for this and the rest of the experiments were programmed in Psychopy (Peirce et al., 2019).

1.1.4. Data analysis

We analyzed accuracy in the three SL streams using a generalized linear mixed model (GLMM). We assumed a binomial distribution and a logit link for the accuracy scores, with Condition, Group, and their interaction as fixed effects of interest. The Condition factor was Helmert coded to contrast the simple against the complex conditions first and then contrast the two complex conditions against each other. Similarly, the Group factor was reverse Helmert coded according to our hypotheses. We first contrasted the two bilingual groups. Then we contrasted both bilingual groups against the monolingual group. These factors were dummy-coded for each different contrast. All contrasts were planned and coded explicitly according to our hypotheses. We decided to adopt a confirmatory approach, rather than an exploratory omnibus approach as used in previous research, to evaluate the effects of bilingual experience in all experiments more robustly through a priori planned contrasts. This approach makes the GLMM analysis directly interpretable without the need for corrected post-hoc tests (Schad et al., 2020). The GLMM model was fitted using the *glmer* function from the "lme4" package (Bates et al., 2015) in R (R Core Team, 2022). Following standard practice, we tried to fit the maximal random-effects structure then reduced it to achieve convergence by eliminating the correlations between random slopes or the random slopes themselves (Barr et al., 2013). The final model achieved convergence using the by-participant and by-item intercepts with by-participant uncorrelated slopes for the Condition factor.

We report both the frequentist and Bayesian versions of the tests where possible. We opted to report the results from the Bayesian framework because these provide robust tests of the differences between the groups while simultaneously testing for the null hypothesis (van Doorn et al., 2020). Exact *p*-values are reported up to the 0.001 level for frequentist tests. For Bayesian tests, we report the exact Bayes Factor (BF_{10}) from 0.001 to 100. BF10 values below 1 indicate more support for the null hypothesis, with values below 0.3 indicating moderate to considerable support for the null hypothesis. BF10 values above 1 support the alternative hypothesis and are considered as substantial support at values above 3 (Kelter, 2020; Wei et al., 2022). In general, BF10 values between 0.3 and 3 are considered inconclusive or very weak evidence for either hypothesis (Kelter, 2020; Wei et al., 2022). We used uninformative priors for all Bayesian analyses. Paired, one-sample and independent-samples frequentist and Bayesian tests were conducted in JASP (Love et al., 2019). However, to obtain an approximate BF_{10} from the GLMM model, we contrasted nested models, including each fixed factor and interaction step-wise, against an intercept-only model with the same random-effects structure (Wagenmakers, 2007). In this case, the BF_{10} was calculated using the Bayesian Information Criterion (BIC) of the intercept-only (m_0) and each fixed-effect nested model (m_1) with the following formula:

$$BF_{10} = exp\left(\frac{BIC(m_0) - BIC(m_1)}{2}\right)$$

1.2. Results and discussion

Two specific methodological control analyses were carried out to exclude possible confounding learning effects. Firstly, although the order of the conditions was counterbalanced across participants (Order 1: Simple-Legal-Illegal; Order 2: Legal- Illegal-Simple; Order 3: Illegal-Simple-Legal), we analyzed possible condition order effects by means of an H Kruskal-Wallis test for each task condition separately (Simple, Complex legal, Complex illegal), using Order (three levels: first, second, third) as the grouping variable and accuracy as the contrast variable. This analysis showed no performance difference according to the order of the conditions, indicating that levels of accuracy were similar whether the condition was carried out in first, second, or third place. Secondly, we compared participants' accuracy between the first and second halves of each 2AFC test by means of Wilcoxon tests to exclude the possibility that they learned to discriminate the target words better during the 2AFC tests. The results showed no differences for Simple and Complex Illegal conditions and decreased performance in the second part of the Complex legal condition ($Z = -2.53$, $p = 0.011$). These results show that participants did not learn to discriminate the words during the performance of the 2AFC tests.

We first tested whether the type of foil (i.e., non-words, part-words) influenced the 2AFC test results. Wilcoxon paired-samples tests over the aggregated accuracy suggested no differences between the non-word and part-word trials in the simple ($p = 0.764$, $BF_{10} = 0.106$), complex legal ($p = 0.567$, $BF_{10} = 0.112$), or complex illegal ($p = 0.992$, $BF_{10} = 0.104$) conditions. Thus, we did not include type of foil as a factor in the GLMM analysis. As a second step, we tested whether participants' average accuracy score was above chance level (50%) in the three conditions using Wilcoxon one-sample tests. Participants in the Spanish monolingual group performed above chance level in the simple ($M_{acc} = 68.4$, $SD = 12.1$, $p < 0.001$, $BF_{10} > 100$), complex legal ($M_{acc} = 58.3$, $SD = 14.4$, $p < 0.001$, $BF_{10} > 100$), and complex illegal ($M_{acc} = 65.0$, $SD = 12.8$, $p < 0.001$, $BF_{10} > 100$) conditions. Participants in the Spanish-English bilingual group also performed above chance level in the simple ($M_{acc} = 62.5$, $SD = 14.8$, $p < 0.001$, $BF_{10} > 100$), complex legal ($M_{acc} = 65.5$, $SD = 11.4$, $p < 0.001$, $BF_{10} > 100$), and complex illegal ($M_{acc} = 60.1$, $SD = 13.8$, $p < 0.001$, $BF_{10} > 100$) conditions. Similarly, participants in the Spanish-Basque bilingual group performed above chance level in the simple ($M_{acc} = 67.3$, $SD = 17.7$, $p < 0.001$, $BF_{10} > 100$), complex legal ($M_{acc} = 61.9$, $SD = 11.9$, $p < 0.001$, $BF_{10} > 100$), and complex illegal ($M_{acc} = 60.6$, $SD = 13.1$, $p < 0.001$, $BF_{10} > 100$) conditions. Fig. 2 presents the accuracy results by group and condition. Notably, participants' average performance on each test (around 60–70%) was in line with previous statistical learning studies reporting an average accuracy in the range of 55–70% (Erickson & Thiessen, 2015; Saffran, 2003).

The GLMM analysis (shown in Table 1) indicated a significant difference between the simple and complex conditions ($p = 0.001$, $BF_{10} = 2.816$), but no significant differences between the two complex conditions ($p = 0.508$, $BF_{10} = 0.017$). Specifically, participants performed better in the simple than the complex conditions, without any differences in performance between the legal and illegal complex conditions. There were no differences between Spanish-Basque and Spanish-English bilingual groups ($p = 0.425$, $BF_{10} = 0.018$). In addition, the monolingual and bilingual groups did not differ in their performance ($p = 0.398$, $BF_{10} = 0.018$). A significant two-way interaction indicated a difference between monolinguals and bilinguals in the complex conditions ($p = 0.010$, $BF_{10} = 0.329$). Importantly, this interaction seemed to be driven by the monolingual group having a lower score in the complex legal compared to the complex illegal conditions and is not indicative of a bilingual experience effect since direct differences between monolinguals and bilinguals in the legal or illegal conditions were not found. So, we can state that both complex conditions (i.e., legal and illegal) were more difficult than the simple condition. However, the pattern of data observed in the monolinguals (i.e., decreased accuracy in the legal compared to the illegal condition) requires additional research before drawing any conclusions about the existence of an increased difficulty level in the legal compared to illegal complex condition and/or suggesting any bilingualism-related feature that may have led to a similar

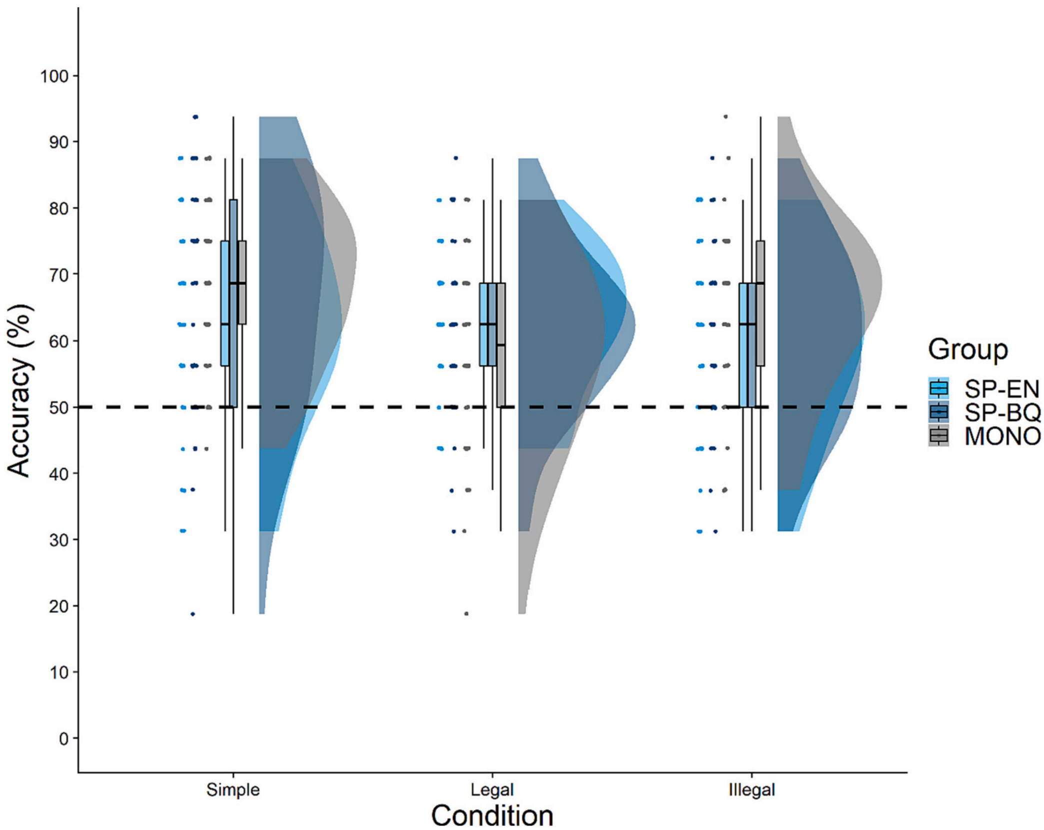

Fig. 2. Average accuracy by condition and group in Experiment 1.
Note. Raincloud plots of accuracy scores. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data) for each group. Dots reflect individual participant scores. The violin plots reflect the probability density of the scores. SP-EN = Spanish-English bilinguals; SP-BQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals.

