Parafoveal processing in bilingual readers: Semantic access within but not across languages

Liv J. Hoversten[1] and Clara D. Martin[2,3]

[1]University of California, Santa Cruz

[2]Basque Center on Cognition, Brain and Language

[3]Ikerbasque, Basque Foundation for Science

Short Title: Bilingual semantic preview benefits

Word Count: 10,737

Author Note

Liv J. Hoversten https://orcid.org/0000-0002-7732-6055

Clara D. Martin https://orcid.org/0000-0003-2701-5045

Data and materials for the project are openly available at https://osf.io/nb78z/. We have no conflicts of interest to disclose. This research was supported by the Basque Government through the BERC 2022-2025 program and by the Spanish State Research Agency through the BCBL Severo Ochoa excellence accreditation CEX2020-001010-S and through a Juan de la Cierva- Incorporación postdoctoral grant to L.J.H. The research was also supported by the Spanish Ministry of Economy and Competitiveness (PID2020-113926GB-I00 to C.D.M.) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programmer (grant agreement No 819093 to C.D.M.).

Please address correspondence to: Liv Hoversten, Department of Psychology, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95604, Email: lhoverst@ucsc.edu

ABSTRACT

Prior research has investigated the quality of information a reader can extract from upcoming parafoveal words. However, very few studies have considered parafoveal processing in bilingual readers, who may differ from monolinguals due to slower lexical access and susceptibility to cross-language activation. This eye-tracking experiment therefore investigated how bilingual readers process parafoveal semantic information within and across languages. We used the boundary technique to replace a preview word in a sentence with a different target word during the first rightward saccade from the pre-target region. We manipulated both preview language (non-switch vs. code-switch) and semantic relatedness (synonym/translation vs. unrelated) between previews and targets. Upon fixation, target words always appeared in the same language as the rest of the sentence to create an essentially monolingual language context. Semantic preview benefits emerged for non-switched synonym previews but not for code-switched translation previews. Furthermore, participants skipped code-switched previews less often than non-switched previews and no more often than previews that were unfamiliar to them. These data suggest that bilinguals *can* extract within-language semantic information from the parafovea in both native and non-native languages, but that cross-language words are not accessible while reading in a monolingual language mode, as per the *partial selectivity hypothesis* of bilingual language control.

Keywords: bilingual reading; bilingual language control; eye movements; parafoveal processing; semantic preview benefit

Public significance: Readers often pre-process the next upcoming word in a text before looking at it directly. However, little is known about whether and how *bilinguals* pre-process upcoming words during reading comprehension. This study provides evidence that young adult bilingual readers *can* efficiently pre-process the meaning of an upcoming word when it belongs to the same language as the rest of the sentence but *not* when it is unexpectedly switched into their other language. This result indicates important limits to cross-language activation while bilinguals read text for comprehension.

INTRODUCTION

Readers typically fixate most words in a text in order to bring each word into clear enough focus to extract the detailed visual information needed for lexical processing. Yet skilled readers routinely skip some words (Rayner, 1998). Most models of eye movement behavior agree that at least some amount of attention can be allocated to the upcoming parafoveal word during a fixation (e.g., starting to process word *n+1* while fixating word *n*; Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle et al., 2009; Snell et al., 2018b). Several prior studies have established that different characteristics of the parafoveal word speed up reading and contribute to skipping behavior (i.e., not fixating a word), suggesting that parafoveal information is processed to some extent prior to, and sometimes in lieu of, direct fixation (Reingold et al., 2012; see Andrews & Veldre, 2019 for a review). Nonetheless, very few studies to date have examined parafoveal processing in bilingual readers, for whom lexical access has been shown to differ from their monolingual counterparts (e.g., Gollan et al., 2008; Kroll et al., 2012). This study therefore set out to assess how bilinguals extract semantic codes from the parafoveal word within and across languages during sentence reading to establish the underlying cognitive mechanisms driving eye movements in bilingual readers.

This introduction will be divided into three parts. First, we will introduce relevant research on parafoveal processing in monolinguals. Second, we will present a summary of the few studies to date on parafoveal processing in bilinguals and then explain why a more systematic investigation of this topic is needed to distinguish among different possible explanations of the prior data. Finally, we will introduce the current study designed to clarify our understanding of the role of parafoveal processing in bilingual reading.

Parafoveal Processing in Monolingual Readers

One way to investigate parafoveal processing is to record a reader's eye movements in the gaze-contingent boundary paradigm, in which a critical preview word embedded in a sentence is replaced with a different target word when the eyes cross an invisible boundary prior to the critical word (Rayner, 1975; see Figure 1 for an example). Previous studies have used this technique to investigate whether the relationship between the preview and target words leads to shorter reading times upon fixation of the target compared to an unrelated word, known as a *preview benefit*. Clear preview benefits have been found when the preview and target share orthographic and/or phonological information (e.g., the preview "cahc" changes to the target word "cake"), suggesting that readers routinely extract orthographic and phonological codes from parafoveal words (Balota et al., 1985). The influence of semantic information in the parafovea has been less straightforward: Although early work failed to find evidence for a semantic preview benefit when previews and targets are semantically related (e.g., "frog" changes to "toad"; Altarriba et al., 2001; Rayner et al., 1986), it has now been demonstrated in several languages, including Chinese, Korean, German, Finnish, and English (Hohenstein et al., 2010; Schotter, 2013; White et al., 2008; Yan et al., 2009; Yan et al., 2019). This effect is generally attributed to a covert shift of attention such that lexical access can begin on the parafoveal word prior to direct fixation (Reichle et al., 2009).

Figure 1. Depiction of the boundary paradigm with a semantic relatedness manipulation between preview and target words. For display purposes, preview and target words are indicated in red and blue, respectively. The asterisk denotes the horizontal position of the

eyes A) before and B) after the boundary change. The gray dotted line represents the position of the invisible boundary that triggers a display change when the eyes cross to the right of it.

One potential source of semantic preview benefits is facilitated integration of information extracted from the parafovea with similar information extracted upon fixation of the target word (Schotter, 2013). Semantic overlap between preview and target words may give target processing a head start prior to direct fixation. Conversely, some recent accounts claim that semantic preview benefits arise based on the plausibility of the preview as a continuation of the sentence rather than the relationship between the preview and target per se (Schotter & Jia, 2016; Veldre & Andrews, 2016; Veldre et al., 2020). According to such an account, plausibility effects should be detectable from the earliest stages of word identification of the parafoveal word, including the measure of skip rate, or the proportion of trials in which the critical word is not fixated before moving to the right of it. However, a large-scale study from Abbott & Staub (2015) contradicted this prediction, showing that preview plausibility did not affect skip rate but did affect fixation durations (cf. Veldre & Andrews, 2020, which found that a condition containing semantic and syntactic violations reduced skip rate). Accounts like the forced fixations hypothesis (Schotter & Leininger, 2016) could explain such a pattern with an early but not immediate effect of plausibility. This hypothesis suggests that some parafoveal information may arrive too late to affect the current saccade plan to fixate the target word but still early enough to initiate a new saccade program away from the target shortly after it is fixated. It is therefore possible that preview plausibility influences fixation times on the target without necessarily affecting the initial stages of word recognition reflected by skip rates. It is also possible that preview plausibility and preview-target integration both contribute to semantic preview benefits (Schotter et al, 2019). Regardless of the source(s), however, these accounts agree that semantic preview effects demonstrate that monolingual readers can indeed activate semantic information from a parafoveal preview prior to fixating the word directly.

Several factors influence the strength of semantic preview benefits, including the nature of the orthography (e.g., alphabetic vs. logographic scripts), the predictability of a word in context, and the strength of the relationship between previews and targets (see Schotter et al., 2012 for a review). For example, semantic preview benefits tend to be more robust in languages like Chinese, in which the orthography has a close relationship to the meaning of a word and the upcoming word often falls within foveal vision due to the high spatial density of information in the script (see Yan et al, 2009 for a discussion). In languages like English, which has a deep alphabetic orthography, semantic preview benefits may require a strong semantic relationship between previews and targets (i.e., synonyms) and/or semantic support for the preview from the preceding context (Schotter, 2013; Schotter & Jia, 2016; cf. López-Pérez et al., 2016 using the fixation-related potential technique with word pair reading), which may explain why early attempts using eye-tracking to find a semantic preview benefit in English failed but later attempts succeeded. Bilingualism may be yet another factor that could influence the presence of semantic preview benefits, for reasons explained below.

Parafoveal Processing in Bilingual Readers

It remains unclear whether bilinguals use parafoveal information in a similar way during reading, both because of established differences between monolingual and bilingual lexico-semantic access and because very few studies to date have investigated parafoveal processing in bilingual readers.

First, many studies have shown a dissociation between native and non-native reading, and reading of mixed language text involves further differences in processing (Grosjean, 1997; Titone, Libben, Mercier, Whitford, & Pivneva, 2011). For example, prior work has established that bilinguals experience subtly slower lexical access than their monolingual counterparts, particularly in a weaker non-native language (Ivanova & Costa, 2008; Martin, Costa, Dering, Hoshino, Wu, & Thierry, 2012; Martin, Thierry, Kuipers, Boutonnet, Foucart, & Costa, 2013). Additionally, a large body of research has shown that bilinguals are susceptible to cross-language activation, whereby both languages are simultaneously active and compete for selection (see Kroll et al., 2012 for a review). However, we do not yet fully understand how these differences impact eye movements during reading in bilinguals, particularly with regard to lexico-semantic processing of parafoveal information. For example, slower lexical access may decrease the amount of time and resources available to extract semantic information from a perceptually degraded parafoveal word while fixating the foveal word. Therefore, it is plausible that bilingual readers may not experience semantic preview benefits, particularly in their weaker language. On the other hand, one might expect cross-language activation to accelerate reading of a target word preceded by its translation as a parafoveal preview, thus producing robust cross-language semantic preview benefits.

