Investigating Phonological and Semantic Cues in Sentence Processing and Prediction among Mandarin Chinese EFL Learners of Different Proficiency Levels: A Visual World Eye-Tracking Study

Qingyang Tang

Master of Science in Applied Linguistics and Second Language Acquisition, 2025

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	Qingyang Tang
Date:	19/08/2025

Investigating Phonological and Semantic Cues in Sentence Processing and Prediction among Mandarin Chinese EFL Learners of Different Proficiency Levels: A Visual World Eye-Tracking Study

University of Oxford

Qingyang Tang

Word Count: 16,227

Thesis submitted in partial fulfilment of the requirements for the Degree of
Master of Science in Applied Linguistics and Second Language Acquisition

August 2025

Table of Contents

Abstract

Predictive mechanisms are central to real-time language comprehension, yet it remains debated how L2 proficiency shapes the use of semantic and phonological cues, and whether L1 phonology is co-activated during L2 processing. This study investigated the time course of prediction in Mandarin Chinese learners of English at two proficiency levels (advanced, intermediate) compared with native English speakers, using a visual world eye-tracking paradigm with cloze sentences omitting the target noun. Sixty-one participants (advanced L2: n = 22; intermediate L2: n = 18; L1 English: n = 21) viewed four-picture arrays that always included one of the following: a target, an English semantic competitor, an English phonological competitor, or a Chinese phonological competitor, with the remaining pictures serving as distractors, while hearing each sentence once. Webcam-based eye-tracking (Gorilla) incorporated automated polling, calibration, and quality control. Across groups, anticipatory looks to the target emerged well before the cloze position, confirming the prediction. Native listeners showed robust, sustained semantic activation; advanced L2 listeners also activated semantic alternatives, but with delayed and more variable timing. English phonological competitors generally attracted low early fixations, though subtle and delayed activation was observed in the L1 and advanced L2 groups. Notably, advanced L2 participants showed a late increase in fixations to the Chinese phonological competitor, surpassing intermediate L2 listeners in the predictive window, which provides evidence for the transient co-activation of L1 phonology during L2 processing among highly proficient bilinguals. Intermediate L2 listeners focused almost exclusively on the target, with no reliable engagement of semantic or phonological competitors. These findings support a graded, proficiency-dependent account of prediction, clarifying when (late vs. early) and for whom (advanced vs. intermediate learners) L1 phonology intrudes during L2 prediction. The results also have implications for bilingual processing models that aim to promote anticipatory, meaning-based comprehension in advanced learners.

1. Introduction

Contemporary theories of language processing increasingly highlight the importance of prediction in guiding comprehension (e.g., Dell & Chang, 2014; Pickering & Garrod, 2013). Within this framework, prediction is considered a core mechanism underlying sentence comprehension. It facilitates incremental processing, in which language is interpreted sequentially (i.e., word by word) as it is encountered (Altmann & Mirkovic, 2009; Kamide, 2008; Levy, 2008).

A variety of experimental methods—most notably visual world eye-tracking, event-related potentials (ERPs), and electroencephalography (EEG)—have been employed to examine predictive processing in real time. These techniques offer direct insight into how predictive mechanisms function during language comprehension. Early investigations into sentence processing primarily focused on whether verb information, when combined with the grammatical subject, could generate anticipatory eye movements. For instance, in Altmann and Kamide's (1999) seminal study, participants viewed a scene featuring a boy, a toy car, a toy train, and a cake while listening to sentences such as "*The boy will eat the cake*." Upon hearing the verb eat, participants shifted their gaze towards the edible object (*cake*) before it was mentioned, suggesting that verb semantics facilitated anticipation of the upcoming word. Moreover, Kamide et al. (2003a) found that participants were more likely to fixate on a motorbike after hearing "The man will ride…" and on a carousel after "The girl will ride…", indicating that subject-verb combinations facilitate predictions about upcoming content. This predictive effect was further supported by research on verb semantics (e.g., Altmann & Kamide, 1999; Kamide et al., 2003b), which demonstrated that the selectional restrictions imposed by verbs help listeners anticipate the next element in a sentence. Additionally, Altmann and Kamide (2007) revealed the importance of tense in referent resolution: participants looked at an empty wine glass when hearing "The man has drunk…" but directed their gaze towards a full glass of beer upon hearing "The man will drink…". Supporting this behavioural evidence, EEG research has revealed that the predictability of an upcoming word influences neural activity, such as modulations in oscillatory patterns before the target word onset (e.g., Molinaro et al., 2013). Likewise, ERP studies have shown that grammatical cues—such as determiners and gender markers—can enhance the predictability of upcoming words by triggering early predictive responses (e.g., DeLong et al., 2005; Rommers et al., 2017).

These findings converge on the view that native speakers construct meaning actively by generating expectations, rather than passively integrating information (e.g., DeLong et al., 2005; Laszlo & Federmeier, 2009; Staub & Clifton, 2006). A growing body of evidence from eye-tracking and ERP research suggests that first language (L1) speakers are capable of making fine-grained predictions across multiple linguistic levels. These include not only grammatical gender (Van Berkum et al., 2005) but also syntactic categories (Lau et al., 2006), object-gap positions in wh-dependencies (e.g., Dickey & Thompson, 2009), and the phonological form of anticipated words (DeLong et al., 2005). This capacity for prediction involves the dynamic integration of linguistic and visual cues to anticipate upcoming input. Within this context, the visual world paradigm has proven particularly effective, as participants' eye movements provide real-time insights into the evolving mental representations that underlie sentence processing and predictive inference (Huettig et al., 2011).

In contrast, while native speakers readily draw on contextual cues to predict upcoming linguistic input, second language (L2) learners typically exhibit more limited predictive processing, particularly in their use of phonological and syntactic cues (e.g., Grüter & Rohde, 2013; Ito et al., 2018; Lew-Williams & Fernald, 2010). This discrepancy has attracted considerable attention in psycholinguistics, given its implications for understanding how real-time processing differs between L1 and L2 users (e.g., Dussias et al., 2013; Martin et al., 2013). A common method for exploring these differences involves comparing L2 learners from varied L1 backgrounds to examine whether typological differences influence predictive abilities (e.g., Grüter et al., 2014; Marinis et al., 2005). Such cross-linguistic comparisons are intended to evaluate the extent to which L1 characteristics influence L2 sentence processing. While some studies have found clear distinctions between native and non-native speakers in their predictive use of contextual cues (e.g., Marinis et al., 2005), others have shown that highly proficient L2 learners can exhibit native-like prediction (e.g., Dussias et al., 2013; Hopp, 2013). These inconsistent findings suggest that the contrast between L1 and L2 processing is not absolute and may be modulated by other factors, such as proficiency or task demands. Moreover, several studies have shown that L2 speakers may pre-activate the phonological forms of their L1 translation equivalents when anticipating upcoming words (e.g. Marian & Spivey, 2003; Mishra & Singh, 2014; Spivey & Marian, 1999). Thus, the precise mechanisms of L2 sentence processing—and the extent to which predictive strategies differ between L1 and L2 speakers—remain to be fully clarified.

Although recent research has expanded our understanding of L2 predictive processing, proficiency in the target language—despite its acknowledged importance—has often been overlooked or underexplored in this context. As Schlenter (2023) observes, much of the existing literature has concentrated on cross-linguistic comparisons, often overlooking how variations in proficiency within a single L2 group influence sentence processing and prediction. This oversight is particularly significant given the unique challenges faced by L2 users, such as difficulties inhibiting L1 activation (Thierry & Wu, 2007). Kaan et al. (2010) argue that such challenges may diminish as learners become more proficient. Nevertheless, much of the existing literature continues to treat L2 speakers as a homogeneous population, restricting our understanding of how individual differences in language background and proficiency shape predictive processing (e.g., Koch et al., 2021). Consequently, there is a lack of clarity regarding how proficiency levels interact with cognitive resources to influence predictive strategies in L2 comprehension (Huettig, 2015). Addressing this gap requires approaches that account for variability in linguistic experience and domain-specific expertise, thereby enabling a more nuanced characterisation of the mechanisms underlying prediction across diverse bilingual profiles.

The current study seeks to address these gaps by examining the influence of L2 proficiency on predictive mechanisms during real-time sentence processing. Beyond comparing native and non-native speakers, the study investigates whether increased proficiency results in more native-like use of predictive strategies. Specifically, it explores how English language proficiency affects the use of semantic and phonological cues during auditory sentence processing among native Mandarin Chinese speakers. Advanced learners of English as a foreign language (EFL) are compared to native English speakers to assess how they differ in their reliance on predictive cues. Importantly, the study distinguishes between intermediate and advanced EFL learners based on their linguistic background and exposure. All advanced learners are either studying English translation and interpreting or are currently working as translators or interpreters. They possess longer durations of English study and, crucially, have an immersive English learning experience, having lived and studied in English-speaking countries. In contrast, intermediate learners are enrolled in English-unrelated academic programmes, have had more limited exposure to English, and have never studied or lived in an English-speaking environment. This distinction allows for a more nuanced analysis of how immersive experience and professional engagement with English influence predictive processing. Moreover, previous research has yielded mixed evidence on whether L2 learners activate the phonological representations of L1 translation equivalents when anticipating upcoming words (e.g. Ito et al., 2018; Marian & Spivey, 2003; Mishra & Singh, 2014; Spivey & Marian, 1999). To investigate this, the present study includes objects whose L1 labels are phonologically related to the L1 translation of the predictable English target. This allows us to determine whether such cross-linguistic phonological activation occurs, and whether it differs between intermediate and advanced EFL learners. Ultimately, the findings aim to clarify how predictive processing and cue integration are shaped by L2 proficiency, offering insights into the mechanisms underlying sentence comprehension in non-native speakers.

The dissertation is organised as follows. Chapter 2 reviews literature on sentence processing and predictive mechanisms in both L1 and L2 speakers, with particular attention to semantic and phonological cues examined through the visual world paradigm. Chapter 3 details the methodology, including participant characteristics, stimulus materials, experimental design, and procedures. Chapter 4 reports the results of the visual world eye-tracking experiment, presenting descriptive statistics, inferential analyses, and time-course visualisations. Chapter 5 discusses these findings in relation to existing theories and empirical evidence, outlining implications for L2 predictive processing, acknowledging limitations, and proposing avenues for future research. Chapter 6 concludes by summarising the study’s principal findings and contributions.

2. Literature Review

2.1 Methodological Foundations of the Visual World Eye Tracking

The visual world paradigm is characterised by a relatively straightforward design, in which participants listen to spoken utterances while viewing a visual display, with their eye movements tracked and later analysed (Huettig et al., 2011). One of the earliest applications of this method was conducted by Cooper (1974), who asked participants to listen to brief narratives while observing displays of familiar objects. Crucially, participants were only informed that their pupil size would be measured and were free to look wherever they wished. Despite this lack of specific instruction, participants consistently fixated on objects that were either directly mentioned or semantically related to the spoken content. For instance, upon hearing "my scatter-brained dog Scotty…", they were more likely to look at an image of a dog, and when hearing "during a photographic safari…", their gaze frequently shifted to a camera.

These eye movements were found to be closely aligned with the timing of the speech, with over 90% of fixations on target objects occurring during the spoken word or within 200 milliseconds of its offset. Based on these findings, Cooper proposed that eye-tracking could serve as a powerful method for exploring perceptual and cognitive processes in real time, particularly those involved in speech perception, memory, and language comprehension. However, his work did not gain immediate traction, largely due to the technical challenges and labour-intensive nature of eye-tracking at the time. It was not until the influential study by Tanenhaus et al. (1995) that the methodology gained broader recognition. Their research provided compelling evidence for a systematic relationship between eye movements and spoken language processing, thereby revitalising interest in eye-tracking and establishing it as a central tool in psycholinguistic research.

2.1.1 Task Design and Display Types in Visual World Paradigm

The visual world paradigm is a widely used method for examining the interaction between spoken language and visual attention. In its common form, participants are presented with scenes containing both objects explicitly mentioned in the spoken utterance and unrelated distractors. These displays can vary in complexity. Semi-realistic arrangements—such as objects spatially organised on a workspace (e.g., Altmann & Kamide, 1999; Hopp, 2015; Kamide et al., 2003b) or positioned in the four corners of a computer screen (e.g., Allopenna et al., 1998; Dijkgraaf et al., 2019; Huettig & McQueen, 2007; Ito & Husband, 2017; Ito et al., 2018)—allow researchers to explore how comprehension is shaped by perceptual cues and listeners' background knowledge (Henderson & Ferreira, 2004). For example, in "The boy will eat the cake" (Altmann & Kamide, 1999), a display may depict both the boy and the cake alongside unrelated items. By contrast, simplified object arrays that minimise contextual richness are more suited to isolating conceptual and lexical activation processes at the word level (Huettig et al., 2011).

A substantial body of research has used the visual world paradigm to investigate whether attention is drawn to objects related to the target word in phonological, semantic, or visual terms (e.g., Huettig & Altmann, 2005; Huettig & McQueen, 2007; Ito et al., 2018; Yee & Sedivy, 2006). One particularly informative approach involves target-absent displays, in which the directly referenced object is omitted from the scene. This design enhances the detection of competition effects by encouraging reliance on associative relationships rather than direct referential matching. For instance, Huettig and Altmann (2005) showed that when the object "piano" was absent, participants looked towards semantically related items such as a trumpet. Removing the direct target heightens the salience of competitors, making it easier to capture subtle predictive or competitive effects in gaze behaviour. This is particularly advantageous when studying the temporal dynamics of prediction, as it prevents immediate fixation on the target—an effect that can obscure early activation patterns when targets and competitors are presented together. Consequently, non-target conditions allow for a clearer assessment of whether phonological or semantic activation emerges before or after the onset of the target word or cloze item.

The present study builds on this logic to examine how L2 speakers use semantic and phonological cues during auditory sentence processing, with a particular focus on predictive contexts. Cloze sentence structures offer a robust methodological and theoretical framework for this purpose, as they allow precise manipulation of contextual constraint and lexical predictability—two key variables in studying predictive mechanisms. Evidence from sentence completion studies shows that both factors influence processing speed; for example, Staub et al. (2015) found that highly predictable contexts facilitate more rapid lexical access. Such predictability can be effectively incorporated into visual world eye-tracking paradigms, aligning fixation patterns with high- or low-cloze sentence continuations.