Table 1
Accuracy GLMM results of Experiment 1.

Fixed Effects		Estimate	SE	z	*p*	BF_{10}
(Intercept)		**0.551**	**0.070**	**7.909**	**< 0.001**	**–**
Simple-Complex		**0.071**	**0.022**	**3.288**	**0.001**	**2.816**
Legal-Illegal		−0.023	0.034	−0.662	0.508	0.017
SPBQ-SPEN		0.036	0.045	0.797	0.425	0.018
MONO-BIL		0.022	0.026	0.845	0.398	0.018
Simple-Complex x SPBQ-SPEN		0.038	0.027	1.411	0.158	0.036
Simple-Complex x MONO-BIL		0.016	0.015	1.029	0.304	0.022
Legal-Illegal x SPBQ-SPEN		−0.012	0.042	−0.287	0.774	0.014
Legal-Illegal x MONO-BIL		**−0.062**	**0.024**	**−2.566**	**0.010**	**0.329**
Random Effects	*Group*	*Variance*	*SD*			
Item	(Intercept)	0.056	0.237			
Participant	(Intercept)	0.063	0.250			
	Simple-Complex	0.008	0.088			
	Legal-Illegal	0.002	0.040			

Note. Significant fixed effects terms are highlighted in bold. SE = standard error; SD = standard deviation; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; BIL = bilinguals.

performance in both complex conditions.

Experiment 1 revealed that more complex sub-lexical phonotactic patterns could modulate the word segmentation difficulty in SL tasks. All three groups, on average, performed the tasks above chance level. However, performance was lower in the complex than the simple condition, with no significant differences between the two complex conditions. More importantly, the SL performance of monolingual and bilingual participants did not significantly differ in any of the conditions, although monolinguals performed better in the illegal than the legal condition whereas no such differences were observed in the bilingual groups. In this context, some studies have shown that bilingual experience influences non-native phonetic learning (Antoniou et al., 2015). A possible explanation for these findings (i.e., no group-related differences) is that our task may not have been challenging enough to elucidate the differences shown in previous studies (Antoniou et al., 2015; Bartolotti et al., 2011; Wang & Saffran, 2014). In other words, bilingual experience might still come into play when the task is sufficiently challenging or produces a very strong interference with previously learned materials.

2. Experiment 2

In the previous experiment, learning the words largely depended on the transitional probabilities of adjacent syllables. Specifically, the patterns presented in these SL streams could not be generalized outside of their respective artificial speech streams. While language learning requires identifying and learning the phonotactic patterns within and across different words, it also encompasses learning the morphological rules that bind them together. Learning these rules allows generalization to instances where the words vary due to their dependency on other constituents (Endress & Bonatti, 2016; Peña et al., 2002). The prime example of this is affixal morphology, where structures are appended to the beginning (prefix) or ending (suffix) of a word to change its class and modify its meaning (e.g., ***un****touch****able****,* ***un****break****able***). These rules can span multiple elements, and in the SL literature, they are commonly known as non-adjacent dependencies (Misyak & Christiansen, 2007).

Some authors posit that similar SL mechanisms to those employed in word segmentation underlie learning these non-adjacent dependencies (Frost & Monaghan, 2016; Misyak & Christiansen, 2007). However, to our knowledge, no prior study has addressed whether bilingual experience can affect the implicit learning and generalization of non-adjacent dependencies.

In Experiment 2, we explored the capacity of monolinguals and bilinguals to learn non-adjacent dependencies from a SL stream and generalize them to novel items. Having observed no differences arising from bilingual experience in the first experiment, the purpose behind this manipulation was to test whether the groups would differ in another aspect of language learning by using this task as a proxy for morphological rule generalization. Experience with a wider range of grammatical and morphological rules could confer a bilingual advantage in morphological rule generalization. Also, the greater morphological complexity of verbs and nouns in Spanish and Basque compared to English (García Mayo & Villarreal Olaizola, 2011) may result in higher SL skills for this task in Spanish-Basque bilinguals compared to Spanish-English bilinguals.

2.1. Methods

2.1.1. Participants

Forty Spanish monolinguals (M_{age} = 21.7, SD = 2.4; 35 females), forty Spanish-Basque bilinguals (M_{age} = 21.8, SD = 2.2; 32 females), and forty Spanish-English bilinguals (M_{age} = 21.0, SD = 2.4; 36 females) participated in Experiment 2. Participants had a similar profile as those in Experiment 1 because the recruitment process was carried out in an analogous way. Appendix B shows the demographic information and statistical contrasts between groups.

2.1.2. Materials

We concatenated nine words with an AXC form, where A and C established a frame with constant syllables, and X was a fill syllable that could vary. Following previous work (Frost & Monaghan, 2016), we used plosive syllables for the A_C frames (*/ke/, /po/, /bi/, /ga/, /du/, /ti/*) and continuants for the X fill syllables (*/mu/*, */fe/*, */li/*). The frame syllables were randomized to create five counterbalanced versions of the SL stream. There were three different frames combined with three fill syllables, creating nine words for each version (that is, each stream included three different types of regularities to be learned but the specific syllables changed across the streams). We manually verified that none of the streams contained words or parts of words that existed in Spanish, English, or Basque. There were 100 repetitions of each word in every version of the stream. We synthetized the streams in the same manner as in Experiment 2, ensuring this time that there were no immediate repetitions of words with the same frame. The overall duration of the speech streams was about 12 min.

2.1.3. Procedure

The procedure was identical to the previous experiment except for the test trials. For the 2AFC test, we used three continuant syllables as generalization fills (*/se/, /ya/, /ni/*) and inserted them in the original frames to produce nine AYC generalization words. We tested participants on their ability to distinguish these words from three types of foils. The first two types were part-words created by concatenating the ending of the target word and the beginning of a different word from the stream. For example, a target word (*/kenipo/*) would be paired with a part-word starting at the fill syllable (*/nipoti/*) or ending with the fill syllable (*/gakeni/*). The third type of foils were non-words created by inverting the word's syllables (e.g., */ponike/*). Due to the randomized nature of the different versions, a target word in one of the lists could potentially be a non-word foil in any other version. The inclusion of three types of tests increased the number of trials to 27, representing a substantial increase from the previous experiment.

2.1.4. Data analysis

The data analysis closely followed that used in Experiment 1, the only difference being that there were no conditions with different difficulty level.

2.2. Results and discussion

As a first step, we tested whether the list or the type of foil influenced average performance on the test. An ANOVA suggested similar performance across all versions of the input stream ($p = 0.069$; $BF_{10} = 0.535$) and types of foil ($p = 0.527$, $BF_{10} = 0.052$), and no interaction between the two ($p = 0.217$, $BF_{10} = 0.009$). Furthermore, Wilcoxon one-sample tests against chance-level indicated that participants in the monolingual (M_{acc} = 59.1, SD = 17.5, $p = 0.011$, $BF_{10} > 100$), Spanish-English (M_{acc} = 57.3, SD = 13.4, $p = 0.005$, $BF_{10} > 100$) and Spanish-Basque (M_{acc} = 60.1, SD = 15.1, $p < 0.001$, $BF_{10} > 100$) groups performed above the 50% chance-level in the test. Fig. 3 depicts the results by group. For this test, the GLMM revealed no differences between the two bilingual groups (Estimate = 0.066, SE = 0.077, $z = 0.858$, $p = 0.391$, BF_{10} = 0.025), or between the monolingual and bilingual groups (Estimate = 0.009, SE = 0.044, $z = 0.216$, $p = 0.829$, $BF_{10} = 0.018$).

Overall, Experiment 2 extended the findings from the previous experiment by testing the generalization of non-adjacent dependencies learned during a SL task as a proxy for morphological rule learning. The main challenge of this experiment was that participants needed to extract the non-adjacent frames from the stream—rather than specific words—and generalize this knowledge to novel items. In line with previous studies, participants could generalize the frames extracted from the artificial language above chance-level, hinting towards similar implicit mechanisms for word and rule learning (Endress & Bonatti, 2016; Frost & Monaghan, 2016). However, the focal point of Experiment 2 was to compare the performance of bilinguals and monolinguals in generalizing information from the SL task. Our results indicated similar performance between the three groups and thus no bilingual experience effects.

