

Research Article

Cite this article: Xie, M., Zheng, B. and Muñoz Martín, R. (2026). A dual route of prediction-by-production and prediction-by-association during simultaneous interpreting: Evidence from the visual world paradigm. *Bilingualism: Language and Cognition* 1–16. <https://doi.org/10.1017/S1366728926101084>

Received: 7 May 2025

Revised: 11 January 2026

Accepted: 15 January 2026

Keywords:

prediction-by-production; prediction-by-association; simultaneous interpreting; visual world paradigm; interpreting expertise

Corresponding author:

Binghan Zheng;

Email: binghan.zheng@durham.ac.uk

 This research article was awarded Open Data badge for transparent practices. See the Data Availability Statement for details.

A dual route of prediction-by-production and prediction-by-association during simultaneous interpreting: Evidence from the visual world paradigm

Mingqing Xie¹, Binghan Zheng¹  and Ricardo Muñoz Martín²

¹School of Modern Languages and Cultures, Durham University, UK and ²Department of Interpreting and Translation, University of Bologna, Italy

Abstract

Despite growing interest in prediction during simultaneous interpreting (SI), the real-time processing mechanisms supporting it remain underexplored. This study employed the visual world paradigm to investigate whether interpreters can predict upcoming content while simultaneously interpreting multi-sentence paragraphs and to examine the mechanisms underlying prediction. Interpreting students and professionals simultaneously interpreted four paragraphs embedded with sentences containing a critical verb that manipulates the predictability of the target noun, while viewing visual displays containing a target object, two semantic competitor objects and one distractor object. Both groups made predictive eye movements to the target objects before hearing the corresponding word, indicating interpreters' ability to predict in a challenging task. The observed fixation patterns further suggest the involvement of both prediction-by-production and prediction-by-association during SI. Crucially, professionals showed more flexible attention shifts and efficient cue use, whereas students shifted attention less and used a more cautious prediction strategy.

Highlights

- Successful prediction observed in complex simultaneous interpreting (SI).
- SI involved a dual route of prediction-by-production and prediction-by-association.
- Professionals showed more flexible attention shifts and efficient cue use.
- Students shifted attention less and used a more cautious prediction strategy.

1. Introduction

The role of prediction in simultaneous interpreting (SI) has been widely recognised in both empirical research and interpreter training. Prediction is assumed to mitigate the high cognitive efforts inherent in managing the incoming speech stream, thereby freeing up resources that interpreters can devote to production and self-monitoring (Chernov, 1994) and helping interpreters maintain a manageable time lag between input and output (Gile, 2009). This function is further supported by broader frameworks of cognitive and bilingual processing (de Groot, 2011). Beyond the potential benefits of successful prediction, research in monolingual processing has found that erroneous predictions do not incur additional cognitive costs than instances where no prediction is made (Frisson et al., 2017). Over time, frequent incorrect predictions may promote learning and adaptation, ultimately reducing the likelihood of mistakes (Dell & Chang, 2013).

Early studies that examined prediction in SI mainly focused on identifying anticipatory production where interpreters produce translated equivalents before the corresponding source utterances are delivered (Jörg, 1995; Seeber, 2001; Wilss, 1978), and assessed prediction effects using output-based metrics such as ear-voice span (EVS; Chmiel, 2021; Hodzik & Williams, 2017). While such measures have provided valuable insights, observed variations – whether in the frequency of anticipatory production or in the length of EVS – may be influenced by factors beyond the prediction effect, such as integration difficulty (Pickering & Gambi, 2018) and individual preferences in interpreting strategy (Timarova et al., 2011). More recently, online prediction has been examined using the visual world paradigm (VWP) (Amos et al., 2022, 2023; Liu et al., 2022). These studies demonstrated successful semantic prediction during SI, as evidenced by anticipatory eye movements towards objects before they were named.

Nonetheless, a critical question remains open, namely, what is the mechanism underlying prediction during SI? Several theoretical accounts of prediction in general language processing have been proposed (e.g., Altmann & Mirković, 2009; Huettig, 2015; Pickering & Gambi, 2018; Pickering & Garrod, 2013). Building on the view that prediction engages multiple cognitive

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

mechanisms (Huettig, 2015; Kuperberg, 2007), the present study focuses on two possible routes: *prediction-by-production* – a top-down simulation of upcoming content, and *prediction-by-association* – a bottom-up activation driven by linguistic associations (Amos & Pickering, 2020; Pickering & Gambi, 2018; Pickering & Garrod, 2013). By tracking predictive eye movements using the VWP, we examine whether interpreting students and professionals can predict upcoming content in coherent, continuous discourse, and seek to investigate the relative contribution of these predictive mechanisms during SI.

1.1. Prediction during language comprehension

Previous studies have provided robust evidence for prediction in language comprehension at multiple linguistic levels, including semantic (Altmann & Kamide, 1999; Mani & Huettig, 2012), syntactic (Kamide et al., 2003; Otten et al., 2007) and phonological levels (Ito et al., 2017, 2018). Prediction is known to be modulated by several factors, such as production ability (Zirnszteina et al., 2018), age (Federmeier & Kutas, 2005; Huang et al., 2012; Wlotko & Federmeier, 2012) and language proficiency (Hopp & Lemmerth, 2018; Lozano-Argüelles & Sagarra, 2022). Critically, language proficiency is a key modulator, with a consistent finding that non-native speakers often display reduced predictive processing compared to native speakers. Although some studies have found that they can engage in semantic prediction in a way comparable to native speakers (Dijkgraaf et al., 2017; Ito et al., 2018), L2 speakers appear less likely to generate predictions at the more demanding syntactic and phonological levels (Ito et al., 2018; Mitsugi & MacWhinney, 2016). This reduced prediction, particularly at finer-grained levels, may stem from less automatic lexical access and less efficient syntactic representation building in L2 (Clahsen & Felser, 2006; Ito & Pickering, 2021; McDonald, 2006; Mitsugi, 2017). Additionally, non-native speakers are often more dominant in their L1, so interference from L1 may influence the processing of lexical and grammatical features in L2 (Dussias et al., 2013; Karaca et al., 2021; Spivey & Marian, 1999), whereas language-specific features in L2 may not be used for prediction (Foucart & Frenck-Mestre, 2011; Hopp, 2013; Lew-Williams & Fernald, 2010).

Interpreting experience is assumed to expand lexical and syntactic representations (Huettig & Pickering, 2019; Özkan et al., 2022) and has been found to facilitate prediction during language comprehension. Lozano-Argüelles et al. (2020) reported the modulating effect of interpreting experience on prediction, using a visual world eye-tracking task where participants heard sentences and selected the corresponding word from a display. This non-interpreting L2 task revealed that L2 learners with interpreting experience made earlier and faster predictions than monolinguals and non-interpreter L2 learners under certain conditions. Subsequently, Lozano-Argüelles et al. (2023) examined whether this advantage was related to higher working memory (WM) capacity, and found an interaction between WM capacity and interpreting experience. They demonstrated that while interpreters enhanced their coordination ability through interpreting experience and deployed WM efficiently like monolinguals, non-interpreter L2 learners with higher WM capacity may have struggled with greater decision-making demands: their increased ability to maintain multiple possibilities could slow the selection among competing alternatives, resulting in slower prediction generation. Similarly, Özkan et al. (2022) compared professional interpreters and trainees using a listening task that did not require verbal production and found more efficient use of case-markers for prediction in individuals with more interpreting experience.

They also observed that higher WM capacity facilitated prediction in professionals, but not in trainees. Collectively, these findings provide evidence for an interpreter advantage in prediction during L2 comprehension. This advantage is likely driven by a combination of factors: higher language proficiency facilitating lexical and syntactic activation; more efficient allocation of cognitive resources enabling flexible prediction updates; and greater resilience to task demands, including time pressure and the cognitive load associated with concurrent production.

1.2. Prediction during simultaneous interpreting

SI from L2 to L1, as commonly practised in international institutions, is considered more cognitively demanding than single-language comprehension. The dual demands of understanding inputs in L2 while producing outputs in L1 place considerable strain on memory and language production systems (Gile, 2009; Sweller, 2011). As a result, the increased cognitive load may constrain the interpreter's engagement in predictive processing during SI. Early empirical studies examined prediction during SI through output-based analyses. Jörg (1995), for instance, observed successful anticipation in approximately 50% of all contextually constraining German sentences. In his study, professional interpreters outperformed interpreting students, and German L1 interpreters outperformed English L1 interpreters. Seeber (2001) examined the effects of intonation patterns (i.e., monotonous versus lively) on anticipation during SI and found minimal difference in the number of anticipations between the two conditions. Participants produced faster and more accurate anticipations, fewer incorrect anticipations and more placeholders in the monotonous condition than in the lively condition, potentially reflecting compensatory cognitive efforts due to the absence of intonation (Moser-Mercer et al., 1998). Hodzik and Williams (2017) measured the latencies between sentence-final verbs in the input and the corresponding output. They found significantly shorter latencies in high-constraining than low-constraining contexts, suggesting a role of contextual semantics in guiding expectations.