ERP research further supports the utility of cloze-based designs. Freunberger and Roehm (2017) demonstrated that high-cloze completions not only reduce semantic processing demands, as reflected in attenuated N400 amplitudes, but also elicit late positivity components, indicative of successful prediction confirmation. Together, these behavioural and neurophysiological findings underscore the value of cloze-based designs for capturing the temporal dynamics of prediction and integration during sentence processing. Building on this evidence, the present study applies high-constraint cloze sentences to the investigation of L2 predictive processing. This allows for fine-grained comparisons of eye movement patterns across language groups, providing a sensitive measure of how L2 proficiency modulates the use of semantic and phonological information. By embedding semantically and phonologically related competitors and distractors within visually controlled displays, cloze-based eye-tracking studies can therefore yield critical insights into both the timing and underlying mechanisms of prediction in L2 users.

2.2 Predictive Mechanisms in L1 and L2

2.2.1 Prediction of Semantic and Phonological Information in L1

Altmann and colleagues argue that predictive processing depends on how closely elements in the visual context match phonetic, semantic, or thematic features represented within an "unfolding mental world" shaped by both language and context (e.g., Altmann & Kamide, 2007; Altmann & Mirković, 2009). This view aligns with the constraint-based framework, which explains anticipation as the product of a forward-looking system integrating multiple sources—such as visual and sentence context, verb-specific thematic constraints, and other cues—to predict upcoming linguistic input (Spivey & Tanenhaus, 1998). Within this framework, semantic and phonological cues are key information sources, processed in parallel, measurable through normed materials (e.g., cloze sentences), and shown to produce reliable time-locked effects in eye movement patterns (Kukona et al., 2011).

Extensive research indicates that eye movements guided by language are sensitive to the semantic connections between visual objects and concurrently spoken words (Huettig & McQueen, 2007). In the typical experimental setting, participants viewed arrays of four common images while listening to spoken sentences. Huettig and Altmann (2005) found that when a semantically related but non-associated word (e.g. "piano") was heard, participants' gaze shifted to the image of a related object (e.g. "trumpet"). Comparable results have since been obtained in multiple studies aiming at L1 sentence processing (e.g. Altmann & Kamide, 1999; Huettig et al., 2006; Ito & Husband, 2017; Kamide et al., 2003a; Yee & Sedivy, 2006). These findings point out that robust semantic activation during spoken-word recognition, together with the unfolding context, can directly guide gaze behaviour in the visual world paradigm among native speakers.

Evidence for phonological prediction during sentence processing in native speakers remains mixed (e.g., DeLong et al., 2005; Huettig & McQueen, 2007; Ito & Husband, 2017; Ito et al., 2018; Nieuwland et al., 2018). Compared with semantic prediction, phonological prediction appears less robust, typically emerging only under conditions such as artificially slowed speech (Ito & Husband, 2017). Such findings suggest that phonological prediction may follow a longer time course. However, it is still unclear whether this reflects an inherently weaker predictive mechanism or merely a delayed onset relative to semantic prediction. Indeed, studies have demonstrated substantial variability in the presence, strength, and timing of phonological effects, indicating that they are highly sensitive to task design, stimulus characteristics, and contextual constraints. For example, Huettig and McQueen (2007) showed that phonological activation can precede semantic activation when visual context is available at sentence onset, but that this advantage disappears when the visual display precedes the critical word by even a short interval. These findings suggest that phonological activation in L1 speakers can be fast but fragile, easily modulated by competing sources of information such as visual-shape or semantic matches. This stands in contrast to semantic prediction, which tends to be more stable and consistently observed across paradigms. As a result, the literature on L1 phonological prediction reflects a high degree of flexibility in outcomes, with effects emerging robustly only under particular temporal and contextual conditions, stressing the need for caution when generalising across studies.

2.2.2 Prediction of Semantic Information in L2

A growing body of evidence suggests that L2 learners are capable of predicting linguistic information during sentence processing by drawing on semantic cues, although the nature of their predictive processes may differ from those of native speakers (e.g., Grüter et al., 2012; Martin et al., 2013). This challenges earlier claims that predictive processing is primarily an L1 ability and largely restricted in L2 speakers (e.g., Dowens & Carreiras, 2006). Notably, Dijkgraaf et al. (2017, 2019) provided strong empirical support for L2 semantic prediction through a series of visual world eye-tracking studies involving Dutch-English bilinguals.

In the 2017 study, 30 Dutch-English L2 learners and 30 native English speakers took part. The L2 participants had learned English in formal classroom contexts without immersion, and all identified Dutch as their dominant language. Participants listened to sentences in both English and Dutch while viewing displays containing four objects. The sentences differed in semantic constraint, comparing predictive contexts (e.g., "Mary knits a scarf") with neutral ones (e.g., "Mary moves a scarf"). Researchers measured anticipatory eye movements towards the target object (e.g., scarf) before its auditory mention. Results showed that both native and non-native speakers demonstrated predictive fixations in semantically constraining conditions, indicating that L2 learners, like L1 speakers, can anticipate upcoming words based on verb semantics. In their 2019 follow-up study, the researchers employed a within-subject design, presenting experimental trials in both Dutch (L1) and English (L2) to enable direct comparisons of predictive processing across the two languages. The results showed that lexico-semantic prediction was qualitatively comparable in the L1 and L2. These findings corroborated earlier evidence, demonstrating reliable semantic prediction in both languages, with stronger effects observed among participants with higher proficiency in English.

Further support for the role of semantic cues in L2 prediction comes from Hopp (2015), who investigated whether late L2 learners of German could anticipate upcoming information in syntactically complex structures. Participants were divided into three groups based on their German proficiency and were presented with sentences in two structures: subject-first and object-first. Each sentence began with a noun phrase (NP1), followed by a lexical verb, an adverb, and a second noun phrase (NP2). Grammatical roles were unambiguous due to clear case markings on both noun phrases. For example, in the subject-first condition: *Der Wolf tötet gleich den Hirsch* (The wolf will soon kill the deer), and in the object-first condition: *Den Wolf tötet gleich der Jäger* (The hunter will soon kill the wolf). This design enabled the researchers to examine participants' anticipatory processing of agent and patient roles. The results indicated that while native speakers integrated both syntactic and semantic cues, L2 learners predominantly relied on lexical-semantic information, regardless of their proficiency level. This pattern suggests that even in the presence of robust morphosyntactic cues, L2 speakers may preferentially adopt semantic-based processing strategies.

Taken together, these studies suggest that L2 predictive processing tends to prioritise semantic over syntactic cues. Because semantic cues are category-based, they can effectively narrow the range of plausible referents—for example, a verb such as eat biases interpretation towards edible entities—thereby enabling the examination of real-time prediction and rapid meaning integration. Such cue-based manipulations have been shown to consistently elicit anticipatory fixations in the visual world paradigm.

2.2.3 Prediction of Phonological Information in L2

Phonological cues provide a means of examining the temporal dynamics of target-word recognition and lexical competition. The inclusion of cohort or rhyme competitors within the display enables researchers to identify and quantify the stages at which phonological information exerts an influence on gaze behaviour. Evidence supporting the prediction of phonological forms primarily comes from ERP studies using highly constraining sentence contexts—contexts that strongly favour a specific upcoming word (Ito et al., 2018). Research with native speakers suggests that predictive mechanisms are sensitive not only to semantic but also to phonological cues, particularly when contextual constraints are strong. For instance, Ito et al. (2016) investigated whether readers pre-activate both the semantic and phonological properties of upcoming words, and whether such predictions depend on the time available for processing. Using ERP recordings, two experiments were conducted in which participants read highly constraining sentences word by word (e.g., "The student is going to the library to borrow a...”). The critical word in each sentence was either the predictable word (e.g., book), a word similar in form (e.g., hook), a semantically related word (e.g., page), or an unrelated word (e.g., sofa). In experiment 1, semantically related words elicited a reduced N400—indicating semantic prediction—whereas phonological distractors did not reduce N400 amplitude but did evoke a Late Positive Component (LPC), reflecting the brain's effort to resolve unexpected input. In experiment 2, reduced N400s were observed for both semantic and phonological conditions in highly constraining contexts, with LPC effects again present for form-related words. These results suggest that while semantic prediction is relatively robust and automatic, phonological prediction depends on sufficient processing time and highly predictive contexts. This supports models of prediction that posit a temporal hierarchy, where semantic activation precedes phonological form activation.

In contrast, research on L2 speakers has produced more variable findings regarding phonological prediction. Martin et al. (2013), adapting DeLong et al.'s (2005) ERP paradigm, compared predictive processing in L1 and L2 English speakers. They found that native speakers showed a clear N400 response to mismatched articles (e.g., an or a preceding an expected versus an unexpected noun, as in “*She has a nice voice and always wanted to be a singer*” versus “*She has a nice voice and always wanted to be an artist*”), indicating phonological anticipation. However, L2 speakers did not exhibit such effects, even though they displayed typical N400 responses to unpredictable nouns and had explicit knowledge of the a/an rule. These findings imply that phonological prediction may be absent or considerably reduced in L2 sentence processing.

Similar results have emerged from eye-tracking studies focused on phonological prediction in L2 speakers. In studies by Ito et al. (2018), native speakers showed anticipatory eye movements toward phonological competitors, whereas L2 speakers did not. This suggests that phonological prediction in L2 comprehension is less robust than semantic prediction. A likely factor contributing to this discrepancy is the moderate cloze probabilities reported by Martin et al. (2013)—that is, the proportion of participants providing a specific word to complete a sentence, serving as an index of its contextual predictability—which were 69% for L1 speakers and 65% for L2 speakers. In contrast, phonological prediction typically requires more highly constraining contexts and therefore more predictable cloze sentences; for example, the L2 cloze probability in Foucart et al. (2014) was 81%. Since L2 speakers generally have less linguistic exposure, lower cloze values may limit their ability to engage in context-based prediction. As a result, they may rely less on contextual cues when processing sentences.

Another methodological issue lies in the degree of phonological overlap between target words and their competitors. In Ito et al. (2018), the average overlap was only 2.9 phonemes (SD = 0.83) out of 4.4 total phonemes (66.2%). The relatively weak overlap may have made it difficult to detect prediction effects, especially in L2 participants. Furthermore, none of these studies treated L2 proficiency as a dependent variable, making it difficult to determine whether the absence of phonological prediction stemmed from general limitations in L2 processing or from differences in proficiency. To address these gaps, the current study adopts a design with stronger phonological overlap between targets and competitors and considers proficiency as a key variable, thereby offering a more nuanced examination of phonological prediction in L2 learners.

2.2.4 Phonological Activation of L1 Equivalents in L2 Processing

Initial evidence for parallel activation in bilinguals comes from a bilingual Stroop task (i.e., the printed words are in one language and colour naming is performed in another language), in which Preston and Lambert (1969) adapted the standard colour-naming paradigm—known to produce interference when ink colour and word meaning are incongruent—for use with bilingual participants. Interference occurred when colour naming and printed words were in different languages, both for bilinguals with similar orthographies (e.g., English-French; Preston & Lambert, 1969) and with entirely different orthographies (e.g., Chinese-English; Chen & Ho, 1986). Later, another set of studies emerged to look at cross-linguistics interference in priming tasks, also in written language materials. In a seminal study, Beauvillain and Grainger (1987) investigated French-English homographs, such as *COIN*, which means "*corner*" in French. In their experiment, French-English bilinguals were shown these homographs embedded within a list of words in one language and subsequently completed a lexical decision task in the other language. For example, the researchers examined whether the word *COIN*, when presented in a French list, would facilitate lexical decision for the semantically unrelated English word which represents money. The results indicated that orthographic input could trigger the simultaneous activation of lexical entries in both languages during the very early stages of word processing. Interestingly, Beauvillain and Grainger (1987) also observed that facilitation effects were more pronounced for high-frequency words than for low-frequency words. For instance, *FOUR* is a high-frequency word in English but a low-frequency word in French, where it means "*oven*." This finding offers valuable insight for future research, particularly in informing the selection and design of experimental materials. Overall, early research established that both L1 and L2 can be activated in parallel during comprehension, laying the groundwork for subsequent investigations into the extent to which such activation persists in the absence of external input from the non-target language. One line of evidence comes from studies using cognates (i.e., words that are identical in both graphemic and semantic form across two languages), which have shown that parallel activation can occur without direct cross-language cues (e.g., De Groot & Nas, 1991; van Hell & de Groot, 1998). Building on these findings, later work has employed eye-tracking paradigms to explore parallel processing during spoken word recognition in bilinguals, offering a more fine-grained temporal account of this phenomenon (e.g., Marian & Spivey, 2003; Spivey & Marian, 1999).

However, in spoken language processing, evidence for the parallel activation of both lexicons in bilinguals has been limited and inconsistent. Specifically, Marian and Spivey (2003) conducted two eye-tracking experiments to explore the nature of lexical processing in late bilinguals, specifically whether language activation is selective or involves cross-linguistic influence. Their study involved Russian-English late bilinguals and aimed to determine the extent to which phonological forms from a speaker's L1 are activated during L2 processing. In experiment 1, fourteen native Russian speakers (mean age: 20), all highly proficient in English and using Russian for approximately one hour daily, completed a task that simulated a monolingual English environment. Participants viewed naturalistic arrays of objects while listening to English sentences, with some displays including phonological competitors from Russian (e.g., card - kartoshka “potato”; shovel - sharik “balloon”). The results revealed significant cross-linguistic activation: participants fixated on Russian competitor objects in 18% of trials, compared to 7% for control fillers. This finding indicates robust activation of L1 phonological forms during L2 comprehension, even in contexts framed as monolingual English.

In contrast, experiment 2 assessed whether L2 (English) phonology would be activated during L1 (Russian) processing. A separate group of fourteen native Russian speakers (mean age: 21) was tested in their L1. Unlike the first experiment, no significant between-language competition was observed, suggesting a directional asymmetry in cross-linguistic activation. The findings suggest that while L1 phonological forms can influence L2 processing, the reverse may not hold under the same experimental conditions. The study thus provides compelling evidence for directional L1-to-L2 phonological interference, stressing the role of language mode and processing direction in bilingual lexical access.

Supporting this view, Mishra and Singh (2014) conducted a visual world eye-tracking study to examine whether Hindi-English sequential bilinguals activate orthographic forms of phonological neighbours from the translation equivalents in their non-target language during spoken word recognition in either their L1 (Hindi) or L2 (English). Forty participants (mean age = 19.9), who had formally learned English from around age 4.7 and were more proficient in Hindi than English, listened to sentences while viewing written word options. One option in each display was a phonological competitor of the translation equivalent from the non-target language. For example, if the English target word was *gun* and its Hindi translation was *bandook*, the item *bandar* ("monkey" in English) would function as its phonological cohort. Early eye movements showed fixations on these competitors, indicating cross-linguistic activation during both L1-to-L2 and L2-to-L1 processing. Crucially, participants demonstrated significantly increased fixations on L1-based competitors when processing L2 sentences, particularly within the 200-400 ms time window, suggesting early predictive activation of L1 phonology. These results support the view of non-selective lexical access and show that predictive use of L1 phonological cues can operate bidirectionally, even across languages with distinct orthographies and no cognate relationships.