Spanish and Basque possess a rich morphology at the verb and noun levels (García Mayo & Villarreal Olaizola, 2011). Although English morphology isn't as rich as that of Spanish and Basque, participants in the Spanish-English group might have benefited from their native knowledge of Spanish. In other words, the three languages may not differ sufficiently in their affixal morphology, such as to affect the non-adjacent dependencies generalization assessed in this SL task. Thus, although the task addressed a different aspect of language compared to the previous experiment, we still did not observe any effect stemming from the bilingual experience of either of the bilingual groups. A possibility raised by prior work is that bilingual experience might mainly influence learning at the lexical level (Kaushanskaya & Marian, 2009; Poepsel & Weiss, 2016). Hence, in Experiment 3, we presented participants with an audio-visual SL task targeting word-referent learning for which we created multiple mappings in order to vary the demands on phonological working memory processes.

3. Experiment 3

Studies have shown that infants and adults can learn mappings between words and visual referents (objects) across multiple scenarios through SL mechanisms, an ability known as cross-situational (CS)SL (Smith & Yu, 2008; Yu & Smith, 2007). The CSSL task differs from other SL tasks because participants learn the names for different visual referents by aggregating information across multiple contexts. In other words, they implicitly "discover" each object's name by seeing it in conjunction with distinct referents. This task was designed to mimic the crowded visual world in which infants initially learn words (Yu & Smith, 2007). It seems to target a crucial implicit mechanism for learning words in native and foreign languages (Benitez et al., 2016; Smith & Yu, 2008).

Perhaps one of the defining characteristics of bilingual experience is

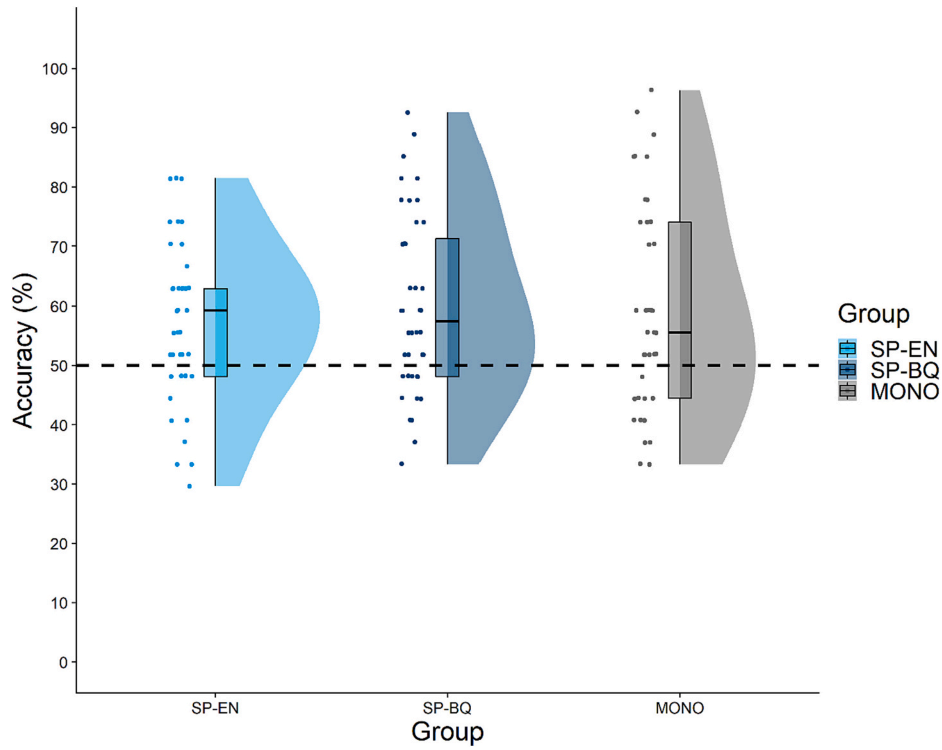

Fig. 3. Average accuracy by Group in Experiment 2.
Note. Raincloud plots of accuracy scores. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data) for each group. Dots reflect individual participant scores. The violin plots reflect the probability density of the scores. SP-EN = Spanish-English bilinguals; SP-BQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals.

learning to map two different words (e.g., *perro* – *dog/txakur*) to one referent, although this phenomenon is not unique to bilingualism as it also happens within languages. For instance, there are concepts with multiple names or *synonyms* (e.g., *paper* – *sheet*) and words (*homonyms)* which have multiple referents (e.g., *bat* can refer to an animal or sports equipment). Bilinguals have the added difficulty of learning these two types of mappings within two languages and across languages, potentially influencing their ability to learn them. Prior research has shown that bilinguals outperform monolinguals in a CSSL task when learning homonyms (i.e., one word to two referents), but not when learning exclusive one-to-one mappings (Poepsel & Weiss, 2016). These authors suggested phonological working memory capacity explained the better performance in bilinguals compared to monolinguals in this task. Nevertheless, other authors reported differences between bilinguals and monolinguals only in the one-to-one mappings (Escudero et al., 2016). Finally, some have found only minimal differences between bilinguals and monolinguals learning one-to-one and synonym (i.e., two words to one referent) mappings (Benitez et al., 2016).

To address these discrepant findings, in Experiment 3, we compared monolinguals and bilinguals in a CSSL task that entailed learning one-to-one, two-to-one (synonym), and one-to-two (homonym) word-referent pairs. A feature distinguishing this task from those carried out in Experiments 1 and 2 is that the CSSL task makes high demands on working memory resources (Weiss et al., 2020). If bilingual experience facilitates overall word learning, participants in the bilingual groups should outperform those in the monolingual group in the one-to-one mappings. However, if bilingual experience only potentiates learning multiple word-referent mappings, they should perform better than monolinguals only in the more challenging multiple mapping conditions.

3.1. Methods

3.1.1. Participants

Forty Spanish monolinguals (M_{age} = 21.7, SD = 2.5; 37 females), forty Spanish-Basque bilinguals (M_{age} = 21.9, SD = 1.9; 31 females), and thirty-seven Spanish-English bilinguals (M_{age} = 21.0, SD = 2.4; 32 females) participated in Experiment 3. Participants had a similar profile as those in Experiments 1 and 2 because the recruitment process was carried out in an analogous way. Appendix C shows the demographic information and statistical contrasts between groups.

3.1.2. Materials

We created 30 words by randomly concatenating letters from a pool of consonants (b, d, g, k, l, m, n, p, s, t) and vowels (a, e, i, o, u) in an alternating manner. The words had either a CVCVCV (e.g., */ninugo/*) or a VCVCV (e.g., */udili/*) form and were synthesized using the Mac OS X system's Text-to-Speech software with the Spanish female voice "Monica". Since these words were presented in isolation, we did not enforce as much control (e.g., constant pitch, vowel duration) as in the previous two experiments, in order to create more natural-sounding stimuli. After compiling an initial list, we eliminated words that existed in or sounded like actual Spanish, English, or Basque words. The final 30 words were paired with 30 color depictions of non-existant objects manually selected from the NOUN database (Horst & Hout, 2016) based on their color saliency and visual complexity scores.

The critical manipulation in this task involved creating the different word-referent mappings. As shown in Fig. 4, twelve words were randomly paired with twelve objects to form the *exclusive* (one-to-one) mappings. For the *homonym* (one-to-two) mappings, we randomly paired six words and twelve objects, producing words with two distinct referents. Conversely, for the *synonym* (two-to-one) mappings, twelve words were randomly paired with six objects, such that each of these objects had two different names. We manually verified that the words

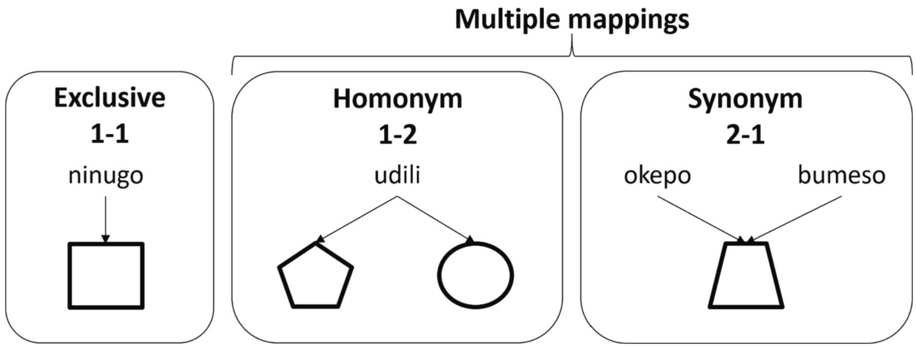

Fig. 4. *Word-referent mappings in Experiment 3.*
Note. The panel on the left depicts an example of the Exclusive one-to-one mappings, where one word was consistently presented with one visual referent. The middle panel illustrates an example of a Homonym one-to-two mapping, where one word was paired with two distinct visual referents. The panel on the right shows an example of a Synonym two-to-one mapping, where two words referred to the same visual referent. The depicted words are actual examples from the experiment, but we replaced the objects from the NOUN dataset with simple geometric figures for illustrative purposes.

and objects in the multiple mapping conditions were not very similar to avoid artificially increasing task difficulty.