Building on these earlier findings, recent studies have adopted the VWP, an online method, to capture predictive processing more directly. Liu et al. (2022) presented Dutch-English bilinguals with predictable or non-predictable sentences mediated by the verb (e.g., *De man schilt/takent op dit moment een appel*, “the man peels/draws at the moment an apple”), while eye movements were tracked to a display containing a target (e.g., *apple*) and three unrelated distractors. They found that participants made predictive eye movements to the target objects when interpreting predictable sentences from Dutch (L1) into English (L2), both consecutively and simultaneously. Amos et al. (2022) extended this line of research but adopted a target-absent design by employing visual scenes with three distractor objects and one critical object. For each experimental sentence (e.g., *The dentist asked the man to open his mouth a little wider*), the critical object is either a target (e.g., *mouth* [French: *bouche*]), an L2 (English) phonological competitor (e.g., *mouse* [French: *souris*]), an L1 (French) phonological competitor (e.g., *cork* [French: *bouchon*]) or an unrelated distractor. This design dissociated semantic from phonological prediction. They found that while semantic prediction took place in both groups, neither group demonstrated prediction of phonological features.

The absence of phonological prediction in Amos et al. (2022) is consistent with the finding of Ito et al. (2018), who found that non-native speakers predicted semantic but not phonological information in a language comprehension task. Meanwhile, Ito et al. (2018)

found successful semantic and phonological prediction in native speakers. The results suggest that, for bilinguals using their L2, predictive processing may operate primarily at the semantic level, without necessarily cascading to the phonological level. Nonetheless, Ito et al.'s (2018) finding that native speakers showed phonological prediction supports a more general cascaded model of prediction (Pickering & Gambi, 2018), in which semantic activation precedes and may facilitate subsequent phonological encoding. The alignment between this cascaded model and the structure of the language production system is consistent with the engagement of the production system in prediction, although the specific stages engaged (semantic versus phonological) may vary with language proficiency and task demands. An alternative interpretation, however, is that some comprehenders may not engage the production system for prediction at all, relying instead on associative mechanisms. Such mechanisms could generate the observed semantic-level predictions without phonological activations.

1.3. Prediction-by-production and/or prediction-by-association?

Pickering and Gambi (2018) proposed that prediction during language comprehension relies on two complementary mechanisms: prediction-by-production and prediction-by-association. Prediction-by-association is an automatic, low-cost process driven by statistical regularities from past experiences (Kuperberg & Jaeger, 2016), in which incoming linguistic input indiscriminately activates a network of associated conceptual and lexical representations. In parallel, prediction-by-production involves recruiting the production system to generate expectations about the speaker's intended meaning. This process is more resource-intensive and entails simulating plausible continuations of the utterance based on internally generated representations across multiple linguistic levels. Evidence for production-based prediction includes VWP studies showing the use of visual context and world knowledge to generate structure predictions (Knoeferle et al., 2005; Knoeferle & Crocker, 2006, 2007) and ERP studies revealing alignment between language prediction and production, progressing from semantics to syntax (Otten et al., 2007; Otten & van Berkum, 2008; van Berkum et al., 2005; Wicha et al., 2004) and form (DeLong et al., 2005; Ito et al., 2017). Meanwhile, evidence for association-based prediction comes from VWP studies demonstrating automatic activation of semantically related concepts during predictive processing (Kukona et al., 2011, 2014).

Amos and Pickering (2020) extended the concept of prediction-by-production to SI, proposing that interpreters utilise the production system to anticipate upcoming content across multiple linguistic levels, from semantics to lexical form in the target language. This mechanism enables interpreters to stay ahead of the speech stream, mitigate processing delays and resolve ambiguities more efficiently. However, their account does not address prediction-by-association, revealing a gap in our understanding of how bottom-up inputs contribute to prediction during SI.

Verb-mediated prediction can be used to investigate the relative contribution of prediction-by-production and prediction-by-association. Verb semantics has been shown to be a powerful trigger of anticipatory eye movements. For instance, Altmann and Kamide (1999) demonstrated that upon hearing a verb such as *eat*, participants shifted their gaze towards edible objects (e.g., *cake*) even before these objects were explicitly mentioned. Extending this work, Kamide et al. (2003) showed that listeners combined subject and verb constraints (e.g., *the girl + ride*) to anticipate the more

plausible upcoming referent (e.g., *carousel* over *motorbike*). Such verb-mediated predictions blur the distinction between prediction-by-production and prediction-by-association: an expectation, such as *cake* upon *eat* or a *carousel* upon *girl* and *ride*, could arise from automatic associative links, but it is also compatible with the engagement of the production system in generating expectations based on thematic roles and communicative intention. This blurring is associated with a characteristic of the paradigm: the anticipatory eye movements do not reveal the representational level at which predictions are generated, nor do they directly verify covert imitation. In the prediction-by-production account, covert imitation is taken to involve the activation of production-related representations at one or more linguistic levels as an initial stage of prediction (Pickering & Gambi, 2018). Consequently, while this paradigm is not suited to adjudicate between accounts at the representational level, it offers an ecologically valid window into the behavioural outcomes of real-time predictive processing, making it a suitable testing ground for examining how these two mechanisms interact during complex tasks such as SI.

1.4. The present study

Previous VWP studies on prediction during interpreting used single sentences as auditory stimuli (Amos et al., 2022; Liu et al., 2022), but in realistic interpreting practices, typical utterances involve several coherent sentences. The extended context can foster the overall representation of the meaning of the text beyond the immediate, language-dependent representation, thereby building a *mental model* (Johnson-Laird, 1983) or *situation model* (Kintsch, 1998). However, retrieving contextual information may undermine prediction during SI as it requires the storage and activation of the inter-sentential information, that is, the *memory effort* in Gile's (2002) Effort model for SI, or the functioning of WM (Dong & Cai, 2015). It, therefore, remains unknown when interpreting full paragraphs, whether prediction is impeded by these increased memory costs or facilitated by the richer contextual information.

With prediction defined as the pre-activation of linguistic information, our study used the VWP to examine whether interpreting students and professionals can predict semantic information in their L2 while simultaneously interpreting extended discourse comprising full paragraphs. By examining this demanding SI task, we aimed to investigate the relative contribution of two predictive mechanisms: prediction-by-production (i.e., top-down contextual analysis) and prediction-by-association (i.e., bottom-up activation based on general linguistic associations) (Amos & Pickering, 2020; Pickering & Gambi, 2018).

Participants listened to and simultaneously interpreted English paragraphs into Chinese while viewing visual displays. The embedded experimental sentence contained a critical verb that created either a high-predictability (e.g., *eat*) or low-predictability (e.g., *buy*) context for a target word (e.g., *In the station store, commuters are eating/buying freshly made bread.*). Meanwhile, the non-experimental sentences served to establish a coherent global context. Each visual display showed four objects, defined by their relationship to the critical verb and the global context:

- 1) The target (e.g., *bread*): the object explicitly mentioned in the source text. It was the only object that was contextually plausible in both conditions.
- 2) Competitor1 (e.g., *juice*): an object semantically and strongly associated with the unpredictable verb (e.g., *buy juice*), and plausible in the low-predictability condition.

- 3) Competitor2 (e.g., *turkey*): an object semantically compatible and strongly associated with the predictive verb (e.g., eat turkey), but implausible within the global context (commuters are unlikely to eat a whole turkey at a train station in the morning).
- 4) The distractor (e.g., *bone*): an object unrelated to both critical verbs and implausible in both conditions.

This design allowed us to test three hypotheses: (1) If interpreters relied primarily on prediction-by-association, they would direct more predictive fixations towards the object with the strongest verb–noun association, that is, Competitor2 in the high-predictability condition (e.g., turkey for eat) and Competitor1 in the low-predictability condition (e.g., juice for buy). (2) If interpreters relied primarily on prediction-by-production, they would direct more predictive fixations towards the most contextually plausible object, that is, the target in the high-predictability condition (e.g., bread for eat) and on both the target and Competitor1 in the low-predictability condition. (3) If interpreters relied on a dynamic interaction between these two mechanisms, they would direct more predictive fixations towards the target in the high-predictability condition, where the context strongly supports it, and towards the plausible competitor in the low-predictability condition, where associative guidance exerts greater influence in the absence of strong contextual cues.

Finally, by testing both interpreting students and professionals, we also examined expertise-related differences. We expected the professionals to demonstrate more efficient predictive processing, characterised by faster prediction of the target and more effective suppression of irrelevant objects than the students.

2. Method

2.1. Source text preparation

Adapted from a travel story, the source text was divided into four paragraphs, comparable in length and readability (see [Supplementary Appendices 1 & 2](#)). The adapted source text was also evaluated for naturalness by two native English speakers. Each stimulus paragraph included five verb-mediated experimental sentences (e.g., *In the station store, commuters are eating/buying freshly made bread.*), three in the high-predictability condition with predictive verbs (e.g., *eating*) and two in the low-predictability condition with unpredictable verbs (e.g., *buying*). All 20 experimental sentences had the same subject–verb–object structure, with adjective phrases (e.g., *freshly made*) separating verbs and objects. These adjective phrases are three-to-six syllables long, thereby creating a time window allowing for potential predictive eye movements upon hearing the critical verbs. Two counterbalanced versions of the source text were created, with the predictive verbs in one version replaced with unpredictable counterparts in the other version, and vice versa.

