Journal of Memory and Language 146 (2026) 104701

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Predicting words across languages depends on language context: Evidence from visual world eye-tracking

Huanhuan Yin *, Martin J. Pickering

Department of Psychology, University of Edinburgh, United Kingdom

ARTICLE INFO

Keywords:
Language prediction
Visual-world paradigm
Cross-language activation
Bilingualism
Selective lexical access

ABSTRACT

There is good evidence that monolingual comprehenders can predict the form of upcoming words, and also that bilinguals often activate words from both languages in parallel during bottom-up language comprehension. But it is unclear whether bilinguals predict the form of upcoming words in the language that they are not hearing, and whether such predictions depend on whether or not they have recently encountered that language. We investigated these questions in two visual-world eye-tracking experiments by asking whether Mandarin Chinese (L1)-English (L2) bilinguals pre-activate Mandarin phonological representations of predictable words during English comprehension. Participants heard English sentences containing a highly predictable word while viewing a display. They fixated more on a competitor object whose Mandarin name was a homophone of the Mandarin translation of the predictable word than an unrelated object when both languages were used (Experiment 2) but not when just English was used (Experiment 1). Our findings suggest that bilinguals predict across languages when both languages are contextually relevant but not otherwise.

Introduction

Language comprehension is not merely a bottom-up passive process of waiting for words, but also involves actively generating top-down predictions about what is likely to occur (Huettig, 2015; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018), including predictions about the form of the upcoming word (Ito, 2024). For bilinguals, this prediction process is made more complicated because they can conduct bottom-up and top-down processing in two languages. There is much evidence that bilingual word recognition often involves the activation of word forms from both the relevant (target) language and the irrelevant (non-target) language (e.g., Dijkstra et al., 1998; Thierry & Wu, 2007), in a language non-selective manner (Green, 1998). But we do not know whether comprehenders predict word forms across languages.

We therefore asked whether bilinguals predict word forms in just the target language that is currently in use (*selective prediction*) or whether they predict in both the target and the non-target languages (*non-selective prediction*). Moreover, we asked whether non-selectivity depends on whether they have recently encountered both languages (in a *two-language context*) or just the target language (in a *one-language context*). To address these questions, we examined whether highly proficient bilingual speakers pre-activate the phonological representations of predictable words in their first language (L1) while comprehending in their second language (L2). We conducted two experiments in which Mandarin Chinese-English bilinguals (i.e., L1 Mandarin) listened to English sentences while looking at pictures and we monitored their eye movements to determine what they were predicting.

Prediction within the target language

A substantial body of research demonstrates that L1 comprehenders can predict many aspects of upcoming words, including meaning (Altmann & Kamide, 1999; Dijkgraaf et al., 2019a; Grisoni et al., 2017), syntax (Van Berkum et al., 2005; Wicha et al., 2004), and form (Ito et al., 2018; Kukona, 2020; Li et al., 2022; Martin et al., 2013). For example, Altmann and Kamide (1999) found that after hearing *The boy will eat the…*, participants were more likely to look at an edible object (a cake) than inedible objects (a ball, a toy train, and a toy car) before the word *cake* was spoken, suggesting that participants pre-activated a component of the meaning of the upcoming word (here, that it is edible). With regard to form, Li et al. (2022) had L1 Mandarin Chinese speakers listen to Mandarin sentences containing a highly predictable word (e.g., *放学了，我把铅笔盒和本子装进了书包里准备回家* "After school, I put my pencil case and notebooks into my schoolbag and get ready to go home") while

* Corresponding author at: Department of Psychology, University of Edinburgh, 7 George Square, EH8 9JZ Edinburgh, Scotland, United Kingdom.
E-mail address: H.Yin-9@sms.ed.ac.uk (H. Yin).

https://doi.org/10.1016/j.jml.2025.104701
Received 22 December 2024; Received in revised form 26 September 2025; Accepted 29 September 2025
Available online 1 October 2025

viewing a scene containing a critical object whose Mandarin name (e.g., *shu*1*zi*5, meaning comb) shared the same initial syllable and tone (i.e., a homophone) as the predictable word (*shu*1*bao*1, meaning schoolbag). Participants fixated more on the Mandarin homophone competitor than an unrelated object before the target word onset, demonstrating pre-activation of the phonological information of the predictable word (see also Ito et al., 2018; Xu et al., 2024; Zhao et al., 2024).

Together, these findings demonstrate that L1 comprehenders can generate rich predictions at multiple levels. One influential account of how such predictions are generated is the prediction-by-production hypothesis (Dell & Chang, 2014; Pickering & Gambi, 2018; Pickering & Garrod, 2013), which proposes that comprehenders use their own production system to predict upcoming words. According to this account, they first derive the speaker's intention by covertly imitating the speaker's utterance, and then pre-activate representations of the predictable word in the order of meaning, syntax, and phonology, mirroring the stages of production, but without articulating.

The evidence for prediction during L2 comprehension is less clear than the evidence for prediction during L1 comprehension. (At this point, we discuss only within-language prediction.) Most studies have found that L2 comprehenders predict semantics (e.g., Alemán Bañón & Martin, 2024; Chun & Kaan, 2019; Dijkgraaf et al., 2019b; Lew-Williams & Fernald, 2010) and syntax (e.g., Fang & Wu, 2024; Hopp & Lemmerth, 2018b), although they often predict less than L1 comprehenders (Hopp, 2015; Lew-Williams & Fernald, 2010; Mitsugi & Macwhinney, 2016; Momenian et al., 2024; Schlenter, 2023; Schlenter & Felser, 2021; see Kaan & Grüter, 2021 for a review).

The clearest difference between L1 and L2 prediction appears to relate to form, as no study has found form prediction in L2 comprehension. For example, in an ERP study, Martin et al. (2013) had Spanish-English bilinguals and L1 English speakers read English sentences with a predictable or unpredictable final noun, which was preceded by an article either congruent or incongruent with the noun (as in DeLong et al., 2005). L1 speakers, but not L2 speakers, showed an increased N400 amplitude for articles incongruent with the predictable nouns, suggesting that L2 speakers did not pre-activate the phonological representation of predictable words. Similarly, Ito et al. (2017b) had Spanish-English bilinguals read highly constraining L2 English sentences that continued with a predictable word, a phonologically related word, a semantically related word, or an unrelated word. Both phonologically and semantically related words elicited increased N400 effects similar to those for non-predictable words but greater than for predictable words, suggesting that L2 speakers did not pre-activate the semantic or form representations of the predictable words (see also Ito et al., 2018, discussed below).

Cross-language activation and the effects of language context

As we have noted, our key questions are whether comprehenders predict across languages (i.e., non-selectively) and whether any such prediction is affected by language context. More specifically, we ask whether they make L1 predictions when comprehending L2 utterances. The main reason to hypothesize non-selective prediction is that there is good evidence for non-selectivity in other aspects of language processing (both comprehension and production), though this non-selectivity is influenced by language context (see below).

Based on evidence that bilinguals show different code-switching patterns in production across contexts (i.e., mixing languages more frequently in two-language contexts than in one-language contexts), Grosjean (1998, 2001, 2008) proposed that the degree to which each language is activated depends on the language context, as shaped by factors such as the interlocutor, the communicative setting, and task demands. When both L1 and L2 are in use (a two-language context), both languages tend to remain active; when only one language is in use (a one-language context), only the target language is typically activated. Much of the evidence for this context-sensitive modulation of cross-language activation comes from two broad types of studies: those using language-ambiguous stimuli, in particular cognates and interlingual homographs; and those using language-unambiguous stimuli.

Cognates

Facilitation for cognates (i.e., words sharing both meaning and form across languages, such as *taxi* in English and Spanish), offers clear evidence of parallel activation. Cognates are processed faster than non-cognate controls in comprehension tasks such as lexical decision (Dijkstra et al., 1998; Duyck et al., 2007; Van Hell & Dijkstra, 2002) and progressive demasking (Dijkstra et al., 1999; Lemhöfer et al., 2008). Similarly, in production, bilinguals name cognates faster than non-cognates in picture naming tasks (Costa et al., 2000; Hoshino & Kroll, 2008) for review, see Costa et al., 2005). This facilitation in comprehension and production is often attributed to the co-activation of the target word's cognate in the non-target language, with converging activation from both languages speeding processing.

There is some evidence that the strength and presence of cognate facilitation depends on language context, although findings are mixed. For example, Titone et al. (2011) found that English–French bilinguals showed cognate facilitation during reading in low- but not high-constraint L1 (English) sentences. But when L2 (French) sentences were also included (thereby creating a two-language context), facilitation emerged even in high-constraint L1 sentences, suggesting that cross-language activation is enhanced in a two-language context. In contrast, Dijkstra et al. (2015) found that Dutch-English bilinguals showed cognate facilitation in an L2 English lexical decision task regardless of whether the cognates were preceded by Dutch or English sentence context. Finally, Vanlangendonck et al. (2020) reported that in an L2 lexical decision task, cognate facilitation turned into inhibition when L1 items were included in the stimulus list. These three studies therefore all showed very different patterns of results. Note, lexical decision tasks involve metalinguistic judgments that may not be present in natural sentence reading or listening.