3.1.3. *Procedure*

As in the previous experiments, participants wore headphones and sat in front of a computer screen in a quiet room. They completed three blocks of alternating familiarization and test phases. In each familiarization phase, they saw thirty-two scenes composed of three objects horizontally aligned at the center of the screen over a white background. Thus, participants viewed a total of 96 objects in each block (i.e., 32 trials × 3 objects in each trial). The participants also heard the name for each object depicted in each scene, starting after one second and with a one-second pause between names. Participants were instructed to pay attention to the different scenes because they would later perform a test based on them. Within each block, each 1–1 mapping (that is, “Word A-Object 1”; 12 words, 12 objects) was presented 4 times giving rise to 48 word-object mappings. Moreover, each 1–2 or homonym mapping (that is, “Word A-Object 1”; “Word A-Object 2”; 6 words, 12 objects) was presented twice giving rise to 24 word-object mappings. In addition, each 2–1 or synonym mapping (that is, “Word A-Object 1”; “Word B-Object 1”; 12 words and 6 objects) was also presented twice giving rise to 24 word-object mappings. The objects’ location (left, center, right) and the order of the audios were fully randomized in each scene so that there was no particular location for a given word-referent (i.e., word-object) mapping. Following previous studies (Poepsel & Weiss, 2016), only one side of the multiple mappings was presented during the first half of the scenes. During the second half, the other side of the multiple mappings was shown, simulating a change of context. This manipulation allowed us to test the effects of primacy and recency on the multiple mappings. We created four counterbalanced lists by varying which side of the two multiple mapping conditions was presented during the first and second halves of the familiarization phase. Thus, the order of presentation of scenes was pseudo-randomized for each familiarization phase so that no word or referent, including two sides of the multiple mappings, appeared in contiguous scenes.

Following each familiarization phase, participants completed a 2AFC test. In each trial, two horizontally-aligned objects (one target and one foil) were presented, and the target object’s name was played. The order of presentation of each trial and the position of target and foil objects were fully randomized. Participants were instructed to select the object they thought corresponded to the heard name using the keyboard (“f” and “j” keys). All the one-to-one mappings were tested during the three test blocks. However, only one side of the multiple mappings was presented during the first two tests blocks (the order of the tested mappings was counterbalanced). In detail, in line with Poepsel and Weiss (2016), if the primacy mapping (i.e., the first label or word for an object encountered by the participant) was tested after the first familiarization phase, then the recency mapping (i.e., the second label or word for an object encountered by the participant) was tested after the second familiarization, and vice versa. Each test phase consisted of 50% primacy and 50% recency mappings. The complete word-referent list was used for the last block. The entire experiment lasted about 45 min, and participants were encouraged to take small breaks between each block.

3.1.4. *Data analysis*

We modeled the exclusive and multiple (synonym/homonym) mappings data separately using GLMMs, similar to the previous two experiments. For the exclusive mappings, we treated block as a continuous linear factor and compared the reverse Helmert-coded groups and their interaction with the blocks as fixed effects. In this case, the GLMM was performed over the proportion of correct data to avoid inflating the estimates for the linearized block (Mirman, 2017). As in the previous experiments, we contrasted the Spanish-Basque against the Spanish-English group first, and then both bilingual groups against monolinguals. We included group, block, and their interactions as fixed effects and the by-participant intercept and uncorrelated block slope as random effects.

For the more challenging multiple mappings, since participants were tested on different mapping sides during the first and second blocks, we only contrasted the groups in the final test at block three. We created a factor to account for the order of presentation in the CSSL task. During the familiarization phase, the mappings displayed first represented the *primacy* mappings, and those displayed later the *recency* mappings. The group contrasts were the same as for the exclusive mappings. We tested the order-of-presentation factor and its interactions with group as fixed effects. We included the by-participant and by-item intercepts and the by-participant slope for order of presentation as random effects in the final GLMMs.

3.2. *Results and discussion*

Before the analysis, we removed one participant from the monolingual group who performed at chance level in all three exclusive blocks. We first confirmed there were no differences between the lists across the three blocks for the exclusive mappings, as the counterbalanced lists only targeted the multiple mappings. A series of ANOVAs showed there were no differences due to the list used in the first ($p = 0.901$, $BF_{10} = 0.058$), second ($p = 0.586$, $BF_{10} = 0.096$), or third block ($p = 0.815$, $BF_{10} = 0.066$). Next, we compared participants’ accuracy on the exclusive mappings to chance level (50%) in the last block. Wilcoxon one-sample tests indicated that Spanish monolinguals ($M_{acc} = 83.1$, SD $= 14.6$, $p < 0.001$, $BF_{10} > 100$), Spanish-English bilinguals ($M_{acc} = 86.9$, SD $= 14.2$, $p < 0.001$, $BF_{10} > 100$), and Spanish-Basque bilinguals ($M_{acc} = 91.0$, SD $= 10.2$, $p < 0.001$, $BF_{10} > 100$) performed significantly above chance level. Fig. 5 depicts the accuracy for each group across the three blocks.

Table 2 shows the results of the GLMM for the exclusive mappings. Unsurprisingly, the analysis revealed a significant effect of block ($p <$ 0.001, $BF_{10} > 100$), indicating that participant's accuracy increased across the three blocks. The results also indicated no significant differences between the two bilingual groups ($p = 0.054$, $BF_{10} = 0.190$). However, there were significant differences between groups (mainly supported by the GLMM), as bilinguals outperformed monolinguals ($p =$ 0.004, $BF_{10} = 1.902$). The interactions between block and group did not reach statistical significance ($p = 0.085$, $BF_{10} = 0.079$; $p = 0.123$, BF10 $= 0.061$; respectively).

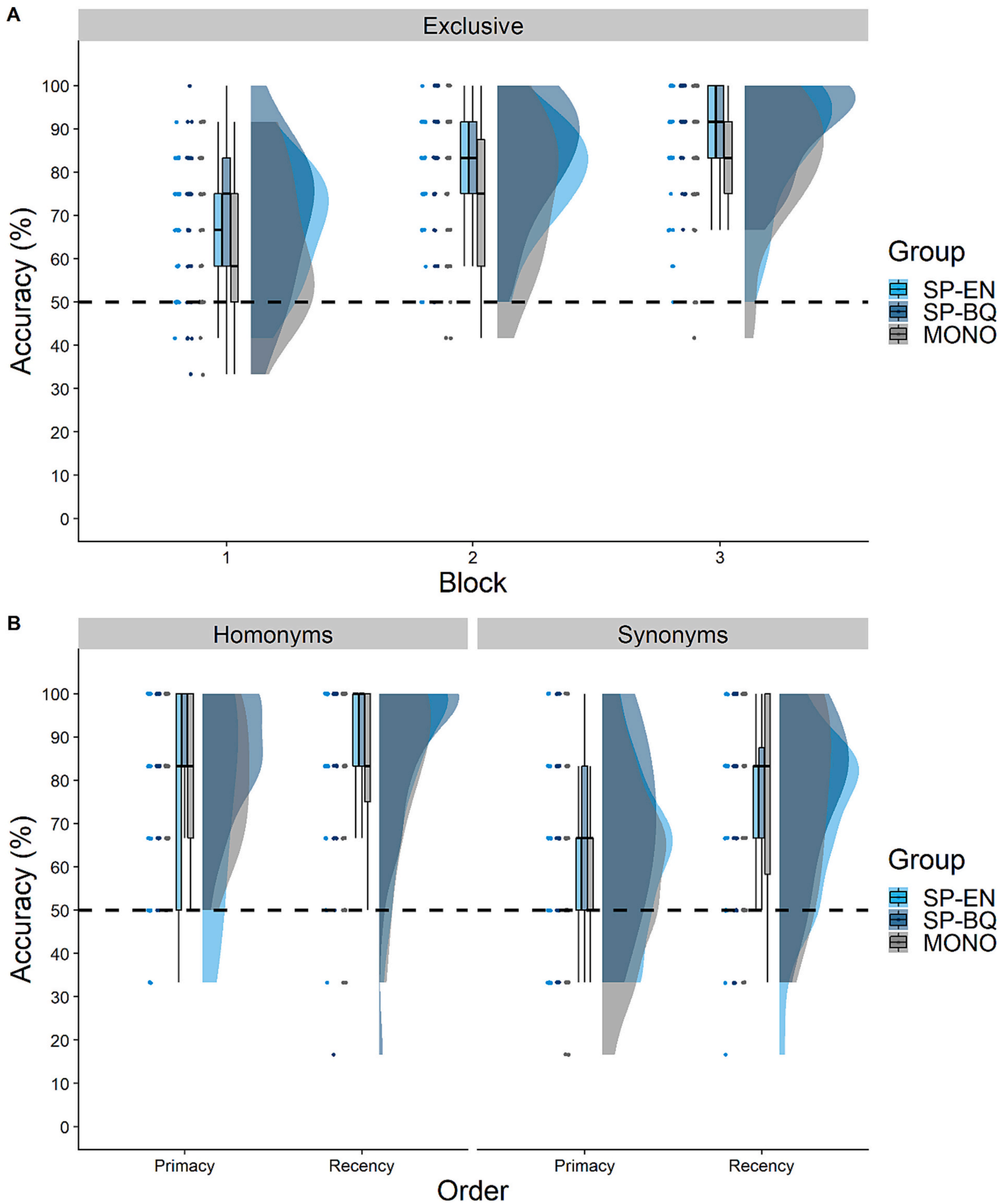

Fig. 5. *Average accuracy by group and block in Experiment 3.*
Note. **(A)** Accuracy by group and block in the one-to-one (Exclusive) mappings. **(B)** Accuracy by group, mapping type, and order of presentation for the multiple (Homonym and Synonym) mappings. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data) for each group. Dots reflect individual participant scores. The violin plots reflect the probability density of the scores. SP-EN = Spanish-English bilinguals; SP-BQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals.