However, the consistency of L1 phonological activation in L2 processing is still debated. Ito et al. (2018) used a visual world eye-tracking design to examine whether Japanese learners of English predictively activate L1 phonological equivalents while listening to predictive English sentences. The study involved native English speakers and Japanese EFL learners (with differing L2 proficiency and English exposure ranging from 4 to 20 years), enabling direct comparisons of predictive language processing between the groups. Each display contained a target word, an English phonological competitor, a phonological competitor based on the Japanese translation of the target, and an unrelated object. For example, in the cloze sentence "The tourists expected rain when the sun went behind the *cloud* (Japanese: kumo), but the weather got better later", the target word is *cloud*, the English phonological competitor is *clown*, and the Japanese phonological competitor is *kuma* ("bear" in English). While native English speakers showed anticipatory fixations towards the phonologically expected English words, Japanese L2 speakers did not demonstrate increased fixations on the Japanese phonological competitors. This lack of activation challenges the assumption of automatic L1 activation in L2 comprehension and implies that cross-linguistic phonological interference may not be universal but instead is subject to variability influenced by language background and typological distance between languages.

Collectively, these studies demonstrate the nuanced and context-sensitive nature of cross-linguistic phonological activation in L2 language processing. While Marian & Spivey (2003) and Mishra & Singh (2014) offer compelling evidence that L1 phonological forms can influence L2 comprehension, the contrasting results from Ito et al. (2018) indicate that such activation may not be a universal phenomenon. One possible explanation concerns participants' L2 proficiency. In Marian and Spivey (2003) and Mishra and Singh (2014), participants had high English proficiency and extended exposure, whereas those in Ito et al. (2018) varied in proficiency and had comparatively shorter exposure to the L2. In addition, the divergent findings point to language typology as a potentially important yet underexplored variable, with more typologically distant language pairs potentially yielding opposite patterns of cross-linguistic activation. However, current research remains limited in its coverage of typologically diverse language combinations, constraining the generalisability of existing models of L2 lexical access and prediction.

These inconsistencies emphasise the necessity for further research that systematically investigates key moderating factors, including language pairing, processing directionality, and L2 proficiency, to elucidate the mechanisms underlying predictive processing in L2 users. Addressing this gap will require extending the scope of visual world eye-tracking studies to incorporate a broader range of language pairings, particularly those involving typologically distant languages. Such an approach would provide a more comprehensive account of whether L1-to-L2 phonological activation represents a universal characteristic of L2 processing or is shaped by specific linguistic and cognitive constraints.

3. Methods

3.1 Participants

Sixty-one individuals participated in the experiment, divided into three language groups. The advanced L2 group comprised 22 Mandarin Chinese EFL learners (mean age = 26.3 years); all members of this group reported IELTS scores (mean = 7.5, range = 7.5-8.5). The intermediate L2 group consisted of 18 Mandarin Chinese EFL learners (mean age = 24.7 years), of whom four provided IELTS scores (mean = 5.5, range = 5.5-6). Finally, the control group was formed by 21 native English speakers (mean age = 30.7 years). Within the advanced L2 group, two participants also speak German, three speak French, and one speaks Korean; in the intermediate group, one speaks French, two speak Korean, and one speaks both Hokkien and Malay; among the native English speakers, one individual also speaks French.

All intermediate-level EFL participants were undergraduate or postgraduate students enrolled in non-English-related programmes at universities in China. They had received six to eight years of English instruction but had never lived or studied in an English-speaking country. By contrast, advanced-level EFL learners possessed an academic grounding in English translation: they were either postgraduate students on accredited translation and interpreting programmes in the UK or certified Mandarin Chinese-English translators or/and interpreters, all of whom had immersed themselves linguistically in an English-speaking environment.

3.2 Stimuli

3.2.1 Sentence Materials

The experimental stimuli comprised 12 cloze sentences—sentences in which a word is omitted and must be inferred by the participant based on contextual cues—each paired with four object images displayed in the four corners of the computer screen. Each sentence featured a highly predictable target word (e.g., *book* in "He turned off his bedside lamp and rolled over to sleep after finishing the last chapter of his *book.*"), which was placed in varying positions, but typically near the sentence-final position. The placement of target words ranged from the 10th to the 27th word (M = 16.69, SD = 4.65). The sentences had an average length of 20.17 words (SD = 3.24, range = 11). Auditory stimuli were recorded by a female native English speaker, with each cloze sentence featuring a ring-ding sound in place of the target word.

The analysis of phoneme overlap between target words and their competitors reveals significant language- and competitor-dependent variations. The target words contained an average of 4.4 phonemes, of which an average of 3.08 (SD = 0.76) were shared with their English phonological competitor words, representing 70% overlap. In contrast, target words shared fewer phonemes with English semantic competitors, at 1.92 (SD = 1.75), or 43.5% of the mean phonemes (4.4). For the Chinese translations of the target words, the overlap with Chinese phonological competitor words was higher, with a mean of 3.92 (SD = 1.54) out of 5 phonemes, accounting for 78.4%. This suggests that for native Chinese speakers, the phonological connection between target words and their Chinese competitors is stronger, potentially facilitating easier recognition and correlation of the target word with its phonological competitor. These findings show the more prominent role of phonological similarity in the Chinese context, supporting the hypothesis that native Mandarin Chinese speakers may exhibit enhanced phonological sensitivity during sentence processing.

Each of the 12 unique sentences in the experiment was paired with four different scenes to create four conditions. Each of these scenes contained one critical image corresponding to the experimental item (i.e., target word, English phonological competitor, English semantic competitor and Chinese phonological competitor) and three serving as distractors. The critical images were counterbalanced across the four quadrants to ensure an even distribution of gaze opportunities. All distractors were unrelated to the experimental item in both semantic and phonological terms. In the target word condition, the critical image corresponding to the target word (e.g., *book* [Chinese: *shū*]) was presented with three other distractors. In the English phonological competitor condition, the English name of the critical image phonologically overlapped with the target word at its onset (e.g., *bush*). In the English semantic competitor condition, the English name of the critical image was semantically related to the context of the target word (e.g., *newspaper*). Lastly, in the Chinese phonological competitor condition, the Chinese name of the critical image overlapped phonologically with the Chinese translation of the target word (e.g., *tree* [*shù*]). Given the tonal nature of Chinese pinyin, in some sentences, the pinyin of the Chinese translation of the target object and the Chinese phonological competitor were identical, differing only in tone (e.g., *shū* [*book*] and *shù* [*tree*]). An example sentence with four conditions is shown in Fig. 1.

An image naming test was conducted as part of the experiment design to assess name agreement for the depicted objects. Nine native English speakers and twelve Chinese-English late bilinguals, who did not participate in the eye-tracking experiment, were involved in the norming process. Participants were shown images on a screen and asked to provide the first word that came to mind upon viewing each picture. Each image in the final set of stimuli was tested by at least eight participants from each group, with some images being revised and retested based on the feedback received. Native English speakers were presented with images corresponding to the target words, English phonological competitors, and semantic competitors, and responded in English. The Chinese-English bilingual participants completed the same task for the target words, English phonological competitors, and semantic competitors in English, followed by an additional set of trials involving Chinese phonological competitors, for which responses were given in Chinese. The results indicated that name agreement was relatively higher in the native English speakers' group compared to the Chinese-English bilingual group.

Condition

The figure originally presented here cannot be made freely available via ORA because of copyright.

Fig. 1. Example visual scene in four conditions for the experimental sentence "*He turned off his bedside lamp and rolled over to sleep after finishing the last chapter of his* ***book***." (target condition: ***book*** **[*shū*]**; English phonological competitor: ***bush***; English semantic competitor: ***newspaper***; Chinese phonological competitor: ***shù*** **[*tree*]**) The object depicted at the top right corner is the experimental object for this trial**.**

3.2.2 Versions

Each of the 12 experimental sentences was paired with four distinct visual scenes, yielding 48 unique conditions. These were organised into four versions of the experiment, all following the same procedure but differing in stimulus sets. Participants were randomly assigned to one of the four versions following successful calibration.

Each version included 12 visual scenes: each paired with a corresponding audio stimulus. Within every version, there were three conditions for each of the following: target word, English phonological competitor, English semantic competitor, and Chinese phonological competitor. To minimise repetition effects, each visual stimulus and sentence was presented only once per participant. The versions were counterbalanced across participants to ensure equal probability of assignment.

3.3 Procedure

The experiment was implemented and administered using the Gorilla Experimental Builder platform. Participants completed the study via an online link provided by the research team. To ensure compatibility with the eye-tracking component, the experiment could only be accessed on laptops or desktop computers; mobile phones and tablets were automatically restricted. Eye movement data were collected using Gorilla's webcam-based eye-tracking system, which estimates gaze position at a sampling rate of approximately 60 Hz. The session began with a bilingual instructional video (approximately 1 minute) introducing task procedures and allowing for repeated viewings. Participants were familiarised with two trial types: one where no image matched the audio (requiring a spacebar press), and one where a matching image was present (requiring a mouse click on the image).

Before the visual world task, a webcam polling test was conducted to evaluate the quality of each participant's webcam feed and the system's ability to detect facial and eye features. If the webcam did not meet the minimum threshold, due to factors such as poor lighting, low resolution, or misalignment, the participant was automatically excluded. Those who passed proceeded to an initial eye-tracking calibration, which required following a sequence of moving dots across the screen to establish accurate gaze mapping. Participants were allowed up to three attempts to complete a successful calibration. Failure to meet the calibration criteria after three attempts resulted in automatic exclusion from the experiment to ensure the integrity of the eye-tracking data.

Within the main experiment, each trial followed a fixed three-phase sequence. The first screen displayed a central fixation cross to centre participants' gaze and minimise any residual influence from the preceding trial. The second screen constituted the primary data collection phase, during which four images were presented in the corners of the screen while an auditory cloze sentence played automatically. Each sentence was played once, and participants' eye movements and fixations were recorded to capture real-time language processing in response to the combined visual and auditory input. Participants could not respond during this period to ensure that they had heard the full auditory stimuli. The third screen served as the response phase, during which participants were instructed to either click on the image corresponding to the sentence or press the spacebar if no suitable image was present. Eye-tracking continued during this phase to monitor gaze behaviour as participants made their selection. Following the submission of a response, the subsequent trial began immediately. The three-phase sequence is shown in Fig.2.

No feedback was provided at any stage of the task. The visual world task took approximately five minutes to complete, while the entire experimental session—including the information sheet, consent form, a pre-experimental questionnaire and the demonstration video—took approximately ten minutes in total.

Three-phase Sequence

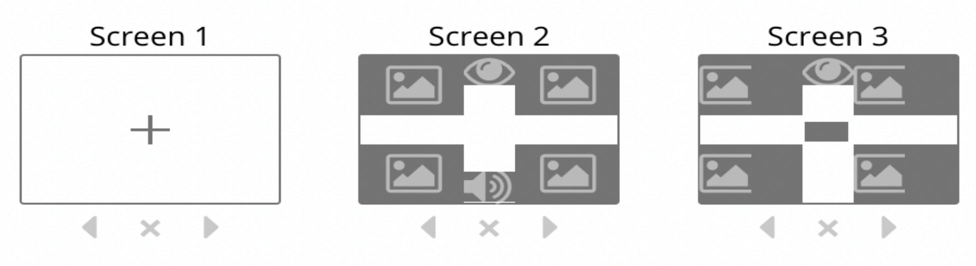


Fig.2. Schematic illustration of the three-phase trial sequence used in the visual world paradigm: Screen 1 (fixation), Screen 2 (stimulus presentation with eye-tracking), and Screen 3 (response phase with continued eye-tracking).

4. Results

Participants from three groups—advanced L2 learners, intermediate L2 learners and native English speakers—listened to sentences containing highly predictable target words while viewing arrays of four objects: the target, an English semantic competitor, an English phonological competitor, and a Chinese phonological competitor. We predicted that anticipatory processing would be evidenced by (1) increased fixations on the target prior to cloze onset; (2) greater fixations on English semantic or phonological competitors when English semantic or phonological cues were predicted or activated; and (3) elevated fixations on the Chinese phonological competitor if L2 speakers pre-activated their L1 phonological equivalents.

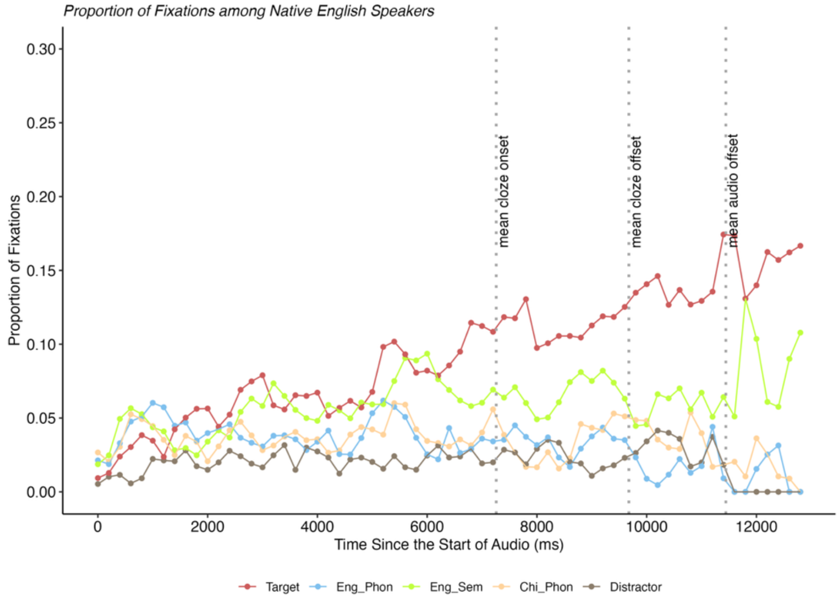
Our results show that native and near-native speakers not only engaged in semantic prediction but also exhibited delayed English phonological activation. Strikingly, advanced L2 learners also activated their L1 phonological equivalents during L2 processing. This chapter is organised as follows: first, we report overall experimental accuracy and present time-course plots of eye movements; next, we detail the statistical analyses that substantiate these effects; and finally, we integrate the quantitative findings with group-specific interpretations.