Interlingual homographs

Interlingual homographs (words with the same orthography but different meanings across languages) provide additional evidence for cross-language activation. Bilinguals take longer to respond to interlingual homographs than non-homographic controls in L2 lexical decision tasks (Dijkstra et al., 1998, 1999; Dijkstra et al., 2000; Schulpen et al., 2003). This delay is often interpreted as evidence for the co-activation of both the form and meaning of homographs in the non-target language (L1).

As with cognates, the interlingual homograph interference effect is influenced by language context. For example, in Dijkstra et al.'s (1998) English lexical decision task, there was no response time difference between interlingual homographs and control words when only L2 English words or nonwords that were orthographically legal in English were used (a one-language context). But when L1 Dutch filler words were added (a two-language context), response times for homographs were longer than for English control words (see also Dijkstra, Bruijn, et al., 2000; Poort & Rodd, 2017). Similarly, Elston-Güttler et al. (2005) presented German-English bilinguals with English sentences (e.g., *The woman gave her friend an expensive…*) ending with either an interlingual homograph (*gift*, meaning "poison" in German) or a control word (*item*), followed by target words corresponding to the homograph's L1 meaning (*poison*). Participants who watched a German movie before the experiment showed faster lexical decisions and N200/N400 modulations for homograph-related targets in the first block—both effects absent for those who watched an English movie, suggesting that cross-language activation is influenced by language context (see also Elston-Güttler & Gunter, 2009).

Evidence from production mirrors this pattern of results. Smits et al.

(2006) had Dutch-English bilinguals name interlingual homographs in English. Naming latencies were longer for homographs than for non-homographic control words when the stimuli included both Dutch and English non-homographs (a two-language context), as well as homographs. However, this interference effect disappeared when the list contained only homographs and English words (a one-language context), suggesting that non-target language activation in production emerges at least primarily in two-language contexts.

Language-unambiguous stimuli

Beyond language-ambiguous stimuli such as cognates and interlingual homographs, evidence for cross-language activation also comes from studies using language-unambiguous stimuli—that is, L2 words that do not have the same (or very similar) form to an L1 word, but which nonetheless appear to trigger L1 activation during comprehension. First, Thierry and Wu (2007) had Mandarin-English bilinguals judge the semantic relatedness of English word pairs (e.g., *train* – *ham*) whose Mandarin translations (e.g., *huo che* – *huo tui*) shared a character, and found reduced N400 amplitudes for these pairs in comparison to English pairs whose Mandarin translations did not share a character. They interpreted these results in terms of unconscious activation of Mandarin translations during English comprehension (see also Wu & Thierry, 2010; Zhang et al., 2011). Second, using the visual-world paradigm, Wang et al. (2017) had Mandarin-English bilinguals listen to English words (e.g., *rain,* "yu3" in Mandarin) while viewing a display containing a target object, a competitor, and two unrelated distractors. Participants were distracted by a competitor object (e.g., a *feather,* also "yu3" in Mandarin) whose Mandarin name was a homophone of the Mandarin translation of the target word in comparison to an unrelated control. However, this effect did not occur when the competitor's Mandarin name shared all segments but differed in tone from the target word (a *fish,* "yu2" in Mandarin), or when it shared rime and tone but not onset (*wheat,* "gu3" in Mandarin). These results suggest that Mandarin–English bilinguals automatically activate L1 translations, including tone information, during L2 word processing.

As with language-ambiguous stimuli, cross-language activation from language-unambiguous stimuli is affected by language context, as shown in code-switching studies. For example, Cheng and Howard (2008) presented Mandarin-Taiwanese bilinguals with pairs of Mandarin sentences that either both included some Taiwanese words (two-language context) or were entirely in Mandarin (one-language context) and had them judge if two sentences were synonymous in Mandarin. Participants showed a switch cost (i.e., were slower to respond) for language-switching pairs only when they expected a monolingual Mandarin context. When they expected that both Mandarin and Taiwanese would be used in the experiment, the switch cost disappeared. In an eye-tracking study, Olson (2017) had Spanish-English bilinguals read a target word in English following a context that was entirely in English, was entirely in Spanish, or consisted of some English words followed by some Spanish words. Relative to the English-only context, participants had more difficulty with a Spanish-only context, but did not have difficulty with a mixed language context, suggesting that the non-target language is activated to a larger extent in a two-language context than in a one-language context (see also Declerck & Grainger, 2017; Kheder & Kaan, 2019).

Hoversten and Traxler (2020) had Spanish-English bilinguals read low-constraint English sentences (e.g., *We saw that his_had a horrible scar*) containing a critical word in English (*hand*), Spanish (*mano*), or a pseudoword (*erva*). Eye movements revealed that overtly presented Spanish words (two-language context) caused early reading disruptions compared to English words, though less than pseudowords, suggesting that code-switched words were less accessible but not fully treated as pseudowords. However, when code-switched words were presented covertly using the gaze-contingent boundary paradigm (arguably a one-language context), they were processed like pseudowords (see also Hoversten & Martin, 2023), and so Spanish did not seem to be activated.

Production data also demonstrate effects of context on cross-language activation. In an fMRI study, Abutalebi et al. (2008) had German-French bilinguals name pictures in either blocked conditions—naming exclusively in L1 or L2 (a one-language context, with respect to the block at least), or in mixed conditions that required switching between L1 and L2 (a two-language context). Naming in L1 during the two-language context elicited greater activation in the left caudate and anterior cingulate cortex (areas associated with inhibition) compared to naming in L1 during the one-language context (see also Guo et al., 2011). This suggests that the language that is not being produced is more strongly activated in a two-language context than in a one-language context.

In another cued picture-naming study, Olson (2016) had bilinguals name pictures either in a context that was nearly one-language (with 95 % of trials in L1 and 5 % in L2, or vice versa) or a two-language context (with 50 % of trials in each language). Asymmetrical switch costs (i.e., greater costs when switching into L1 than into L2) were observed in the nearly one-language context but disappeared in the two-language context. This pattern suggests that the non-target language is less strongly inhibited in two-language contexts compared to nearly one-language contexts. In summary, the evidence from cognates, interlingual homographs, and language-unambiguous stimuli suggest that bilingual speakers often activate both languages during bottom-up comprehension and language production. Moreover, such cross-language activation is affected by language context: Lexical access tends to be language non-selective in a two-language context but more selective in a one-language context.

Evidence relating to cross-language prediction

Given the evidence for prediction in both monolinguals and bilinguals (using their L2), and for bottom-up cross-language activation in bilinguals, we can ask whether bilinguals predict cross-linguistically, and therefore whether prediction is language non-selective (Green, 1998). Such prediction could explain the reduced prediction effects for bilinguals within L2 compared to monolinguals discussed above (Hopp & Lemmerth, 2018b; Kaan, 2014; Schlenter, 2023). More importantly, there is evidence that L2 speakers co-activate their L1 during L2 prediction (i.e., non-selective prediction), including the activation of L1 semantic (e.g., Ito et al., 2024, 2025; van Bergen & Flecken, 2017) and syntactic information (e.g., Alemán Bañón & Martin, 2021; Dussias et al., 2013; Hopp & Lemmerth, 2018b).

But there is currently no evidence that bilinguals using L2 predict word forms in L1 (i.e., the non-target language). For example, in an ERP study, FitzPatrick and Indefrey (2010) had Dutch-English bilinguals listen to high-constraining English sentences ending with semantically incongruent words, predictable words, or words phonologically related to the Dutch translations of predictable words (e.g., *My Christmas present came in a bright-orange doughnut*"; *doughnut* had initial overlap with *doos*, meaning "box"). There was no difference in the N400 peak or onset latency between phonologically related words and incongruent words (though incongruent words did differ from predictable words), suggesting that the L1 phonological representations of the predictable words were not activated during L2 comprehension.

Using the visual world paradigm, Ito et al. (2018) had Japanese-English bilinguals listen to English sentences containing a highly predictable word (e.g., *The tourists expected rain when the sun went behind the cloud*) while viewing a display with a critical object and three distractors. The critical object either corresponded to the predictable word (e.g., *cloud*, "kumo" in Japanese), an English competitor (e.g., *clown*) whose English name was phonologically related to the predictable word, a Japanese competitor (*bear*; "kuma" in Japanese) whose Japanese name was phonologically related to the Japanese translation of the predictable word, or an unrelated object (e.g., *globe*). While participants showed more anticipatory fixations to the target object, there was no difference in predictive looks between the Japanese competitor and the unrelated object, nor between the English competitor and the unrelated object. Thus, there was no evidence that L2 English speakers pre-activated phonological representations of either the predictable English word or its L1 Japanese translation.

This study used a one-language context (i.e., entirely English). Using the same design, Amos et al. (2022) had L1 French translators and interpreters listen to English sentences and simultaneously interpret them into French while viewing a display containing a target object, an English phonological competitor, a French phonological competitor, or an unrelated object. This task of course involves use of both languages (i.e., is a two-language context). Although both groups showed more predictive looks to the target object than to the unrelated object, neither group did so for the English or French phonological competitors.