Table 2
Accuracy GLMM results of exclusive condition in Experiment 3.

Fixed Effects		Estimate	SE	z	p	BF_{10}
(Intercept)		**1.411**	**0.064**	**22.141**	**< 0.001**	–
Block		**0.938**	**0.076**	**12.371**	**< 0.001**	**> 100**
SPBQ-SPEN		0.149	0.078	1.927	0.054	0.190
MONO-BIL		**−0.124**	**0.043**	**−2.886**	**0.004**	**1.902**
Block x SPBQ-SPEN		0.163	0.095	1.720	0.085	0.079
Block x MONO-BIL		−0.079	0.051	−1.542	0.123	0.061
Random Effects	*Group*	*Variance*	*SD*			
Participant	(Intercept)	0.250	0.500			
	Linear	0.059	0.243			

Note. Significant fixed effects terms are highlighted in bold. SE = standard error; SD = standard deviation; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; BIL = bilinguals.

In the homonym condition (Fig. 5B, left), Wilcoxon one-sample tests indicated that participants in the monolingual ($M_{acc} = 82.3$, SD = 14.6, $p < 0.001$, $BF_{10} > 100$), Spanish-Basque bilingual ($M_{acc} = 87.1$, SD = 12.1, $p < 0.001$, $BF_{10} > 100$), and Spanish-English bilingual ($M_{acc} = 80.6$, SD = 17.9, $p < 0.001$, $BF_{10} > 100$) groups performed above the 50% chance level on average. The GLMM results for the homonym mappings indicated a significant effect of the order of presentation (Estimate = −0.359, SE = 0.113, z = −3.173, $p = 0.002$, $BF_{10} = 4.874$), with better performance in the recency mappings, on average, than in the primacy mappings. There were no differences between the two bilingual groups (Estimate = 0.236, SE = 0.145, z = 1.627, $p = 0.104$, $BF_{10} = 0.148$), or between the monolingual and bilingual groups (Estimate = −0.063, SE = 0.082, z = −0.769, $p = 0.442$, $BF_{10} = 0.034$). There were no significant interactions between the order of presentation and the group contrasts (all $p > 0.05$, $BF_{10} < 0.1$).

In the synonym condition (Fig. 5B, right), Wilcoxon one-sample tests indicated that participants in the monolingual ($M_{acc} = 67.7$, SD = 16.0, $p < 0.001$, $BF_{10} > 100$), Spanish-Basque ($M_{acc} = 73.5$, SD = 14.2, $p < 0.001$, $BF_{10} > 100$), and Spanish-English ($M_{acc} = 66.9$, SD = 13.5, $p < 0.001$, $BF_{10} > 100$) groups performed above the 50% chance level on average. The GLMM for this condition revealed an effect of order of presentation in the same direction as for the homonym mappings (Estimate = −0.286, SE = 0.068, z = −4.211, $p < 0.001$, $BF_{10} > 100$). However, in this case, the analysis revealed a significant difference between Spanish-Basque and Spanish-English bilinguals (Estimate = 0.174, SE = 0.086, z = 2.023, $p = 0.043$, $BF_{10} = 0.211$), and no difference between monolinguals and bilinguals (Estimate = −0.037, SE = 0.079, z = −0.752, $p = 0.452$, $BF_{10} = 0.037$). As in the homonym condition, there were no significant interactions between the order of presentation and the group contrasts (all $p > 0.05$, $BF_{10} < 0.1$). Additionally, a Wilcoxon paired-samples test confirmed that accuracy was overall higher in the homonym than in the synonym mappings ($p < 0.001$, $BF_{10} > 100$).

Overall, Experiment 3 indicated that participants could learn the exclusive mappings and the two types of multiple mappings from the same audio-visual CSSL task. Contrary to our two previous experiments, this design also allowed us to test participants' learning at three stages, which showed a progressive increase in their performance for the exclusive mappings. Both bilingual groups outperformed the monolingual group on the exclusive one-to-one mappings. Thus, our results align with prior findings suggesting bilingual experience effects for these types of mappings in CSSL and other explicit word-learning tasks (Escudero et al., 2016; Kaushanskaya & Marian, 2009). However, these findings should be interpreted in the context of the more challenging multiple mappings. Therefore, it is unclear whether the observed bilingual experience effect is constrained to the exclusive mappings or emerged due to the presence of conflicting multiple mapping conditions, which presumably increased the executive and working memory demands of the task.

Despite a similar design, our results diverge from prior work reporting bilingual experience effects in the homonym but not in the exclusive mappings condition (Poepsel & Weiss, 2016). The homonym mappings were easier to acquire for our participants than the synonym mappings, and there were no differences between monolinguals and bilinguals on these multiple mapping conditions. A possible reason for these findings is that pure synonyms—especially those referring to concrete objects with the same meaning—are significantly rarer than homonyms in most natural languages (Hurford, 2003). Moreover, it is often necessary to introduce additional cues, such as speaker identity or spacing between presentations, to avoid direct competition from multiple mappings during learning (Benitez et al., 2016). Out of these two factors, we only spaced the presentations of multiple mappings during the learning phase, leading to the observed difference between primacy and recency mappings extensively reported in associative learning research (Pineño & Miller, 2005). These results suggest exciting avenues for investigating the role of other cues in how bilinguals and monolinguals learn multiple mappings.

4. General discussion

The present study examined the effects of bilingual experience across three SL experiments. We opted to use three well-established SL tasks to investigate representative aspects of foreign language learning; namely: word segmentation from continuous speech (Experiment 1), morphological rule generalization (Experiment 2), and word-referent learning (Experiment 3). The inclusion of two bilingual groups allowed us to disentangle effects from specific language pairs from the overall effects of bilingual experience. Additionally, our confirmatory analytical approach and large sample size facilitated a reliable evaluation of the raised research issues.

Overall, in all experimental manipulations, the average performance of all groups was significantly above chance and within the range reported by previous studies (Erickson & Thiessen, 2015; Perruchet & Pacton, 2006; Saffran, 2003), which is consistent with SL as an implicit mechanism for foreign language learning. Regarding group-related differences, Experiment 1 showed that, while manipulating the sub-lexical phonotactics of the SL stream affected participants' performance, there were no overall differences between any of the three groups for word segmentation despite a different pattern of performance in bilinguals compared to monolinguals. That is, whereas there were no differences between complex legal and complex illegal conditions in bilinguals, the monolingual group showed an enhanced performance in the complex illegal compared to complex legal condition. In Experiment 2, participants generalized the learned non-adjacent dependencies during the test phase, but again, we did not find any bilingual experience effects. Lastly, in Experiment 3, we tested participants' ability to learn exclusive and multiple word-referent pairs across three blocks. The results revealed that bilingual participants outperformed their monolingual peers when learning the exclusive but not the multiple word-referent mappings.

In order to interpret our findings, in the following paragraphs we will consider them from the perspective of SL as a cognitive mechanism, an experimental task, and a proxy for foreign language learning.

4.1. SL as a cognitive mechanism

Considering SL as a cognitive mechanism, other individual differences could perhaps be more important for auditory SL performance than bilingual experience, such as spontaneous synchronization to speech (Assaneo et al., 2019). The tasks in Experiments 1 and 2 were designed to isolate as much as possible speech segmentation and morphological rule generalization processes; that is, those tasks, and particularly the word segmentation task, did not demand a high amount of executive control or other higher order cognitive processes. The results showed no bilingual experience effects and, therefore, also no evidence that the degree of dissimilarity between language pairs is a possible variable enhancing SL skills. This lack of effect is not in line with some predictions derived from some mentioned theoretical accounts or frameworks, such as the perceptual assimilation model (Best & Tyler, 2007) and the structural sensitivity hypothesis (Weiss et al., 2020). As discussed in the first two experiments, the use of phonemes present in Spanish, Basque, and English and the fact that all the participants were native Spanish speakers (a quite complex language from a grammatical standpoint) may have contributed to the absence of group-related differences in Experiments 1 and 2, respectively.

In contrast, Experiment 3 elicited differences between monolinguals and bilinguals. Such differences could have emerged due to the multiple mapping conditions, which probably impose higher executive control demands (e.g., monitoring ongoing information and working memory maintenance of previous word-referent pairs) than those required in the other tasks. Prior studies seem to support this idea, showing bilingual experience effects only when participants learn from interfering SL streams (Bartolotti et al., 2011; Onnis et al., 2018). It is possible that enhanced working memory abilities contributed to a certain extent to the differences observed in Experiment 3. On the other hand, similarly to Experiments 1 and 2, the results of Experiment 3 did not provide evidence that specific language pairs may modulate performance in SL tasks. In addition, the absence of differences between both bilingual groups indicates that differences in the bilingual context –in this case, bilinguals learning a non-native L2 while living in a monolingual society (Spanish-English bilinguals) vs. bilinguals learning a native L2 due to living in a bilingual society (Spanish-Basque bilinguals)- does not modulate the performance in any SL task.