Secondly, very few studies to date have examined whether bilinguals can extract semantic information from the parafovea during sentence reading. In a seminal study, Altarriba et al. (2001) presented Spanish-English cognates (words with shared form and meaning; e.g. , *crema* means "cream"), pseudocognates (words with shared form but different meaning; e.g., *grasa* means "fat," not "grass"; also known as false friends or interlingual homographs), non-cognate translations (words with the same meaning but different form; e.g., *dulce* means "sweet"), and unrelated words in the two languages as previews to test orthographic and semantic preview benefits in bilingual readers. Both cognate and pseudocognate previews conferred significant benefits of the same size upon target fixation relative to unrelated words, showing robust effects of orthographic and phonological similarity between previews and targets (see also Jouravlev & Jared, 2016, and Cong & Chen, 2022, for similar evidence of orthographic and phonological preview benefits in Russian-English and Chinese-English bilinguals, respectively). Critically, however, non-cognate translation equivalents provided no semantic preview benefit, suggesting that a semantic relationship between previews and targets in the absence of orthographic and phonological similarity did not facilitate reading (see also Snell et al., 2018a for similar evidence with French-English bilinguals; cf. Wang et al., 2016 with Korean-Chinese bilinguals). In other words, shared orthography and/or phonology produced preview benefits but shared semantics across previews and targets had no observable effect. The authors concluded that readers either did not process semantic information from the parafovea in the first place, or that any semantic information extracted from the preview was discarded upon fixation of an orthographically unrelated target word.

It is important to note that evidence for semantic preview benefits in monolingual readers has only been established in the decades after the original Altarriba et al. (2001) study, raising the question as to why semantic preview benefits have not been observed across languages. Such results are particularly curious given the strong semantic relationship between translation equivalents (Altarriba, 1992) and the prominent view in the bilingual literature that both languages are always active (Kroll et al., 2012) and that translation equivalents are automatically activated even when using only one language (Thierry & Wu, 2007; Wu & Thierry, 2010). According to these accounts, preview-target translation equivalents should yield particularly robust semantic preview benefits.

Several possible interpretations of the absence of cross-language semantic preview benefits exist. First, slower lexical access in bilinguals compared to monolinguals suggests that monolinguals and bilinguals may process parafoveal information differently during reading (Ivanova & Costa, 2008; Gollan et al., 2008; Martin et al., 2012; Martin et al., 2013). Bilingual readers may not extract information from the written text as efficiently as their monolingual counterparts, particularly in a weaker non-native language in which lexical access proceeds more slowly. According to a *bilingual difference account*, therefore, bilinguals may not process semantic information from the parafovea quickly enough to produce semantic preview benefits, particularly in their weaker language. Thus, bilinguals may not exhibit semantic preview benefits from *either* within- or cross-language previews. Indeed, one study found no evidence of within-language semantic preview benefits for native Korean speakers reading in their weaker second language (L2) Chinese, with both previews and targets appearing in Chinese (Wang, et al., 2014), whereas another study did find within-language semantic preview benefits in highly proficient Tibetan-Chinese bilinguals reading in their L2 Chinese (Xiao et al., 2021). Several studies have also demonstrated smaller preview benefits in less skilled compared to more skilled monolingual readers (Chace et al., 2005; Veldre & Andrews, 2015), demonstrating that language proficiency may be a key factor influencing whether or not readers exhibit semantic preview benefits. It is therefore important to test within-language semantic preview benefits in both native and non-native languages of bilingual readers to assess this hypothesis.

Alternatively, an *integration account* of the absence of cross-language semantic preview benefits would suggest that even if bilingual readers can process semantic information from the parafovea, semantic codes extracted from a preview in one language may be difficult to integrate with a target word in the other language (Schotter et al., 2012). Any semantic information extracted from a cross-language preview may be discarded upon fixation of the target word in the other language.[1] This account would be supported by

[1]Alternatively, following the hypothesis that semantic preview benefits emerge due to the relationship between the sentence context and the preview (e.g., Schotter & Jia, 2016; Veldre & Andrews, 2016; Veldre et al., 2020), it may be that semantic information

language-nonselective models of bilingual word recognition such as BIA+ (Dijkstra & van Heuven, 2002; Dijkstra et al., 2019). According to this model, lexical access is driven by bottom-up sources of information from the text without regard to the language membership of a word. This model would therefore predict that bilinguals process semantic information from the parafovea regardless of the language of the preview. However, a post-lexical task/decision system may then detect a language switch later during processing, and post-lexical switch costs would counteract any potential semantic preview benefit upon fixation of the target word. Thus, the BIA+ model would predict that bilingual readers extract semantic information from a code-switched preview but only exhibit a semantic preview benefit when the sentence context, preview, and target belong to the same language such that the reader does not experience any switch costs that could counteract the semantic preview benefit. It is therefore important to test both initial extraction of semantic information from the parafovea *and* semantic preview benefits upon fixation of the target word. Respectively, these would map onto measures of skip rate, which reflects the earliest observable stages of word recognition in the eye movement record (Reichle et al., 2009), and fixation times such as gaze duration and total reading time that index the effect of the preview on target word processing.

A third possibility is that the alternate language is less accessible when bilinguals read in a single language. Work supporting this view suggests that switch costs may arise from dynamics within the lexico-semantic processing system itself rather than post-lexical decision processes (Declerck & Philipp, 2015; Dijkstra & van Heuven, 1998; Hoversten, Brothers, Swaab, & Traxler, 2015, 2017; Hoversten & Traxler, 2020). In one study, code-switched words, non-switched words, and pseudowords were presented as parafoveal previews and

---

extracted from the preview may be difficult to integrate with the prior sentence context due to the language mismatch. Note that predictions of the *integration account* would remain the same regardless of the underlying source of semantic preview benefits.

masked with an unrelated target word in the same language as the rest of the sentence upon fixation, creating an essentially monolingual language context (Hoversten & Traxler, 2020, Experiment 2). Bilingual readers skipped code-switched words less often than non-switched words but no more often than pseudowords, indicating that a parafoveal preview in the other language was treated like a pseudoword from the outset. This result suggests that readers regarded parafoveal words in the alternate language as inaccessible early during lexical access. It is thus possible that bilingual readers suppress the alternate language such that they do not process semantic information from a code-switched parafoveal preview word in the first place.[2] By this *partial selectivity account* of the absence of cross-language semantic preview benefits, bilinguals *can* extract semantic codes from the parafovea when the preview belongs to the same language as the rest of the sentence, but *not* when the preview is code-switched to the alternate language. Thus, semantic preview benefits would emerge within but not across languages according to this account.

Current Study

It is important to test these differing accounts of bilingual semantic preview benefits to gain a more complete understanding of the cognitive and oculomotor processes involved in bilingual reading. **To distinguish among these accounts and to systematically assess** bilinguals' use of parafoveal information during reading, **the current study investigated** both skip rates and semantic preview benefits on fixation times within and across languages in a single group of participants. It is also important to test these effects in both switch directions, since the strength of the preview language may influence the emergence of semantic preview effects.

[2] It is also possible that passive decay of activation of the alternate language while reading in the target language would render the alternate language less accessible and produce the same effect. Research has not yet clearly disentangled the contribution of active suppression versus passive decay to accessibility differences between languages, which remains an important open question for future studies to disentangle.

We first tested the *bilingual difference account* by assessing within-language semantic preview benefits in bilingual readers in both native and weaker non-native language contexts. For example, the target word "toad" would be preceded either by the synonym preview "frog" (as shown in Figure 1) or the unrelated preview word "hill." This allowed us to assess whether bilingual reading differs qualitatively from monolingual reading in terms of parafoveal processing (i.e., if bilinguals simply do not exhibit the within-language semantic preview benefits that monolinguals have been shown to exhibit, especially in the weaker language), or if it involves similar coordinative processes among the language, attention, and eye movement control systems (i.e., that bilinguals do exhibit these within-language semantic preview benefits).

Next, we assessed cross-language semantic preview benefits in the same group of participants. For example, the target word "toad" would be preceded either by the translation preview "sapo" or the unrelated preview word "loma," which means "hill" in Spanish. This allowed us to follow up on prior studies to provide additional evidence concerning the absence of cross-language semantic preview benefits in alphabetic languages previously found. We extended these studies by examining both switch directions to further test whether slower access in a weaker non-native language may contribute to the absence of cross-language preview benefits previously demonstrated.

Finally, we distinguished between the *integration* and *partial selectivity* accounts by measuring skip rates of non-switched and code-switched previews. While both accounts predict an absence of cross-language semantic preview benefits, the key difference between the two accounts lies in the accessibility of code-switched previews during initial stages of word recognition. Skip rates reflect these early stages of word recognition since skipping decisions are made based on the spatially degraded parafoveal preview of the upcoming word before it is directly fixated in foveal vision (Rayner, 1998). If code-switched previews are skipped equally often as non-switched previews, this would suggest that early stages of access proceeded similarly regardless of the language membership of the parafoveal preview. Such a result would support the *integration account* that semantic information *is* extracted from the preview but difficulty arises upon integration with the target word across languages due to post-lexical switch costs (i.e., switch costs cancel out any potential semantic preview benefit from pre-processing the target word's translation parafoveally).

If, on the other hand, code-switched previews are skipped less often than non-switched previews, this would suggest that these words were less accessible from the outset due to deactivation of the alternate language. Furthermore, if code-switched previews are skipped no more often than words with which participants are unfamiliar, similar to the pseudowords in Hoversten & Traxler (2020), this would suggest that words in the alternate language were not recognized in time to affect skipping decisions. In combination, a low skip rate of code switches and a lack of cross-language semantic preview benefits would suggest that semantic information was not extracted from parafoveal previews in the alternate language, as per the *partial selectivity account*.