However, we cannot be certain that L1 phonological representations of predictable words are not pre-activated during L2 comprehension, especially given that phonological prediction effects are often small and therefore difficult to detect (see Ito, 2024). For example, in Ito et al. (2018), Japanese competitors and Japanese translations of predictable words shared only 2.6 out of 4.9 phonemes on average (53 %), which may not have been sufficient to trigger activation of phonologically related words. Additionally, the study had low power (96 observations per condition in Ito et al., 2018), as did Amos et al. (2022) (192 per condition), though note that visual-world studies typically involve far fewer observations per condition (e.g., 192 in Altmann & Kamide, 1999, Experiments 1 and 2; and only 18 in Tanenhaus et al., 1995 and Spivey et al., 2002) than response-time experiments.

In sum, existing research does not provide clear evidence that bilingual speakers pre-activate L1 phonological representations of predictable words during L2 comprehension. Moreover, since most experiments concerning prediction have been conducted in one-language contexts, we do not know whether lexical access during predictive processing is language selective, non-selective, or context-dependent.

The present study

Our study investigated whether bilinguals comprehending sentences in their L2 make lexical predictions in their L1. To do so, we conducted two visual-world eye-tracking experiments. We examined whether Mandarin-English bilinguals listening to English predict in both Mandarin and English or whether they restrict predictions to English. Also, we explored whether such predictions occur when they heard only English (one-language context, Experiment 1) or both Mandarin and English (two-language context, Experiment 2).

Our experiments were related to Ito et al. (2018), but differ in both research focus and methodology. Theoretically, our research aims were to investigate whether L2 speakers pre-activate L1 form representations of predictable words during L2 comprehension, and whether any such cross-language prediction depends on language context. Therefore, we included only three conditions, excluding the English phonologically related condition from Ito et al. (2018), and in addition, we conducted Experiment 2 using a two-language context.

Methodologically, our study differed from Ito et al. (2018) in two ways. First, to enhance the likelihood of detecting L1 phonological activation, we used (intralingual) homophones in Mandarin as L1 phonological competitors. Mandarin has an abundance of homophones—words with identical pronunciations but different meanings and spellings. Compared to the L1 phonological competitors used by Ito et al. (2018), our use of homophones greatly increased the phonological overlap with the Mandarin translations of predictable words. We assumed that this would maximize the likelihood of detecting cross-language L1 phonological activation, provided that L1 Mandarin translations of predictable words are indeed activated during L2 English comprehension. The second difference from Ito et al. (2018) was an increased sample size. As noted, Ito et al. (2018) had 96 observations per condition and Amos et al. (2022) had 192, but neither study observed cross-language L1 phonological activation. In contrast, studies using Mandarin homophones have reliably detected L1 phonological activation during L1 comprehension (Li et al., 2022; Xu et al., 2024). Xu et al. (2024) had 288 observations per condition and Li et al. (2022) had 450. We decided to have 360 observations per condition, with each participant viewing 12 items per condition. (We acknowledge that our studies have low power and supplemented our analyses with Bayesian modelling that quantified relative evidence for the alternative versus the null model to assess the strength of evidence for the alternative over the null hypothesis; see Methods for details).

Our participants heard English sentences containing a highly predictable word (e.g., *rain* in *You should take an umbrella with you, because there will be heavy rain at three o'clock this afternoon*) while viewing a display which contained three distractors and a critical object (see Fig. 1). The critical object corresponded to the target word (e.g., *rain* [Mandarin: *yu*3]), a Mandarin homophone competitor whose Mandarin name was a homophone (e.g., *feather* [Mandarin: *yu*3]) of the Mandarin translation of the predictable word, or a word (e.g., *comb* [Mandarin: *shu*1]) that was unrelated to the predictable word.

We assume that participants predict components of the meaning of highly predictable words in L2 sentences, and therefore hypothesize that they will fixate the target objects more than unrelated objects before the target word onset. The critical question is whether they pre-activate L1 phonological representations of predictable words in L2 sentences. If they do, they should fixate more on Mandarin homophone objects than on unrelated objects before the target word onset. Experiment 1 investigated these issues using a one-language context (where filler sentences were in English), and Experiment 2 did so using a two-language context (where filler sentences were in Mandarin).

Experiment 1

In Experiment 1, the whole experimental session was conducted in English. The only use of Mandarin was in a familiarization session (see below).

Methods

Participants

Thirty Mandarin Chinese students at the University of Edinburgh (5 males and 25 females), aged between 21 and 28 years ($M = 23.37$, $SD = 1.61$) participated in this experiment. All participants were native Mandarin Chinese speakers with English as their L2. They had all lived in the UK for at least 5 months and were currently studying in the UK at the time of the experiment. Each participant received a payment of £9 for their participation. After the main experiment, participants were asked to indicate their scores on International English Language Testing System, the age at which they began learning English, the daily duration of their English use, and their immersion length in an English-speaking country. Additionally, they were asked to self-rate their English proficiency in reading, writing, speaking, and listening separately on a scale from 1 (not good at all) to 7 (very good); see Table 1.

Stimuli

We created 36 pairs of English sentences (see Appendix 1) involving a highly predictable target noun (underlined), for example *You should take an umbrella with you because there will be heavy rain at three o'clock this afternoon* and *Many birds, like parrots and peacocks, have beautiful and colorful feathers that make them very exotic and eye-catching.* The Mandarin translation of the predictable word from one member of the pair (here, *rain*, with translation *yu3*) was homophonous with the Mandarin translation of the predictable word from the other member of the pair (here, *feathers*, also with translation *yu3*). The experimental sentences had a mean length of 23.9 words ($SD = 3.1$, range = 17–31 words), and

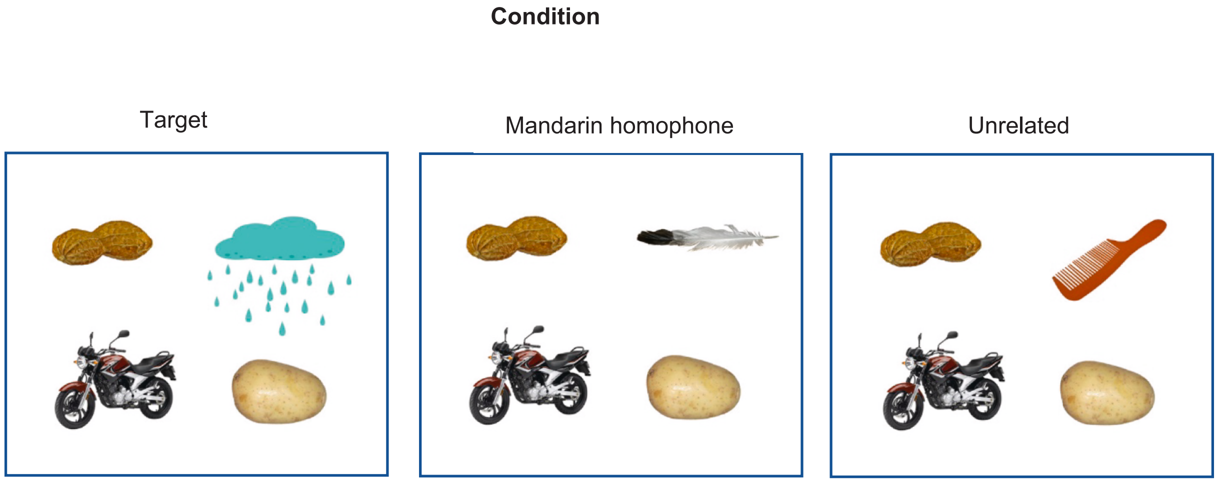

Fig. 1. *Example Visual Scene in the Three Conditions for the Experimental Sentence "You should take an umbrella with you, because there will be heavy rain at three o'clock this afternoon". Note.* The object depicted at the top right corner is the critical object for this item. The visual stimuli were also paired with the filler sentence "When my brother and I were playing in the street, we lost 200 lb, and we don't want our mum to know that".

Table 1
The Language History and Proficiency of the Mandarin-English Participants in Experiment 1 and 2.

Measure	One-language context (Experiment 1)	Two-language context (Experiment 2)
	Mean (*SD*)	Mean (*SD*)
English self-rating reading (7-point scale)	5.45 (1.09)	5.93 (0.74)
English self-rating listening (7-point scale)	5.26 (1.24)	5.53 (1.00)
English self-rating speaking (7-point scale)	4.71 (1.22)	4.77 (1.25)
English self-rating writing (7-point scale)	4.68 (0.98)	5.07 (0.94)
English proficiency test (IELTS score)	7.03 (0.44)	7.19 (0.42)
Daily English usage (hours)	6.50 (4.67)	5.17 (4.35)
Age of English acquisition (years)	7.71 (2.67)	6.95 (2.66)
Length of immersion (months)	13.87 (13.11)	10.57 (11.02)

the predictable word appeared at different positions in each sentence (range = 10th-25th word, M = 16.6, SD = 3.0) but never sentence-finally.