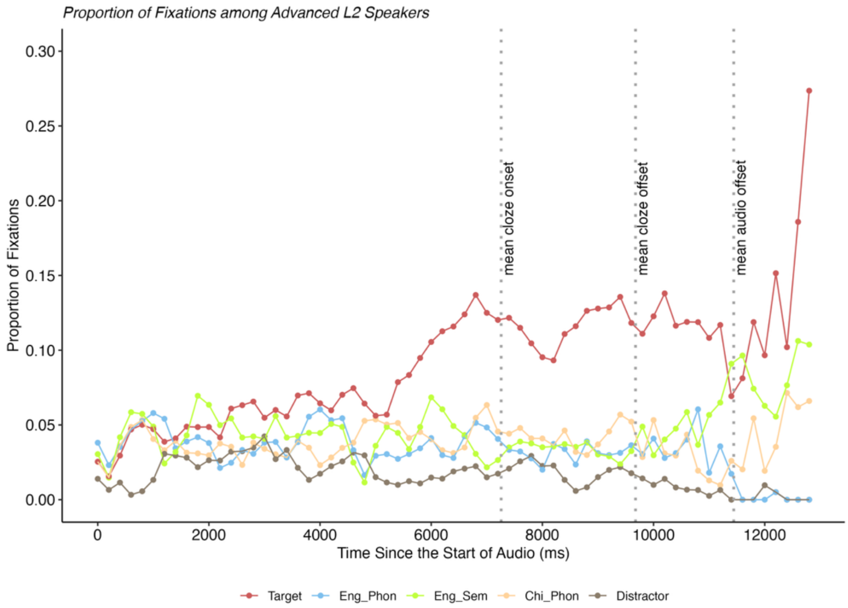
4.1 Cloze Sentence Comprehension Accuracy

The mean accuracy on the comprehension questions in the experimental trials was 98.8% (SD = 0.36%) for L1 speakers, 96.4% (SD = 3.1%) for advanced-level L2 speakers, and 89.2% (SD = 5.6%) for intermediate-level L2 speakers. Trials in which questions were answered incorrectly were excluded from the eye-tracking analyses.

4.2 Eye-Tracking Data

Fig. 3 illustrates the time course of fixation proportions to the target, English-phonological competitor, English semantic competitor, Chinese phonological competitor and distractor for the three participant groups. On inspection: all groups exhibited a marked increase in fixations on the target object well before audio offset, beginning around 6,000 ms—approximately 1,258 ms before the average cloze onset (vertical line 1 at 7,258 ms)—indicating successful predictive processing. Native speakers, however, displayed a particularly pronounced orientation towards the English semantic competitor: their fixations to this region rose modestly between 2,000 ms and 4,000 ms and thereafter paralleled the upward trajectory of target fixations, remaining the most attended competitor throughout, surpassed only by the target itself. In contrast, advanced L2 listeners did not exhibit a substantial increase in semantic competitor fixations until after the mean cloze offset (vertical line 2 at 9,675 ms), with this preference peaking close to the mean audio offset (vertical line 3 at 11,442 ms). Both phonological competitors attracted relatively low fixation proportions across all groups, hovering near the distractor baseline. Although advanced L2 participants exhibited a modest rise in looks to the Chinese phonological competitor after audio offset, suggesting a potential late phonological activation of L1 equivalents, neither phonological condition approached the level of semantic activation observed in native listeners. Intermediate L2 learners devoted the greatest proportion of their gaze to the target, while fixations to all three competitor types remained close to the distractor baseline, in contrast to both native speakers and advanced L2 learners. Notably, between the mean cloze onset and audio offset (7,258–11,442 ms), the English semantic competitor garnered more fixations than either phonological competitor; however, even during this interval, their overall attention to semantic items remained markedly lower than that observed in the native and advanced L2 groups.


Proportion of Fixations among Native English Speakers
Proportion of Fixations
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
2000
4000
6000
8000
10000
12000
Time Since the Start of Audio (ms)
mean cloze onset
mean cloze offset
mean audio offset
Target
Eng_Phon
Eng_Sem
Chi_Phon
Distractor


Proportion of Fixations among Advanced L2 Speakers
Proportion of Fixations
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
2000
4000
6000
8000
10000
12000
Time Since the Start of Audio (ms)
mean cloze onset
mean cloze offset
mean audio offset
Target
Eng_Phon
Eng_Sem
Chi_Phon
Distractor

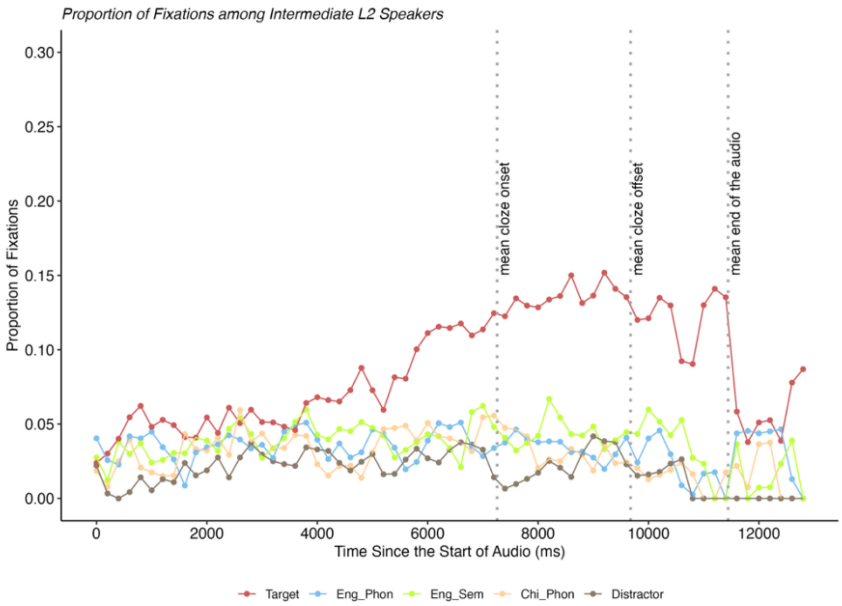


Fig.3. Eye-tracking results in English L1 group (top), Advanced L2 group (middle), and Intermediate L2 group (bottom). Time-course plots showing fixation proportion on target, English phonological competitor, English semantic competitor, Chinese phonological competitor and distractors. Three dashed vertical lines represent mean cloze onset (x = 7,258 ms), mean cloze offset (x = 9,675 ms), and mean audio offset (x = 11,442 ms).

4.3 Eye-Tracking Data Analysis

For our inferential analyses, we first conducted a one-way repeated-measures ANOVA to establish whether fixation proportions varied systematically across the five areas of interest—target, English semantic competitor, English phonological competitor, Chinese phonological competitor and distractor—within each participant group. Next, we performed a series of pairwise t-tests between the language groups for each condition. These t-tests allow us to compare fixation differences directly between groups—insights that the omnibus ANOVA, which only indicates that a difference exists somewhere, cannot provide. Finally, we employed growth curve analysis (GCA) as an auxiliary, more sensitive and temporally dynamic method. GCA characterises the overall shape and points of inflexion of gaze trajectories over time, yielding a more nuanced account of gaze behaviour across our three listener groups and enhancing our ability to detect subtle or delayed effects that might not reach significance in mean-based ANOVA models.

We applied ANOVA to the entire audio stream as well as two discrete epochs: from audio onset to the mean cloze-offset (0–7,258 ms), which we label the predictive window, and from the mean cloze-offset to the mean audio-offset (7,258–11,442 ms), which we label the late-activation window. By contrast, our pairwise t-tests and GCA were restricted to these two shorter windows rather than the full audio period. Because our stimuli were cloze sentences with the target noun omitted, the interval leading up to the mean cloze-offset corresponds to the predictive stage, while the subsequent interval constitutes the later-processing stage. In line with Kamide et al. (2003), this effectively mirrors the division into a verb-region window (predictive) and a post-cloze-offset window (integration), demonstrating that—even though the noun itself never occurs—participants' fixations nonetheless reveal that they have updated their referential encoding as soon as the critical noun position is reached.

Fixation proportions were log-transformed to stabilise variance and approximate normality before analysis. We employed the ezANOVA function from the ez package (Lawrence, 2016) in R (R Development Core Team, 2015), with participant identifier as the subject variable and condition as the within-subjects factor. To evaluate the assumption of sphericity, we conducted Mauchly's test; upon detecting a violation, we applied the Greenhouse-Geisser correction to adjust the degrees of freedom and control the Type I error rate. Finally, to probe specific pairwise differences against the distractor baseline while safeguarding against inflated alpha levels, we performed Bonferroni-adjusted paired t-tests. This combination of procedures ensures a robust framework for determining whether and how visual attention varies across our experimental conditions.

4.3.1 Native English Speakers

Across the entire time window, native speakers showed a robust main effect of Condition, $F (4, 80) = 13.61$, $p < .001$, $\eta^2\_G = .31$. Mauchly's test indicated a sphericity violation ($W = .25$, $p = .003$), and Greenhouse–Geisser correction ($\varepsilon = .65$) upheld the effect, $F (2.60, 52.02) = 13.61$, $p < .001$. Bonferroni-adjusted pairwise t-tests against the distractor baseline revealed significantly longer fixations for the target ($p < .001$) and English semantic competitor ($p = .004$), but not for the English phonological ($p = .187$) or Chinese phonological ($p = .208$) competitors.

In the first-time window (0-9675 ms), there was a highly significant effect of fixation location, $F (4, 80) = 12.42$, $p < .001$, generalised $\eta^2 = .268$. Pairwise contrasts against the distractor reveal that looks to the target were substantially elevated ($p < .001$), as were looks to the semantic competitor ($p = .003$) and the English-phonological competitor ($p = .042$); by contrast, the Chinese-phonological competitor did not differ from the distractor ($p = .128$). By the later window (9675-11442 ms), fixation location remained significant, $F (4, 80) = 6.13$, $p < .001$, $\eta^2 = .216$, but now no non-target competitor differed reliably from the distractor (Chinese-phonological: $p = 1.000$; English-phonological: $p = 1.000$; semantic: $p = .510$). The data showed that the native listeners quickly co-activate semantic information alongside the target, accompanied by a subtle and early English phonological activation.

4.3.2 Advanced L2 Group

In the whole audio window analysis, a highly significant effect of condition was also found in the advanced L2 group, (4,84) = 23.75, $p <.001$, generalised eta squared = 0.437. Mauchly's test indicated a violation of sphericity, = 0.280, $p = 0.003$, so Greenhouse-Geisser correction ($\varepsilon = 0.599$) was applied, yielding $p <.001$ for the main effect. Bonferroni-adjusted paired-t comparisons against the distractor baseline showed that fixations to the target ($p <.001$), English semantic competitor ($p <.001$), the English phonological competitor ($p =.0086$) and the Chinese phonological competitor ($p =.0497$)

were all significantly greater than baseline. Furthermore, fixation proportions to the target region exceeded those to each competitor—versus Chinese phonological ($p < .001$), versus English phonological ($p < .001$), and versus English semantic ($p = .021$).

During the early window (0–9,675 ms), advanced L2 listeners exhibited a pronounced main effect of fixation location, $F(4, 84) = 24.603$, $p < .001$, $\eta^2\_G = .411$. Relative to the distractor, they fixated the target ($p < .001$) and semantic competitor ($p < .001$) significantly more, showed a modest increase for the English-phonological competitor ($p = .011$), but did not differ for the Chinese-phonological competitor ($p = .103$). In the later window (9,675–11,442 ms), fixation location remained significant, $F(4, 84) = 8.387$, $p < .001$, $\eta^2\_G = .254$: target fixations continued to exceed baseline ($p = .002$), and Chinese-phonological fixations also rose significantly ($p = .041$), reflecting a delayed L1 phonological activation, while English-phonological fixations did not ($p = .405$). These findings imply that advanced L2 listeners, like native speakers, rapidly co-activate semantic and phonological neighbours—albeit less strongly—and uniquely display late L1 phonological activation consistent with their overall time-course.

4.3.3 Intermediate L2 Group

A robust effect of condition was also observed among intermediate L2 speakers throughout the entire audio window, $F(4,68) = 21.64$, $p < .001$, generalised eta squared $= .411$, confirming that fixation behaviour varied markedly across the five display conditions. Mauchly's test indicated that the sphericity assumption was met ($W = .796$, $p = .940$), obviating the need for correction. Subsequent Bonferroni-adjusted paired-t comparisons against the distractor baseline revealed that only the target region elicited significantly longer fixations ($p < .001$); neither the Chinese phonological ($p = .949$), English phonological ($p = .084$), nor English semantic ($p = .086$) competitors differed significantly from baseline. Direct contrasts among conditions further showed that fixation durations to the target exceeded those to the Chinese phonological ($p < .001$), English phonological ($p < .001$) and English semantic ($p = .003$) regions, whereas no significant differences emerged among the three competitor types (all $p > 1.00$).

When split into two intervals, intermediate L2 listeners showed a similar profile. In the early window (0-9,675 ms), there was a significant effect of fixation location, $F(4, 68) = 18.22$, $p < .001$, $\eta^2\_G = .344$; only the target attracted more fixations than the distractor ($p < .001$), while semantic ($p = .183$), English-phonological ($p = .135$) and Chinese-phonological ($p = 1.000$) competitors did not. In the later window (9,675–11,442 ms), the effect persisted, $F(4, 68) = 13.12$, $p < .001$, $\eta^2\_G = .399$, with the target still exceeding baseline ($p = .001$) but no competitor showing elevated fixations (semantic: $p = .170$; English-phonological: $p = .627$; Chinese-phonological: $p = 1.000$). This indicates that intermediate L2 learners concentrate on the intended referent from the outset and exhibit minimal semantic or phonological co-activation throughout.

4.3.4 Cross-Group Comparison

We assessed cross-group differences in fixation behaviour by performing a series of pairwise ttests in R, comparing each participant group across all experimental conditions.