By systematically testing each of these hypotheses in this manner, we will build a more complete picture of the eye movement behavior involved in bilingual reading, specifically with regard to how bilingual readers use parafoveal sources of semantic information within and across languages.

METHODS

Transparency and Openness

Below we report how we determined sample size, the demographic information about our participants, all manipulations and measures, data exclusion procedures, and analyses performed. Data were collected between 2018-2019 and analyzed using R, version 3.6.2 (R Core Team, 2019). Data and materials are publicly available at https://osf.io/nb78z/. The design and analysis of the study were not pre-registered but were approved by the institutional ethics committee prior to data collection.

Participants

We first conducted a power analysis using the tools introduced by Judd, Westfall, & Kenny (2017) to determine the sample size needed to yield a power of .8 to detect typical gaze duration preview benefits (see explanation of this measure below) of 15-20 ms reported in the literature reviewed above. This analysis determined that a sample of sixty participants would yield 80% power to detect a gaze duration effect of at least 13ms using the design described below. We therefore aimed to test a total of seventy participants to account for any exclusionary criteria, including insufficient English proficiency (defined as lack of familiarity with more than one-third of the critical words used in the experiment, as measured by the post-experiment questionnaire), detection of boundary changes as measured by the post-experiment questionnaire, or poorly calibrated data based on researcher inspection prior to analysis. Seventy Spanish-English speakers from the Basque Country, Spain provided written informed consent and were paid to take part in the study according to the institution's ethical guidelines. Fourteen participants were excluded from analysis due to the exclusion criteria described above.

The final sample thus included 56 participants (age = 25 years (*SD* = 5); 39 female). All participants were native Spanish speakers with intermediate to high proficiency in English as either the second or third language (L2 = 11 participants; L3 = 45 participants). A majority of participants had some proficiency in Basque (L2 = 33 participants; L3 = 4 participants), and a few had knowledge of Catalan, French, German, or Italian (6 participants).[3] As part of participation in the institution's participant database outside of the context of the experiment, participants completed a lexical decision task (LexTALE: Lemhöfer & Broersma, 2012), picture naming task (BEST; De Bruin, Carreiras, & Duñabeitia, 2017), verbal language assessment interview, and language history questionnaire to assess proficiency and usage of each language. Table 1 presents language background and proficiency characteristics of the participant sample. English was clearly the weaker language compared to Spanish in all participants, allowing us to test whether unbalanced proficiency produces asymmetries in semantic preview benefits between languages and switch directions.

Table 1. Language proficiency scores or self-ratings, and standard deviations.

Measure	**Spanish (L1)**	**Basque (L2/3)**	**English (L2/3)**
Age of Acquisition	0.1 (0.6)	3.1 (5.6)	6.2 (2.5)
Current Exposure (%)	60.2 (16.8)	24.9 (17.5)	16.7 (9.3)
Overall Level (1-10)	9.4 (0.8)	7.5 (2.1)	7.3 (1.2)
Reading (1-10)	9.6 (0.7)	8.3 (1.9)	8.0 (1.2)
Writing (1-10)	9.3 (1.0)	7.6 (2.2)	7.2 (1.5)
Speaking (1-10)	9.5 (0.7)	7.5 (2.5)	6.9 (1.5)
Listening (1-10)	9.6 (0.6)	8.5 (1.8)	7.9 (1.3)
Interview (1-5)	5.0 (0.0)	3.7 (1.4)	3.8 (0.7)
LexTALE (%)	94.2 (5.7)	80.5 (15.5)	69.8 (8.9)
BEST Naming (%)	99.8 (0.5)	69.0 (33.3)	83.6 (7.9)

Materials

We selected 125 pairs of non-cognate Spanish-English translation equivalents of the same length (*M* = 5.7 letters; range: 4 - 9 letters). Each of these words served as a target word

[3] We have no reason to believe that multilingualism systematically influenced the results, since critical words had no systematic relationship to any other language known by any participants. Additionally, the sample was homogenous in terms of the relative strength of the two languages.

in the experiment. Two different low constraint sentences were constructed and translated across languages for each of the translation equivalent pairs. To verify that the stimuli were not predictable in context, a separate group of nineteen English monolinguals at the University of California, Santa Cruz (population selected for feasibility) completed the English sentence fragments up to but not including the critical word. Very few participants guessed any of the critical words or their English translation ($< 2\%$), and no participant guessed any word in the unrelated condition. A one-way ANOVA revealed significant differences in cloze probability across conditions ($F(3,992) = 4.08$, $p < .01$), which were entirely driven by the unrelated condition (Post Hoc Tukey HSD ps $< .05$; comparisons across the other three conditions: ps $> .34$).

In the main eye-tracking experiment, each participant saw one of each sentence pair in one language and the other sentence in the other language in a separate session, such that participants saw both translation equivalents as target words but in different sentence contexts and in separate sessions. There was no semantic overlap in the sentences across sessions other than the critical words themselves. We then used the gaze-contingent boundary paradigm to manipulate the parafoveal preview of target words. With this technique, a critical word embedded in a sentence (the preview) was replaced with a different word (the target) when the reader's eyes crossed to the right of an invisible boundary prior to the critical word. The target word then remained on the screen for the remainder of the trial.

We created five preview conditions in each language: a valid control condition in which the preview and target words were identical (*identical*), a within-language synonym preview (*synonym*), a within-language semantically unrelated preview (*unrelated non-switch*), a cross-language translation preview (*translation*), and a cross-language semantically-unrelated preview (*unrelated code switch*). The identical condition should lead to a preview benefit relative to the other 4 conditions, since those conditions all differ from the target word (i.e., they impose a preview cost; Hutzler et al., 2019). Critically, however, all four of the non-identical conditions in this design should impose the same degree of preview cost due to the fact that they each masked the identity of the target word to the same extent. In this way, we then examined the amount of preview benefit gained from the semantic relationship between previews and targets in the synonym and translation conditions relative to the unrelated conditions. With this design, we obtained 25 stimuli per condition within each language (Spanish/English). Example sentences and previews for a translation pair appear in Table 2, with the target word underlined.

Table 2. Example sentences and preview conditions for a pair of translation equivalents. The target word is underlined for display purposes. NS = non-switch, CS = code switch.

English	**Identical NS**	**Synonym NS**	**Unrelated NS**	**Translation CS**	**Unrelated CS**
They were asking the student questions about his toad because they were interested. The strong odor was coming from his toad and the entire class could smell it.	toad	frog	hill	sapo	loma
Spanish					
Le estaban haciendo preguntas al estudiante sobre su sapo porque estaban interesados. El olor fuerte venía de su sapo y toda la clase podía olerlo.	sapo	rana	loma	toad	hill

We selected within-language synonyms for the translation pairs in one of two ways. Of the 125 translation pairs, 76 also had a within-language synonym of the same length and minimal orthographic overlap within the list, creating 38 quadruplets of synonyms and translations. For example, the translation pairs *toad/sapo* and *frog/rana* are synonyms of one another and form a quadruplet. In the four semantically unique sentences written for this quadruplet, each of these words served as a target word in one sentence. Each word also served as a synonym or translation preview for one of the other three sentences (Table 3 top).

Table 3. Example preview conditions for four pairs of translation equivalents, two of which make up a fully counterbalanced quadruplet (76 of these were included in the experiment, making 38 quadruplets) and two of which do not make a quadruplet (49 of these were included in the experiment). NS = non-switch, CS = code switch

	Base language	**Target NS**	**Identical NS**	**Synonym NS**	**Unrelated NS**	**Translation CS**	**Unrelated CS**
Pai	English	toad	toad	frog	hill	sapo	loma
	Spanish	sapo	sapo	rana	loma	toad	hill
Pai	English	frog	frog	toad	moat	rana	foso
	Spanish	rana	rana	sapo	foso	frog	moat
Pai	English	fault	fault	blame	crowd	culpa	junta
	Spanish	culpa	culpa	falla	junta	fault	crowd
Pair	English	hill	hill	hump	frog	loma	rana
	Spanish	loma	loma	cima	rana	hill	frog

The other 49 pairs of translation equivalents did not form synonym quadruplets. For these words, separate synonym previews were chosen for each language. For example, for the translation pair *fault/culpa*, *blame* was chosen as a within-language synonym of *fault*, and *falla* was chosen as a within-language synonym of *culpa* (note that *blame* and *falla* are not translation equivalents of one another and hence never appeared as target words). The translation equivalent pair *fault/culpa* each served as a target word in one of the two semantically unique sentences written for this pair. They also appeared as cross-language translation previews in the other sentence or were replaced by the within-language synonym to form the synonym condition (Table 3 bottom).

To create the unrelated conditions, the same preview words from the synonym condition were shuffled and re-paired with a semantically and orthographically unrelated target word. In other words, the same words were used as related and unrelated previews across subjects but were matched with different pairs of target words to manipulate relatedness in the sentence context. Unrelated previews were pseudorandomly shuffled to maximize their semantic implausibility while still being syntactically licensed by the sentence context in order to both maximize detection of any semantic preview benefits and to test the time course of the effect of semantic implausibility (Veldre et al., 2020). A separate norming experiment with twenty English monolingual participants from the University of California, Santa Cruz (population selected for feasibility) confirmed this plausibility manipulation. Stimuli were presented in English up to and including the critical word (or its translation in English). Participants related the valid, synonym, and translation conditions as quite plausible (M = 4.0, SD = .50) and the unrelated (M = 2.7, SD = .73) condition as rather implausible on a scale of one to five. A one-way ANOVA demonstrated significant differences in plausibility across conditions ($F(3,76) = 27.6$, $p < .001$), which were entirely driven by the unrelated condition (Post Hoc Tukey HSD ps < .001; comparisons across the other three conditions: ps > .67). These results confirm that the unrelated condition was indeed less plausible than the valid and related conditions.