Target-word predictability was assessed using a cloze-probability test. Fourteen further Mandarin-English bilinguals who were studying in the UK were presented with sentence fragments which were truncated immediately before the target word, and were asked to complete each sentence with the first word that came to mind. The cloze probability of a word was defined as the percentage of participants who used the exact word to complete the sentence. The mean cloze probability of the target word was 88.08 % (SD = 12.54, range = 61.54 %–100 %). There were an additional 36 English filler sentences (e.g., *When my brother and I were playing in the street, we lost 200 lb, and we don't want our mum to know that*) with a similar length to the experimental sentences (M = 24.0, SD = 3.5, range = 20–34 words) but were not designed to have a highly predictable word.

The sentences were generated using Murf AI voice generator, a text to speech software which can produce human-like and natural-sounding voice. These sentences were generated with the synthesized voice which was set as "English-UK-Edward (M), middle-aged, general tone, pitch (0 %), speed (−50 %)" (https://murf.ai/studio/project/2/P017026364560496CY?workspaceId=WORKSPACEID017026364543209BE&) at 48 kHz with a format of 32-bit float and a bit rate of 1,536 kbps, and was downloaded in high-quality WAV format. The resulting speech rate was 3.67 syllables per second. This configuration was chosen because it produced audio that closely resembled a native British English speaker, as confirmed by three further Mandarin-English bilinguals from the University of Edinburgh.

Each of the sentences was paired with a scene containing four objects: a critical object and three distractors. In the target condition, the critical object corresponded to the target word (e.g., *rain* [Mandarin: *yu*3]). In the Mandarin homophone condition, the Mandarin name of the critical object was a homophone of the Mandarin translation of the target word (e.g., *feather* [Mandarin: *yu*3]). English and Mandarin names of the Mandarin homophone objects were both unrelated to any of the English names of the target and unrelated objects. In the unrelated condition, the name of the critical object was neither phonologically nor semantically related to the predictable word (e.g., *comb* [Mandarin: *shu*1*zi*0]). The unrelated condition served as a baseline. English and Mandarin names of each critical object had no phonemes in common. An example item with four conditions is shown in Fig. 1.

For the two versions of sentence contexts in each homophone pair, the target object in the target condition for one sentence context was used as the Mandarin homophone object for the other sentence context in the Mandarin homophone condition, and vice versa. In the unrelated condition, these two sentence contexts were paired with two different unrelated objects, which were also used as a target or homophone object in other trials. This design ensured counterbalancing, and allowing each critical object to alternate between roles across conditions for a more controlled comparison.

A further 28 Mandarin-English bilinguals from the University of Edinburgh participated in a name agreement test for the critical objects: Of these, 14 provided the first Mandarin name that came to their mind for each picture, and the other 14 provided the English name. In the English-naming norming task, the instructions were in English throughout the test. In the Mandarin-naming norming task, the instructions were in Mandarin throughout the test. The Mandarin and English naming agreement for the retained critical objects was 86.71 % (SD = 15.90, range = 50 %–100 %) and 88.89 % (SD = 14.48, range = 50 %–100 %) respectively.

All the visual stimuli were shown twice, once in an experimental trial and once in a filler trial. The matched visual stimuli for an experimental and filler recording contained the same four objects, but the quadrant in which each of the four objects appeared was varied. Each experimental list comprised 72 sentences: 36 experimental sentences and 36 filler sentences. Each list was divided into two half lists, each containing 36 visual arrays paired with 18 experimental and 18 filler sentences. Visual arrays that were paired with experimental items in one half-list were paired with fillers in the other half-list, and vice versa. Experimental pictures were counterbalanced across the full lists, resulting in three different sets of items and a total of 12 experimental lists (six lists for each version of sentence context in a homophone pair). The order of the sentences within each block was randomized for each participant and the whole experiment took approximately 35 min, with participants having four breaks during the session.

Critical objects appeared in each quadrant equally frequently. Filler sentences mentioned one of the three distractor objects in the visual display 2/3 of the time. Thus, in total 50 % of sentences mentioned an object in the visual display.

Procedure

Participants were tested in a laboratory at the University of Edinburgh, and the whole session had three parts: the picture-familiarization task, the eye-tracking experiment, and the language-background questionnaire.

In the picture-familiarization task, participants were shown 180 objects (72 critical objects and 108 distractor objects) one by one on a screen, each accompanied by its English and Mandarin names displayed horizontally below. The name order was counterbalanced: For half the pictures, English names appeared on the left and Mandarin names on the right, and vice versa for the other half. Participants were told to memorize both the Mandarin and English names for each object. Once they indicated that they had remembered all the names, the pictures were shown again without the names, and participants were asked to name each object in both Mandarin and English using the previously given names. To balance positional effects, participants first articulated the Mandarin names for pictures with English names on the left, followed by the English names, and vice versa for pictures with Mandarin names on the left. The mean picture naming accuracy on the first presentation was 92.41 %. The high accuracy suggests that the pictures were familiar to our participants and were relatively easy to associate with the intended names. If they did not respond correctly with both the English and Mandarin names for a picture, that picture was presented again in another picture-familiarization session (until all pictures were named correctly).

The experimenter used solely English for all communication, including greetings and instructions, and participants were also told to respond only in English. In addition, all instructions displayed on the screen and filler sentences were in English. Thus, the experiment provided an entirely one-language context, except for the familiarization session.

In the eye-tracking experiment, participants were seated in front of a computer screen, at a distance of about 60 cm. Participants' right eye movements were recorded using an SR Research EyeLink® 1000 mount eye-tracker with a sample rate of 500 Hz. Participants were instructed to listen carefully to each sentence and judge whether it mentioned any of the objects on the screen after the sentences ended. Before the experiment, participants put their chin on a chin rest, and the eye-tracker was calibrated using a nine-point calibration grid. The experiment began with two practice trials, followed by a break in which participants could ask questions. The pictures were displayed on a screen with a resolution of 1024 × 768 pixels. Each trial started with a drift correction, followed by a 500 ms blank screen. On experimental trials, the visual scene was presented 1000 ms before the onset of predictable words in experimental trials. In filler trials, the scene appeared 1000 ms before the onset of a word referring to a distractor object or at an arbitrary mid-sentence position if the sentence did not mention any of the objects. The picture remained on the screen for 750 ms after sentence offset, at which point it was replaced with the question "Did the sentence mention any of the pictures?". Participants indicated their answers by pressing the left button for "No" and the right button for "Yes", using a button box, and the next trial began immediately after the response. No feedback was given during the experiment. The session took about 40 min. At the end, we included an open-ended awareness check inviting comments about anything. No participant reported noticing the L1 homophone manipulation in either Experiment 1 or 2.

Data analyses

Two items were excluded from the eye-tracking analyses in Experiment 1, because the target object for these items attracted more looks than the unrelated objects within 1000 ms after the picture onset when the pictures were presented with a neutral sentence that was unrelated to the target objects in filler trials (see the analysis for filler trials below; Ito et al. 2018 used similar ways). This left 34 items for the analyses. Following Ito et al. (2018), we began by analyzing the fixation on the critical objects for the filler sentences (without having analyzed the experimental sentences) and two items showed many more looks to the target object than the unrelated object. Indeed, the analysis of all the filler items showed that participants looked at the target object more than the unrelated object. When we removed the two items, the analysis now showed no difference in fixations toward the target objects and the unrelated objects.

In both Experiment 1 and 2, we first analyzed the eye-tracking data using generalized linear mixed-effects models (Baayen et al., 2008) with the lme4 package (Bates et al., 2015) in R (R Core Team, 2020). The fixations on critical objects (target, Mandarin homophone, and unrelated objects) were coded binomially (fixated = 1, not fixated = 0) in each 50 ms bin from 1000 ms before to 1000 ms after the target word onset. If fixations fell in the area of 300 × 300 pixels around a picture, they were regarded as falling on it. Blinks and fixations outside the interest areas were coded as 0 and were included in the data. The model used binomially coded fixation as the dependent variable, predicted by condition. The condition variable was dummy-coded, with the unrelated condition set as the reference level. Thus, we could test effects of each critical condition relative to the unrelated (baseline) condition (target vs. unrelated, Mandarin homophone vs. unrelated). The model included random intercepts by participants and by items (Baayen et al., 2008). Random slopes were not included because the models with them did not converge for several of the time bins. As in Ito et al. (2018), we determined whether a critical condition significantly differed from the baseline condition by assessing if there was a period with at least three consecutive bins showing significant differences.

But this approach has two limitations. First, fixations in adjacent time bins are highly correlated, because the eye-tracker records a fixation at every millisecond, but fixations usually last for a few hundreds of milliseconds (Ito & Knoeferle, 2023; Rayner, 1998). Second, the binned analysis was conducted in many time bins and therefore involved multiple comparisons, and so it increases the chance of Type 1 error (Hochberg, 1987). To address these issues, we additionally conducted a cluster-based permutation analysis for the time period from −1000 ms before until 1000 ms after target word onset, which controls for autocorrelation of eye-movement across neighboring time bins and overcomes the multiple comparisons issue.