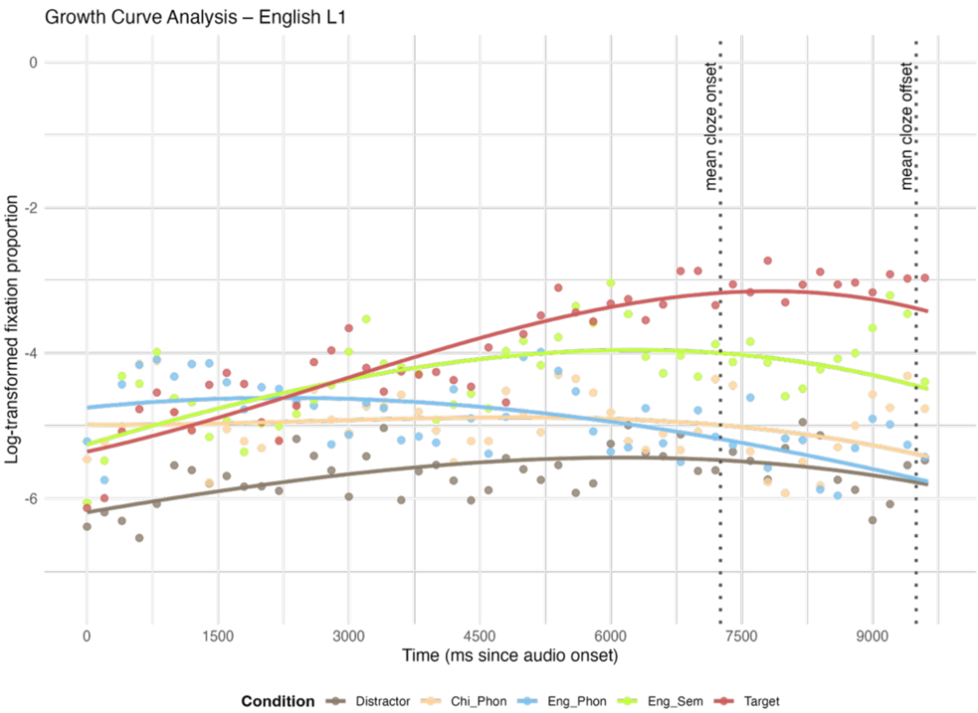
In the early processing window, fixation patterns to the target were broadly similar across all three groups. Surprisingly, intermediate L2 speakers showed the highest average proportion of fixations to the target ($M = 0.0861$), followed by advanced L2 speakers ($M = 0.0787$), and native speakers ($M = 0.0765$). However, none of the group differences reached statistical significance: L1 vs advanced L2, $t(40.5) = –0.16$, $p = .871$; L1 vs intermediate L2, $t(36.9) = –0.72$, $p = .478$; and advanced vs intermediate L2, $t(37.8) = –0.58$, $p = .567$. These results suggest that attention to the target object was relatively stable and unaffected by language background in the early time course. For the English semantic competitor, native speakers exhibited significantly more fixations ($M = 0.0579$) than advanced L2 speakers ($M = 0.0402$) and intermediate L2 speakers ($M = 0.0395$), with a statistically significant group difference in L1 vs advanced L2: $t\ (28.5) = 2.30$, $p = .0291$, and advanced L2 vs intermediate L2: $t\ (28) = 2.12$, $p = .043$. This indicates that native speakers and advanced L2 speakers were more sensitive to semantic competitors, likely due to stronger or faster activation of lexical-semantic associations.

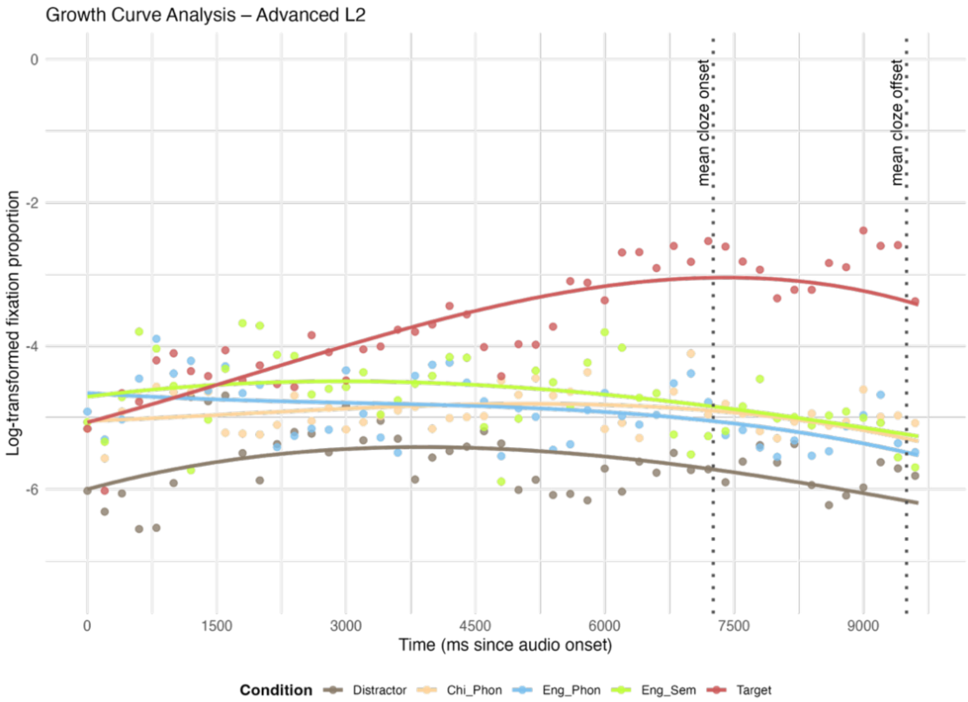
In the English phonological condition, fixation proportions were low across all groups. Native speakers ($M = 0.0369$) looked marginally more at the phonological competitor than advanced ($M = 0.0353$) and intermediate L2 speakers ($M = 0.0329$), but none of the comparisons were significant: L1 vs advanced L2, $t(38.7) = 0.33$, $p = .746$; L1 vs intermediate L2, $t(32.6) = 0.61$, $p = .545$; advanced vs intermediate L2, $t(28.3) = 0.39$, $p = .701$. This suggests that phonological competitor activation was weak or rapidly resolved across all groups in the early phase. For the Chinese phonological condition, Advanced L2 speakers showed slightly higher proportions ($M = 0.0364$) than intermediate L2 speakers ($M = 0.0352$). Again, all differences were non-significant.

In the later processing window, all groups again showed strong fixations to the target object. Native speakers had the highest fixation proportion ($M = 0.125$), followed by advanced L2 speakers ($M = 0.120$) and intermediate L2 speakers ($M = 0.117$). This confirms that all participants, regardless of language background, successfully identified and fixated on the target during the main decision phase. For the English semantic competitor, native speakers continued to show the highest fixation proportions ($M = 0.0587$), followed by advanced L2 speakers ($M = 0.0463$) and then intermediate L2 speakers ($M = 0.0374$). This consistent ranking across both windows suggests that native speakers engaged more robustly and persistently with semantically related alternatives. In the late window, a statistically significant difference emerged in the Chinese phonological condition between advanced and intermediate L2 speakers, $t\ (32.1) = 2.11$, $p = .0428$. Advanced L2 speakers exhibited a higher mean fixation proportion ($M = 0.0351$), whereas intermediate L2 speakers showed markedly lower engagement (M = 0.0143). This result suggests a proficiency-related increase in sensitivity to cross-language phonological interference, with more advanced bilinguals demonstrating greater activation of phonologically similar distractors from their L1.

4.3.5 Growth Curve Analysis

In these growth curve analyses (GCA), the linear term (T1) captures the overall direction and rate of change in fixation probability, the quadratic term (T2) identifies where the trajectory bends into a peak or trough, and the cubic term (T3) highlights more complex inflexion and asymmetry in the rise-and-fall profile. Unlike repeated-measures ANOVA, which summarises mean differences at fixed time points, GCA fits these polynomial components continuously across the trial, revealing how gaze patterns emerge, peak, and subside in each condition. We restricted our cubic GCA to the early, predictive window for two reasons: first, reliable group-level fixation differences appeared only in the trial's first half—the latter half showed no consistent effects, with the sole exception of one significant finding in the intermediate L2 group; and second, applying a single third-order polynomial to the entire trial would smooth over subtle, late-trial fluctuations, reducing sensitivity to the dynamics of interest.


Growth Curve Analysis – English L1
Log-transformed fixation proportion
0
-2
-4
-6
mean cloze onset
mean cloze offset
0
1500
3000
4500
6000
7500
9000
Time (ms since audio onset)
Condition
Distractor
Chi_Phon
Eng_Phon
Eng_Sem
Target


Growth Curve Analysis – Advanced L2
Log-transformed fixation proportion
0
-2
-4
-6
mean cloze onset
mean cloze offset
0
1500
3000
4500
6000
7500
9000
Time (ms since audio onset)
Condition
Distractor
Chi_Phon
Eng_Phon
Eng_Sem
Target

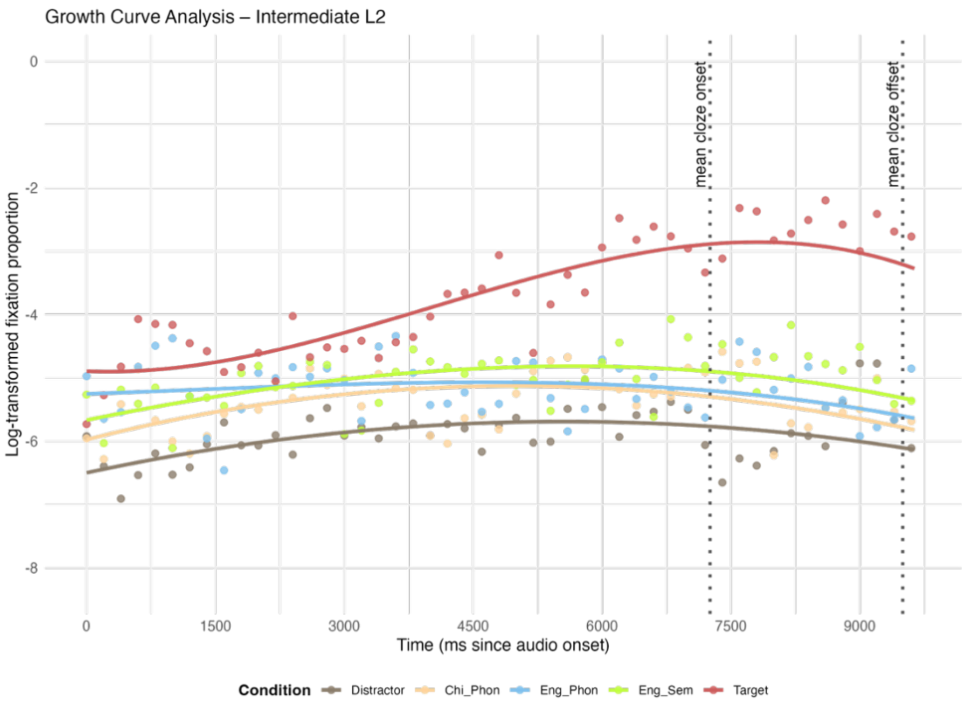


Fig.5. Growth curve analysis model in English native speakers (top), advanced L2 group (middle) and intermediate L2 group (bottom). The dots on the plots represent the mean log-transformed fixation proportions for each condition at each 200 ms time bin, averaged across participants, and are used to visualise the empirical data points underlying the modelled growth curves.

Among native speakers, only the target (β = 29.37, $p < .001$) and the English semantic competitor (β = 12.94, $p = .04$) exhibited significant positive linear trends (T1) in the predictive window, reflecting sharp increases in fixations to these items; neither quadratic (T2) nor cubic (T3) terms reached significance, indicating an essentially linear shift in gaze. Thus, native listeners swiftly directed—and then sustained—their attention on both the target and the semantic competitor, revealing a stable, group-wide fixation pattern.

Among advanced L2 listeners, in the predictive window (0-7,258 ms), only the target condition showed a significant positive linear effect (T1: β = 42.53, SE = 6.27, t = 6.78, $p < .001$), meaning a steep and sustained rise in fixations over time. Notably, this linear slope was steeper in advanced L2 speakers than in native speakers. While native speakers maintained an upward gaze trajectory throughout the window, advanced L2s shifted their attention to the target somewhat later (around 6,000 ms) but with greater intensity. None of the other conditions—including the English semantic, English phonological, or Chinese phonological competitors—exhibited significant linear trends in this window. However, a significant positive quadratic term (T2) emerged for the Chinese phonological competitor (T2: $\beta = 15.35$, $SE = 6.27$, $t = 2.45$, $p = .014$) and English phonological competitor (T2: $\beta = 17.53$, $p = .005$), suggesting a slightly curved upward trajectory, indicating a slightly delayed attention to both phonological items. No significant effects were observed in the later time window, suggesting potentially flat or inconsistent gaze patterns.

In the predictive-time-window analysis for intermediate L2 participants, the linear time term (T1) revealed a pronounced increase in target looks ($\beta = 23.86$, $SE = 6.04$, $t = 3.95$, $p < .001$), revealing a quick and steep upward trend in their gaze to target items. In the subsequent window, that linear slope flipped ($\beta = –22.03$, $p < .001$), reflecting a marked decrease in target fixations.

In sum, native speakers showed strong early fixations on the target and English semantic competitor, with a consistent gaze pattern over time. Advanced L2 listeners demonstrated a sharp, linear increase in target fixations early on, while delayed, curved increases were observed for phonological competitors, suggesting slower or less consistent attention to the phonological cues. In contrast, intermediate L2 participants fixated heavily and exclusively on the target early but showed a rapid decline in the later time window, showing focused but short-lived attention to the target with minimal interference from competitors.

4.4 Synthesis of Findings

4.4.1 Native English Speakers

L1 listeners displayed the most robust and temporally structured gaze behaviour. ANOVA results showed significantly longer fixations on both the target and English semantic competitor, confirming their central role in sentence processing. GCA further revealed significant positive linear trends for these two conditions during the predictive window, reflecting an early, rapid, and sustained increase in fixations. T-tests supported this, showing that native participants fixated significantly more on the semantic competitor than L2 groups. Notably, the English phonological competitor received only minimal or brief attention. Although the initial ANOVA indicated some early activation, the t-tests revealed that fixations to this item were relatively weak. This may help explain why the GCA failed to detect any statistically significant effects—possibly due to a lack of consistent timing in gaze patterns or variability in individual fixation behaviour among native speakers. Overall, native participants engaged rapidly and reliably with the target and semantically related items. In contrast, phonological activation was subtle and less stable, with the English phonological competitor eliciting brief and inconsistent attention during sentence processing.

4.4.2 Advanced L2 Listeners

Advanced L2 participants also demonstrated significant fixation preferences for the target and English semantic competitor in the ANOVA, with additional significant fixations to both English and Chinese phonological competitors. However, t-tests indicated that native speakers still showed more engagement with the semantic competitors. GCA results further clarified this pattern: the target item alone showed a steep and consistent linear increase in fixations during the predictive window, mirroring native-like engagement. However, neither the semantic nor phonological competitors yielded significant linear effects, suggesting that while these items were fixated, their activation was more variable and less systematic. Instead, both phonological competitors exhibited significant positive quadratic effects, indicative of a delayed and slightly curved rise in attention, possibly reflecting slower or more effortful lexical access. Importantly, during the later time window, ANOVA revealed a significant rise in fixations to the Chinese phonological competitor, reflecting delayed L1 activation.

The absence of significant temporal trends in GCA for the semantic competitor implies inconsistent engagement timing across participants—activation occurred, but not in a uniform or predictive way.