Since approximately two-thirds of the stimuli formed quadruplets that could be fully counterbalanced across conditions, most words appeared in each preview condition across subjects (see Table 4 for stimuli characteristics). This allowed us to control for any potential effects of preview and target frequency (see Schotter & Leinenger, 2016), as average log frequency per million (SUBTLEX databases; Cuetos, Glez-Nosti, Barbon, & Brysbaert, 2012; New, Brysbaert, Veronis, & Pallier, 2007) was similar across preview and target conditions (previews: $M = 1.08$; targets: $M = 1.18$; $p > .05$). In addition, orthographic overlap between all non-identical previews and targets was low (mean number of overlapping letters < .15). A 2x2 ANOVA on the overlap between invalid previews and targets with the factors semantic Relatedness (Related vs. Unrelated) and Switching (Within vs. Across) revealed a main effect of Switching ($F(1,996) = 41.6$, $p <. 001$), showing that within-language previews shared more orthographic overlap with targets than cross-language previews did. Critically, however, we found no main effect of Relatedness or interaction between the two factors (*p*s > .23), suggesting that any preview benefit for related previews, either within or across languages, could not be accounted for based on the minimal orthographic overlap present between previews and targets. In other words, this analysis confirms that we indeed tested semantic preview benefits as opposed to orthographic preview benefits.

Table 4. Means and standard deviations of stimuli characteristics and English norming data by condition. NS = non-switch, CS = code switch

	Valid	**Synonym NS**	**Unrelated NS**	**Translation CS**	**Unrelated CS**
Frequency (log/million)	1.18 (.70)	1.04 (.68)	1.04 (.68)	1.18 (70)	1.04 (.68)
Length	5.7 (1.2)	5.7 (1.2)	5.7 (1.2)	5.7 (1.2)	5.7 (1.2)
Proportion of letter overlap with target	1.00 (.00)	.14 (.12)	.13 (.09)	.09 (.09)	.08 (.08)
Plausibility (1-5)	4.1 (.5)	3.9 (.5)	2.7 (.7)	4.1 (.5)	2.7 (.7)
Cloze probability	.02 (.09)	.01 (.06)	.00 (.00)	.02 (.08)	.00 (.00)

Apparatus and Procedure

Two separate sessions were conducted between three and seven days apart. Each session was conducted in a different language, with the order counterbalanced across participants. All sentences were presented in the same language in a single session. Although previews could appear in the same language (non-switches) or in the alternate language (code-switches), all target words in a session appeared in the same language as the sentence contexts. Additionally, separate experimenters conducted each session entirely in a single language so as to minimize any influence of past linguistic experience with a particular interlocutor. This design created an essentially monolingual language mode in which participants read in a single language context within a session without explicit awareness of the presence or relevance of the alternate language (Grosjean, 2001).

During each session, an Eye Link 1000 (SR Research, Ontario, Canada) monitored and recorded eye movements while participants read sentences one at a time on a CRT ViewSonic monitor with a refresh rate of 90 Hz. Sentences were presented using Experiment Builder software in black size 14 Consolas font on a white background. Participants were seated with their chin on a chin rest approximately 80cm from the monitor. Three characters subtended 1 degree of visual angle. Calibration and validation were performed with a 9-point grid, and the tracker was recalibrated as needed throughout the experiment if drift check error exceeded 0.3 degrees of visual angle, or the width of one character. A comprehension question appeared after 20% of the sentences to ensure attentiveness during reading. Participants completed a post-experiment questionnaire at the end of the second reading session that probed their awareness of display changes or code switches during the experiment. Participants also indicated whether or not they were familiar with each of the English critical words in order to both verify sufficient English proficiency and to separate trials based on familiarity as described below.

Data Analysis

Fixation durations less than 80 ms were either merged with any adjacent fixation within one character or else discarded. Fixation durations greater than 2.5 standard deviations above the mean for a condition for an individual were trimmed (2.1% of the data). To ensure the critical word appeared in the parafovea, we discarded all trials in which the fixation immediately after the display change landed prior to the critical word (7.7% of the data) or the pre-target region was not fixated prior to the display change (5.0% of the data). Because the word prior to the critical word tended to be rather short (*M* = 2.9 characters, *SD* = 1.3), we used the two-word region prior to the critical word (*M* = 8.4 characters, *SD* = 2.8) to perform this calculation. Note that the way in which stimuli were created and counterbalanced means that the length of the pre-target word(s) was exactly the same across all conditions, so we did not expect it to systematically influence results.

Finally, for each participant's data, all trials that contained a critical word that the participant reported as unfamiliar were discarded from the main analyses (7.8% of the data). These trials were then aggregated into a separate condition for a planned comparison to the code switch conditions on skip rate to determine the relative accessibility of words belonging to the alternate language. While these words were all English words, they appeared in both English contexts (as an intended NS) and Spanish contexts (as an intended CS), so the effect of familiarity was analyzed by base language context, similar to the main dataset.[4]

Standard measures of eye movement data were analyzed, including a) *skip rate*- the proportion of trials in which the critical word did not receive a first pass fixation, b) *gaze duration*- the amount of time the eyes spent fixated on first pass of the critical word, including all refixations before exiting the region, c) *regression path duration* (also known as *go-past duration*)- the amount of time beginning with the first fixation on the critical word until the eyes crossed the right-hand boundary of the region, and d) *total time*- the total amount of time the eyes spent fixated on the critical word throughout the duration of the trial.

[4] The relative paucity of data in this condition does indeed mean that we did not have as high power as for the main analyses. However, this analysis can still provide valuable information, particularly as a conceptual replication attempt of the high-powered Hoversten & Traxler (2020) demonstration of difference between code-switched and pseudoword preview conditions. If the patterns remain the same in the current study, particularly with support of the Bayes Factor analysis, we can increase our confidence in the finding.

We will refer to gaze duration, regression path duration, and total time collectively as "fixation duration measures," as compared to skip rate, which is a measure of fixation probability. Measures such as skip rate and gaze duration are considered earlier measures of lexical access, whereas regression path duration and total time reflect later stages of lexical integration and discourse processing (Rayner, 1998). These four measures were chosen to assess the entire time course of lexico-semantic influence on eye movement behavior while minimizing the number of measures analyzed. To correct for multiple comparisons (von der Marlsburg & Angele, 2017), we divided the significance criterion at the .05 level by the four measures such that an effect was considered significant if the *p* value fell below .0125.

Linear mixed-effects models were constructed with the lme4 package in 'R' statistical software (Bates et al., 2015; R Development Core Team, 2019). We used the maximal random effect structure justified by the design with crossed random slopes and intercepts for participants and items (Barr et al., 2013).[5] Binomial general linear mixed-effects models with a logit link function were fitted to skip rate data. Fixation time data was log-transformed to correct for skew (Vasishth & Nicenboim, 2016). Wald *Z* tests and Sattherwaite approximations from the lmTest package were used to obtain *p*-values for skip rate and fixation time data, respectively (Hothorn et al., 2019). We followed up effects of interest by calculating the Bayes Factor using the generalTestBF function in the BayesFactor package with the default JZS priors and the default comparison against the null model (Morey et al., 2015). For theoretically important null effects, we fit these models using a range of prior

[5] For models that failed to converge or resulted in a singular fit, we reduced the complexity of the model until it converged following recommendations from Barr et al. (2013) to progressively remove random correlation parameters at the boundaries (+/-1 or 0) and/or random effects that explained the smallest amount of variance one by one until convergence was achieved without a singularity. Thus, we report the maximal effects supported by the data that could be modeled. These adjustments have been shown to preserve the intended Type I error rate (Barr et al., 2013).

scaling factors, from the default narrow priors ($\sqrt{2}/4$) to medium ($\sqrt{2}/2$) to ultrawide ($\sqrt{2}$).

For the sake of comparison to prior and future research, we first tested the effect of the boundary change itself with a model of each dependent measure with the factors preview Validity (Identical vs. Synonym) and Base Language (Spanish vs. English sentences). Skip rates should not differ based on preview validity since both identical and synonym conditions were plausible continuations of the low-constraint sentences. In contrast, we expected the three fixation duration measures to differ across these conditions due to the disruption of reading caused by a display change to a different word upon fixation of the target, also known as a *preview cost* (i.e., a preview cost; Brothers & Traxler, 2016; Hutzler et al., 2019). Note that preview validity does not address the research aims of the study but is rather a nuisance variable that we report for completeness.

We then split the data into within- and cross-language subsets and fit a series of a priori specified 2x2 models to test each of our research questions in turn (please see data analysis script on OSF page for the full R syntax of each model):

1) We tested the *bilingual difference account* by fitting a 2x2 model of preview Relatedness (Synonym vs. Unrelated Non-switch) for each Base Language (Spanish vs. English sentence) on each measure in the within-language dataset. This model tests for the presence of within-language semantic preview benefits that have been found in monolingual readers in prior studies as well as a potentially smaller within-language semantic preview benefit in the weaker language.
2) We tested the replicability of prior studies that have not found cross-language semantic preview benefits in alphabetic languages by fitting a 2x2 model of preview Relatedness (Translation vs. Unrelated Code switch) for each Base Language (Spanish vs. English

sentence) on each measure in the cross-language dataset. This model also allowed us to extend prior studies by testing both switch directions (i.e., whether preview language affects the presence of cross-language semantic preview benefits).

3) We distinguished between the *integration* and *partial selectivity accounts'* predictions concerning the initial accessibility of the alternate language by comparing skipping behaviors between within- and cross-language previews. We fit a 2x2 model of language Switching (Non-switch vs. Code switch) for each Base Language (Spanish vs. English sentence) on skip rate. We then used the post-experiment questionnaire data to compare the effect of Familiarity (Code switch vs. Unfamiliar) in each Base Language (Spanish vs. English sentence) on skip rate, similar to the pseudoword effects examined by Hoversten & Traxler (2020).