Experiment 3 was the only experiment that revealed differences between monolinguals and bilinguals and, even so, the relatively low BF value (1.902) suggests that this difference should be interpreted with caution. Below, we suggest three factors that could have contributed to these small or null bilingualism-related differences in the three experiments. The first two factors are related to the characteristics of the samples: 1) although our sample size was similar to (or even greater than) that in previous studies, it is possible that the small differences observed between monolinguals and bilinguals were related to a sample size that was not large enough to detect small effect sizes that are usually associated with bilingual advantages in cognition (Bialystok, 2017). 2) Our monolinguals were not "pure monolinguals" and so may have been benefited from residual knowledge of English, which may have been sufficient to improve their performance in SL. Nevertheless, we consider that improvement in the ability to detect regularities would probably require exposure and practice in the L2 on a regular basis. 3) SL may play a less important role in learning a foreign language in adulthood than it does in learning a native language in childhood (as in early bilingualism) or a second language in immersion contexts (such as in late bilingualism). In classroom situations, students learn a significant part of a foreign language through explicit processes, and implicit learning might only start to play a more prominent role after individuals have achieved a certain proficiency level in a new language (Ellis, 2015).

A possible explanation for the lack of differences between both bilingual groups is that both speaking similar and dissimilar language pairs conveys advantages. Specifically, although speaking two similar languages involves a reduced range of linguistic input compared to speaking two very dissimilar languages, such similarity between languages could involve a greater training of the executive system to prevent interference between them (Bialystok, 2017). If so, this training could enhance executive functions and compensate for the relatively reduced range of linguistic input, leading to equivalent performance on SL tasks. In this regard, a recent theoretical framework considered SL and executive functions as mechanisms that interact to support foreign language acquisition (Hirosh & Degani, 2018). Nevertheless, it is important to take into account that although several studies have found positive relationships between executive functions and SL (Park et al., 2020; Petok et al., 2022) others have reported negative relationships (Pedraza et al., 2023; Virag et al., 2015). Increasing the dialogue between research fields focusing on the relationship between bilingualism and executive functioning and bilingualism and SL could provide insights into how bilingualism modulates cognition (Donnelly et al., 2015; Janacsek & Nemeth, 2013; Janacsek & Nemeth, 2015; Park et al., 2020).

4.2. SL as an experimental task

Considering SL simply as an experimental task, participants are sensitive to specific and controlled manipulations of the probabilities between and within words that might target properties of a specific language (Finn & Hudson Kam, 2008; Onnis & Thiessen, 2013). Consequently, any bilingual experience effects may primarily rely on the properties of participants' known languages and not their overall bilingual experience, although the present study did not provide evidence supporting this view. Moreover, bilingual experience effects might emerge progressively through learning, and one-shot 2AFC tests may not be sensitive enough to capture them. Most SL studies base their results on a single familiarization and test phase, constraining their findings to what is learned upon first exposure to a foreign language (Romberg & Saffran, 2010). A common criticism with this approach is that performance, as measured by a one-shot 2AFC test, is highly variable and noisy, partly due to the low number of trials generated from these artificial streams (Siegelman & Frost, 2015). In other words, there are usually only a small number of target and foil words presented during the test phase, and using a single test introduces additional variability in participants' scores. Thus, while these experimental tasks could be adequate for measuring learning at the group level, they might not be sensitive enough to detect individual differences (Siegelman & Frost, 2015). We partially addressed these limitations in Experiment 3 by adding three familiarization and test phases for the twelve exclusive mappings, and here we found differences between monolinguals and bilinguals. However, the potential to observe any effects in other SL tasks might be reduced due to the variability in responses from one-shot 2AFC tests. Therefore, an issue that could have contributed to the modest group-related differences in SL is the low reliability of the SL measures. Nonetheless, recent research found that this problem had a bigger impact with samples of children than adults (Arnon, 2020). In addition to increasing the number of trials (which would improve the reliability of these tasks and enable neurophysiological techniques to be used to study the neural processes underlying performance), future research could further address these issues by using continuous measures to focus on the learning process rather than the learning outcome. In addition, it is worth noting that recent studies (de Bruin, 2019; Marian & Hayakawa, 2021; Surrain & Luk, 2019), have suggested future research should adopt a complementary approach by considering all participants as one group and analyzing bilingualism as a continuous variable. This approach would require recruiting a heterogeneous sample regarding their knowledge of an L2 (i.e., from no L2 knowledge to high L2 proficiency, and covering intermediate levels of L2).

4.3. SL as a proxy for foreign language learning

Lastly, considering SL as a proxy for foreign language learning, bilingual experience effects might arise mainly at the lexical level. Most of the literature supporting a bilingual advantage primarily focuses on explicit vocabulary learning tasks, highlighting that bilinguals might more easily establish links between new word forms (orthographic or phonological) and their meanings within the mental lexicon than monolinguals (Antoniou et al., 2015; Kaushanskaya & Marian, 2009). In this regard, the present study showed no bilingual experience effects using SL tasks as proxies for word segmentation and morphological levels. However, there were significant differences at the lexical level in the word-referent learning task. Our results extend prior findings by introducing synonyms and homonyms and showing that participants can learn these multiple mappings from the same implicit task, although no differences were found between monolinguals and bilinguals. Since experimental work seldom targets other aspects of language (e.g., grammar, syntax), it would be interesting to explore whether bilingual experience effects are constrained to the lexical level or extend to more higher-level aspects of a foreign language. To make additional progress in this field, we believe studies should focus on how life experiences modulate cognitive processes supporting implicit learning of linguistic and non-linguistic material.

5. Conclusion

Statistical learning is a robust cognitive process supported by a plethora of experimental findings. At the same time, decisions about stimuli selection, experiment design, participant selection, data collection, and statistical analysis can significantly influence the outcome and interpretation of SL experiments. This study explored three contrasting manipulations that could have potentially elicited bilingual experience effects that may have arisen through experience with different phonotactic patterns, the ability to generalize learned knowledge, and/or the capacity to learn exclusive and multiple word-referent mappings. The results of the current research support the following three conclusions: 1) SL improvement due to bilingualism is not a general finding but may occur under specific experimental conditions; specifically, in the present study, bilinguals outperformed monolinguals in an audio-visual SL task targeting word-referent learning but not in tasks involving word segmentation and morphological rule generalization; 2) the degree of dissimilarity between two spoken languages did not have an effect on SL tasks performance, suggesting that experience with a larger variety of linguistic input does not contribute to a bilingual advantage in SL tasks. Thus, the findings of the present study do not provide support for the predictions that had been formulated according to the structural sensitivity hypothesis and the perceptual assimilation model (i.e., that increased variety of linguistic input in the repertory of an individual leads to enhanced capability to learn a new language); and 3) tasks with high working memory and executive demands may contribute to some differences between monolinguals and bilinguals in SL tasks, possibly due to the strengthening of some working memory processes related to SL tasks, irrespective of the specific pairs of languages spoken by bilinguals.

Credit author contribution statement

JAA: Conceptualization, Methodology, Software, Investigation, Resources, Data Curation, Formal analysis, Writing – Original Draft, Visualization, Project administration, and Funding acquisition **JC**: Conceptualization, formal analysis, methodology, writing – review & editing **MC:** Conceptualization, Methodology, Writing – Review & Editing, Supervision, Project administration, and Funding acquisition.

Open practices statement

Anonymized participant information, experimental data, materials, and annotated analyses scripts for all experiments are freely available at https://osf.io/86sr4/. This study was not preregistered.

Declaration of Competing Interest

The authors declare no financial, commercial, or institutional competing interests.

Data availability

Data will be made available on request.

Acknowledgments

This work has received funding from "la Caixa" Foundation and from the European Union's Horizon 2020 research and innovation programme under the Marie-Skłodowska-Curie grant agreement no. 713673, personal fellowship code LCF/BQ/IN17/116200154004, awarded to JAA. MC was supported by the Basque Government through the BERC 2022-2025 program, Funded by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010/AEI/10.13039/501100011033, and by Agencia estatal de investigación (Reference number: PID2021-122918OB-I00)

Appendix A. Appendix

Linguistic profile and contrasts per group in Experiment 1

	Spanish Monolinguals		Spanish-Basque Bilinguals		Spanish-English Bilinguals		ANOVA			Helmert contrasts p-value	
	M	*SD*	*M*	*SD*	*M*	*SD*	$F_{(2, 114)}$	*p*	BF_{10}	MONO-BIL	SPBQ-SPEN
Age	21.8	2.6	21.2	1.9	20.9	2.3	1.610	0.204	0.303	0.088	0.586
Non-verbal IQ	101.5	8.3	100.7	8.6	99.8	6.6	0.481	0.619	0.120	0.406	0.592
Age of Acquisition L1 [a]	0.0	0.0	0.0	0.0	0.0	0.0	–	–	–	–	–
Total Exposure L1 (%)	91.4	8.1	60.3	11.8	64.5	9.9	111.685	**< 0.001**	**> 100**	**< 0.001**	0.070
Self-rated proficiency L1 (1–10)	9.8	0.4	9.8	0.2	9.9	0.3	0.876	0.419	0.166	0.411	0.294
BEST L1 (0–65)	64.6	0.6	64.6	0.7	64.8	0.5	1.425	0.245	0.260	0.217	0.243
Age of Acquisition L2	5.3	1.8	3.0	1.6	3.5	1.7	20.508	**< 0.001**	**> 100**	**< 0.001**	0.237
Total Exposure L2 (%)	7.7	7.2	32.5	13.1	33.5	9.4	80.766	**< 0.001**	**> 100**	**< 0.001**	0.664
Self-rated proficiency L2 (1–10)	4.3	1.8	8.6	0.7	8.1	0.8	149.854	**< 0.001**	**> 100**	**< 0.001**	0.074
BEST L2 (0–65)	25.9	8.2	55.9	6.9	55.9	4.2	264.291	**< 0.001**	**> 100**	**< 0.001**	0.991
LexTALE L1 (%)	–	–	92.9	5.6	93.8	6.3	–	–	–	–	0.506[b]
LexTALE L2 (%)	–	–	90.7	6.1	88.8	6.0	–	–	–	–	0.176[b]

Note. Significant contrast terms are highlighted in bold. SD = standard deviation; MONO = Spanish monolinguals; BIL = bilinguals; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals.