4.4.3 Intermediate L2 Listeners

Intermediate L2 participants presented the most restricted fixation pattern. ANOVA revealed significantly increased fixations only for the target condition, with no evidence of enhanced attention to semantic or phonological competitors. GCA confirmed this highly focused profile: fixations on the target rose rapidly during the predictive window (positive linear effect) but declined sharply in the later window (negative linear effect), reflecting a quick orientation to the referent followed by disengagement. Neither quadratic nor cubic effects emerged for any competitor, and t-tests showed no significant between-group differences in competitor fixation proportions for this group. This suggests a limited or underdeveloped ability to process or co-activate lexical alternatives, likely due to lower English proficiency or a reduced parallel activation of competitors. In short, intermediate L2 listeners prioritised the target and showed minimal semantic or phonological preference, relying more on direct integration than on anticipatory mechanisms.

5. Discussion

We investigated the real-time prediction of semantic and phonological information for highly predictable English words in both native and non-native listeners using the visual world paradigm. Our design comprised four critical experimental objects—target, English semantic competitor, English phonological competitor and Chinese phonological competitor—each presented among unrelated distractors per trial. Analysis of fixation proportions within groups and between-group comparisons across conditions revealed that all three cohorts directed their gaze increasingly towards the target early in the predictive window. The onset of target activation occurred at a comparable time across groups, and fixation proportions reached similar levels, with no statistically significant differences observed between the three language groups. Moreover, native English speakers and advanced L2 learners both exhibited early semantic activation, accompanied by a more modest and subtle English phonological activation. In contrast, intermediate L2 listeners focused almost exclusively on the target, showing no reliable gaze shifts towards any competitors. Surprisingly, the advanced L2 group showed a delayed yet significant rise in fixations to the Chinese phonological competitor, while the same pattern was not found among intermediate L2 participants. This suggests that highly proficient L2 speakers momentarily activate their L1 phonological equivalents when processing L2 sentences.

5.1 Semantic Prediction in L1 and Advanced L2 Speakers

Both native speakers and advanced L2 learners exhibited strong semantic prediction: fixations to semantic competitors were consistently the second highest, trailing only the target and significantly outpacing phonological competitors. Although the precise timing of these gaze shifts varied more among advanced L2 participants, their elevated fixation proportions confirm that semantic alternatives were actively considered in real time. This finding converges with previous research demonstrating that proficient L2 learners can predict upcoming meaning during processing (e.g., Chee et al., 2001; Dijkgraaf et al., 2017, 2019; Kaan & Grüter, 2021; Schlenter & Felser, 2021; Yang et al., 2018), confirming that advanced learners are capable of accessing and integrating semantic content on-the-fly.

The prominence and persistence of semantic-over-phonological effects likely reflect the inherently broader scope of semantic prediction: listeners anticipate contextually relevant categories (e.g. “food” or “animal”) rather than specific phonological forms, thereby maintaining attention on semantic competitors for longer (Hintz & Huettig, 2015; Huettig & McQueen, 2007; Ito et al., 2018). In the present study, native speakers exhibited a highly synchronised temporal profile of semantic activation, whereas the advanced L2 group demonstrated comparatively weaker and slower activation, accompanied by substantial inter-individual variability. Such variation may reflect differences in processing speed, proficiency, and attentional strategies (Dijkgraaf et al., 2019). This pattern is consistent with interactive models of bilingual word recognition (e.g., BIA+; Dijkstra & Van Heuven, 2002), which propose that the lower subjective frequency of L2 lexical items slows both form recognition and subsequent semantic retrieval, even if the overall activation trajectory remains broadly similar to that of native listeners. Furthermore, the disparity between semantic prediction in L1 and L2 supports the view that semantic mappings—defined as the mental connections between a word’s form and its meaning—are less robust in L2 than in L1 (Gollan et al., 2005, 2008). It also aligns with the notion that slower word form activation in L2 delays the spread of semantic activation. Collectively, these findings offer further support for the claim that L2 processing may be less proactive than L1 processing (Grüter et al., 2012; Grüter & Rohde, 2013).

In summary, although advanced L2 learners’ semantic predictions are relatively delayed and less temporally uniform than those of native speakers, they nevertheless demonstrate clear, real-time access to and integration of semantic representations during sentence comprehension.

5.2 Subtle and Later English Phonological Activation in L1 and Advanced L2 Speakers

While semantic prediction was clearly the dominant pattern, modest but statistically reliable within-group increases in fixations to the English phonological competitor were observed among native speakers and, to a lesser extent, advanced L2 learners. It should be noted that the absolute fixation proportions to phonological competitors, however, remained low across all groups. Nevertheless, the significant within-group effects for the L1 and advanced L2 cohorts prove that phonological information was elicited in a subtle and variable manner across individuals, yet still detectable by the statistical models. By contrast, the intermediate L2 group showed no statistically reliable phonological activation. This null result, however, should be interpreted cautiously given the modest sample size and possible variability in gaze patterns, rather than the complete absence of an effect. Because only the native and advanced L2 groups yielded statistically robust phonological activation within our data, the detailed analyses presented below focus on these two groups to examine the temporal dynamics of their predictive gaze behaviour in greater depth.

The modest phonological activation observed in both L1 and advanced L2 groups from our study contrasts with the stronger effects reported for native speakers in Ito et al. (2018) with Japanese EFL learners and in ERP studies (Ito et al., 2016; Martin et al., 2013). However, this divergence is likely attributable to key methodological differences between the studies. We employed a separate picture-norming phase—using an independent sample to verify object recognisability and eliminate visual confounds—whereas Ito and colleagues used an immediate picture-familiarisation task, in which the same participants matched images to their labels immediately prior to the visual world experiment. Because phonological prediction depends on anticipating an exact word form, this pre-exposure likely enhanced sensitivity to phonological features and thus amplified the phonological anticipation effects in their native speaker sample. Moreover, Ito et al. (2018) restricted their design to phonological competitors alone, whereas our inclusion of semantic competitors may have attenuated phonological activation by introducing stronger meaning-based cues. Indeed, Ito and Husband (2017), who tested both semantic and phonological cues in L1 sentence processing, found no evidence in phonological prediction in predictable contexts, concluding that form-based anticipation is markedly weaker than semantic prediction, even for native English listeners.

These findings point out how competitor composition and cue salience shape predictive processing. When only phonological cues are presented, listeners readily predict form overlap; yet in more intricate paradigms that also feature semantic competitors, the more diagnostic, category-based information appears to be dominant, suppressing or delaying activation of phonological alternatives. In other words, prediction is not uniform but highly flexibly modulated by task demands and cue salience: semantic cues, offering broader contextual relevance, appear to take precedence over sound-based cues, thereby altering the distribution of visual attention and explaining the absence or weaker phonological effects we discovered in the present study when semantic competitors are also in play.

Finally, the modality of the display itself can facilitate phonological prediction. Huettig and McQueen (2007) demonstrated that presenting printed words—rather than pictures—markedly increases fixations on phonological competitors during sentence processing. This suggests that the nature of the display—whether pictorial or orthographic—can either facilitate or constrain gaze preferences. Huettig and McQueen (2007) argue that listeners' eye movements emerge from the interplay of three core factors. First, the format of the visual input—whether participants are viewing pictorial scenes or written words—sets up different initial processing pathways, since picture recognition relies on holistic shape cues while word recognition unfolds over individual letters and graphemes. Second, the relative timing of these recognition systems matters: the word-recognition machinery may become engaged sooner or later than the picture-recognition machinery, depending on task demands and stimulus complexity. Finally, the dynamic unfolding of the auditory signal continuously updates lexical hypotheses as speech unfolds in time. It is the combination of these elements—visual format, system-specific processing latencies and the temporal structure of the spoken input—that jointly determines where and when listeners shift their gaze.

5.3 Chinese L1 Phonological Activation in English Sentence Processing

Early eye-tracking research has demonstrated that late bilinguals may engage their L1 phonological representations even when processing an L2 (e.g., Marian & Spivey, 2003; Mishra & Singh, 2014; Blumenfeld & Marian, 2007). These findings support non-selective lexical access models, which propose that bilinguals automatically and in parallel activate entries from both languages upon encountering a word (Thierry & Wu, 2007; Wu et al., 2013). However, in our study, we did not find evidence to support the L1 activation among intermediate L2 speakers. At the same time, the advanced L2 learners displayed a modest yet statistically significant increase in fixations to the Chinese phonological competitor during the predictive window. Although this phonological activation emerged slightly later and was less pronounced than the peaks for target and semantic looks, it nevertheless elevated fixation proportions above the distractor baseline and constituted the most robust response to Chinese phonological response of any group. These results reveal that highly proficient L2 speakers reliably co-activate their native phonological equivalents during L2 sentence processing, in line with interactive bilingual-recognition frameworks (e.g. BIA+), which posit early, parallel activation followed by top-down inhibition as context accrues. In this framework, greater proficiency in the target language (in our study, English) is associated with stronger and more immediate cross-linguistic competition (e.g., Tatsuno & Sakai, 2005). The advanced L2 participants in the present study, all of whom were qualified Chinese-English translators and/or interpreters with near-native phonological proficiency, demonstrated rapid and automatic parallel activation of both languages.

Such activation reflects the potential for pronounced cross-linguistic interference and aligns with the cognitive advantages frequently attributed to sustained bilingual practice.

Moreover, our findings are likely to provide methodological guidance on how the design of experimental materials shapes phonological activation. For example, although Ito et al. (2018) incorporated Japanese phonological equivalents into their eye-tracking study of English sentence processing, they did not observe any corresponding phonological activation in Japanese EFL learners. Three procedural differences may explain this divergence. First, our experiment was conducted in a bilingual context: instructions, consent forms, demonstration videos and calibration prompts were presented in both Chinese and English, while Ito and colleagues used English exclusively in their experimental setting. Second, the degree of phoneme overlap in our materials was markedly higher: the Chinese translations of our target words shared, on average, 3.92 out of 5 phonemes (78.4 %) with their Chinese phonological competitors, compared to only 2.6 out of 4.9 phonemes (53.8 %) in Ito et al.'s Japanese materials. Hence, the close phonological correspondence between the L1 equivalents chosen and the target in our materials likely amplified the prominence of Chinese phonological cues, thereby facilitating the transient co-activation observed in advanced L2 speakers. Third, the present study employed longer, more naturalistic sentences, with the cloze item generally positioned in the latter half (and in some cases at the very end) of the sentence. This design provided richer contextual information and a longer predictive interval in which anticipatory processing could occur. In our materials, target words were placed between the 10th and 27th word ($M = 16.69$, $SD = 4.65$), with sentences averaging 20.17 words in length. By contrast, in Ito et al. (2018), target words occurred earlier, between the 9th and 20th position ($M = 13.7$, $SD = 2.6$), and were never sentence-final. Their sentences averaged 17.6 words in length, a comparatively shorter structure that may have resulted in a shorter predictive window and thus reduced reaction time, potentially producing more variable gaze patterns that were not consistently captured in the analysis.

Although a significant phonological effect was observed, the paucity of evidence among intermediate L2 speakers and the heterogeneity of fixation patterns within the advanced cohort—reflecting substantial inter-individual variability—suggest that L1 phonological activation is not invariably robust. Building on Huettig and McQueen's (2007) emphasis on the critical role of orthographic presentation in eliciting phonological activation, we assume that orthographic overlap may underlie both between-group and within-group variability. Specifically, in English, target words and their phonological competitors often share orthographic features (e.g., beer - bee; belt - bell), thereby reinforcing phonological activation. In Chinese, however, targets and their phonological competitors may share similar pinyin yet differ only in tone— shū ('book", 书) versus shù ('tree", 树) —creating purely phonological, rather than orthographic, overlap. This distinction may account for the observed cross-group differences and the heterogeneity in gaze behaviour among advanced L2 speakers.

5.3.1 Challenging the Inhibitory Control Assumption in Bilingualism

Although this remains a contentious issue, it is often argued that when bilinguals reached certain proficiency levels, one of the "bilingual advantages" is the enhanced inhibitory control: as learners become more proficient, they demonstrate superior attentional regulation (Van Hell & Tanner, 2012) and broader executive-function gains (Bialystok & Craik, 2010; Elston-Güttler et al., 2005), including the suppression of irrelevant stimuli and the inhibition of dominant but inappropriate responses. Moreover, even L2 acquired later in life, once mastered at a high level, can modulate the established L1 system (Baus et al., 2013; Linck et al., 2019). The Inhibitory Control Model (Green, 1998) and similar frameworks posit that as bilingual proficiency grows, individuals become increasingly adept at suppressing their non-target language. However, our results argue that L1 activation is not wholly inhibited even among highly proficient L2 speakers.

Multiple studies have demonstrated that highly proficient bilinguals continue to exhibit activation of L1 representations during L2 processing tasks, particularly at the phonological and lexical levels (e.g., Blumenfeld & Marian, 2007; Spivey & Marian, 1999). Rather than indicating a failure of inhibitory control, such co-activation is increasingly interpreted as evidence of an integrated and highly interactive bilingual lexicon (Kroll et al., 2014). In this view, proficient bilinguals engage in dynamic, context-sensitive language management, deploying inhibitory control only when necessary, rather than operating through a blanket suppression of the non-target language (e.g., Hopp, 2013). This suggests that higher proficiency does not necessarily equate to stronger inhibition of L1, but rather to greater flexibility in managing cross-linguistic activation.

Moreover, recent large-scale meta-analyses have cast doubt on the broader claim of a consistent “bilingual advantage” in domain-general inhibitory control. Lehtonen et al. (2018), reviewing over 150 studies, found no reliable bilingual advantage in inhibitory control when methodological confounds such as socio-economic status and language context were controlled. Similarly, Paap et al. (2015) argue that observed differences in inhibition between bilinguals and monolinguals are likely due to extraneous variables rather than bilingualism itself.

Collectively, our results and those of prior studies speak against the notion that higher L2 proficiency simply equates to stronger inhibition of the L1. Instead, they endorse a more nuanced account of bilingual processing in which both languages remain concurrently active, and highly proficient speakers cultivate strategies for managing this dual activation rather than relying solely on suppression.