Finally, for the sake of completeness, we also report an omnibus model comparing all invalid conditions in the supplementary materials.

RESULTS

Accuracy on comprehension questions was high (92%). Display changes were completed an average of 8.9 ms after the eyes crossed the boundary. Debriefing data demonstrated that most participants were at least somewhat aware of display changes throughout the experiment. Of the 54 participants, 39 reported noticing changes less than 5% of the time, 7 reported noticing changes 5-10% of the time, and 8 reported noticing 30-40% of the time. Additionally, 34 participants reported seeing a code-switched word less than 5% of the time, and 5 participants reported seeing code-switches between 10-50% of the time. These rates are likely attributable to the comparatively low refresh rate of 90 Hz on the testing monitor. However, an inspection of the data suggests that the 8 participants who noticed a substantial number of display changes and/or code-switches displayed the same general patterns as reported below for the full dataset.

Table 5 presents condition means and standard deviations for the critical word on the four different measures broken down by Base Language. Figures 2 and 3 present graphical representations of the gaze duration data and skip rate data, respectively, collapsed across Base Language since this factor did not interact with any of our critical factors of interest. Table 6 presents a summary of the results, including *z/t* values, *p* values, and Bayes Factors in favor of the alternative hypothesis.

Preview validity

In initial models including the factors preview Validity (Identical vs. Synonym) and Base Language (Spanish vs. English), the main effect of Base Language was significant on all measures, with more skips and faster reading in Spanish (skip rate: z = -3.31, p = < .001, d = .39; gaze duration: t = 8.77, p < .001, d = .21; regression path: t = 8.98, p < .001, d = .21; total time: t = 8.98, p < .001, d = .27). As expected, Validity did not have a significant effect on skip rate (p = .33) but did significantly affect all fixation duration measures (gaze duration: t = -4.62, p < .001, d = .09; regression path: t = -6.77, p < .001, d = .12; total time: t = -6.85, p < .001 d = .14), confirming that the plausible synonym preview fit equally well in context compared to the identical preview but that the display change to a different target word disrupted reading compared to a valid preview. Base Language and Validity did not interact significantly on any measure (ps > .66, ds < .02).

Table 5. Means and standard deviations for each measure on the critical word for each preview condition split by base language. NS = non-switch, CS = code switch.

	Identical NS	Synonym NS	Unrelated NS	Translation CS	Unrelated CS	Unfamiliar

Skip rate (%)	6.7 (24.9)	5.9 (23.5)	7.0 (25.5)	2.7 (16.2)	3.8 (19.1)	2.5 (15.6)
Gaze duration	294 (152)	325 (181)	341 (177)	352 (194)	354 (187)	
Reg. path	361 (235)	433 (317)	433 (289)	486 (318)	501 (356)	
Total time	397 (260)	460 (277)	479 (283)	491 (289)	500 (308)	
Skip rate (%)	2.6 (15.9)	3.4 (18.2)	2.5 (15.6)	2.4 (15.2)	1.6 (12.6)	1.7 (12.9)
Gaze duration	359 (196)	392 (210)	407 (211)	409 (219)	417 (226)	
Reg. path	459 (358)	516 (355)	565 (422)	564 (403)	594 (431)	
Total time	523 (351)	610 (422)	621 (453)	619 (393)	652 (444)	

Within-language comparisons: Testing the *bilingual difference* account

In the model of skip rate including the factors Relatedness (Synonym + Identical[6] vs. Unrelated Non-switch) and Base Language (Spanish vs. English), a main effect of Base Language (z = -5.08, p < .001, d = .40) demonstrated that Spanish critical words were skipped more often than English critical words. However, no effects of Relatedness or interaction with Base Language emerged (ps > .10, ds < .09). A Bayes Factor analysis on the main effect of Relatedness demonstrated that the null hypothesis is strongly favored given the data, which remained robust across a range of priors (narrow BF_{01} =34.9, medium BF_{01} = 69.8, wide BF_{01} = 139.6). In the fixation time models including the factors Relatedness (Synonym vs. Unrelated Non-switch) and Base Language (Spanish vs. English), the main effect of Base Language was significant in all models (gaze duration: t = 8.26, p < .001, d = .20; regression path: t = 8.86, p < .001, d = .22; total time: t = 7.93, p < .001, d = .26) with longer fixation times on English compared to Spanish words. A main effect of Relatedness emerged in gaze duration (t = -2.93, p = .003, d = .05) and regression path time (t = -2.62, p = .010, d = .04) and was significant in total time before correction for multiple comparisons (t = -2.10, p

[6] We combined the identical and synonym conditions for within-language skip rate to increase power since these conditions did not differ on this measure, as expected.

= .035, d = .03). The two factors did not interact in any model (ps > .20, ds < .02), showing that the effect of Relatedness was consistent in both Base Language contexts.

Cross-language comparisons: Testing replicability

In the model of skip rate including the factors Relatedness (Translation vs. Unrelated code switch) and Base Language (Spanish vs. English), a main effect of Base Language (z = -2.19, p = .028, d = .18) emerged but did not survive correction for multiple comparisons, suggesting that critical words may have been skipped marginally more often when reading Spanish than English. No main effect of Relatedness or interaction with Base Language emerged (ps > .10, ds < .12). A Bayes Factor analysis on the main effect of Relatedness demonstrated that the null hypothesis is strongly favored given the data, which remained robust across a range of priors (narrow BF_{01} =28.6, medium BF_{01} = 57.1, wide BF_{01} = 114.11). In the fixation time models including the factors Relatedness (Translation vs. Unrelated code switch) and Base Language (Spanish vs. English), the main effect of Base Language was significant on all measures (gaze duration: t = 7.08, p < .001, d = .16; regression path: t = 6.02, p < .001, d = .16; total time: t = 7.79, p < .001, d = .23), with longer fixation times for English compared to Spanish target words. However, the critical effect of Relatedness did not reach significance on any measure, either before or after correction for multiple comparisons (gaze duration: t = -.98, p = .33, d = .02; regression path: t = -1.97, p = .051, d = .03; total time: t = -1.45, p = .15, d = .02). Again, Bayes Factor analyses suggested that a null effect of Relatedness is more likely given the data (gaze duration: narrow BF_{01} =13.9, medium BF_{01} = 27.7, wide BF_{01} = 55.3; regression path: narrow BF_{01} = 5.5, medium BF_{01} = 10.9, wide BF_{01} = 21.8; total time: narrow BF_{01} = 9.3, medium BF_{01} = 18.4, wide BF_{01} = 36.8).

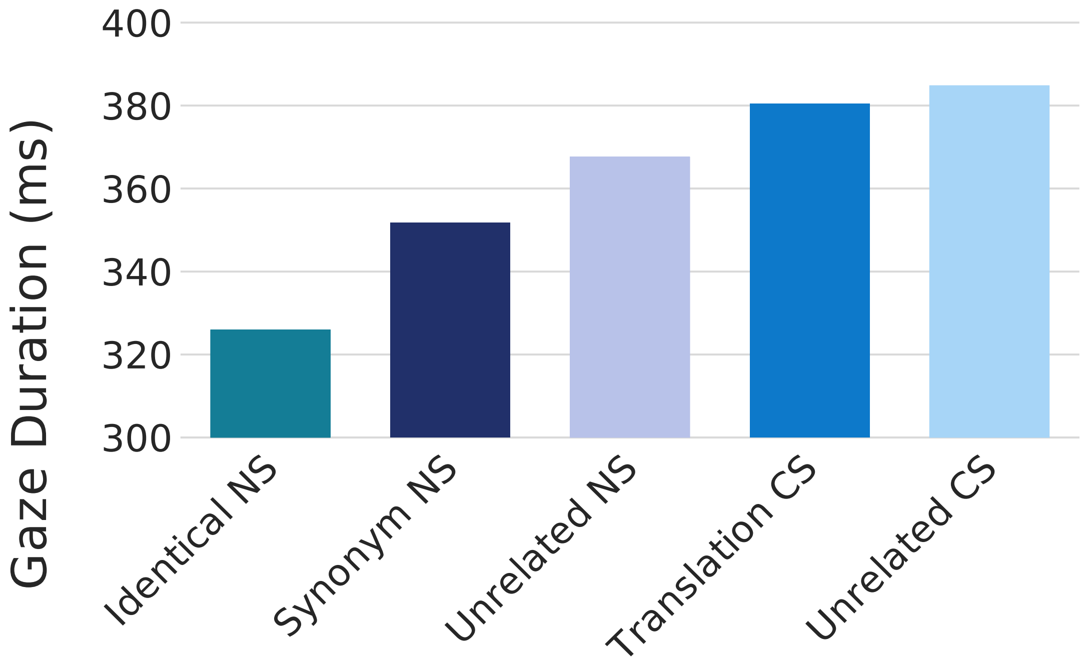
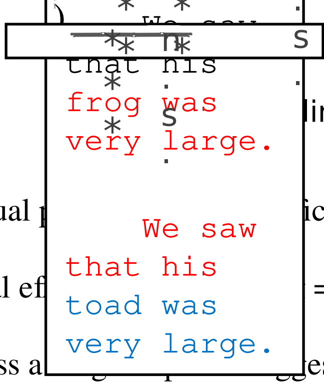


Figure 2. Gaze durations on the target word for each preview condition collapsed across Base Language context, which did not interact with factors of interest in any model. Error bars represent within-subject standard errors of the mean (Morey, 2008)**.** NS = non-switch, CS = code switch. n.s. = non-significant, *** $p < .001$.