2.1.1. Word length and frequency

The critical verbs were controlled for word length and lexical frequency to ensure comparable difficulties in recognising spoken critical verbs across conditions. The mean numbers of syllables of predictive and unpredictable verbs were 1.85 ($SD = 0.86$) and 1.85 ($SD = 0.81$), respectively, with no significant difference between the two groups ($p > .1$). Word frequencies were log-transformed from the Corpus of Contemporary American English (COCA). The mean word frequency for the predictive critical verbs was 4.37 ($SD = 0.58$), significantly lower than the mean for the unpredictable critical verbs (Mean = 4.83; $SD = 0.58$; $t(38) = -2.536$; $p = .015$).

This difference does not undermine the predictability effect, as predictive eye movements are expected in the high-predictability condition but not in the low-predictability condition. On the contrary, even though unpredictable verbs are of higher frequencies and thus are expected to be comprehended and integrated into the context more quickly, this does not necessarily facilitate prediction for the target.

2.1.2. Cloze test

To measure the predictability of the target words, a cloze test was administered to 48 students (Mean age = 23.73 years, $SD = 1.12$) enrolled in a Master's programme in Translation and Interpreting (MTI), all of whom were native speakers of Mandarin Chinese and second-language speakers of English. Participants were presented with the source texts in which the target words from the experimental sentences were omitted. For each sentence, four object options were provided, and participants were instructed to select the most likely object to appear in the context and to supply a label for the chosen item. After excluding invalid responses (0.5%), the average predictability score for target words in the high-predictability condition was .88 ($SD = .11$), significantly higher than in the low-predictability condition (Mean = .63, $SD = .24$; $t(27) = 2.84$, $p = .008$). The cloze probability in the low-predictability condition was above chance level (.25). This is likely because the experimental sentences were embedded in a coherent discourse, which created a global context that differentially constrained expectations for the four objects. Specifically, while the target remained the most plausible referent overall (e.g., *bread* for *eat*), the discourse rendered Competitor2 (e.g., *turkey*) and the distractor (e.g., *bone*) highly implausible, thereby elevating the relative salience of the target and Competitor1, both of which remained contextually plausible.

2.2. Visual stimuli preparation

Each visual display consisted of four objects representing the target, the distractor and the two competitor words, matched to their corresponding experimental sentence (example given in [Figure 1](#)). Objects were depicted with monochrome line drawings, and their positions were randomised using a Latin-square design within a (virtual) 2×2 grid. The four objects in each display were unrelated perceptually or linguistically, and they were not mentioned in the preceding non-experimental sentences, thereby preventing any

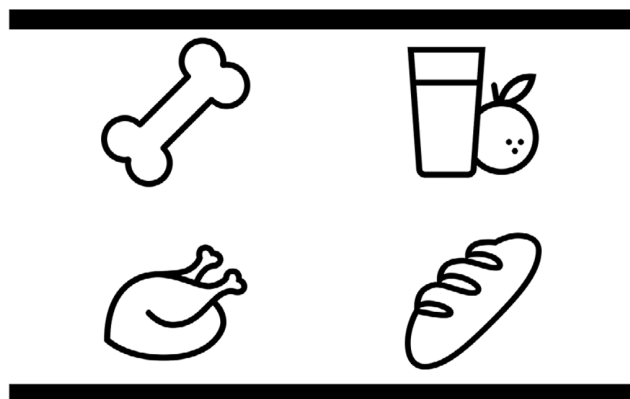


Figure 1. Example display for the experimental sentence: *In the station store, commuters are eating/buying freshly made bread.* The target, bread, on the lower right; Competitor1, juice, on the upper right; Competitor2, turkey, on the lower left; and the distractor, bone, on the upper left.

priming of the experimental objects prior to the critical trial. A series of validation procedures was taken to prepare and evaluate the visual stimuli.

2.2.1. Free association test

A free association test, adapted from Hintz et al.'s (2017) free verb–noun association test, was used to select the semantic competitor and distractor words. An independent group of MTI students were recruited to participate in this test ($N = 48$; mean age = 22.83; $SD = 1.26$). Participants were provided with a list of all the critical verbs and asked to generate the first three nouns that came to mind upon seeing these verbs. The average association strength of the target words with the predictive critical verbs (Mean = .14) is slightly higher than that with unpredictable critical verbs (Mean = .07), but the difference is not significant ($t(34) = 1.652$; $p = .108$). Competitor and distractor words were selected based on the association test results (Table 1). Competitor1 exhibited similar or higher association strengths with corresponding unpredictable verbs than the target words and unpredictable verbs did. Competitor2 had similar or higher association strengths with predictive verbs than the target words and predictive verbs. Distractors had almost no association with either predictive or unpredictable verbs.

2.2.2. Word frequencies of visual objects

The mean log-transformed word frequency of the target words was 4.12 ($SD = 0.47$), significantly lower than that of Competitor1 (Mean = 4.53; $SD = 0.39$; $t(36) = -2.937$; $p = .006$) and Competitor2 (Mean = 4.82; $SD = 0.56$; $t(37) = -4.192$; $p < .001$), but not significantly different from that of the distractor words (Mean = 4.33; $SD = 0.47$; $t(37) = -1.383$; $p = .175$). There was no significant difference among the two competitor words and the distractor words. Given the limited power of corpus frequency norms on L2 lexical processing, a subjective frequency rating was adopted as a complementary measure (Brysbaert & New, 2009; Chen & Dong, 2019). MTI students ($N = 48$; mean age = 23.12; $SD = 1.33$) were instructed to assess word frequencies on a 7-point Likert scale with 7 being the highest frequency level and 1 being the lowest. The results show that the subjective ratings were generally similar across the four types of words (Mean: the target = 5.42 ($SD = 1.01$), Competitor1 = 5.38 ($SD = 1.24$), Competitor2 = 5.67 ($SD = 0.84$), Distractor = 4.98 ($SD = 0.98$); $F(3, 76) = 1.55$, $p = .21$).

2.2.3. Visual similarity test and naming test

A visual similarity test, adapted from Hintz et al. (2017), and a naming test were conducted to evaluate (a) the extent to which the typical visual representations of the target, the competitor and the distractor words resembled the intended referents, and (b) whether these visual objects could invoke the intended concepts of the words. The participants of the visual similarity test were Chinese

students studying at British universities ($N = 11$; mean age = 25.64; $SD = 4.15$). They were instructed to rate the extent to which each printed word was represented by a corresponding visual object on a rating scale from 0 (no similarity) to 10 (identical). The average visual similarity rating across all objects was 9.65 ($SD = 0.42$), indicating that the visual objects accurately depicted their corresponding printed words.

Another group of Chinese students from UK universities ($N = 13$; mean age = 25.70; $SD = 3.35$) were recruited to provide a name for each of 144 randomised objects. The naming consistency was calculated by dividing the number of participants who produced the intended name for each object by the total number of participants. The overall mean naming consistency was .96 ($SD = .08$). The conceptual representations invoked by the given objects could vary across cultures, but the participants in the naming test shared the same cultural background as those in the formal experiment. Therefore, it was assumed that the visual stimuli would invoke similar conceptual representations in both groups.

2.3. Formal experiment

2.3.1. Participants

22 professional interpreters and 44 MTI students participated in the formal experiment. This sample size was determined by recruiting the maximum number of participants feasible within the project's constraints, thereby balancing statistical considerations with the practical realities of working with a specialist and limited population. In fact, our resulting sample is consistent with, and in several cases exceeds, established norms within the field (e.g., Amos et al., 2022; Chmiel, 2021; Hodzik & Williams, 2017; Liu et al., 2022)¹. All participants were based in mainland China at the time of testing. They were native speakers of Mandarin Chinese who had acquired English (L2) through formal education and had received at least one year of professional interpreting training. Data from nine student participants were excluded from analyses due to excessive eye-tracking data loss: that is, in over 75% of trials, their eye-tracking loss rates were higher than the average eye-tracking loss rate of all participants by more than one standard deviation. The final sample consisted of 22 professionals and 35 students.

Table 2 presents a detailed profile of the participants' language background. The professionals were significantly older than the students and had a later age of L2 acquisition and greater exposure to the L2. The student group reported no professional interpreting experience at the time of testing. There were no significant differences between the two groups in their L2 proficiencies. The significant age difference is a common and often unavoidable confound in expertise studies comparing students to established professionals (Chmiel, 2021; Özkan et al., 2022). In our analytical approach, we, therefore, treated expertise as our primary theoretical variable of interest, while interpreting the findings with an awareness that its effect may be entangled with age and life experience in the present design. All participants reported normal or corrected-to-normal vision and no history of language disorders or hearing impairments. Prior to participation, they were informed about the purpose and procedures of the experiment and provided written informed

Table 1. Association strengths between critical verbs and names of the non-target objects

Predictability	Competitor1		Competitor2		Distractor	
	High	Low	High	Low	High	Low
Max	.21	.94	.72	.38	0	.02
Min	0	0	0	0	0	0
Mean	.02	.16	.33	.04	0	.00
SD	.06	.29	.25	.10	0	.00

¹A post hoc simulation-based power analysis was conducted using the final growth curve analysis models with the *simr* package in R (Green & MacLeod, 2016). Based on 100 simulations per group, the estimated power to detect the fixed effect of predictability was 100% (95% CI: 96.38%, 100%) for both groups.