Given the possibility that our study may be underpowered, we additionally analyzed our data using Bayesian generalized mixed-effects models in 50 ms time bins from −1000 ms to 1000 ms relative to target word onset (in both Experiment 1 and 2). These models were implemented using the brms package (version 2.20.4) in R. We used weakly informative priors to regularize estimates. Fixed-effect coefficients for condition were given Normal (0,1) priors on the log-odds scale, group-level standard deviations were given Normal (0,1) priors, and the model intercept was assigned Student-t (3,0,2.5). These choices follow common practice for logistic regression models in brms. To quantify the evidence for our specific hypotheses (Target vs. Unrelated, and Homophone vs. Unrelated), we conducted model comparison using Bayes Factors (BF_{10}) calculated via bridge sampling. For each comparison, we contrasted a full model that included the effect of interest (M1) against a null-hypothesis model (M0) where that specific effect was absent. The resulting BF_{10} quantifies the evidence for M1 over M0. A BF_{10} less than 1 provides evidence in favor of M0; values between 1 and 3 suggest weak evidence for M1; values above 3 indicate moderate evidence for M1; and values above 10 indicate strong evidence for M1 (Lee & Wagenmakers, 2014).

Results

Comprehension question accuracy

The mean accuracy for the comprehension questions in the experimental trials was 98.19 % (SD = 13.33 %). Incorrectly answered trials were excluded from the eye-tracking analyses.

Eye-tracking data

Fig. 2 shows the mean fixation proportions on the target objects, the Mandarin homophone objects, and unrelated objects in the time window from −1000 ms before to 1000 ms after target word onset. Time was synchronized to target word onset.

The binned generalized linear mixed-effects models showed that bilinguals were more likely to fixate on target objects than on unrelated objects from −650 ms before they heard the target word until (at least) 1000 ms after the target word onset. However, from −1000 ms before to at least 1000 ms after the target word onset, there was no fixation bias toward Mandarin homophone objects compared to unrelated objects. In fact, participants fixated more on unrelated objects than on homophone objects during two time bins: −100 ms to −50 ms and −50 ms to 0 ms.

Consistent with the results from the generalized mixed-effects models, the Bayesian generalized mixed-effects models showed that the BF_{10} for the contrast between target and unrelated objects exceeded 1 after −750 ms, 3 after −700 ms, and was consistently above 10 from −650 ms until 1000 ms after target onset (when recording finished), therefore indicating strong evidence for the alternative hypothesis (a

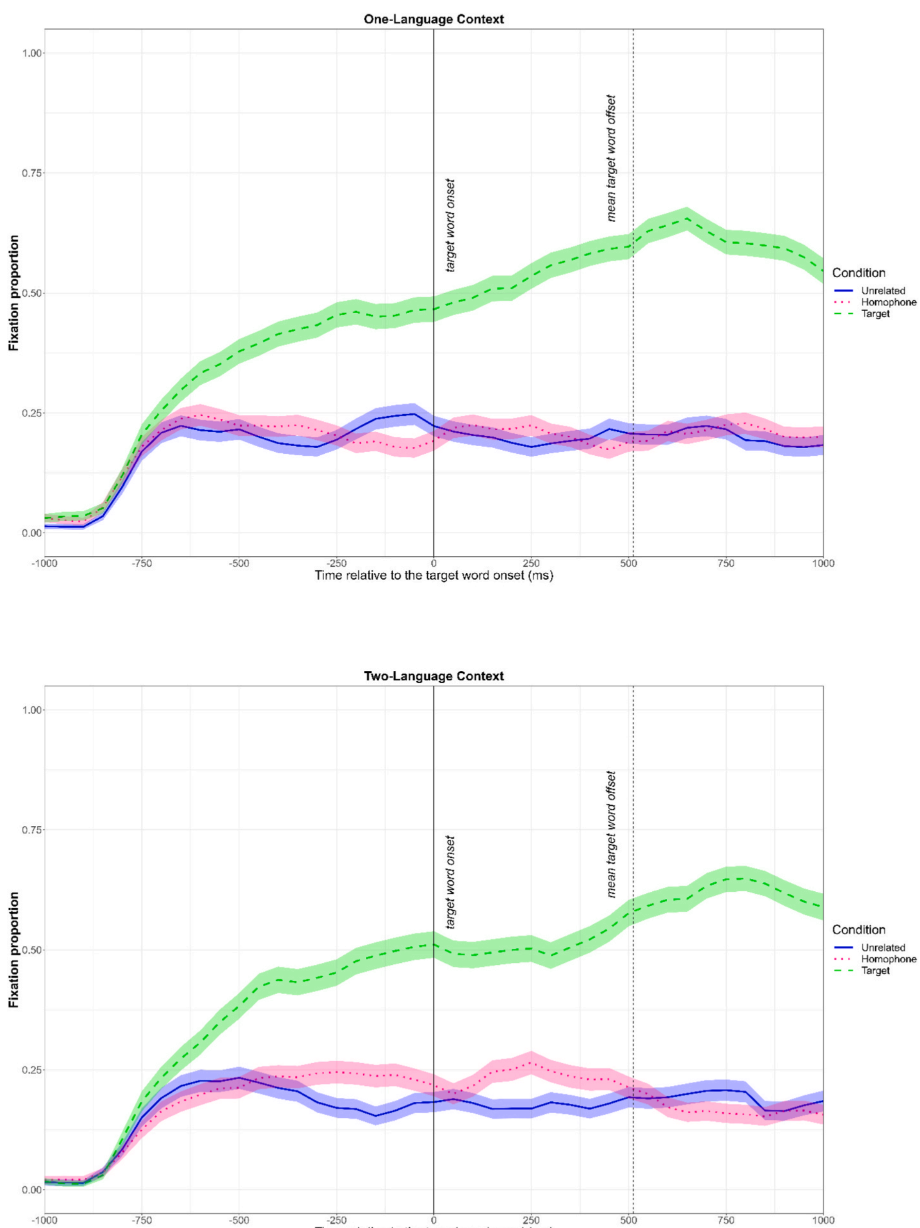

Fig. 2. *Mean Fixation Proportions on Critical Objects. Note.* Eye-tracking results for the one-language context in Experiment 1 (top) and two-language context in Experiment 2 (bottom). Time-course graphs show the mean fixation proportions on target, Mandarin homophone, and unrelated objects. Time 0 ms shows target word onset. The dashed vertical line (y = 512 ms) indicates the mean target word offset. Transparent thick lines are error bars representing standard errors.

fixation preference toward the target) over the null. In contrast, the BF_{10} for the homophone versus unrelated comparison remained below 1 (see Appendix 2 for exact values) both before and after the target word onset (except at −100 to 0 ms, where BF_{10} favored models for more fixations to the unrelated than to the homophone object), showing evidence for the null model (no fixation bias toward Mandarin homophones).

The cluster-based permutation analysis showed that target objects attracted more fixations than unrelated objects in the time window from −700 ms before to 1000 ms[1] after the target word onset (cluster-mass = 6917, $p < 0.001$). Mandarin homophone objects received more fixations than unrelated objects from −900 ms to −800 ms before the target word onset (cluster-mass = 26.4, $p < 0.001$), which, based on the visual plot (see Fig. 2) and the fact that it occurred very soon after picture presentation and before the target-object effect began, was interpreted as fixation noise. From time −100 ms to −50 ms, participants fixated more on unrelated objects than on homophone objects (cluster-mass = 11.8, $p = 0.045$).

We also analyzed the filler trials in order to examine whether there was any visual bias towards critical objects irrespective of the predictive contexts. As reported above, two items were removed. Both the generalized mixed-effects model analysis and the cluster-based permutation analysis showed no significant fixation bias toward target objects or Mandarin homophone objects than unrelated objects.

Discussion

Experiment 1 investigated whether bilinguals pre-activate L1 phonological representations of predictable words during L2 comprehension when only L2 was needed throughout the experiment (except for the familiarization session). Mandarin-English participants made predictive fixations on target objects more than unrelated objects before hearing the target words. This suggests that participants predicted the L2 target words or their meanings, aligning with prior research demonstrating L2 speakers' ability to predict target words or their meaning during L2 comprehension (e.g., Dijkgraaf et al., 2017; Ito, Pickering, et al., 2018). Crucially, however, they did not show more fixations on Mandarin homophone objects than unrelated objects. This suggests that while participants predicted the L2 target words or their meanings, they did not pre-activate L1 representations of predictable words in a one-language (L2) context.

It is also worth noting that participants showed no fixation difference between Mandarin homophone objects and unrelated objects even after hearing the target word. This is consistent with evidence suggesting that lexical access during bottom-up language comprehension is language selective (restricted to the target language) in a one-language context (e.g., Dijkstra et al., 1999; Hoversten & Traxler, 2020).