Within- vs. cross-language skip rates: Testing *integration* vs. *partial selectivity* accounts

In the model of skip rate comparing effects of non-switched versus code-switched previews in each language, main effects of both Base Language ($z = -3.70$, $p < .001$, $d = .35$) and language Switching ($z = -4.67$, $p < .001$, $d = .23$) emerged, with no interaction between them ($p = .18$, $d = .08$).[7] Participants skipped the critical word more often when reading in Spanish than in English and when the critical word appeared in the same language as the rest of the sentence compared to when it was switched to the alternate language. In the model comparing skip rate of code-switched translation previews and the English words marked as

[7] The factor Relatedness was not included in this skip rate model, as it was not significant in either of the core models described above.



unfamiliar by each individual p icant effects emerged (all $ps > .15$, $ds < .17$), including the critical ef = .264, $p = .79$, Figure 3, $d = .02$) A Bayes Factor analysis across a ests that this null effect of Familiarity is strongly favored by the data (narrow BF_{01} =26.9, medium BF_{01} = 53.8, wide BF_{01} = 107.5).

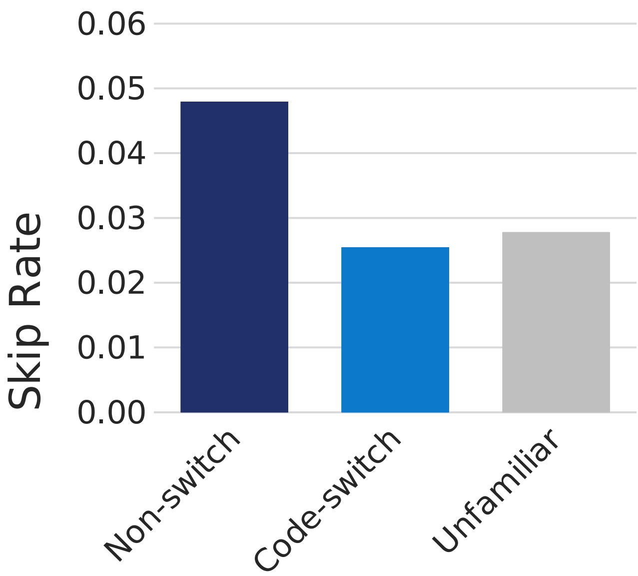


Figure 3. Skip rates for Non-switch (Identical + Synonym + Unrelated Non-switch), Code switch (Translation + Unrelated Code switch), and Unfamiliar previews. Error bars represent within-subject standard errors of the mean. n.s. = non-significant, *** $p < .001$.

Table 6. Summary of results for each model of interest. Significant results in favor of the alternative hypothesis appear in bold black ink, and non-significant results in favor of the null hypothesis appear in gray ink. SK = skip rate, GD = gaze duration, RP = regression path duration, TT = total time; ***p < .001, **p < .01, *p < .05, ns = non-significant.

	Validity	Base Language	Interaction
SK	$z = .97, p = .33$, ns $BF_{10} = .049$	**$z = -3.31, p < .001$***** **$BF_{10} = 870$**	$z = .10, p = .92$, ns $BF_{10} = .007$
GD	**$t = -4.62, p < .001$***** **$BF_{10} = 1737$**	**$t = 8.77, p < .001$***** **$BF_{10} = 9.36 \times 10^{26}$**	$t = -.38, p = .70$, ns $BF_{10} = .066$
RP	**$t = -6.77, p < .001$***** **$BF_{10} = 5.64 \times 10^{8}$**	**$t = 8.98, p < .001$***** **$BF_{10} = 5.52 \times 10^{23}$**	$t = .12, p = .90$, ns $BF_{10} = .062$

TT	**t = -6.84, p < .001*** $BF_{10} = 6.10 \times 10^{11}$**	**t = 8.98, p < .001*** $BF_{10} = 2.64 \times 10^{39}$**	t = .44, p = .66, ns $BF_{10} = .058$
	Within-language SPB	Base Language	Interaction
SK	z = 1.02, p = .31, ns $BF_{10} = .014$	**z = -5.08, p = <.001*** $BF_{10} = 2.06 \times 10^{8}$**	z = 1.67, p = .10, ns $BF_{10} = .006$
GD	**t = -2.93, p < .01** $BF_{10} = 2.07$**	**t = 8.26, p < .001*** $BF_{10} = 5.77 \times 10^{23}$**	t = 87, p = .38, ns $BF_{10} = .085$
RP	**t = -2.62, p < .01** $BF_{10} = .727$**	**t = 8.86, p < .001*** $BF_{10} = 3.65 \times 10^{28}$**	t = -1.29., p = .19, ns $BF_{10} = .134$
TT	**t = -2.10, p = .035*** $BF_{10} = .178$	**t = 7.93, p < .001*** $BF_{10} = 6.66 \times 10^{36}$**	t = 1.16., p = .25, ns $BF_{10} = .032$
	Cross-language SPB	Base Language	Interaction
SK	z = .008, p = .99, ns $BF_{10} = .035$	**z = -2.19, p = .028* $BF_{10} = 1.02$**	z = 1.63, p = .10, ns $BF_{10} = .025$
GD	t = -.98, p = .32, ns $BF_{10} = .072$	**t = 7.08, p < .001*** $BF_{10} = 5.64 \times 10^{15}$**	t = -.13, p = .90, ns $BF_{10} = .007$
RP	t = -1.97, p = .051, ns $BF_{10} = .182$	**t = 6.02, p < .001*** $BF_{10} = 4.14 \times 10^{14}$**	t = -.12, p = .91, ns $BF_{10} = .021$
TT	t = -1.45, p = .15, ns $BF_{10} = .108$	**t = 7.79, p < .001*** $BF_{10} = 2.06 \times 10^{31}$**	t = -.52, p = .60, ns $BF_{10} = .011$
	Switching	Base Language	Interaction
SK	**z = -4.67, p = <.001*** $BF_{10} = 5.09 \times 10^{5}$**	**z = -3.70, p = <.001*** $BF_{10} = 3.73 \times 10^{11}$**	z = 1.34, p = .18, ns $BF_{10} = 5.19$
	Familiarity	Base Language	Interaction
SK	z = .27, p = .79, ns $BF_{10} = .037$	z = -1.42, p = .15, ns $BF_{10} = .920$	z = .58, p = .56, ns $BF_{10} = .004$

DISCUSSION

This study aimed to examine how bilingual readers access semantic information from the parafovea within and across native and non-native languages. Using the gaze-contingent boundary paradigm, we manipulated semantic relatedness between preview and target words. We also manipulated the language of the preview while keeping the target in the same language as the rest of the sentence, maintaining an essentially monolingual reading context while probing for accessibility of the alternate language. This design allowed us to systematically assess why previous studies have not found semantic preview benefits across languages despite reports of these benefits within a language in monolinguals. Three potential explanations include 1) a *bilingual difference account*, which posits that lexical access proceeds too slowly for bilinguals to extract semantic codes from a parafoveal preview, particularly in a weaker non-native language (Gollan et al., 2008; Ivanova & Costa, 2008; Martin et al., 2012; Martin et al., 2013), 2) an *integration account*, which posits that semantic codes extracted from a code-switched parafoveal preview are difficult to integrate with the preceding contex and/or with the subsequent target word in a different language (Dijkstra & van Heuven, 2002; Kroll et al., 2012; Schotter, 2013; Thierry & Wu, 2007), or 3) a *partial selectivity account*, which posits that reduced activation of the alternate language renders code-switched previews harder to access from the earliest stages of word recognition (Hoversten & Traxler, 2020). While the *bilingual difference account* predicted no semantic preview benefits either within or across languages, at least not in the weaker language, both the *integration* and *partial selectivity accounts* predicted within- but not cross-language semantic preview benefits. Moreover, the *integration account* predicted evidence of initial semantic processing of cross-language previews, whereas the *partial selectivity account* predicted none.

We found evidence of within-language semantic preview benefits in both languages, with no difference in the effect size or time course between languages, thus ruling out the *bilingual difference account*. In contrast, we did not find evidence of cross-language semantic preview benefits in either switch direction in the same participants. The fact that we did not find any evidence that semantic preview benefits depended on the base language of the sentence suggests that the relative balance between the two languages was not a critical factor driving the pattern of results. Furthermore, code-switched words were skipped less often than non-switched words and no more often than words that participants marked as unfamiliar, thereby ruling out the *integration account* and supporting the *partial selectivity account*.

Based on these results, we propose that bilinguals can indeed extract semantic information from a parafoveal preview even in their weaker language, but that access is constrained to previews belonging to the same language as the rest of the sentence context. We will first briefly discuss the source of the within-language preview benefits found and then discuss within- and cross-language effects in turn before addressing the implications for models of bilingual reading and oculomotor control.

Preview plausibility

Although some researchers have suggested that preview plausibility affects eye movement behavior on the early measure of skip rate (Veldre et al., 2020; Schotter & Jia, 2016), the results of the current study did not support this hypothesis, in line with the results of Abbott & Staub (2015). Unrelated previews were chosen to maximize implausibility in context and hence any potential relatedness effects upon fixation of the target. However, we found no evidence of a plausibility effect on skipping decisions within or across languages in either language. One possible explanation is that all previews in the current study were syntactically acceptable in context, whereas several prior studies that have reported plausibility effects on skipping have included syntactic violations in the unrelated condition (see Brothers & Traxler, 2016, for evidence of early effects of grammatical violations on skipping behavior, c.f. Veldre & Andrews, 2018). Another possibility is that bilingual readers do not process the contextual fit of the preview in time to affect skipping decisions even if they do access semantic information about the preview itself, as indicated by the within-language semantic preview benefits discussed below. Future research is required to disentangle these and other potential explanations.

While the present data indicate that semantic plausibility alone does not affect the earliest stages of word recognition enough to influence skipping decisions, a related hypothesis is that semantic preview benefits on fixation times arise due to the plausibility of the preview in context rather than semantic relatedness between the preview and target. As mentioned in the introduction, the forced fixations account could explain such a divergence across measures in that preview plausibility may arrive too late to influence skipping decisions but still early enough to affect fixation durations on the target word (Schotter & Leininger, 2016). With the current data, we cannot distinguish between this account, a trans-saccadic integration account, or a hybrid of the two (Schotter et al., 2019), but we can be confident that the fixation time effects found do indeed reflect the extraction of semantic codes from the parafoveal preview.