Table 2. Background information of the two groups and t-test comparison results

	Professional (<i>n</i> = 22)	Student (<i>n</i> = 35)	Comparison
Age (years)	32.27 (<i>SD</i> = 5.15)	23.71 (<i>SD</i> = 2.56)	$p < .001$ ***
Age (years) of acquisition of L2	10.00 (<i>SD</i> = 1.66)	7.69 (<i>SD</i> = 2.59)	$p < .001$ ***
Time (years) of exposure to L2	22.26 (<i>SD</i> = 4.25)	15.97 (<i>SD</i> = 2.88)	$p < .001$ ***
Time (years) in profession	7.77 (<i>SD</i> = 4.30)	–	–
Days of conference interpreting per year	74.18 (<i>SD</i> = 45.90)	–	–
TEM-8	77.88	78.74	$p = .531$
Self-rated English proficiency			
Listening	6.00 (<i>SD</i> = 0.76)	5.69 (<i>SD</i> = 0.87)	$p = .168$
Reading	6.23 (<i>SD</i> = 0.75)	6.00 (<i>SD</i> = 0.73)	$p = .262$
Speaking	5.86 (<i>SD</i> = 0.84)	5.54 (<i>SD</i> = 0.78)	$p = .147$
Writing	5.50 (<i>SD</i> = 0.80)	5.46 (<i>SD</i> = 0.78)	$p = .842$

Note: * $p < .05$, ** $p < .01$, *** $p < .001$. TEM-8 stands for Test for English Major – Grade 8, an examination for university students majoring in English. The self-rated English proficiency was assessed on a 7-point Likert-type scale (from 1 = “very low” to 7 = “very high”).

consent. Ethical approval for the study was obtained from the Ethics Committee of Durham University.

2.3.2. Stimuli

The auditory stimuli were recorded by a male native English speaker with a standard American accent. The speaker read the source text at a consistent rate of approximately 2.5 syllables per

second, or 110 words per minute. A 5-second pause was inserted after every two-to-four sentences. Each paragraph lasted approximately 4 minutes long, including the interspersed 5-second pauses. Two versions of auditory stimuli were created, corresponding to the two versions of the source text. Each visual display appeared on the screen at least 1.5 s prior to the critical verb onset, allowing participants with sufficient time to preview and identify all displayed objects (Huetting *et al.*, 2011), and disappeared 1 s after the target word offset (see Figure 2). Between trials, a fixation cross was displayed at the screen’s centre to control the starting point of eye movements before each visual stimulus was presented.

2.3.3. Apparatus

The visual stimuli were presented on a 23.8-inch EIZO FlexScan EV2451 monitor at a resolution of 1920 × 1080 pixels (refresh rate 55–76 Hz). The source text recordings were played through a Sennheiser PC8USB headset. Eye movements were registered using a Tobii Pro Spectrum eye tracker with a 600 Hz sampling rate. The viewing distance was around 65 cm.

2.3.4. Procedure

One day before the formal experiment, participants received a glossary of potentially challenging words and phrases, which they were instructed to review in advance. None of the words and phrases included in the glossary appeared in the experimental sentences. The formal experiment started with a brief orientation outlining the procedure and consisted of a warm-up session followed by four study sessions. The warm-up paragraph, approximately 1 minute in duration, provided experiment instructions and introduced the narrative context of the travel story to be interpreted. The warm-up material did not include any of the experimental sentences.

Each session started with a 5-point calibration of the eye tracker. The participants were asked to keep their heads still following calibration and throughout each session, and to focus on the central fixation cross whenever no visual stimulus was presented. Participants were informed that objects presented in the visual displays might or might not be relevant to the spoken sentences, and were

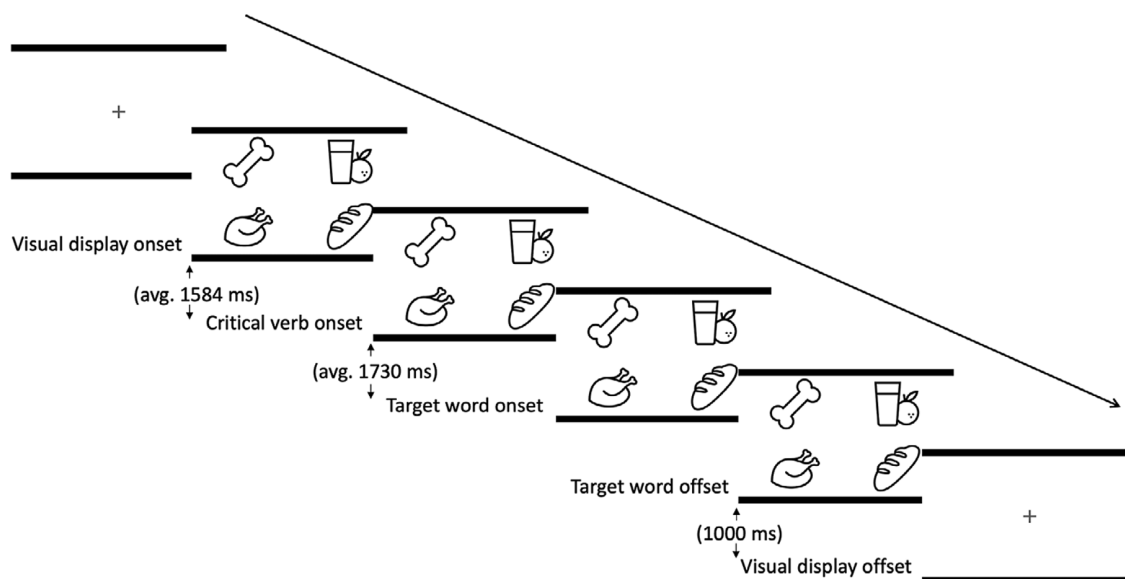


Figure 2. Timeline of a visual display for a single experiment sentence. The critical verb onset was at about 1600 ms (Mean = 1584 ms, *SD* = 231 ms) after the visual display onset, and the target word onset was at about 1750 ms (Mean = 1730 ms, *SD* = 337 ms) after the critical verb onset. The visual display disappeared 1000 ms after the target word offset.

encouraged, but not required, to attend to the displays. Their visual search behaviours, therefore, remained autonomous. Participants were assigned to one of two versions of the auditory stimuli. A retrospective interview after each interpreting task was conducted, during which participants assessed the task's difficulty and the speech rate, and were invited to describe their interpreting processes. A 5-minute break followed each interview, during which participants were permitted to review the glossary. The total duration of the experiment ranged from 45 to 60 minutes.

3. Data analyses and results

The raw eye-tracking data were processed with Tobii Pro Lab 1.162. Fixations were identified through the Tobii I-VT filter. Four areas of interest (AOIs) were created corresponding to the four objects. Each trial was marked with a critical verb onset and a target word onset. The eye-tracking data were processed and analysed with the *eyetrackingR* package (version 0.2.0, Dink & Ferguson, 2015) in R Studio (version 2024.04.2 + 764, RStudio Team, 2024). A 180-ms forward shift was added to account for saccade programming latency – the delay between the reception of a visual signal and the initiation of an eye movement – which typically ranges from 150 ms to 200 ms (Findlay, 1997; Salverda et al., 2014). The proportion of fixations on each AOI was calculated separately for each 50 ms bin from –1550 ms (180 ms after the critical verb onset) to 1180 ms after the target word onset. Following Ito et al. (2018) and Amos et al. (2022), blinks and fixations outside the four AOIs were included in the calculation of the fixation proportions.

3.1. By-group analysis

We first examine each group with cluster-based permutation analyses (CPAs; Maris & Oostenveld, 2007; Ito & Knoeferle, 2023) to compare fixation proportions on the four AOIs across the entire time course. This extended window enables the detection of both predictive and post-target processes, providing a comprehensive behavioural validation of speech-to-visual mapping. The CPA utilised one-sided t-tests to identify time bins during which Elog fixation proportions on the target were significantly higher than those on each of the other objects, under each condition. Three contrast labels were created: Target_vs_Competitor1 (contrasting the target against Competitor 1), Target_vs_Competitor2 (contrasting the target against Competitor 2) and Target_vs_Distractor (contrasting the target against the distractor). Statistical significance was determined using a dual criterion: t-values exceeding 2 and p-values below .05. Neighbouring significant bins were grouped into clusters, and cluster-mass statistics were computed as the sum of t-values within each cluster. Permutations ($n = 1000$) were performed by shuffling data within participants for each time bin. For each permutation, the cluster detection and cluster-mass statistic calculations were repeated, and the maximum cluster-mass statistics were stored to create a null distribution. Statistical significance of clusters was evaluated using Monte Carlo p-values derived from this null distribution. Following Amos et al. (2022), we only report clusters longer than 150 ms here.