5.3.2 Challenging the L1 Regression Assumption in Highly Proficient L2 Speakers

A long-standing concern in bilingualism research is that sustained exposure to and use of an L2 may lead to L1 attrition or regression, particularly in contexts where the L2 is dominant. This “L1 regression” hypothesis proposes that increased reliance on L2 can recalibrate the bilingual’s lexical and grammatical systems, resulting in slower lexical retrieval, greater cross-language interference, or reduced grammatical accuracy in the native language (Köpke & Schmid, 2006; Pavlenko, 2000; Schmid & Köpke, 2017). Within the framework of the Inhibitory Control Model (Green, 1998), such attrition can be conceptualised as the long-term outcome of sustained suppression of the L1 during L2 use. Prolonged inhibitory control may progressively diminish the baseline accessibility of L1 representations. This effect is likely to be especially salient among highly proficient L2 users in L2-dominant contexts, where continuous inhibition of the non-target language is necessary to support fluent L2 production and comprehension.

However, the results of the present study do not support the notion that higher L2 proficiency in our participant group equates to diminished L1 accessibility. Notably, the advanced L2 participants possessed distinctive linguistic profiles, being qualified Chinese-English interpreters or translators. In such professions, high proficiency in both languages is not only sustained but essential for professional performance; successful interpreting and translation necessitate frequent and often simultaneous access to both lexicons, sometimes within the same communicative exchange. Empirical evidence from studies on professional interpreters (e.g., Christoffels et al., 2006; Ibáñez et al., 2010; Hervais-Adelman et al., 2015) indicates that their work involves continuous, rapid switching between languages and the precise regulation of activation and inhibition. Rather than engaging in prolonged or global suppression of the non-target language, these individuals employ task-specific inhibitory control, activating suppression only when interference is likely (Green & Abutalebi, 2013).

Furthermore, longitudinal research indicates that the training and sustained practice involved in professional interpreting and translation enhance dynamic control mechanisms, strengthening selective attention, conflict monitoring, and cognitive flexibility (Yudes et al., 2011). These skills are directly relevant to regulating cross-linguistic activation while avoiding long-term detriment to either language. For individuals with extensive professional experience, such cognitive flexibility may mitigate potential L1 attrition, allowing them to maintain balanced proficiency despite intensive and frequent L2 use. Thus, the L1 co-activation observed in the present study is more appropriately viewed as indicative of a highly interactive and effectively regulated bilingual lexicon, shaped and optimised by the participants' professional linguistic expertise.

5.4 No Evidence for Non-Target Activation in Intermediate L2 Speakers

5.4.1 Weaker or Slower Lexical Activation

One plausible explanation for the absence of a competitor gaze preference in the intermediate L2 group is that their lexical activation in L2 is weaker and/or slower compared to that of native speakers and advanced L2 users. Research on bilingual word recognition has consistently shown that L2 words take longer to reach activation thresholds sufficient to influence online processing, particularly in less proficient speakers (e.g., Gollan et al., 2008). In the visual world paradigm, where eye fixations and durations typically happen within milliseconds, even slight delays in activation can mean that competitors fail to attract measurable visual attention before the listener has shifted focus to the target or other elements of the scene (Allopenna et al., 1998).

For intermediate learners, weaker lexical representations in L2 may lead to less and more limited activation of phonological and semantic information. This reduced activation strength can be attributed to a lower frequency of exposure, smaller lexical neighbourhoods, and less entrenchment of L2 phonological and semantic mappings (Nation, 2001; Perfetti, 2007). Consequently, even if competitor activation does occur, it may not surpass the visual salience threshold needed to register as a fixation bias in the eye-tracking data. This is consistent with findings from both monolingual and bilingual populations showing that low-frequency or less familiar words elicit weaker and slower competitor effects (e.g., Magnuson et al., 2003; Marian & Spivey, 2003).

In our study, this explanation would account for the absence of gaze preference in both phonological and semantic competitor conditions for intermediate L2 participants. Rather than indicating an absence of cross-linguistic activation per se, the data may instead reflect activation that occurs too late or too weakly to be detected in the time windows used for analysis, thereby producing the appearance of no effect.

5.4.2 Higher Reliance on Bottom-Up Processing

A second, complementary explanation is that intermediate L2 speakers may rely more heavily on bottom-up acoustic processing rather than on top-down predictive mechanisms during L2 comprehension. Skilled listeners, including native speakers and highly proficient bilinguals, are known to integrate contextual cues and lexical knowledge to anticipate upcoming words before they are fully articulated (Huettig & Mani, 2016). In contrast, less proficient L2 users often adopt a more conservative comprehension strategy, waiting for more of the phonetic signal before committing to lexical activation (McQueen & Viebahn, 2007).

Such a strategy prioritises accuracy over speed but comes at the cost of reduced anticipatory processing. In the visual world paradigm, this means that competitor activation—particularly in target-absent experimental settings—may be delayed until the critical word is nearly or fully recognised. By that point, opportunities for measurable competitor-driven fixations have passed. This is consistent with evidence that lower proficiency listeners exhibit attenuated or absent anticipatory eye movements in predictive processing tasks (e.g., Grüter et al., 2012).

In the present study, such a bottom-up processing orientation may account for the absence of consistent gaze preferences towards non-target competitors among intermediate participants. If these participants relied predominantly on incremental, input-driven processing, competitor activation would likely arise only after the target had been recognised. As a result, it would not manifest within the anticipatory time windows used for gaze analysis, or it may have occurred too late and too sporadically—towards the end of the auditory stimulus—to be reliably detected by our statistical models. This interpretation is consistent with restricted processing capacity accounts, yet it reframes the issue as one of strategic allocation of processing resources rather than a uniform cognitive limitation. In doing so, it is worth discussing the qualitative distinction in processing strategies between intermediate and advanced L2 users.

5.4.3 Limitations in Processing Capacity Allocation

The explanations above point to two related but distinct mechanisms: slower lexical activation and a preference for bottom-up processing, which may, in combination, reflect a broader constraint on the allocation of processing capacity in intermediate L2 comprehension. For the intermediate L2 group, no evidence emerged of predictive processing or co-activation for non-target competitors, with gaze patterns remaining predominantly fixed on the target throughout the predictive window. This pronounced attentional narrowing suggests a processing strategy in which attention was restricted to the most probable referent at each moment, rather than distributed across potential alternatives.

This trajectory may reflect a proficiency-driven transition from effortful, capacity-intensive control to more automatic, resource-efficient processing. Highly proficient L2 users automate lower-level operations—such as phonological decoding and lexical access—thereby conserving cognitive bandwidth for anticipatory mechanisms and yielding faster, more fluent responses (Chee et al., 2001; Tatsuno & Sakai, 2005). In contrast, less proficient bilinguals depend on deliberate, strategic control processes (Abutalebi & Green, 2007; Horiba & Fukaya, 2015), which impose substantial working-memory demands (Dussias & Piñar, 2010). Consequently, intermediate learners must allocate the majority of their limited resources to basic decoding tasks, leaving insufficient capacity for activating, evaluating, and fixating on potential competitors, whether semantic or phonological. This explains why their gaze patterns differ so markedly from those of native speakers and advanced L2 learners, who process input more rapidly and can sustain parallel activation of multiple candidates. In this view, the absence of gaze preference in the intermediate group does not reflect a fundamental inability to activate competitors but rather a processing strategy shaped by resource constraints and proficiency level.

Taken together, the lack of evidence of competitor effects in the intermediate L2 group can be understood as the outcome of multiple, interacting factors. Slower and weaker lexical activation may delay or diminish the emergence of competitor representations, whereas a greater reliance on bottom-up processing limits the extent to which predictive mechanisms are deployed in real-time. These tendencies, in turn, reflect a broader restriction in the allocation of processing capacity, whereby the cognitive resources of intermediate bilinguals are primarily devoted to accurate target recognition at the expense of parallel activation of alternative candidates. By contrast, having automatised lower-level operations, advanced L2 users and native speakers are able to sustain broader lexical activation and to engage in context-driven prediction with minimal cost to overall processing efficiency. This pattern suggests that proficiency plays a central role in influencing both the timing and strength of lexical activation, as well as the balance between effortful control and more automatic processes in bilingual comprehension.

5.5 Limitations

5.5.1 Sample and Stimulus Set Sizes

Most studies comparing L1 and L2 language processing, whether using eye-tracking or

ERP, have employed relatively moderate sample sizes (e.g., 30 L2 participants in Dijkgraaf et al., 2017; 45 L2 participants in Hopp, 2015; 14 participants in Marian & Spivey, 2003; 40 in Mishra & Singh, 2014; 24 L2 participants in Ito et al., 2018). A primary reason for this lies in the need to control participants' linguistic profiles, as variables such as language typology, exposure duration, length of residence in the L2-speaking country, and language test scores are critical in studies comparing L1 and L2 processing abilities. The present study adopted a similar approach, with a moderate sample size to ensure precise control over the proficiency of our L2 participants. All advanced L2 speakers were qualified translators and/or interpreters, and recruitment was conducted under close supervision to ensure their backgrounds reflected high L2 proficiency. Consequently, the intermediate L2 group and the native English control group were matched in size to the advanced group, resulting in approximately 20 participants per group (40 L2 participants in total). Although the sample size in the present study was modest, statistically significant differences were nonetheless observed between the language groups. It is noteworthy that the apparent absence of effect among the intermediate L2 speakers may partly be attributable to the limited statistical power resulting from the modest sample size, especially given the relatively brief duration of the visual world task. Under such conditions, more subtle patterns of predictive processing may remain undetected at the data analysis stage. Therefore, increasing the sample size would enhance statistical power, enabling more robust conclusions to be drawn—especially regarding differences in processing timing and cross-group gaze behaviour. Consequently, a high-powered replication of the present study, while retaining strict controls over the linguistic background, would represent an important avenue for future research.

Beyond sample size, the range of stimuli used in the present study is another aspect that may have influenced the findings. The present study employed 12 unique cloze sentences, resulting in 12 trials in the eye-tracking task, which was a comparatively short design relative to other visual world studies (e.g., 18 trials in Dijkgraaf et al., 2017; 16 trials in Hopp, 2015; 40 trials in Marian & Spivey, 2003; 56 sentences in Ito &

Husband, 2017; 14 sentences in Ito et al., 2018). This reduced length was a deliberate methodological choice to mitigate participant fatigue, given that the experiment was conducted entirely online and the testing environment was beyond the researchers' control. By prioritising a concise, highly focused design, the study sought to minimise cognitive load and thereby capture participants' eye movements with greater precision. Such brevity likely supported consistent engagement and stable performance across trials; however, it also meant that the range of target words and competitor types was necessarily restricted, which may limit the generalisability of our findings. Future research could address this limitation by employing longer experimental designs that incorporate a broader range of lexical categories and sentence structures, thereby broadening the scope of inferences that can be drawn.

5.5.2 Bilingual Profile

Another limitation of the present study concerns the considerable variability observed within the advanced L2 group. Although this group, as a whole, demonstrated statistically significant evidence of predictive mechanisms and robust linguistic activation, closer examination revealed notable inconsistencies in reaction timing and gaze patterns across individual participants. This variability suggests that factors beyond overall proficiency, such as the depth and quality of language immersion, may have influenced real-time processing behaviour. The study sought to differentiate the advanced and intermediate groups partly on this basis: all advanced L2 participants either had ongoing or prior immersive language experience, whereas the intermediate group's exposure to English was predominantly classroom-based. It is plausible that among the advanced participants, those with prolonged and continuous immersion in English-speaking environments developed faster lexical access and more stable predictive patterns than those with less sustained exposure. However, these experiential factors were not systematically quantified in our study. Incorporating a post-experimental survey to document a more comprehensive language background of each participant (e.g., age of onset and cumulative length of L2 immersion) would have allowed for a more precise characterisation of bilingual profiles. Such measures could then be statistically related to gaze behaviour, offering deeper insight into how individual differences in language experience shape predictive processing in highly proficient L2 speakers.

5.6 Implications for Future Studies

The limitations outlined above point to several productive directions for future research. First, increasing sample sizes and composition of the stimulus set while maintaining strict controls over proficiency would strengthen statistical power, potentially revealing subtler patterns of predictive processing—especially in intermediate L2 speakers where effects may be small or emerge late in the time course. Likewise, more detailed profiling of bilingual experience (e.g., age of onset, immersion history, domains of use) could clarify the role of experiential factors in shaping real-time processing efficiency.

Building on this, the marked divergence in processing strategies between intermediate and advanced L2 speakers observed in the present study stresses the value of investigating how cognitive resource allocation evolves with proficiency. Intermediate learners appear to prioritise lower-level decoding, leaving limited capacity for predictive fixations, while advanced speakers can sustain broader lexical activation with minimal cost. Longitudinal or training-based studies could track this strategic shift, identifying the point at which automatised lower-level skills free resources for higher-order anticipatory mechanisms.

Another methodological consideration relates to the L2 activation of L1 phonological equivalents. In the present study, the relatively high phoneme overlap between the Chinese names of the target words and their Chinese phonological competitors may have contributed to the observed activation patterns. This reduced cross-linguistic phonological distance contrasts with Ito et al.'s (2018) findings, where a similar design yielded no such activation. This discrepancy points to a potentially important methodological factor that warrants further investigation. Although beyond the primary scope of the current research, future studies could systematically manipulate semantic and phonological distance when selecting experimental materials or directly examine how these variables influence both the timing and strength of activation during real-time language processing.

Finally, the absence of consistent non-target activation in intermediate L2 speakers raises further questions about the interplay between lexical activation speed, language predictive mechanisms, and real-time sentence processing capacity. Combining eye-tracking with high-temporal-resolution techniques such as EEG and incorporating individual difference measures (e.g., working memory, inhibitory control) would enable a more nuanced understanding of processing variability across proficiency levels.

6. Conclusion

The current study set out to examine how phonological and semantic information is activated and recruited for prediction during real-time sentence comprehension by Mandarin Chinese learners of English at different proficiency levels, using native English speakers as a reference group. Leveraging the visual world eye-tracking paradigm, we analysed fine-grained gaze dynamics under carefully controlled display conditions to assess whether, and to what extent, listeners activate non-target lexical competitors as they process spoken sentences in real time.

The results revealed a clear proficiency-driven divergence in processing patterns. Advanced L2 speakers displayed predictive fixations on both semantic and phonological competitors, with the time course and distribution of looks closely matching those observed for native speakers. This supports that highly proficient bilinguals can sustain broader lexical activation and engage in anticipatory processing without incurring significant cognitive costs. In contrast, intermediate L2 speakers exhibited no reliable evidence of competitor-driven looks: their gaze remained firmly anchored to the target referent, with minimal spread of activation to related items. The absence of competitor effects at this level is most naturally interpreted as heavier reliance on bottom-up decoding, slower or weaker lexical activation, and a more limited allocation of attentional resources to predictive routines during comprehension.