Within-language effects

One of the major goals of the study was to assess whether bilingual readers can extract semantic codes from the parafovea in native and non-native languages in a qualitatively similar way as their monolingual counterparts. Because many models posit that lexical access proceeds more slowly in bilinguals, particularly in a weaker language (Gollan et al., 2008; Ivanova & Costa, 2008; Martin et al., 2012; Martin et al., 2013), it is conceivable that bilinguals would not reach semantic stages of processing of a parafoveal word before it is fixated. Yet we found robust evidence of within-language semantic preview benefits in bilingual readers in both a dominant and a weaker language, with no difference in the size or time course of the effect between languages. This result demonstrates that bilinguals can and do obtain semantic information from the parafovea during natural reading, similar to results found in prior studies with monolingual participants (Andrews & Veldre, 2019). These data therefore refute a *bilingual difference account* that parafoveal processing is fundamentally different in bilinguals and monolinguals.

The data also demonstrate that participants extracted semantic information from the parafovea to the same extent in both dominant and weaker languages. Even though participants read slower and skipped less often in their non-native L2/L3 English compared to their native L1 Spanish, they still experienced significant semantic preview benefits in both languages on all fixation time measures. We account for this result with a simple explanation: The oculomotor control system compensates for slower access in the weaker language by slowing down the progression of saccades through the text accordingly. Increased fixation time on word *n* in the weaker language not only allows more time for the reader to identify word *n* but also allocates extra time to enable access to begin on word *n+1* before the next saccade is executed. The additional time available to begin identifying word *n+1* increases the likelihood of activating the corresponding semantic information, producing the semantic preview benefits found, even in the weaker language. We therefore conclude that parafoveal information contributes to reading fluency in a qualitatively similar way in native and non-native languages.

Cross-language effects

In contrast to the within-language results, we found no evidence of cross-language semantic preview benefits in either switch direction on any measure. Although we designed the stimuli to maximize the detection of any potential semantic effects — with low plausibility of the unrelated condition, high plausibility of the translation condition, and a strong semantic relationship between previews and targets (using translation equivalents) — we nevertheless found no reliable evidence for extraction of semantic codes from cross-language previews. This result replicates Altarriba et al. (2001) and Snell et al. (2018a), who also found no evidence of cross-language semantic preview benefits in bilingual readers of alphabetic languages.

We not only replicated but also extended these prior studies by measuring both switch directions and by determining the source of the absence of cross-language semantic preview benefits. By measuring effects across both languages and in both switch directions, we demonstrated that the presence or absence of semantic preview benefits did not depend on language dominance either within or across languages. By measuring within-language preview benefits in bilingual participants, we refuted the *bilingual difference account,* as discussed above. Finally, by measuring the pattern of skipping behavior between non-switches and code-switches, we dismissed the *integration account* and supported the *partial selectivity account*. Specifically, we found that code-switched previews were skipped significantly less often than non-switched words but no more often than words that participants marked as unfamiliar. This result confirms a similar pattern of effects found in a prior study using a pseudoword baseline with a different population of bilinguals (Hoversten & Traxler, 2020). Considering that unfamiliar words are, by definition, inaccessible in terms of activation of lexical representations stored in long-term memory, this result suggests that code-switched words were treated as inaccessible during the earliest stages of word recognition. In other words, semantic information from a parafoveal preview in the alternate language was never sufficiently activated to significantly influence eye movement behavior on any measure. We therefore conclude that the absence of cross-language semantic preview benefits resulted from reduced accessibility of words in the alternate language when reading in a monolingual language context.

Nonetheless, it is possible that other methods could potentially reveal small cross-language semantic preview benefits that we and prior studies did not detect by measuring eye movements. For example, Antunez et al. (2021) co-registered eye-tracking and electroencephalography to compute fixation-related potentials (FRPs) and found evidence of parafoveal processing of semantic information for cross-language translations. However, one key difference in the design of the study could explain the different results across methods: whereas Antunez et al. (2021) presented their stimuli as isolated word pairs, the current study presented full sentences with rich semantic content within a carefully constructed global language context that clearly indicated the relevant language. Indeed, Snell et al. (2018a) conducted two eye-tracking experiments and found that isolated word pairs produced evidence of cross-language semantic preview benefits while sentence reading did not in the same bilingual population. This pattern suggests that the deactivation of the alternate language may heavily depend on the richness of the surrounding semantic and language context for the current communicative setting. Future research will be needed to directly test this hypothesis and the influence of the method used to measure semantic preview benefits (i.e., eye-tracking alone vs. FRPs). The co-registration technique in particular may thus open new avenues of investigation of cross-language semantic preview benefits.

Even if we cannot rule out a small cross-language semantic effect that was not strong enough to influence eye movement behavior, the current pattern of results does clearly indicate that the alternate language was down-regulated relative to the target language during sentence reading. Based on the pattern of results across studies of cross-language preview benefits, we suggest that this deactivation occurred at the lexico-semantic level. As discussed in the introduction, prior studies have shown robust sublexical preview benefits despite the absence of semantic preview benefits (Altarriba et al., 2001; Cong & Chen, 2022; Jouravlev & Jared, 2016; Snell et al., 2018a). As discussed in the introduction, Altarriba et al. (2001) demonstrated that cognates and pseudowords produced similar preview benefits, presumably due to orthographic overlap despite pseudocognates' divergence at the semantic level. At the same time, non-cognate translations did not show any preview benefits since it did not contain form-level overlap between previews and targets. In concert with our corroborative results showing no cross-language semantic preview benefits for non-cognate translations, it appears that activation of orthographic and/or phonological codes proceeds regardless of language membership but that deactivation of the alternate language in a monolingual reading context impedes processing of parafoveal semantic information. Future research will be required to establish the source of this deactivation in terms of active suppression, passive decay of the alternate language, and/or excitation of the target language.

Implications for models of bilingual reading

We found evidence supporting two key conclusions concerning the cognitive and oculomotor processes underlying bilingual reading. First, the processes driving eye movement behavior in bilingual readers are akin to those in monolingual readers when the text remains entirely in a single language. Although bilinguals may read more slowly, at least in their weaker language, linguistic representations are activated and drive eye movements through the text in both dominant and weaker languages in a qualitatively similar way as in monolingual readers. It therefore seems reasonable to expect that models of oculomotor control – such as E-Z Reader, SWIFT, and OB1 Reader (Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle et al., 2009; Snell et al., 2018b) – that have been built to simulate monolingual reading data could be extended to bilingual reading data with minimal modifications when the text remains exclusively in one language.

Secondly, the processes driving eye movement behavior differ when the text includes previews in the alternate language (i.e., code switches). We found evidence supporting the conclusion that the alternate language was deactivated during reading such that participants did not access semantic information from cross-language parafoveal previews in either the native or non-native language. Such a result may seem surprising given the prevalent view that bilingual word recognition is language-nonselective in nature (Kroll et al., 2012). For example, the BIA+ model would predict that semantic representations should be equally accessible whether the preview matches or mismatches the language membership of the rest of the sentence (Dijkstra & Van Heuven, 2002). Some researchers have even extended this nonselective view to suggest that translation equivalents are automatically and unconsciously activated, even in a monolingual language context (Thierry & Wu, 2007; Wu & Thierry, 2010). This account would predict that semantic representations should be extracted from a cross-language preview and automatically activate the translation equivalent, which would generate semantic preview benefits since the target word in the related condition was the translation of the preview. However, recent work has called these assumptions into question (Costa et al., 2017), and the present study provides direct evidence against such a ubiquitous nonselective system of bilingual language processing.

The data can instead be explained by the *partial selectivity* hypothesis, which posits that bilinguals control their languages to the extent warranted by the current communicative context (Hoversten et al., 2015; Hoversten & Traxler, 2020). When reading in a monolingual language context, bilinguals may deactivate the alternate language to the extent that it appears to be completely inaccessible, as in the current study. When reading in a more bilingual mode, in which there is evidence that the alternate language is relevant to the communicative setting, activation of the alternate language would increase to the appropriate degree to maximize the cost-benefit tradeoff of increased accessibility with increased competition between languages. Future research might therefore investigate whether cross-language semantic preview benefits emerge in an increasingly bilingual language context. From the current dataset alone, however, we can conclude that models of bilingual word recognition need to be modified to allow for language membership information to directly affect ongoing lexico-semantic access during reading.

Finally, we propose that extending models of oculomotor control to bilingual reading would also require consideration of the role of language membership information in driving eye movements through a text. Most current models do not implement any such mechanism that could be used to account for the cross-language results found in the current study (Engbert et al., 2005; Reichle et al., 2009; Snell et al., 2018b). These models would need to implement a factor of language control to achieve a reduced accessibility of the alternate language when reading in a monolingual language context, potentially in a manner similar to the BIMOLA model of spoken word recognition, which explicitly allows for the influence of language mode on lexical access (Lewy, 2008). Future research will need to assess how best to implement this feature in order to simulate the results of this and other bilingual reading studies and to generate new predictions concerning the cognitive processes underlying bilingual reading.

Constraints on generality

The absence of cross-language semantic preview benefits in our participants converge with similar prior evidence from bilingual readers (Altarriba et al., 2001; Snell et al., 2018), and our results did not depend on language dominance for either within- or cross-language effects. As such, we expect the results to generalize to other populations of young adult skilled bilingual readers of alphabetic languages. However, it is possible that non-alphabetic scripts would yield different results, since scripts like Chinese have specific characteristics that may yield stronger semantic preview benefits (e.g., Wang et al., 2016). It should also be noted that the results of an omnibus model including all invalid conditions (see Supplementary Materials) did not reveal a robust statistical difference in the size of within-and cross-language effects. It is thus possible that participants experienced a small cross-language effect that simply did not reach significance in our a priori-specified models. Future research will need adequate design and power to directly compare the size of within- and cross-language semantic preview benefits as well as the influence of orthographic script.