Given the evidence that cross-language activation is more likely to occur in two-language contexts than one-language contexts, the absence of L1 lexical pre-activation during L2 comprehension in Experiment 1 is not surprising. We now ask whether participants pre-activate L1 lexical representations of predictable words during L2 comprehension in a two-language context (where both languages are encountered)? To address this, Experiment 2 used the same experimental materials and procedure as Experiment 1 but now included two-thirds of fillers and all instruction and conversation in L1 Mandarin.

Experiment 2

In Experiment 2, we tested whether Mandarin-English bilinguals pre-activate Mandarin phonological representations of predictable words during English comprehension in a two-language context where both Mandarin and English were used.

Method

Participants

Thirty further participants from the same population as Experiment 1 (3 males and 29 females) aged 21–28 ($M = 23.37$ years, $SD = 1.47$) took part in this experiment. After the main experiment, participants were also asked to indicate their language history and English proficiency; see Table 1. There were no significant differences in language history or proficiency between the participants in Experiment 1 and Experiment 2 (all $p > .05$).

Stimuli and procedure

To ensure a two-language context, the experimenter greeted participants and talked to them only in Mandarin throughout the whole experiment; the instructions for the experiment were in Mandarin; we translated 2/3 of the filler sentences from Experiment 1 into Mandarin, with half of them containing a "YES" answer and the other half containing a "NO" answer. The remaining 1/3 filler sentences from Experiment 1 were still in English, with all of them containing a "YES" answer. Like the English experimental and filler sentences, the Mandarin filler sentences were generated using the Murf AI voice generator. The synthesized voice was set to "Mandarin-China-Zhang (Male), middle-aged, general tone, pitch (0 %), speed (−20 %)", resulting in a speech rate of 4.98 syllables per second. Two additional Mandarin-English bilinguals from the University of Edinburgh rated these Mandarin filler sentences as both comprehensible and natural. All audio files were downloaded in high-quality WAV format.

Overall, filler sentences mentioned one of the three distractor objects in the visual display 2/3 of the time, so together with the experimental sentences (which mentioned one of the four objects 1/3 of the time), 50 % of sentences mentioned an object in the visual scene. The experimental materials and other procedures remained the same as Experiment 1. The mean picture naming accuracy on the first presentation was 94.32 %.

Results

Comprehension question accuracy and eye-tracking data

The mean accuracy for the comprehension questions in the experimental trials was 97.93 % ($SD = 14.22$ %). The bottom part of Fig. 2 shows bilinguals' proportions of fixations on target objects, Mandarin homophone objects, and unrelated objects in a two-language context. The eye-tracking data were analyzed using the same procedures as in Experiment 1. Two items were excluded from the eye-tracking analyses in Experiment 2, because in the filler trials that used the same target and Mandarin homophone objects, those objects attracted more looks than the unrelated object within 1000 ms before the picture onset (see filler trial analysis below). This left 34 items for the analyses. We analyzed fixations on the critical objects for the filler sentences using the same procedure as in Experiment 1. After we removed two items, the analysis showed no difference in fixations toward the target objects, homophone objects, and the unrelated objects.

The binned generalized mixed-effects model showed that participants fixated target objects more than unrelated objects from −650 ms before they heard the target word until at least 1000 ms after the target word onset. Critically, participants also fixated more on Mandarin homophone objects than on unrelated objects from −250 ms before to −100 ms (three consecutive time bins) before the target word onset and from 150 ms to 300 ms (three consecutive time bins) after the target word onset.

The Bayesian generalized mixed-effects models supported these

[1] Note that the cluster-based results do not reveal the onset, offset, or duration of these effects (Matuschek et al., 2017). Rather they indicate a significant effect somewhere within these time windows (Ito & Knoeferle, 2023).

findings. Consistent with the results from the generalized mixed-effects models, the Bayesian generalized mixed-effects models showed that the BF_{10} for the contrast between target and unrelated objects exceeded 1 after −700 ms, 3 after −650 ms, and was consistently above 10 from −600 ms before target-word onset until recording finished at 1000 ms after target-word onset, therefore indicating strong evidence for a fixation preference toward the target. For the homophone vs. unrelated comparison (see Appendix 3 for BF_{10} values at all time bins), the BF_{10} exceeded 3 at − 250 ms ($BF_{10} = 4.54$) and − 200 ms ($BF_{10} = 4.72$), and remained above 1 until − 100 ms before target word onset. (A BF_{10} above 1 also occurred at 200 ms and 300 ms, and above 3 occurred at 250 ms after target word onset). These patterns suggest that participants showed a fixation bias toward the homophone object relative to the unrelated object before target word onset (−250 ms to − 100 ms), as well as after target-word onset (200 ms to 300 ms). But note that this homophone fixation bias was smaller, emerged later, and was shorter-lived than the robust target fixation bias.

Finally, the cluster-based permutation analysis further confirmed these effects. Target objects attracted more fixations than unrelated objects in the time window from −750 ms before to 1000 ms after the target word onset (cluster-mass = 6577, $p < 0.001$). Importantly, Mandarin homophone objects received more fixations than unrelated objects in the time window from −300 ms to −50 ms before the target word (cluster-mass = 61.3, $p < 0.001$), and also in the time window from 150 ms to 450 ms after the target word onset (cluster-mass = 51.5, $p < 0.001$).

We also analyzed the filler trials in order to examine whether there was any visual bias towards critical objects irrespective of the predictive contexts. As reported above, two items were removed. Both the generalized mixed-effects model analysis and the cluster-based permutation analysis showed no significant fixation bias toward target objects or Mandarin homophone objects than unrelated objects.

Discussion

Experiment 2 investigated whether bilinguals pre-activate L1 phonological representations of predictable words during L2 comprehension when both their L1 and L2 were used. We found that Mandarin-English participants fixated more on target objects than unrelated objects before hearing the target words. Critically, they were more likely to fixate on Mandarin homophone objects than unrelated objects, both before and after the target word onset. These findings suggest that in a two-language context, participants not only pre-activated the highly predictable target words (or their meaning), but also pre-activated the L1 translations of predictable words during L2 comprehension.

Comparison of Experiment 1 and 2

Experiment 1 and Experiment 2 tested whether bilingual speakers pre-activated L1 representations of predictable words during L2 comprehension in a one-language and two-language contexts, respectively. Participants showed a cross-language L1 prediction effect in Experiment 2 but not in Experiment 1, suggesting an effect of language context on cross-language prediction. To verify the role of language context, we combined data from both experiments and conducted a cluster-based permutation analysis from 1000 ms before to 1000 ms after the target word onset, with the effect of language context sum-coded and the condition dummy-coded, using the unrelated condition as the reference level.

The analysis showed that participants fixated more on the target objects than the unrelated objects in the time window from −800 ms before to at least 1000 ms after the target word onset, in both Experiment 1 and 2 (cluster-mass = 13303, $p < 0.001$). They also fixated more on the Mandarin homophone object than the unrelated object in the time windows from −350 to −200 ms (cluster-mass = 30.4, $p < 0.001$), and from 150 to 400 ms (cluster-mass = 55.9, $p < 0.001$). Crucially, in Experiment 2 (the two-language context) participants made significantly more fixations on the homophone objects than in Experiment 1 (the one-language context) during the –200 to –50 ms window before the target word onset (cluster-mass = 53.8, $p < 0.001$). The combined analysis suggests that bilinguals in a two-language context pre-activated L1 translations of predictable words to a larger extent than those in a one-language context.

General Discussion

We investigated whether L1 phonological representation of predictable words are pre-activated during L2 comprehension in a one-language context (Experiment 1, where only L2 English was used except for the familiarization session) and a two-language context (Experiment 2, where both English and Mandarin were used throughout). In both experiments, participants fixated more on target objects than on unrelated objects before the target word onset. In the two-language context, participants also fixated more on objects whose Mandarin names were homophones of the Mandarin translations of predictable words than unrelated objects, both before and after the target word onset. However, this effect was not observed in the one-language context.

Evidence for cross-language prediction during L2 comprehension

In the two-language context, participants made predictive fixations toward Mandarin homophone objects before the target word onset. This provides evidence that bilingual speakers can pre-activate L1 translation equivalents of predictable words during L2 comprehension.

Our findings contrast with those of Amos et al. (2022) and Ito et al. (2018), both of whom reported that L2 speakers did not make predictive eye movements toward L1 phonological competitors before the target word onset. Several methodological differences might explain the discrepancy. Compared with Ito et al. (2018), we used Chinese homophones that provided complete phonological overlap with L1 picture names (rather than partial overlap in Ito et al., 2018), included a picture familiarization phase that presented participants with both L1 (Mandarin) and L2 (English) names for each picture (whereas Ito et al., 2018 presented only L2 names), collected considerably more observations per condition, and manipulated the language context. Although any one of these factors could have contributed to the difference between our results and those of Ito et al. (2018), we consider language context the most compelling explanation. Our Experiment 1, which used a one-language context similar to Ito et al. (2018), produced no evidence of cross-language prediction. Crucially, cross-language prediction emerged only in Experiment 2, where we introduced additional L1 elements (i.e., the instructions, conversation, and two-thirds of the fillers), suggesting that the use of two-language contexts promoted predictive activation of L1.