Taken together, these findings contribute to the broader debate on the nature of bilingual predictive processing by demonstrating that high L2 proficiency—particularly when coupled with intensive professional experience—enables more flexible, resource-efficient strategies that facilitate parallel activation of multiple candidates. At the same time, the data temper strong interpretations of inhibitory-control accounts that treat proficiency growth as tantamount to global suppression of the non-target language. Instead, the present pattern accords with models that posit dynamic, context-dependent regulation of cross-linguistic activation: non-target representations are not eradicated but are selectively attenuated or allowed to surface depending on processing speed, cognitive capacity and task workload.

Methodologically, the present study provides insight into the decisive influence of design choices on what predictive effects are recovered. Competitor composition, the degree and locus of phoneme overlap between targets and competitors, and the broader configuration of task parameters all shaped both the detectability and magnitude of anticipatory responses. In visual world sentence processing research, particularly when examining predictive mechanisms, the use of highly constraining contexts and rigorously controlled materials is essential for generating robust and interpretable findings. Even minor modifications to linguistic stimuli or visual display properties may influence the timing, magnitude, and character of predictive responses. Given that reaction times are recorded in milliseconds, such subtle variations in experimental materials or task configuration can exert a substantial impact on both the temporal onset and the strength of observed effects, thereby shaping statistical outcomes. It is therefore imperative that future visual world studies exercise rigorous control over these parameters in order to capture the fine-grained temporal dynamics of prediction with precision. The findings further underline the importance of incorporating a narrowed and carefully controlled bilingual profile into predictive accounts. Dimensions such as language immersion experience and sustained professional engagement with language appear to shape both the adoption and efficiency of predictive strategies. Although the study's moderate sample size necessarily limits the breadth of generalisation, the sharp group-level contrasts documented here furnish a solid foundation for targeted follow-up work on how proficiency, cognitive resource allocation, and predictive mechanisms interact during bilingual comprehension.

In sum, this dissertation extends the visual world literature on Mandarin–English bilingual processing and clarifies how L2 proficiency reorganises the real-time integration of phonological and semantic cues. The evidence indicates that language comprehension and the deployment of prediction are highly flexible: performance reflects not a fixed capacity but a developing skill, sculpted by gains in proficiency, accumulated linguistic experience, and the strategic allocation of cognitive resources. The results also point to the need for researchers to recognise the sensitivity of visual world paradigms to experimental design choices. Consequently, meticulous attention to stimulus construction and selection is essential to ensure valid interpretations and methodological robustness.

Looking ahead, longitudinal studies, larger cohorts, and multimodal methodologies would be valuable for charting the developmental progression from more resource-dependent, bottom-up processing towards the automatic, anticipatory strategies characteristic of highly proficient bilinguals and native speakers. Such work would help specify how proficiency growth, usage experience, and domain-general control interact over time to cultivate the predictive skill evidenced by advanced L2 listeners.

Appendix A. Experimental Cloze Sentences

No.	Cloze Sentence	Target	English Phonological	English Semantic	Chinese Phonological	Distractors
1	To make sushi, the chef went to the market in the morning to buy some ***fish***.	fish (yú)	fist	elephant	yǔ (rain)	hammer, mirror, clock
2	She ground her beans every day, so every morning when she was preparing breakfast, the smell of fresh ***coffee*** filled the kitchen.	coffee (kā fēi)	coffin	water	Jiā fēi (Garfield the cat)	screwer, gloves, camera
3	She added flour, sugar, and eggs to the bowl, mixing them into a smooth mixture. An hour later, a perfect ***cake*** had been baked!	cake (dàn gāo)	cape	turkey	dàn gōng (slingshot)	scooter, brush, pencil
4	He turned off his bedside lamp and rolled over to sleep after finishing the last chapter of his ***book***.	book (shū)	bush	newspaper	shù (tree)	chair, mug, apple
5	People can now walk to the island with ease because the local government built a ***bridge*** last year.	bridge (qiáo)	brick	ferry	qiáng (wall)	fork, eraser, pear
6	It was his first visit to Paris, with excitement, he took a photo of the Eiffel Tower with his new ***camera***.	camera (xiàng jī)	camel	CCTV camera	xiàng quān (dog collar)	spoon, juice box, notebook
7	It is hunting season, Philip woke up at 6 in the morning, carefully polishing his ***gun*** for good luck.	gun (qiāng)	gum	bullet	qiáng (wall)	water bottle, book, cloud

8	After an exhausting day at work, he decided to have a drink with some colleagues. He went to a nearby pub and ordered a pint of ***beer***.	beer (pí jiǔ)	bee	tea	pí qiú (ball)	hairdryer, carjack, lighter
9	He washed his hands thoroughly after gardening, using a bar of ***soap*** to remove the dirt.	soap (féi zào)	sofa	sanitising spray	féi liào (fertiliser)	ring, bowl, socks
10	Florence, Italy, has long been famous for its high-quality leather goods. Before leaving, Sara bought a ***belt*** as a souvenir for her dad.	belt (pí dài)	bell	sash	pí dàn (preserved egg)	ginger, milk, scissors
11	Instead of sending an email, she preferred to write a handwritten ***letter*** to tell her grandpa about her life.	letter (xìn)	lettuce	stamp	xīn (heart)	avocado, cat, tweezers
12	After stepping out of the shower, he grabbed a ***towel*** from the rack to dry himself off.	towel (máo jīn)	tower	napkin	máo yī (jumper)	glasses, laptop, cheese

Appendix B. CUREC Approval

Applicant:
Principal Investigator:
Department: Education

Study title: Investigating Phonological and Semantic Cues in Sentence Processing and Prediction among Mandarin Chinese EFL Learners of Different Proficiency Levels: A Visual World Eye-Tracking Study

(Version: 2.0)

Ethics reference: Education (Educ) DREC - 1260557

Dear,

On behalf of the Committee, I confirm that the above research study described in the application and other supporting documentation submitted to the committee has been carefully considered on behalf of the Education (Educ) DREC in accordance with the University's regulations and policy for ethics approval of research involving human participants, human tissue and/or personal data. The opinion is as follows:

Opinion of Research Ethics Committee: Favourable Opinion

Subject to the following conditions:

Decision Date: 2 Jul 2025, 01:19

Opinion End Date: 19 Oct 2026

If favourable, insurance-provided indemnity arrangements will be in place between the decision date and opinion end date and you may now commence your study activities. Should you plan to continue the research beyond the end date above, it is your responsibility to ensure that you request, and receive, an extension (via amendment) from the committee for indemnity to remain in place. You may be required to provide a justification.

Please note the following:

Amendments: Should there be any subsequent changes to the reviewed study, applications for amendments can be made via the Oxford Ethics Application System (Worktribe Ethics).

Reports: Studies considered by OxTREC are expected to submit an *annual progress report* on each anniversary of study approval, until the study is completed. An end of study report is also required.

Audit: This study may be selected for audit at the discretion of the Research Governance, Ethics and Assurance Team.

Data safety: It is the responsibility of the PI to ensure that all data collected during the course of the study is stored and transferred safely and securely in accordance with University requirements. Further guidance and advice are available from the Research Data Team. Additional information is available at https://researchsupport.web.ox.ac.uk/governance/ethics

Yours Sincerely

Education Ethics Officer

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. Journal of Neurolinguistics, 20(3), 242-275.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419-439.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502-518.

Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583-609.

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

Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language*, *26*(6), 658-672.

Bialystok, E., & Craik, F. I. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science*, *19*(1), 19-23.

Blumenfeld, H. K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*(5), 633-660.

Blumenfeld, H. K., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals.*Journal of Cognitive Psychology*, *25*(5), 547-567.

Chee, M. W., Hon, N., Lee, H. L., & Soon, C. S. (2001). Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. *Neuroimage*, *13*(6), 1155-1163.

Chen, H. C., & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 397.

Christoffels, I. K., De Groot, A. M., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*(3), 324-345.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84-107.

De Groot, A. M., & Nas, G. L. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, *30*(1), 90-123.

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), 20120394.

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), 1117-1121.

Dickey, M. W., & Thompson, C. K. (2009). Automatic processing of wh-and NP-movement in agrammatic aphasia: Evidence from eyetracking.*Journal of Neurolinguistics*, *22*(6), 563-583.

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.

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.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision.*Bilingualism: Language and Cognition*, *5*(3), 175-197.

Dowens, M. G., & Carreiras, M. (2006). The shallow structure hypothesis of second language sentence processing: What is restricted and why?. *Applied Psycholinguistics*, *27*(1), 49-52.

Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. *Second Language Research*, *26*(4), 443-472.

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.

Elston-Güttler, K. E., Paulmann, S., & Kotz, S. A. (2005). Who's in control? Proficiency and L1 influence on L2 processing. Journal of Cognitive Neuroscience, 17(10), 1593-1610.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading.*Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461.

Freunberger, D., & Roehm, D. (2017). The costs of being certain: Brain potential evidence for linguistic preactivation in sentence processing. *Psychophysiology*, *54*(6), 824-832.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification.*Memory & Cognition*, *33*(7), 1220-1234.

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*(3), 787-814.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. Bilingualism: Language and Cognition, 1(2), 67-81.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515-530.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem?. *Second Language Research*, *28*(2), 191-215.

Grüter, T., & Rohde, H. (2013). L2 processing is affected by RAGE: Evidence from reference resolution. Inthe *12th conference on Generative Approaches to Second Language Acquisition (GASLA)*.

Grüter, T., Rohde, H., & Schafer, A. (2014). The role of discourse-level expectations in non-native speakers' referential choices. In*Proceedings of the Annual Boston University Conference on Language Development*.

Henderson, J. M., & Ferreira, F. (2013). Scene perception for psycholinguists. In *The interface of language, vision, and action* (pp. 1-58). Psychology Press.

Hervais-Adelman, A., Moser-Mercer, B., & Golestani, N. (2015). Brain functional plasticity associated with the emergence of expertise in extreme language control. *NeuroImage*, *114*, 264-274.

Hintz, F., & Huettig, F. (2015). The complexity of the visual environment modulates language-mediated eye gaze. In *Attention and Vision in Language Processing* (pp. 39-55). New Delhi: Springer India.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33-56.

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching*, *53*(3), 277-306.

Horiba, Y., & Fukaya, K. (2015). Reading and learning from L2 text: Effects of reading goal, topic familiarity, and language proficiency. 27(1), 22-46.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23-B32.

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica*, 121(1), 65-80.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460-482.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151-171.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118-135.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19-31.

Ibáñez, A. J., Macizo, P., & Bajo, M. T. (2010). Language access and language selection in professional translators. *Acta psychologica*, *135*(2), 257-266.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157-171.

Ito, A., & Husband, E. (2017). How robust are effects of semantic and phonological prediction during language comprehension? A visual world eye-tracking study. *Technical Committee on Thought and Language*, *117*(149).

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.

Kaan, E., Dallas, A., & Wijnen, F. (2010). Syntactic predictions in second-language sentence processing. In *Structure preserved: Studies in syntax for Jan Koster* (pp. 207-214). John Benjamins Publishing Company.

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.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003a). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133-156.

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003b). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37-55.

Kamide, Y. (2008). Anticipatory processes in sentence processing.*Language and Linguistics Compass*, *2*(4), 647-670.

Koch, E., Bulté, B., Housen, A., & Godfroid, A. (2021). Using verb morphology to predict subject number in L1 and L2 sentence processing: A visual world eye-tracking experiment. *Journal of the European Second Language Association*, *5*(1), 115-132.

Köpke, B., & Schmid, M. S. (2006). Language attrition: The next phase. In *First language attrition* (pp. 1-43). John Benjamins Publishing Company.

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, mind, and brain. *Annu. Rev. Linguist.*, *1*(1), 377-394.

Kukona, A., Fang, S. Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23-42.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighbourhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326-338.

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, *98*(1), 74-88.

Lawrence, M. A., & Lawrence, M. M. A. (2016). Package 'ez'. *R package version*, *4*(0), 1620-1631.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., De Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, *144*(4), 394.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

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.

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

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and cognition*, 6(2), 97-115.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*(1), 53-78.

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.

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, *60*(5), 661-671.

Mishra, R. K., & Singh, N. (2014). Language non-selective activation of orthography during spoken word processing in Hindi–English sequential bilinguals: An eye tracking visual world study. *Reading and Writing*, *27*, 129-151.

Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *NeuroImage*, *72*, 120-132.

Nation, I. S., & Nation, I. S. P. (2001). *Learning vocabulary in another language* (Vol. 10, pp. 126-132). Cambridge: Cambridge University Press.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., ... & Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, *7*, e33468.

Paap, K. R., Johnson, H. A., & Sawi, O. (2016). Should the search for bilingual advantages in executive functioning continue?. *Cortex*, *74*, 305-314.

Pavlenko, A. (2000). L2 influence on L1 in late bilingualism.*Issues in applied linguistics*, *11*(2).

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension.*Scientific Studies of Reading*, *11*(4), 357-383.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences*, *36*(4), 329-347.

Preston, M. S., & Lambert, W. E. (1969). Interlingual interference in a bilingual version of the Stroop color-word task. *Journal of Verbal Learning and Verbal Behavior*, 8(2), 295-301.

Rommers, J., Dickson, D. S., Norton, J. J., Wlotko, E. W., & Federmeier, K. D. (2017). Alpha and theta band dynamics related to sentential constraint and word expectancy. *Language, Cognition and Neuroscience*, *32*(5), 576-589.

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.

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.

Schmid, M. S., & Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism*, *7*(6), 637-667.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, 10(3), 281-284.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: modelling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1521.

Staub, A., & Clifton Jr, C. (2006). Syntactic prediction in language comprehension: evidence from either... or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 425.

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, *82*, 1-17.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

Tatsuno, Y., & Sakai, K. L. (2005). Language-related activations in the left prefrontal regions are differentially modulated by age, proficiency, and task demands. *Journal of*

Neuroscience, *25*(7), 1637-1644.

Team, R. C. (2015, February). *Language and Environment for Statistical Computing: R Foundation for Statistical Computing, Vienna, Autriche*.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*(30), 12530-12535.

Van Berkum, J. J., 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.

Van Hell, J. G., & De Groot, A. M. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*(3), 193-211.

Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, *62*, 148-171.

Wu, Y. J., Cristino, F., Leek, C., & Thierry, G. (2013). Non-selective lexical access in bilinguals is spontaneous and independent of input monitoring: Evidence from eye tracking. *Cognition*, 129(2), 418-425.

Yang, C. L., Perfetti, C. A., Tan, L. H., & Jiang, Y. (2018). ERP indicators of L2 proficiency in word-to-text integration processes. Neuropsychologia, 117, 287-301.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *32*(1), 1.

Yudes, C., Macizo, P., & Bajo, T. (2012). Coordinating comprehension and production in simultaneous interpreters: Evidence from the articulatory suppression effect. *Bilingualism: Language and cognition*, *15*(2), 329-339.