To isolate the predictive processing from post-target integration and inhibition, growth-curve analyses (GCAs) were performed for each group to model the non-linear fixation dynamics during the prediction window (i.e., –1550 ms to 180 ms relative to the target word onset). This time window ensures that the analysed gaze behaviour was initiated before the processing of the target word

information. GCA models were constructed for each group using the *lmer()* function from the *lme4* package (version 1.1–31.1, Bates et al., 2015). The model evaluated the fixation proportions predicted by fixed effects of condition (high-predictability versus low-predictability) and AOI (Target versus Competitor1, Competitor2 or Distractor), and the interaction of the two on all time terms. The fixation proportions were empirical logit (Elog) transformed, calculated as $\log((Proportion + 0.5)/(1 - Proportion + 0.5))$ (Barr, 2008). The condition effect was sum-contrast coded (0.5 = high-predictability condition; –0.5 = low-predictability condition), and the AOI effect was treatment-coded with the target as the reference. Polynomial terms were determined by comparing model fits, considering visual inspection of the fixation curves, and balancing statistical fit with the risk of Type I error (Huang & Snedeker, 2020) and interpretability (Mirman et al., 2008). Based on these criteria, linear, quadratic and cubic terms were retained in the final models.

Following the recommendation by Barr et al. (2013) to include all theoretically justified random effects, model building began with a maximal random-effects structure, incorporating by-participant and by-trial random intercepts and slopes for all orthogonal polynomial terms. The “bobyqa” optimiser was employed to improve model convergence. When convergence issues arose, an iterative model reduction procedure (Bates et al., 2015) commenced with the removal of item-level correlations, followed by random slopes, until warnings ceased without significant loss of fit. The final model for the professionals specified random effects as $(1 + ot1 + Predictability || Trial) + (1 + Predictability | Participant)$, and for the students as $(1 + ot1 + Predictability || Trial) + (1 + ot1 + Predictability | Participant)$.

3.1.1. The professional interpreter group

The CPA for the AOI effect in the professional group (Figure 3) identified significant clusters where fixation proportions on the target exceeded those on the other three objects before the target word onset in the high-predictability condition but not in the low-predictability condition (full results in Supplementary Appendix 3). In the high-predictability condition, the significant clusters before the target word onset demonstrate that the professionals preferentially fixated on the target before it was explicitly mentioned, suggesting successful target word prediction. The fact that significant clusters for Target_vs_Competitor2 and Target_vs_Distractor appeared earlier than those for Target_vs_Competitor1 suggests that the professionals excluded Competitor2 and distractor objects before Competitor1. After the target word onset, the significant clusters reflect confirmatory target fixations and integrate the target word into their comprehension and/or production process.

In the low-predictability condition, significant clusters appeared only after the target word onset for Target_vs_Competitor1 and Target_vs_Competitor2. The absence of significant clusters for Target_vs_Competitor1 and Target_vs_Competitor2 before target word onset indicates a delay in fixating on the target. This also supports that fixations were more evenly distributed among the AOIs in the low-predictability condition. The two significant clusters for Target_vs_Distractor before the target word onset are likely driven by random fluctuations or subtle biases, as they do not consistently favour the target.

The GCA for the professional interpreter group revealed significant interactions between condition and AOI, reflecting differential temporal dynamics in fixation proportions across the four objects as a function of predictability (model fit in Supplementary Appendix 4). Specifically, fixations on Competitor1 were significantly reduced in the

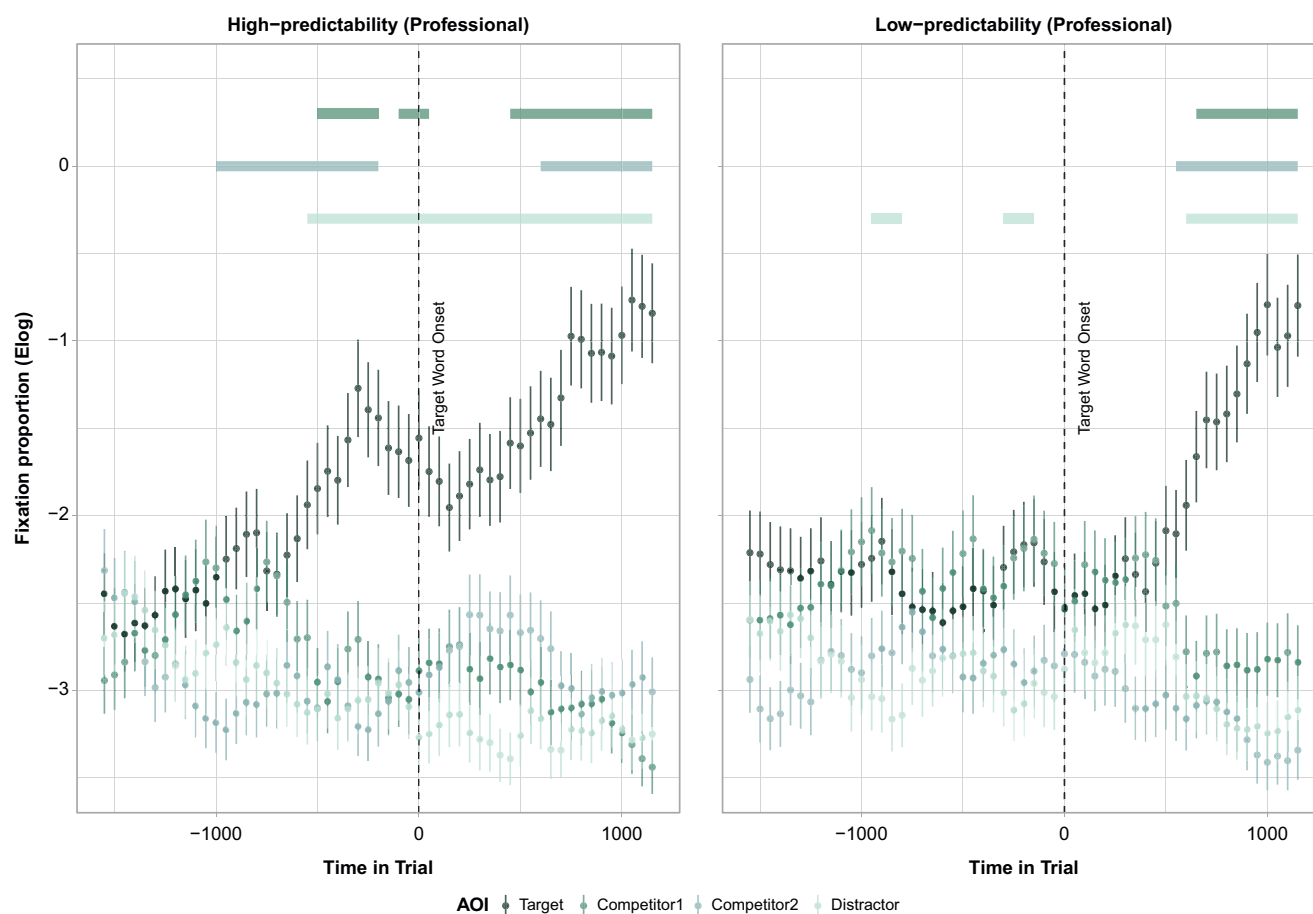


Figure 3. Time course of the fixation proportions of the four objects under each condition, with the results of CPAs for the AOI effect for the professional interpreter group. The lines on the top ($y = c(-0.3, 0, 0.3)$) indicate the clusters where fixation proportions on the target were significantly higher than each of the non-target objects respectively.

high-predictability condition compared to the low-predictability condition ($\beta = -0.643$, $SE = 0.074$, $t = -8.729$, $p < .001$). Similar reductions were observed for Competitor2 ($\beta = -0.372$, $SE = 0.074$, $t = -5.052$, $p < .001$) and the distractor ($\beta = -0.342$, $SE = 0.074$, $t = -4.647$, $p < .001$). Post hoc analysis of estimated marginal means further showed that in the low-predictability condition, the target and Competitor1 attracted comparable fixation proportions ($p = .999$), both significantly higher than those for Competitor2 and the distractor ($ps < .001$; see [Supplementary Appendix 5](#)). These findings indicate anticipatory fixations on the target in the high-predictability condition, and a more diffuse pattern in the low-predictability condition, with attention distributed primarily between the target and Competitor1.

In terms of temporal dynamics, significant interactions between predictability and AOIs emerged mainly on the linear term: Competitor1 ($\beta = -3.286$, $SE = 0.436$, $t = -7.538$, $p < .001$); Competitor2 ($\beta = -3.218$, $SE = 0.436$, $t = -7.382$, $p < .001$); and the distractor ($\beta = -3.037$, $SE = 0.436$, $t = -6.968$, $p < .001$). On the quadratic term, only Competitor2 exhibited a significant interaction ($\beta = 1.725$, $SE = 0.436$, $t = 3.953$, $p < .001$), while on the cubic term, only Competitor1 showed a significant interaction ($\beta = 1.226$, $SE = 0.436$, $t = 2.813$, $p = .005$). The robustness of the findings was confirmed by a supplementary analysis on a randomly subsampled dataset, which yielded the same pattern of results (see [Supplementary Appendix 6](#)). The significant interactions on the linear term indicate that, in the high-predictability condition, the

professionals initiated target fixations and suppressed non-target fixations rapidly during the prediction window, whereas in the low-predictability condition, the between-AOI differences remained relatively stable until approximately 500 ms after the target word onset.