[a] Statistic and *p*-value undefined due to zero variance.

[b] Difference calculated using Welch *t*-tests.

Appendix B. Appendix

Linguistic profile and contrasts per group in Experiment 2

	Spanish Monolinguals (n = 40)		Spanish-Basque Bilinguals (n = 40)		Spanish-English Bilinguals (n = 40)		ANOVA			Helmert contrasts *p*-value	
	M	*SD*	*M*	*SD*	*M*	*SD*	$F_{(2, 114)}$	*p*	BF_{10}	MONO-BIL	SPBQ-SPEN
Age	21.7	2.4	21.8	2.2	21.0	2.4	1.405	0.250	0.252	0.478	0.132
Non-verbal IQ	99.8	6.6	101.1	8.8	100.1	6.6	0.320	0.727	0.103	0.578	0.567
Age of Acquisition L1 [a]	0.0	0.0	0.0	0.0	0.0	0.0	–	–	–	–	–
Total Exposure L1 (%)	91.4	8.6	61.3	8.1	64.8	9.9	136.974	**< 0.001**	**> 100**	**< 0.001**	0.081
Self-rated proficiency L1 (1–10)	9.9	0.3	9.8	0.4	9.9	0.3	1.084	0.342	0.194	0.925	0.144
BEST L1 (0–65)	64.7	0.6	64.6	0.7	64.8	0.4	1.784	0.172	0.344	0.829	0.063
Age of Acquisition L2	5.6	1.5	3.8	1.0	3.5	1.7	25.839	**< 0.001**	**> 100**	**< 0.001**	0.435
Total Exposure L2 (%)	8.6	8.6	34.0	9.6	33.4	9.3	99.097	**< 0.001**	**> 100**	**< 0.001**	0.762
Self-rated proficiency L2 (1–10)	4.2	1.7	8.5	1.0	8.1	0.8	146.243	**< 0.001**	**> 100**	**< 0.001**	0.145
BEST L2 (0–65)	25.6	9.6	56.3	6.5	55.9	4.3	243.908	**< 0.001**	**> 100**	**< 0.001**	0.802
LexTALE L1 (%)	–	–	93.4	5.6	93.3	6.8	–	–	–	–	0.986[b]
LexTALE L2 (%)	–	–	87.1	7.9	88.9	6.0	–	–	–	–	0.262[b]

Note. Significant contrast terms are highlighted in bold. SD = standard deviation; MONO = Spanish monolinguals; BIL = bilinguals; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals.

[a] Statistic and *p*-value undefined due to zero variance.

[b] Difference calculated using Welch t-tests.

Appendix C. Appendix

Linguistic profile and contrasts per group in Experiment 3

	Spanish Monolinguals (n = 40)		Spanish-Basque Bilinguals (n = 40)		Spanish-English Bilinguals (*n* = 37)		ANOVA			Helmert contrasts p-value	
	M	*SD*	*M*	*SD*	*M*	*SD*	$F_{(2, 114)}$	*p*	BF_{10}	MONO-BIL	SPBQ-SPEN
Age	21.7	2.5	21.9	1.9	21.0	2.4	1.712	0.185	0.328	0.696	0.072
Non-verbal IQ	102.2	8.0	101.5	6.2	100.6	6.5	0.477	0.622	0.120	0.313	0.583
Age of Acquisition L1 [a]	0.0	0.0	0.0	0.0	0.0	0.0	–	–	–	–	–
Total Exposure L1 (%)	90.7	7.5	62.5	10.8	65.0	9.4	111.191	**< 0.001**	**> 100**	**< 0.001**	0.242
Self-rated proficiency L1 (1–10)	9.9	0.3	9.8	0.3	9.9	0.2	1.54	0.219	0.286	0.962	0.082
BEST L1 (0–65)	64.7	0.5	64.7	0.5	64.8	0.4	1.618	0.203	0.304	0.145	0.284
Age of Acquisition L2	5.5	2.0	2.9	1.0	3.6	1.7	26.798	**< 0.001**	**> 100**	**< 0.001**	0.067
Total Exposure L2 (%)	8.9	7.2	30.5	9.3	33.1	8.9	95.487	**< 0.001**	**> 100**	**< 0.001**	0.183
Self-rated proficiency L2 (1–10)	3.9	1.6	8.5	0.9	8.1	0.9	180.184	**< 0.001**	**> 100**	**< 0.001**	0.082
BEST L2 (0–65)	26.9	9.2	56.2	7.3	55.7	4.3	210.094	**< 0.001**	**> 100**	**< 0.001**	0.766
LexTALE L1 (%)	–	–	91.1	7.6	93.9	5.9	–	–	–	–	0.077[b]
LexTALE L2 (%)	–	–	88.7	8.9	88.9	6.1	–	–	–	–	0.900[b]

Note. Significant contrast terms are highlighted in bold. SD = standard deviation; MONO = Spanish monolinguals; BIL = bilinguals; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals.

[a] Statistic and *p*-value undefined due to zero variance.

[b] Difference calculated using Welch t-tests.

References

Alonso, R. A. (2016). Cross-linguistic influence in second language acquisition. *Multilingual Matters*. https://doi.org/10.21832/9781783094837

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Bilingualism, 18*(4), 683–695. https://doi.org/10.1017/S1366728914000777

Antovich, D. M., & Graf Estes, K. (2020). One language or two? Navigating cross-language conflict in statistical word segmentation. *Developmental Science, 23*(6). https://doi.org/10.1111/desc.12960. Article e12960.

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods, 52*, 68–81. https://doi.org/10.3758/s13428-019-01205-5

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience, 22*(4), 627–632. https://doi.org/10.1038/s41593-019-0353-z

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartolotti, J., Marian, V., Schroeder, S. R., & Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00324. Article 324.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bengtson, J. (2011). The Basque language: History and origin. *International Journal of Modern Anthropology, 1*(4), 43–59. https://doi.org/10.4314/ijma.v1i4.3

Benitez, V. L., Bulgarelli, F., Byers-Heinlein, K., Saffran, J. R., & Weiss, D. J. (2020). Statistical learning of multiple speech streams: A challenge for monolingual infants. *Developmental Science, 23*(2). https://doi.org/10.1111/desc.12896. Article e12896.

Benitez, V. L., Yurovsky, D., & Smith, L. B. (2016). Competition between multiple words for a referent in cross-situational word learning. *Journal of Memory and Language, 90*, 31–48. https://doi.org/10.1016/j.jml.2016.03.004

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 13–34). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.17.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development, 57*(2), 498–510. https://doi.org/10.2307/1130604.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin, 143*(3), 233–262. https://doi.org/10.1037/bul0000099

Bialystok, E., Poarch, G., Luo, L., & Craik, F. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychology and Aging, 29*(3), 696–705. https://doi.org/10.1037/a0037254

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience and biobehavioral reviews, 112*, 279–299. https://doi.org/10.1016/j.neubiorev.2020.01.032.

Cromdal, J. (1999). Childhood bilingualism and metalinguistic skills: Analysis and control in young Swedish–English bilinguals. *Applied Psycholinguistics, 20*(1), 1–20. https://doi.org/10.1017/S0142716499001010.

de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Science, 9*(3). https://doi.org/10.3390/bs9030033

de Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology, 8*, 522. https://doi.org/10.3389/fpsyg.2017.00522

Bulgarelli, F., & Weiss, D. J. (2016). Anchors aweigh: The impact of overlearning on entrenchment effects in statistical learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 42*, 1621–1631. https://doi.org/10.1037/xlm0000263

Calvo, N., Ibáñez, A., & García, A. M. (2016). The impact of bilingualism on working memory: A null effect on the whole may not be so on the parts. *Frontiers in Psychology, 25*(7). https://doi.org/10.3389/fpsyg.2016.00265. Article 265.

Donnelly, S., Brooks, P. J., & Homer, B. D. (2015). Examining the bilingual advantage on conflict resolution tasks: A meta-analysis. *Cognitive Science*. Article 467400.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., … Wagenmakers, E. J. (2020). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review, 28*, 1–14. https://doi.org/10.3758/s13423-020-01798-5

Dutoit, T., Pagel, V., Pierret, N., Bataille, E., & van der Vrecken, O. (1996). MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *, 1393–1396. Proceeding of fourth international conference on spoken language processing, ICSLP*. https://doi.org/10.1109/icslp.1996.607874

Ellis, N. C. (2015). *Implicit AND explicit language learning.* John Benjamins publishing company. https://doi.org/10.1075/sibil.48.01ell

Endress, A. D., & Bonatti, L. L. (2016). Words, rules, and mechanisms of language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 7*(1), 19–35. https://doi.org/10.1002/wcs.1376

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37*, 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Escudero, P., Mulak, K. E., Fu, C. S. L., & Singh, L. (2016). More limitations to monolingualism: Bilinguals outperform monolinguals in implicit word learning. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01218. Article 1218.