Nonetheless, we believe that the current results will replicate in strong monolingual language contexts in which the presence of the alternate language is minimal and unknown to participants. We hypothesize that the results would differ in a bilingual language context in which both languages are overtly present and relevant, although this remains to be tested empirically. We also expect the results to generalize to other low-constraint sentences containing synonyms or translation equivalents, as long as they are carefully controlled for lexical confounds such as length, frequency, and Levenshtein distance. We lack evidence specifying whether the within-language semantic preview benefits emerged due to relatedness between previews and targets, plausibility of previews in context, or some combination of the two factors. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

CONCLUSION

This study investigated parafoveal processing within and across languages to build a more complete picture of the oculomotor and cognitive mechanisms involved in bilingual reading. The results demonstrate that bilinguals extract semantic information from the parafovea within but not across languages. These data support qualitatively similar mechanisms of within-language reading behavior in monolinguals and bilinguals and across native and non-native languages, with an additional consideration of language membership information during bilingual reading. Furthermore, the data indicate that the absence of cross-language semantic preview benefits in this and prior studies results from reduced accessibility of the alternate language while reading in a monolingual language context, as per the *partial selectivity hypothesis* of bilingual language control.

REFERENCES

Abbott, M. J., & Staub, A. (2015). The effect of plausibility on eye movements in reading: Testing EZ Reader's null predictions. *Journal of Memory and Language*, *85*, 76-87.

Altarriba, J. (1992). The representation of translation equivalents in bilingual memory. In *Advances in Psychology* (Vol. 83, pp. 157-174). North-Holland.

Altarriba, J., Kambe, G., Pollatsek, A., & Rayner, K. (2001). Semantic codes are not used in integrating information across eye fixations in reading: Evidence from fluent Spanish-English bilinguals. *Perception & Psychophysics, 63*, 875–890.

Andrews, S., & Veldre, A. (2019). What is the most plausible account of the role of parafoveal processing in reading?. *Language and Linguistics Compass*, *13*(7), e12344.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*, 364-390.

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

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, *12*, 2.

Brothers, T., & Traxler, M. J. (2016). Anticipating syntax during reading: Evidence from the boundary change paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 1894.

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: effects of reading skill. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *59*, 209.

Cong, F., & Chen, B. (2022). Parafoveal orthographic processing in bilingual reading. *International Journal of Bilingual Education and Bilingualism, 25*, 3698-3710.

Costa, A., Pannunzi, M., Deco, G., & Pickering, M. J. (2017). Do bilinguals automatically activate their native language when they are not using it? *Cognitive Science, 41*, 1629–1644.

Cuetos, F., Glez-Nosti, M., Barbon, A., & Brysbaert, M. (2012). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica, 33*, 133–143.

De Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology*, *8*, 522.

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

Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger, & A. M. Jacobs (Eds.). *Localist connectionist approaches to human cognition* (pp. 189–225). Mahwah, NJ: Erlbaum.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*, 175–197.

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: a computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, *22*, 657-679.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777.

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

Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. *Tutorials in bilingualism: Psycholinguistic perspectives,* 225-254.

Grosjean, F. (2001). The bilingual's language modes. *One mind, two languages: Bilingual language processing*, *vol. 122*.

Grosjean, F. (2013). Bilingualism: A short introduction. *The Psycholinguistics of Bilingualism, 2,* 5.

Hohenstein, S., Laubrock, J., & Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1150.

Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C., Millo, G., Mitchell, D., & Zeileis, M. A. (2019). Package 'lmtest'.

Hoversten, L. J., Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Language membership identification precedes semantic access: Suppression during bilingual word recognition. *Journal of Cognitive Neuroscience, 27*, 2108–2116.

Hoversten, L. J., Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Early processing of orthographic language membership information in bilingual visual word recognition: Evidence from ERPs. *Neuropsychologia, 103*, 183–190.

Hoversten, L. J., & Traxler, M. J. (2020). Zooming in on zooming out: Partial selectivity and dynamic tuning of bilingual language control during reading. *Cognition*, *195*, 104118.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production?. *Acta Psychologica*, *127*(2), 277-288.

Jouravlev, O., & Jared, D. (2016). Cross-script orthographic and phonological preview benefits. *The Quarterly Journal of Experimental Psychology*, 1-10.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601-625.

Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Kroff, J. R. V. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In *Psychology of learning and motivation* (Vol. 56, pp. 229-262). Academic press.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325-343.

Léwy, N. (2008). The Léwy and Grosjean BIMOLA model. In F. Grosjean (Ed.). *Studying bilinguals* (pp. 201-210). Oxford University Press.

Martin, C. D., Costa, A., Dering, B., Hoshino, N., Wu, Y. J., & Thierry, G. (2012). Effects of speed of word processing on semantic access: The case of bilingualism. *Brain and Language*, *120*, 61-65.

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*, 574-588.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*, 61–64.

Morey, R. D., Rouder, J. N., Jamil, T., & Morey, M. R. D. (2015). Package 'bayesfactor'. *http://cran/r-projectorg/web/packages/BayesFactor/BayesFactor pdf*.

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics, 28*, 661–677.

R Development Core Team. (2014). R: A language and environment for statistical computing (Version 3.1.0). Vienna, Austria: R Foundation for Statistical Computing.

Rayner K (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7,* 65–81.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372.

Rayner, K., Balota, D. A., & Pollatsek, A. (1986). Against parafoveal semantic preprocessing during eye fixations in reading. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *40*(4), 473.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using EZ Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*, 1-21.

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology*, *65*, 177-206.

Schotter, E. R. (2013). Synonyms provide semantic preview benefit in English. *Journal of Memory and Language*, *69*(4), 619-633.

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics, 74*, 5-35.

Schotter, E. R., & Jia, A. (2016). Semantic and plausibility preview benefit effects in English: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1839.

Schotter, E. R., & Leinenger, M. (2016). Reversed preview benefit effects: Forced fixations emphasize the importance of parafoveal vision for efficient reading. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 2039.

Schotter, E. R., von der Malsburg, T., & Leinenger, M. (2019). Forced fixations, trans-saccadic integration, and word recognition: Evidence for a hybrid mechanism of saccade triggering in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 677-688.

Slattery, T. J., Angele, B., & Rayner, K. (2011). Eye movements and display change detection during reading. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1924.

Snell, J., Declerck, M., & Grainger, J. (2018a). Parallel semantic processing in reading revisited: Effects of translation equivalents in bilingual readers. *Language, Cognition and Neuroscience*, *33*, 563-574.

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018b). OB1-reader: A model of word recognition and eye movements in text reading. *Psychological Review*, *125*, 969.

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

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and

L1–L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1412.

Van Hell, J. G., Litcofsky, K. A., & Ting, C. Y. (2015). Intra-sentential code-switching: Cognitive and neural approaches. In Schwieter, J. W. (Ed.) *The Cambridge Handbook of Bilingual Processing*, Cambridge University Press.

Vasilev, M. R., & Angele, B. (2017). Parafoveal preview effects from word N+ 1 and word N+ 2 during reading: A critical review and Bayesian meta-analysis. *Psychonomic Bulletin & Review*, *24*, 666-689.

Vasishth, S., & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational ideas–Part I. *Language and Linguistics Compass*, *10*(8), 349-369.

Veldre, A., & Andrews, S. (2015). Parafoveal preview benefit is modulated by the precision of skilled readers' lexical representations. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 219.

Veldre, A., & Andrews, S. (2016). Is semantic preview benefit due to relatedness or plausibility?. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 939.

Veldre, A., Reichle, E. D., Wong, R., & Andrews, S. (2020). The effect of contextual plausibility on word skipping during reading. *Cognition*, *197*, 104184.

Von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119-133.

Wang, A., Yeon, J., Zhou, W., Shu, H., & Yan, M. (2016). Cross-language parafoveal semantic processing: Evidence from Korean–Chinese bilinguals. *Psychonomic Bulletin & Review*, *23*, 285-290.

Wang, A., Zhou, W., Shu, H., & Yan, M. (2014). Reading proficiency modulates parafoveal processing efficiency: Evidence from reading Chinese as a second language. *Acta Psychologica*, *152*, 29-33.

White, S. J., Bertram, R., & Hyönä, J. (2008). Semantic processing of previews within compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 988.

Wu, Y. J., & Thierry, G. (2010). Chinese–English bilinguals reading English hear Chinese. *Journal of Neuroscience, 30*, 7646–7651.

Xiao, X. Z., Jia, G., & Wang, A. (2021). Semantic Preview Benefit of Tibetan-Chinese Bilinguals during Chinese Reading. *Language Learning and Development*, 1-15.

Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review*, *16*, 561-566.

Yan, M., Wang, A., Song, H., & Kliegl, R. (2019). Parafoveal processing of phonology and semantics during the reading of Korean sentences. *Cognition*, *193*, 104009.

AUTHOR CONTRIBUTIONS

Liv Hoversten: conceptualization (lead), methodology, formal analysis, funding acquisition, project administration, resources, software, writing- original draft, writing- review and editing **Clara Martin:** conceptualization (supporting), funding acquisition, supervision, writing- review and editing

ACKNOWLEDGEMENTS

This research was supported by the Basque Government through the BERC 2022-2025 program and by the Spanish State Research Agency through the BCBL Severo Ochoa excellence accreditation CEX2020-001010-S and through a Juan de la Cierva- Incorporación postdoctoral grant to L.J.H. The research was also supported by the Spanish Ministry of Economy and Competitiveness (PID2020-113926GB-I00 to C.D.M.) and the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 819093 to C.D.M.).