However, the discrepancies between our findings and those of Amos et al. (2022) likely stem from different reasons. Like our study, Amos et al. (2022) presented bilingual participants with both L1 and L2 picture names during familiarization. More importantly, their participants were also placed in a two-language context by constantly interpreting or translating from L2 to L1. Yet, they found no evidence for the pre-activation of L1 translation equivalents of predictable words. This divergence may be due to the high cognitive load experienced by participants in their study. The interpreting task might have taxed participants so much that insufficient resources remained for L1 form prediction, considering that cognitive resources are crucial for prediction (Pickering & Gambi, 2018). Alternatively, the overlap between the L1 French phonological competitors and L1 French translations of predictable words in their study may have been too small to affect participants' eye movements, even if some pre-activation of L1 translations did occur. It is also possible that their experiment had too few observations per condition to detect any effect.

Note that participants' greater fixations to target than to unrelated objects before target word onset indicate prediction of the L2 target words or their meanings, but do not, by themselves, demonstrate pre-activation of L2 phonological forms. Our finding that L2 speakers pre-activate the L1 phonological representations is, therefore, compatible with the possibility that previous failures to observe L2 form prediction (Amos et al., 2022; FitzPatrick & Indefrey, 2010; Ito et al., 2018) may partly reflect interference from L1. However, because L1 form co-activation in our data occurred only in the two-language context but not in a one-language context, L1 interference cannot fully account for the absence of L2 form prediction in purely one-language studies (e.g., Ito et al., 2018). An additional explanation is that pre-activating L2 form representations is especially resource-demanding: reduced automaticity in L2 comprehension and production may leave fewer cognitive resources available for L2 form prediction (Ito & Pickering, 2021). But as our experiments did not include an L2 phonological competitor (as in Ito et al. 2018), we cannot be sure whether L2 phonological forms are pre-activated. Future work could include both L1 and L2 phonological competitors and manipulate context and cognitive load to establish whether L2 form prediction occurs, and if not, whether its absence reflects L1 interference or resource limitations.

Cross-language prediction and language context

Our study revealed a clear influence of language context on predictive processing: Participants made predictive fixations toward L1 Mandarin homophone objects only in a two-language context, but not in a one-language setting. This strongly suggests that phonological activation during predictive processing is highly sensitive to the language context. Bilinguals in two-language contexts appear to activate representations of predictable words in both languages, whereas in one-language contexts, they restrict activation to the target language. This prediction pattern is consistent with evidence that lexical access is more language non-selective in two- than one-language contexts, across both comprehension and production.

The similarity in cross-language activation patterns between prediction and production suggests that these processes may share mechanisms governing how lexical access is modulated by language context. This pattern is broadly consistent with the prediction-by-production hypothesis, which proposes that prediction during comprehension engages the production system (Dell & Chang, 2014; Pickering & Gambi, 2018; Pickering & Garrod, 2013). On this account, comprehenders use their production systems to pre-activate representations of predictable word in the order of meaning, syntax, and phonology, much like in production, but without articulation. If lexical selectivity in production is context-dependent, and prediction recruits the production system, then lexical selectivity during prediction would likewise be expected to vary with language context—a pattern compatible with our results. We stress, however, that this interpretation is tentative, as our data do not directly demonstrate production system involvement.

So, how exactly does language context influence lexical access in prediction? In our study, the two-language context in which L1 filler sentences were included, participants encountered L1 and L2 sentences apparently randomly, creating uncertainty about the language of upcoming words. This led to non-selective predictions, with participants activating L1 representations of predictable words while using their production system for prediction. In contrast, in the one-language context, where only L2 sentences were included, participants were quite certain that they would only hear L2 sentences, leading them to expect that the upcoming predictable word would be in L2. This language-specific certainty led to selective predictions, with the prediction system activating only L2 representations, thus avoiding the co-activation of L1 representations.

Cross-language lexical prediction mechanism

Although we found evidence that bilinguals pre-activate L1 representations of predictable words during L2 comprehension, it remains unclear whether they directly pre-activated L1 lexical representations of predictable words, or they first pre-activated L2 lexical representations, which then activated their L1 translation equivalents. If bilinguals directly pre-activate L1 lexical representations, potentially alongside L2 representations, this would suggest that predictive processing is inherently language-nonselective. In this case, bilinguals make predictions in both their L1 and L2 simultaneously. Within the framework of the prediction-by-production account (Pickering & Gambi, 2018), bilingual speakers would use their production systems to pre-activate lexical forms in both languages, mirroring the lexical access process in bilingual speech production. In contrast, if L2 lexical representations are pre-activated first and L1 representations are subsequently activated (either via shared semantic links or direct translation connections), this would suggest that predictive processing is initially language-selective. If so, L1 activation would reflect post-lexical co-activation, a by-product rather than a core component of prediction.

A third possibility is that bilinguals initially pre-activate L1 lexical representations, which then activate their L2 translations. This route may be favored when L1 words are more easily accessible or less cognitively demanding to retrieve, which may be especially relevant when L2 lexical prediction is resource-intensive. On this account, pre-activating L1 words may serve as a cognitive shortcut to facilitate L2 comprehension. Future research might distinguish among these possibilities by comparing the time course of fixations to L1 and L2 phonological competitors. If bilinguals directly pre-activate L1 lexical representations, fixations on the L1 competitor might emerge earlier than those on the L2 competitor, or L2 fixations may be absent altogether. If L1 is activated via L2, however, fixations on the L1 competitor might emerge later than those on the L2 competitor.

Conclusion

We reported two visual-world eye-tracking experiments that investigated whether highly proficient L2 speakers pre-activate L1 translation equivalents of predictable words during L2 comprehension. We found that bilingual speakers pre-activated L1 phonological representations of predictable words during L2 comprehension, but such cross-language lexical predictions only occurred in a two-language context where both L1 and L2 were relevant for the task, but not in a one-language context where only L2 was needed for the task. These findings suggest that whether lexical predictions are language-selective depends on language context.

CRediT authorship contribution statement

Huanhuan Yin: Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Martin J. Pickering:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Funding

This work was supported by the Research Support Grants for PPLS PhD students at the University of Edinburgh.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2025.104701.

Data availability

The code and data associated with this paper are available from https://osf.io/4w6hy/.

References

Abutalebi, J., Annoni, J.-M., Zimine, I., Pegna, A. J., Seghier, M. L., Lee-Jahnke, H., Lazeyras, F., Cappa, S. F., & Khateb, A. (2008). Language control and lexical competition in bilinguals: An event-related fMRI study. *Cerebral Cortex, 18*(7), 1496–1505. https://doi.org/10.1093/cercor/bhm182

Alemán Bañón, J., & Martin, C. (2021). The role of crosslinguistic differences in second language anticipatory processing: An event-related potentials study. *Neuropsychologia, 155*, Article 107797. https://doi.org/10.1016/j.neuropsychologia.2021.107797

Alemán Bañón, J., & Martin, C. D. (2024). Lexicosemantic prediction in native speakers of English and Swedish-speaking learners of english: An event-related potential study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 50*(12), 1982–2007. https://doi.org/10.1037/xlm0001421

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Amos, R. M., Seeber, K. G., & Pickering, M. J. (2022). Prediction during simultaneous interpreting: Evidence from the visual-world paradigm. *Cognition, 220*, Article 104987. https://doi.org/10.1016/j.cognition.2021.104987

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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

Cheng, Y.-L., & Howard, D. (2008). The time cost of mixed-language processing: An investigation. *International Journal of Bilingualism, 12*(3), 209–222. https://doi.org/10.1177/1367006908098568

Chun, E., & Kaan, E. (2019). L2 Prediction during complex sentence processing. *Journal of Cultural Cognitive Science, 3*(2), 203–216. https://doi.org/10.1007/s41809-019-00038-0

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283.

Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language, 94*(1), 94–103. https://doi.org/10.1016/j.bandl.2004.12.002

Declerck, M., & Grainger, J. (2017). Inducing asymmetrical switch costs in bilingual language comprehension by language practice. *Acta Psychologica, 178*, 100–106. https://doi.org/10.1016/j.actpsy.2017.06.002

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), Article 20120394. https://doi.org/10.1098/rstb.2012.0394

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), Article 8. https://doi.org/10.1038/nn1504

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition, 20*(5), 917–930. https://doi.org/10.1017/S1366728916000547

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2019). Prediction and integration of semantics during L2 and L1 listening. *Language, Cognition and Neuroscience, 34*(7), 881–900. https://doi.org/10.1080/23273798.2019.1591469

Dijkstra, T., Bruijn, E. D., Schriefers, H., & Brinke, S. T. (2000). More on interlingual homograph recognition: Language intermixing versus explicitness of instruction. *Bilingualism: Language and Cognition, 3*(1), 69–78. https://doi.org/10.1017/S1366728900000146

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of Cognates and Interlingual Homographs: The Neglected Role of Phonology. *Journal of Memory and Language, 41*(4), 496–518. https://doi.org/10.1006/jmla.1999.2654

Dijkstra, T., Hell, J. G. V., & Brenders, P. (2015). Sentence context effects in bilingual word recognition: Cognate status, sentence language, and semantic constraint. *Bilingualism: Language and Cognition, 18*(4), 597–613. https://doi.org/10.1017/S1366728914000388

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42*(4), 445–464. https://doi.org/10.1006/jmla.1999.2697

Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition, 1*(1), 51–66.