3.1.2. The interpreting student group

The CPA for the AOI effect in the interpreting student group ([Figure 4](#)) was largely consistent with the patterns observed in the professional interpreter group (full results in [Supplementary Appendix 3](#)). In the student group, significant clusters also appeared prior to the target word onset in the high-predictability condition but not in the low-predictability condition, indicating a modulating effect of contextual predictability on anticipatory eye movements to the target. Compared to the professionals, the students exhibited fewer but longer clusters, suggesting less frequent shifts of visual attention.

The GCA results for the interpreting student group revealed patterns that differed slightly from those of professional interpreters. In the student group, the high-predictability condition was associated with reduced fixation proportions on the non-target objects (model fit in [Supplementary Appendix 4](#)). Significant interactions were observed between the high-predictability condition and Competitor1 ($\beta = -0.463$, $SE = 0.054$, $t = -8.604$, $p < .001$), Competitor2 ($\beta = -0.363$, $SE = 0.054$, $t = -6.745$, $p < .001$) and the distractor ($\beta = -0.339$, $SE = 0.054$, $t = -6.314$, $p < .001$). Post hoc

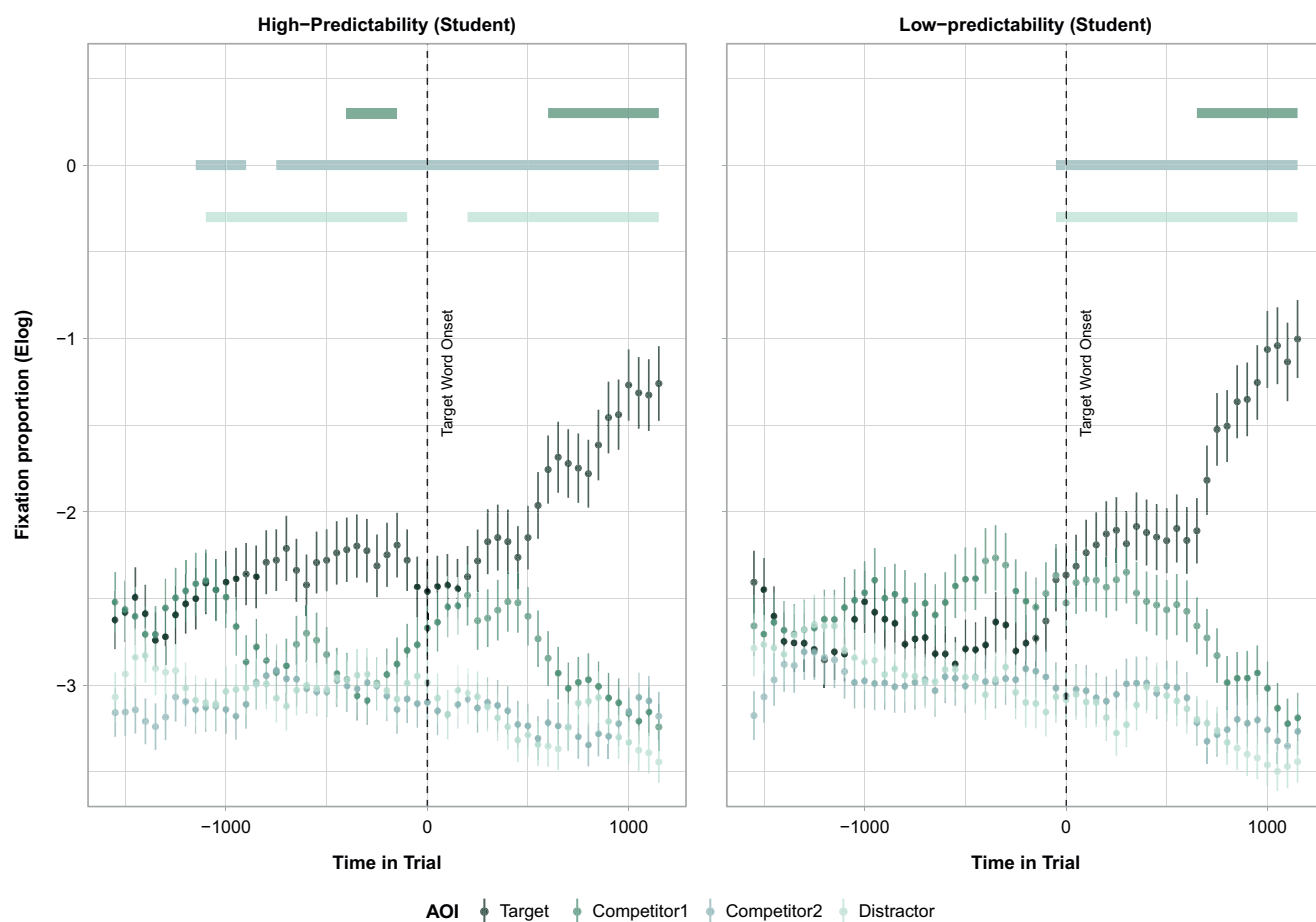


Figure 4. Time course of the fixation proportions of the four objects under each condition, with the results of CPAs for the AOI effect for the interpreting student group. The lines on the top ($y = c(-0.3, 0, 0.3)$) indicate the clusters where fixation proportions on the target were higher than each of the non-target objects respectively.

analysis further showed that in the low-predictability condition, Competitor1 attracted more fixation proportions than the target ($p = .002$), and both were fixated significantly more than Competitor2 and the distractor ($ps < .001$; see [Supplementary Appendix 5](#)). These results indicate that, unlike the professionals, the students tended to prioritise Competitor1 over the target and the other non-target objects in the low-predictability condition.

The major differences between the two groups were observed in temporal interactions. For the professionals, significant three-way interactions mainly involved the linear term. By contrast, for the students, these were observed mainly on the quadratic term: Competitor1 ($\beta = 1.902$, $SE = 0.318$, $t = 5.987$, $p < .001$), Competitor2 ($\beta = 1.028$, $SE = 0.318$, $t = 3.235$, $p = .001$) and the distractor ($\beta = 1.306$, $SE = 0.318$, $t = 4.111$, $p < .001$). These significant quadratic interactions align closely with the CPA findings, which demonstrated significant clusters appearing before the target word onset in the high-predictability condition. On the linear and cubic terms, only Competitor1 showed significant interactions with the high-predictability condition (linear: $\beta = -1.208$, $SE = 0.318$, $t = -3.802$, $p < .001$; cubic: $\beta = 1.148$, $SE = 0.318$, $t = 3.606$, $p < .001$). Results from the robustness check confirmed these main findings (see [Supplementary Appendix 6](#)). This pattern suggests that the students' suppression of non-target objects was not as immediate as that of the professionals. Instead, the significant quadratic effects reflect a delayed change in the students' gaze patterns, consistent with a less efficient inhibitory response to predictive cues.

3.2. By-trial analysis

A by-trial CPA was conducted to estimate the continuous influence of the core mechanisms underlying our hypotheses. This analysis examined how trial-by-trial variations in cloze probability (reflecting production-based mechanisms) and verb–noun association strength (reflecting associative mechanisms) shaped the time course of eye movements during the prediction window, thereby complementing the condition-based and AOI-based comparisons. The CPA was based on linear regression models, with trial-aggregated Elog fixation proportions on each object in 50-ms time bins as the dependent variable, predicted by cloze probability and verb–noun association. Given the relatively small sample size (i.e., 40 data points per time bin), which might limit the statistical power of linear regression, the model was fitted on the aggregated data from every two consecutive time bins. Clusters were identified for each predictor as groups of consecutive time bins where the regression estimates were statistically significant and had the same direction (either positive or negative). The cluster mass was calculated by summing the t-statistics within each identified cluster. The data were permuted 1000 times for each predictor by shuffling that predictor across trials, and the regression analysis and cluster detection procedures were repeated for each permuted dataset. Statistical significance of clusters was evaluated following the same method as described in [Section 3.1](#).

For the professional group, the by-trial CPA revealed significant effects of cloze probability on all objects except the distractor

Table 3. By-trial CPA for the effects of cloze probability and verb–noun (VN) association

Professional					
AOI	Predictor	Cluster	Estimate sign	<i>p</i>	Cluster mass
Target	Cloze probability	–1550 to –1250	Positive	< .001***	8.639
		–950 to –650	Positive	< .001***	11.665
		–350 to –150	Positive	< .001***	15.057
	VN association	–1450 to –1250	Negative	< .001***	7.356
Competitor1	Cloze probability	–1550 to –1250	Negative	< .001***	11.726
		–950 to –650	Negative	< .001***	10.517
		–250 to 50	Negative	< .001***	13.404
Competitor2	Cloze probability	–1050 to –750	Negative	< .001***	13.746
Student					
AOI	Predictor	Cluster	Estimate sign	<i>p</i>	Cluster mass
Target	Cloze probability	–1550 to –150	Positive	< .001***	41.961
Competitor1	Cloze probability	–1050 to 150	Negative	< .001***	48.064
Competitor2	Cloze probability	–1550 to –1050	Negative	< .001***	21.554
Distractor	Cloze probability	–50 to 150	Positive	< .001***	7.082

Note: **p* < .05, ***p* < .01, ****p* < .001.