Estes, K. G., Gluck, S. C. W., & Grimm, K. J. (2016). Finding patterns and learning words: Infant phonotactic knowledge is associated with vocabulary size. *Journal of Experimental Child Psychology, 146*, 34–49. https://doi.org/10.1016/j.jecp.2016.01.012

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition, 108*(2), 477–499. https://doi.org/10.1016/j.cognition.2008.04.002

Foursha-Stevenson, C., & Nicoladis, E. (2011). Early emergence of syntactic awareness and cross-linguistic influence in bilingual children's judgments. *International Journal of Bilingualism, 15*(4), 521–534. https://doi.org/10.1177/1367006911425818.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences, 19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition, 147*, 70–74. https://doi.org/10.1016/j.cognition.2015.11.010

García Mayo, M. D. P., & Villarreal Olaizola, I. (2011). The development of suppletive and affixal tense and agreement morphemes in the L3 english of basque-spanish bilinguals. *Second Language Research, 27*(1), 129–149. https://doi.org/10.1177/0267658310386523

Hirosh, Z., & Degani, T. (2018). Direct and indirect effects of multilingualism on novel language learning: An integrative review. *Psychonomic Bulletin & Review, 25*(3), 892–916. https://doi.org/10.3758/s13423-017-1315-7

Horst, J. S., & Hout, M. C. (2016). The novel object and unusual name (NOUN) database: A collection of novel images for use in experimental research. *Behavior Research Methods, 48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Hurford, J. R. (2003). Why synonymy is rare: Fitness is in the speaker. In W. Banzhaf, J. Ziegler, T. Christaller, P. Dittrich, & J. T. Kim (Eds.), *Advances in artificial life* (pp. 442–451). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-39432-7_47.

Janacsek, K., & Nemeth, D. (2013). Implicit sequence learning and working memory: Correlated or complicated? *Cortex, 49*(8), 2001–2006. https://doi.org/10.1016/j.cortex.2013.02.012

Janacsek, K., & Nemeth, D. (2015). The puzzle is complicated: When should working memory be related to implicit sequence learning, and when should it not? (Response to Martini et al.). *Cortex, 64*, 411–412. https://doi.org/10.1016/j.cortex.2014.07.020

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*(3–4), 159–207. https://doi.org/10.1006/cogp.1999.0716

Kaufman, A. S., & Kaufman, N. L. (2014). Kaufman brief intelligence test. In *Encyclopedia of special education* (2nd ed.). https://doi.org/10.1002/9781118660584.ese1325

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin and Review, 16*(4), 705–710. https://doi.org/10.3758/PBR.16.4.705

Kelter, R. (2020). Analysis of Bayesian posterior significance and effect size indices for the two-sample t-test to support reproducible medical research. *BMC Medical Research Methodology, 20*, 88. https://doi.org/10.1186/s12874-020-00968-2

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*(2), 35–42. https://doi.org/10.1016/S0010-0277(02)00004-5

Kuipers, J. R., & Westphal, K. H. (2021). Auditory processing and high task demands facilitate the bilingual executive control advantage in young adults. *Journal of Neurolinguistics, 57*. https://doi.org/10.1016/j.jneuroling.2020.100954. Article 100954.

Kuo, L. J., & Anderson, R. C. (2010). Beyond cross-language transfer: Reconceptualizing the impact of early bilingualism on phonological awareness. *Scientific Studies of Reading, 14*, 365–385. https://doi.org/10.1080/10888431003623470

Kuo, L. J., & Anderson, R. C. (2012). Effects of early bilingualism on learning phonological regularities in a new language. *Journal of Experimental Child Psychology, 111*, 455–467. https://doi.org/10.1016/j.jecp.2011.08.013

Kuo, L. J., & Kim, T. J. (2014). Effect of early bilingualism on metalinguistic development and language processing: Evidence from Chinese-speaking bilingual children. In X. Chen-Bumgardner, Q. Wang, & Y. Luo (Eds.), *Reading development and difficulties in monolingual and bilingual Chinese children* (pp. 171–190). New York, N.Y: Springer.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., … Wagenmakers, E. J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software, 88*(1), 1–17. https://doi.org/10.18637/jss.v088.i02

Marian, V., & Hayakawa, S. (2021). Measuring bilingualism: The quest for a "bilingualism quotient". *Applied PsychoLinguistics, 42*(2), 527–548. https://doi.org/10.1017/s0142716420000533

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), 101–111. https://doi.org/10.1016/S0010-0277(01)00157-3

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science, 12*(3), 369–378. https://doi.org/10.1111/j.1467-7687.2009.00822.x

Mirman, D. (2017). *Growth curve analysis and visualization using R*. CRC press. https://doi.org/10.1201/9781315373218

Misyak, J., & Christiansen, M. H. (2007). Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language. *Proceedings of the Annual Meeting of the Cognitive Science Society, 29*(29).

Nemeth, D., Janacsek, K., Csifcsak, G., Szvoboda, G., Howard, J. H., Jr., & Howard, D. V. (2011). Interference between sentence processing and probabilistic implicit sequence learning. *PLoS One, 6*(3). https://doi.org/10.1371/journal.pone.0017577. Article e17577.

Onnis, L., Chun, W. E., & Lou-Magnuson, M. (2018). Improved statistical learning abilities in adult bilinguals. *Bilingualism, 21*(2), 427–433. https://doi.org/10.1017/S1366728917000529

Onnis, L., Frank, S., Yun, H., & Lou-Magnuson, M. (2016). Statistical learning bias predicts second-language reading efficiency. *Annual Meeting of the Cognitive Science Society*.

Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition, 126*(2), 268–284. https://doi.org/10.1016/j.cognition.2012.10.008

Park, J., Yoon, H. D., Yoo, T., Shin, M., & Jeon, H. A. (2020). Potential and efficiency of statistical learning closely intertwined with individuals' executive functions: A mathematical modeling study. *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-020-75157-8. Article 18843.

Pedraza, F., Farkas, B. C., Vékony, T., Haesebaert, F., Janacsek, K., Anders, R., … Nemeth, D. (2023). *Evidence for a competitive relationship between executive functions and statistical learning*. bioRxiv. https://doi.org/10.1101/2023.01.19.524710.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*(5593), 604–607. https://doi.org/10.1126/science.1072901

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Petok, J. R., Dang, L., & Hammel, B. (2022). Impaired executive functioning mediates the association between aging and deterministic sequence learning. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, 1–17. https://doi.org/10.1080/13825585.2022.2153789

Pineño, O., & Miller, R. R. (2005). Primacy and recency effects in extinction and latent inhibition: A selective review with implications for models of learning. *Behavioural Processes, 69*(2), 223–235. https://doi.org/10.1016/j.beproc.2005.02.006

Poepsel, T. J., & Weiss, D. J. (2016). The influence of bilingualism on statistical word learning. *Cognition, 152*, 9–19. https://doi.org/10.1016/j.cognition.2016.03.001

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org.

Rivera, J. L. (2019). A study conception about language similarities. *Open Journal of Modern Linguistics, 9*, 47–58. https://doi.org/10.4236/ojml.2019.92005

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 906–914. https://doi.org/10.1002/wcs.78

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*(4), 110–114. https://doi.org/10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language, 110*. https://doi.org/10.1016/j.jml.2019.104038. Article 104038.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 372*(1711). https://doi.org/10.1098/rstb.2016.0059. Article 20160059.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Surrain, S., & Luk, G. (2019). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition, 22*(2), 401–415. https://doi.org/10.1017/S1366728917000682

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39*(4), 706–716. https://doi.org/10.1037/0012-1649.39.4.706

Tsui, A. S. M., Erickson, L. C., Mallikarjunn, A., Thiessen, E. D., & Fennell, C. T. (2021). Dual language statistical word segmentation in infancy: Simulating a language-mixing bilingual environment. *Developmental Science, 24*(3). https://doi.org/10.1111/desc.13050. Article e13050.

Virag, M., Janacsek, K., Horvath, A., Bujdoso, Z., Fabo, D., & Nemeth, D. (2015). Competition between frontal lobe functions and implicit sequence learning: Evidence from the long-term effects of alcohol. *Experimental Brain Research, 233*(7), 2081–2089. https://doi.org/10.1007/s00221-015-4279-8

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00953. Article 953.

Wei, Z., Yang, A., Rocha, L., Miranda, M. F., & Nathoo, F. S. (2022). A review of Bayesian hypothesis testing and its practical implementations. *Entropy, 24*(2), 161. https://doi.org/10.3390/e24020161

Weiss, D. J., Schwob, N., & Lebkuecher, A. L. (2020). Bilingualism and statistical learning: Lessons from studies using artificial languages. *Bilingualism: Language and Cognition, 23*(1), 92–97. https://doi.org/10.1017/S1366728919000579

Yim, D., & Rudoy, J. (2013). Implicit statistical learning and language skills in bilingual children. *Journal of Speech Language and Hearing Research, 56*, 310–322. https://doi.org/10.1044/1092-4388(2012/11-0243)

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x