Dussias, P. E., Kroff, J. R. V., Tamargo, R. E. G., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition, 35*(2), 353–387. https://doi.org/10.1017/S0272263112000915

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 663–679. https://doi.org/10.1037/0278-7393.33.4.663

Elston-Güttler, K. E., & Gunter, T. C. (2009). Fine-tuned: Phonology and Semantics Affect First- to Second-language Zooming In. *Journal of Cognitive Neuroscience, 21*(1), 180–196. https://doi.org/10.1162/jocn.2009.21015

Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2005). Zooming into L2: Global language context and adjustment affect processing of interlingual homographs in sentences. *Cognitive Brain Research, 25*(1), 57–70. https://doi.org/10.1016/j.cogbrainres.2005.04.007

Fang, S., & Wu, Z. (2024). Syntactic prediction in L2 learners: Evidence from English disjunction processing. *International Review of Applied Linguistics in Language Teaching, 62*(2), 429–456. https://doi.org/10.1515/iral-2021-0223

FitzPatrick, I., & Indefrey, P. (2010). Lexical competition in nonnative speech comprehension. *Journal of Cognitive Neuroscience, 22*(6), 1165–1178. https://doi.org/10.1162/jocn.2009.21301

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural correlates of semantic prediction and resolution in sentence processing. *Journal of Neuroscience, 37*(18), 4848–4858.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1*(2), 131–149. https://doi.org/10.1017/S136672899800025X

Grosjean, F. (2001). The bilingual's language modes. In J. L. Nicol (Ed.), *One Mind, Two Languages: Bilingual Language Processing, Malden*. USA: Blackwell. S.

Grosjean, F. (2008). *Studying Bilinguals*. Oxford University Press.

Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: FMRI evidence from Chinese-English bilinguals. *NeuroImage, 56*(4), 2300–2309. https://doi.org/10.1016/j.neuroimage.2011.03.049

Hochberg, Y. (1987). Multiple comparison procedures. *Wiley Series in Probability and Statistics*.

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching, 53*(3), 277–306. https://doi.org/10.1515/iral-2015-0014

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition, 40*(1), 171–199.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition, 106*(1), 501–511. https://doi.org/10.1016/j.cognition.2007.02.001

Hoversten, L. J., & Martin, C. D. (2023). Parafoveal processing in bilingual readers: Semantic access within but not across languages. *Journal of Experimental Psychology: Human Perception and Performance, 49*(12), 1564–1578. https://doi.org/10.1037/xhp0001161

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*, Article 104118. https://doi.org/10.1016/j.cognition.2019.104118

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Ito, A. (2024). Phonological prediction during comprehension: A review and meta-analysis of visual-world eye-tracking studies. *Journal of Memory and Language, 139*, Article 104553. https://doi.org/10.1016/j.jml.2024.104553

Ito, A., Bautista, A., & Martin, C. (2025). Cognate facilitation effect on verb-based semantic prediction in L2 is modulated by L2 proficiency. *Bilingualism: Language and Cognition, 28*(3), 684–696. https://doi.org/10.1017/S1366728924000968

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods, 55*(7), 3461–3493. https://doi.org/10.3758/s13428-022-01969-3

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(4), 635–652. https://doi.org/10.1037/xlm0000315

Ito, A., Nguyen, H. T. T., & Knoeferle, P. (2024). German-dominant Vietnamese heritage speakers use semantic constraints of German for anticipation during comprehension in Vietnamese. *Bilingualism: Language and Cognition, 27*(1), 57–74. https://doi.org/10.1017/S136672892300041X

Ito, A., & Pickering, M. (2021). *Automaticity and prediction in non-native language comprehension.*. https://doi.org/10.1075/bpa.12.02ito

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different?*. *Linguistic Approaches to Bilingualism, 4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa

Kaan, E., & Grüter, T. (2021). Prediction in second language processing and learning: Advances and directions. In Prediction in second language processing and learning (pp. 2–24). John Benjamins.

Kheder, S., & Kaan, E. (2019). Lexical selection, cross-language interaction, and switch costs in habitually codeswitching bilinguals. *Bilingualism: Language and Cognition, 22*(3), 569–589. https://doi.org/10.1017/S1366728918000500

Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(11), 2153. https://doi.org/10.1037/xlm0000935

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Lee, M. D., & Wagenmakers, E.-J. (2014). Bayesian cognitive modeling: A practical course. Cambridge university press.

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 12–31. https://doi.org/10.1037/0278-7393.34.1.12

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language, 63*(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003

Li, X., Li, X., & Qu, Q. (2022). Predicting phonology in language comprehension: Evidence from the visual world eye-tracking task in Mandarin Chinese. *Journal of Experimental Psychology: Human Perception and Performance, 48*, 531–547. https://doi.org/10.1037/xhp0000999

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*(4), 574–588. https://doi.org/10.1016/j.jml.2013.08.001

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315.

Mitsugi, S., & Macwhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition, 19*(1), 19–35. https://doi.org/10.1017/S1366728914000881

Momenian, M., Vaghefi, M., Sadeghi, H., Momtazi, S., & Meyer, L. (2024). Language prediction in monolingual and bilingual speakers: An EEG study. *Scientific Reports, 14*(1), 6818. https://doi.org/10.1038/s41598-024-57426-y

Olson, D. J. (2016). The gradient effect of context on language switching and lexical access in bilingual production. *Applied Psycholinguistics, 37*(3), 725–756. https://doi.org/10.1017/S0142716415000223

Olson, D. J. (2017). Bilingual language switching costs in auditory comprehension. *Language, Cognition and Neuroscience, 32*(4), 494–513. https://doi.org/10.1080/23273798.2016.1250927

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*, 1002–1044. https://doi.org/10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Poort, E. D., & Rodd, J. M. (2017). The cognate facilitation effect in bilingual lexical decision is influenced by stimulus list composition. *Acta Psychologica, 180*, 52–63. https://doi.org/10.1016/j.actpsy.2017.08.008

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372.

Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition, 26*(2), 253–267. https://doi.org/10.1017/S1366728922000736

Schlenter, J., & Felser, C. (2021). Chapter 3. Second language prediction ability across different linguistic domains: Evidence from German. In Prediction in second language processing and learning (pp. 47–68). John Benjamins Publishing Company.

Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 29*(6), 1155–1178. https://doi.org/10.1037/0096-1523.29.6.1155

Smits, E., Martensen, H., Dijkstra, T., & Sandra, D. (2006). Naming interlingual homographs: Variable competition and the role of the decision system. *Bilingualism: Language and Cognition, 9*(3), 281–297. https://doi.org/10.1017/S136672890600263X

Thierry, G., & Wu, Y. J. (2007). From the Cover: Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 104*(30), 12530. https://doi.org/10.1073/pnas.0609927104

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1412–1431. https://doi.org/10.1037/a0024492

van Bergen, G., & Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language, 92*, 26–42. https://doi.org/10.1016/j.jml.2016.05.003

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review, 9*(4), 780–789. https://doi.org/10.3758/BF03196335

Vanlangendonck, F., Peeters, D., Rueschemeyer, S.-A., & Dijkstra, T. (2020). Mixing the stimulus list in bilingual lexical decision turns cognate facilitation effects into mirrored inhibition effects. *Bilingualism: Language and Cognition, 23*(4), 836–844. https://doi.org/10.1017/S1366728919000531

Wang, X., Wang, J., & Malins, J. G. (2017). Do you hear 'feather' when listening to 'rain'? Lexical tone activation during unconscious translation: Evidence from Mandarin-English bilinguals. *Cognition, 169*, 15–24. https://doi.org/10.1016/j.cognition.2017.07.013

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*(7), 1272–1288. https://doi.org/10.1162/0898929041920487

Wu, Y. J., & Thierry, G. (2010). Chinese–english bilinguals reading english hear chinese. *Journal of Neuroscience, 30*(22), 7646–7651.

Xu, X., Ji, C., Li, T., & Pickering, M. J. (2024). The prediction of segmental and tonal information in Mandarin Chinese: An eye-tracking investigation. *Language, Cognition and Neuroscience*, 1–15. https://doi.org/10.1080/23273798.2024.2395549

Zhang, T., van Heuven, W. J. B., & Conklin, K. (2011). Fast Automatic Translation and Morphological Decomposition in Chinese-English Bilinguals. *Psychological Science, 22*(10), 1237–1242. https://doi.org/10.1177/0956797611421492

Zhao, Z., Ding, J., Wang, J., Chen, Y., & Li, X. (2024). The flexibility and representational nature of phonological prediction in listening comprehension: Evidence from the visual world paradigm. *Language and Cognition, 16*(2), 481–504. https://doi.org/10.1017/langcog.2023.38