(Table 3). Cloze probability was positively associated with fixation proportions on the target but negatively associated with the two competitors, highlighting its critical role in guiding predictive processing – enhancing fixation on the anticipated target while suppressing competitors. Three significant clusters were identified for both the target and Competitor1, indicating a dynamic and flexible use of cloze probability in prediction strategies. Contrary to our expectation that verb–noun association would be positively associated with fixation proportion, the effect of verb–noun association was significantly negative on the target in the early stage following the critical verb onset (from –1450 ms to –1250 ms).

For the student group (Table 3), only one significant cluster of the cloze probability effect was identified for each object. Compared to the professionals, these clusters tended to have a longer duration, suggesting that cloze probability triggered less frequent eye movements in the students. Interestingly, a positive cluster for the distractor was observed near the target word onset (from –50 ms to 150 ms), indicating an increase in fixation proportion on the distractor during this period. No effect of verb–noun association was observed for any of the four objects.

The transitional probability (TP) has been shown to serve as a predictor of prediction during reading (Frisson et al., 2005; McDonald & Shillcock, 2003), so we conducted a supplementary analysis to examine the role of forward TP in guiding anticipatory eye movements. As an objective measure of lexical co-occurrence

frequency between verbs and nouns, forward TP is conceptually similar to general verb–noun associations measured through free association tasks (Hintz et al., 2017). Forward TP was computed following the approach of Hodzick and Williams (2017), using frequency counts from COCA to calculate the probabilities of the names of the four objects given the critical verb (Table 4). In the high-predictability condition, the forward TPs for the target and for Competitor2 were similar (*p* = .683) and significantly higher than those for Competitor1 and the distractor (*ps* < .01), resulting in the pattern: target ≈ Competitor2 > Competitor1 ≈ distractor. This contrasts with the verb–noun association pattern in the same condition, where Competitor2 exhibited the highest association strength. In the low-predictability condition, the forward TP was similarly high for the target and for Competitor1 (*p* = .871), and significantly higher than those of Competitor2 and the distractor (*ps* < .05): target ≈ Competitor1 > Competitor2 ≈ distractor. Again, this diverged from the verb–noun association pattern, where Competitor1 showed the strongest association. These discrepancies reflect the conceptual distinction between objective distributional statistics (forward TP) and subjectively rated associative strength.

A separate CPA was then conducted in which forward TP replaced verb–noun association as a predictor in the linear regression models, while keeping all other analytical procedures identical. For the professional group (Table 5), forward TP was positively associated only with the target, indicating a facilitation role in

Table 4. Transitional probabilities between critical verbs and names of the four visual objects

	Target	Competitor1	Competitor2	Distractor
High predictability	.016 (<i>SD</i> = .023)	.001 (<i>SD</i> = .002)	.019 (<i>SD</i> = .028)	.000 (<i>SD</i> = .000)
Low predictability	.017 (<i>SD</i> = .029)	.016 (<i>SD</i> = .024)	.002 (<i>SD</i> = .008)	.000 (<i>SD</i> = .001)
Comparison	<i>p</i> = .871	<i>p</i> = .010 *	<i>p</i> = .016 *	<i>p</i> = .479

Note: **p* < .05, ***p* < .01, ****p* < .001.

Table 5. By-trial CPA for the effects of cloze probability and forward transitional probability (TP)

Professional					
AOI	Predictor	Cluster	Estimate sign	<i>p</i>	Cluster mass
Target	Forward TP	–1550 to –450	Positive	< .001***	37.298
		–250 to –50	Positive	< .001***	6.876
Competitor1	Cloze probability	–1550 to –1250	Negative	< .001***	11.812
		–950 to –650	Negative	< .001***	10.571
		–250 to 50	Negative	< .001***	13.360
Competitor2	Cloze probability	–1150 to –750	Negative	< .001***	14.284
Student					
AOI	Predictor	Cluster	Estimate sign	<i>p</i>	Cluster mass
Target	Cloze probability	–1250 to –1050	Positive	< .001***	6.479
		–850 to –450	Positive	< .001***	11.687
	Forward TP	–1250 to –850	Positive	< .001***	12.541
		–550 to 50	Positive	< .001***	18.319
Competitor1	Cloze probability	–1050 to 150	Negative	< .001***	48.129
Competitor2	Cloze probability	–1550 to –1050	Negative	< .001***	19.823
	Forward TP	–250 to 50	Positive	< .001***	9.572

Note: **p* < .05, ***p* < .01, ****p* < .001.

guiding predictive looks to the target. However, the significant clusters of cloze probability on the target observed in the original CPA involving verb–noun associations were absent in this analysis. Patterns for the competitors were largely consistent with those of the original CPA: Competitor1 showed three significant negative clusters for cloze probability, and Competitor2 showed one. No significant effects of forward TP were observed for either competitor, and no clusters reached significance for the distractor.

The student group exhibited a more complicated pattern (Table 5). For the target, both cloze probability and forward TP showed two significant positive clusters, suggesting the combined effects of contextual constraints and associative links on anticipatory eye movements. For Competitor1, cloze probability again showed a significant negative cluster. Competitor2, however, displayed a significant negative cluster for cloze probability alongside a significant positive cluster for forward TP, indicating opposing influences of contextual predictability and statistical association. No significant effects were observed for the distractor.

Furthermore, to rule out the possibility that the adjective phrase (e.g., *freshly made*) preceding the target word facilitates fixations towards a specific object, we conducted a post-hoc analysis on forward TPs as an effect of the adjective phrase on all four objects. The results showed no significant differences across objects ($F(3, 76) = 1.043, p = .379$), indicating that the adjective phrase did not confound the observed effects of the critical verb.

3.3. Between-group analysis

To compare the results from the professional and student groups, a GCM model was run on both groups and specified an interaction by the expert group in the prediction window (model fit in Supplementary Appendix 3). The expertise effect was sum-contrast coded (0.5 = professional; –0.5 = student). The model fitting followed the procedure described in Section 3.1. The final model

specified random effects as $(1 + \text{Predictability} \parallel \text{Trial}) + (1 + \text{Predictability} \parallel \text{Participant})$. The interactions between expertise and condition were significant for the linear ($\beta = 2.232, SE = 0.378, t = 5.908, p < .001$) and quadratic terms ($\beta = 0.816, SE = 0.378, t = 2.160, p = .031$), reflecting faster and more pronounced fixation shifts towards the target in the professionals than in the students, especially under the high-predictability condition. The expertise effect for Competitor1 was evident in a significant interaction with the high-predictability condition on the linear time term ($\beta = -2.078, SE = 0.534, t = -3.888, p < .001$) and on the quadratic terms ($\beta = -1.756, SE = 0.534, t = -3.284, p = .001$), indicating that the professionals demonstrated a steeper decline in fixations on Competitor1 in the prediction window compared to students, particularly in the high-predictability condition.

For Competitor2, expertise effects appeared in interaction with the high-predictability condition only on the linear term ($\beta = -3.336, SE = 0.534, t = -6.244, p < .001$). Specifically, when predictability was high, the professionals showed a sharper reduction in fixations on Competitor2 over time. Significant three-way interactions between the expertise, the high-predictability condition and the distractor were significant on the linear ($\beta = -3.488, SE = 0.534, t = -6.529, p < .001$) and quadratic terms ($\beta = -1.264, SE = 0.534, t = -2.365, p = .018$). These significant interactions demonstrate that the professionals rapidly disengaged from the distractor more rapidly, particularly under high-predictability conditions, while the students displayed a slower and more variable pattern of disengagement, underscoring differences in suppression efficiency between the groups. Results from the robustness check confirmed these main findings (see Supplementary Appendix 6).

4. Discussion

The present study investigated whether professional and student interpreters engage in prediction during SI of multi-sentence

paragraphs, examined its underlying mechanisms, specifically, prediction-by-production and/ or prediction-by association and explored differences related to interpreting expertise. The findings provide compelling evidence for successful semantic prediction during SI in both groups. In the high-predictability condition, both groups demonstrated significantly higher fixation proportions on the target objects than on the non-targets before the target word onset. This anticipatory effect was confirmed by significant pre-target clusters for between-AOI comparisons in the by-group CPAs, and also by the significant interactions between the predictability and non-target objects on the linear and the quadratic terms in the by-group GCAs. By contrast, this anticipatory fixation pattern was absent in the low-predictability condition. Instead, both groups showed delayed shifts of visual attention from the non-target to the target objects, emerging approximately 600 ms after the target word onset. This temporal lag suggests that participants were unable to predict the target in the low-predictability condition.

4.1. A dual route of prediction-by-production and prediction-by-association

Our findings support a dual-route model involving both prediction-by-production and prediction-by-association. The eye movement patterns revealed a dynamic interaction between these two mechanisms. In the high-predictability condition, both groups fixated more on the target than on Competitor2 before the target word onset, even though the critical verb was compatible with both objects and was more strongly associated with Competitor2 than with the target. This pattern suggests that global contextual plausibility overrode local lexical associations, supporting top-down, production-based prediction (Federmeier, 2007; Huettig, 2015; Pickering & Gambi, 2018; Pickering & Garrod, 2013). This finding, centred on the demonstration of context-driven referential anticipation, supports the broader conception of prediction-by-production, which posits that the production system can generate expectations across linguistic levels, from semantics (concept) to lexical forms. Thus, our evidence for this mechanism remains valid irrespective of whether the underlying representation was conceptual or lexical. Conversely, in the low-predictability condition, while both the target and Competitor1 were compatible with the critical verb and plausible in the context, Competitor1 attracted comparable or even more fixations than the target, particularly among student interpreters. The increased fixation on Competitor1 may reflect their stronger association with the unpredictable critical verb, highlighting a role of prediction-by-association in guiding predictive eye movements.

The by-trial CPAs further support this dual-route interpretation. Cloze probability significantly enhanced fixations to the target and suppressed non-target objects, aligning with the production-based account. In contrast, general verb–noun association strength exhibited only a limited contribution to predictive eye movements, showing a short-lived negative effect on target fixations immediately after the critical verb onset in the professional group. This may reflect their “watchful waiting” strategy, a form of anticipatory readiness for unexpected developments of events (Özkan et al., 2022). Specifically, the professionals, with greater cognitive flexibility and more efficient L2 processing abilities than the students (Ito & Pickering, 2021; Kaan & Grüter, 2021), may have remained strategically open to less expected outcomes despite early contextual support.

Forward TP exerted a more consistent facilitatory effect, increasing anticipatory fixations on the target and, among the students, also enhanced looks to the Competitor2. This dissociation between

forward TP and verb–noun association suggests that they capture related but distinct dimensions of association-based prediction. Forward TP, as an objective corpus-based measure of lexical co-occurrence frequency, reflects distributional regularities that support automatic, probabilistic expectations (Frisson et al., 2005; Hodzík & Williams, 2017; McDonald & Shillcock, 2003). On the other hand, verb–noun association strength, as measured through free association tasks, may reflect cognitive salience and conceptual relatedness as represented in semantic memory (Hutchison, 2003; Nelson et al., 2004). This subjective, conceptually grounded measure is sensitive to diverse associative links (Hintz et al., 2017) and may, therefore, vary across language tasks and contexts (e.g., free association task versus SI), whereas forward TP indexes environmental statistical structure and provides a more stable predictor across different language processing scenarios.

This interpretation also aligns with the findings of Hintz et al. (2017). They reported that functional associations were stronger predictors of anticipatory eye movements than general verb–noun associations, suggesting that the association effect may depend on the specific types or directions of associations activated by the processing context. In the current study, the processing context was more complex; thus, it is possible that the effect of general verb–noun associations was overridden by more specific associations that were most relevant to prediction. Future research should aim to disentangle the specific types of associations most relevant during SI and their roles in facilitating prediction.

Additionally, the present study observed higher fixation proportions on the target and Competitor1 than on Competitor2 and the distractor prior to the target word onset in both conditions. The early exclusion of Competitor2 and the distractor likely reflects a thematic priming effect from the global context, which occurred even earlier, before the critical verb onset. Rather than being a result of deliberate strategic prediction, the mental model constructed from prior discourse may have operated more as an associative backdrop, which activated semantically or visually related representations. Items that were congruent with this contextual mental model, for example, *bread* (the target) and *juice* (Competitor1) at a train station in the early morning, received a processing advantage. Such processing advantage guided visual attention in an associative yet context-sensitive manner. By contrast, unlikely objects, for example, *bone* (the distractor), may be suppressed as quickly as possible to minimise interference and conserve processing resources, according to the utility view of prediction proposed by Kuperberg and Jaeger (2016). This interpretation aligns with Ferretti et al. (2001), who showed that verbs can immediately prime typical thematic role fillers (e.g., agents or instruments), indicating that conceptual schemes, such as common situations or events, are rapidly activated during comprehension. These findings suggest that prediction-by-association can operate not only at the lexical level but also at broader discourse levels.

Taken together, our findings demonstrate an interactive, layered model of predictive processing in which both prediction-by-production and prediction-by-association operate in parallel during SI of coherent discourse. Notably, the prediction-by-association is not a unitary construct as initially hypothesised. Instead, it may encompass lexically driven, contextually driven and distributionally based forms of expectation, which are differentially engaged depending on contextual constraints, task demands and interpreter expertise. These findings call for more refined measures of associative strength in future research to better disentangle the contributions of different predictive mechanisms in real-time language processing.

4.2. Expertise-related differences in the predictive processing during SI

Overall, the group-level analysis revealed similarities in prediction between the interpreting students and professionals. Both groups showed robust anticipatory effects in the high-predictability condition, as evidenced by pronounced fixation shifts towards the target object relative to the other three objects. However, in the low-predictability condition, increases in fixations on the target were delayed until the target word onset, with fixations more evenly distributed across the four objects. Consistent with our hypothesis, the professionals were more likely to predict the target than the students, as evidenced by the professionals' higher fixation proportion on the target in the high-predictability condition and the significant quadratic interaction between expertise and condition in the between-group GCA. The professionals also exhibited faster suppression of unrelated objects than the students, especially in the high-predictability condition, supported by the significant interactions between expertise and conditions on the linear term across all three non-target objects.

Meanwhile, the professionals demonstrated greater flexibility in visual attention patterns and strategic use of prediction cues, whereas the students exhibited more static gaze patterns. Specifically, the professionals demonstrated faster and more frequent attention shifts across the four objects in both conditions, as evidenced by multiple shorter temporal clusters in the cloze probability effect observed in the by-trial CPA. Even after forming a prediction in the high-predictability condition, they continued to monitor other objects (see Figure 3). These results are consistent with the findings of Özkan et al. (2022) that professional interpreters returned their gaze to the baseline following initial predictive fixations. Furthermore, the professionals also exhibited more pronounced disengagement from Competitor2 and the distractor, particularly under the high-predictability condition, as indicated by significant linear and quadratic interactions in the between-group GCA. These dynamic patterns of visual attention shifts likely reflect that, as the source text unfolds, professional interpreters continuously update their prediction by integrating prior knowledge and new bottom-up inputs (Kuperberg & Jaeger, 2016). Such rapid and efficient adjustment, including the ability to flexibly shift towards relevant targets and suppress distractors, highlights enhanced cognitive flexibility and superior inhibitory control in professionals. These cognitive enhancements are unlikely to be fully explained by general age-related cognitive changes (Federmeier & Kutas, 2005; Huang et al., 2012) and are more parsimoniously accounted for by the development of task-specific cognitive skills acquired through extensive interpreting experience (Lozano-Argüelles et al., 2020; Lozano-Argüelles & Sagarra, 2022; Özkan et al., 2022).

In contrast, the students made less pronounced and less frequent visual attention shifts, as indicated by the smaller absolute values of estimates on the quadratic terms in the GCA and the presence of a single, longer cluster of the cloze probability effect identified in the by-trial CPA. This aligns with Liu et al. (2022), who found that some participants did not move their eyes at all. This suggests that the extreme cognitive demands of SI may have hindered the students' ability to actively update prediction by integrating prior contextual knowledge and new incoming bottom-up inputs, an ability more readily observed in the professionals. Instead, when encountering a potential continuation for the speaker's utterances, especially one that strongly aligned with their internal representation of context, the students appeared more likely to settle for a "good enough" interpretation (Ferreira, 2003; Kuperberg, 2007;

Kuperberg & Jaeger, 2016) to conserve cognitive and metabolic resources for cognitive sub-processes. As a result, the students relied more on a shallow processing of broader contextual cues, leading to less precise target identification.

5. Conclusion

This study adopted the VWP to examine predictive processing during SI. Consistent with previous VWP studies on prediction during interpreting (Amos et al., 2022; Liu et al., 2022), the eye-tracking data revealed predictive eye movements towards the target before the target word onset in the high-predictability condition, providing robust evidence for successful prediction during SI involving extended linguistic materials. The observed fixation patterns – characterised by increased attention to the target in the high-predictability condition and to Competitor1 in the low-predictability condition – support a dual-route model of prediction during SI, incorporating both prediction-by-production and prediction-by-association mechanisms. Furthermore, this study uncovered expertise-related differences in visual attention allocation and prediction strategies. While the professionals demonstrated more frequent and dynamic attention shifts – indicating active updating of prediction based on evolving linguistic input – the students displayed more static eye movement patterns and appeared to have employed a more cautious, less adaptive prediction strategy, relying on a shallower processing of contextual information.

Several limitations should be acknowledged. First, while the paradigm used in the present study is ideal for capturing real-time anticipatory processing during SI, it does not allow direct inference about the representational level of predictions or the operation of covert imitation as proposed in the prediction-by-production account (Pickering & Gambi, 2018). An important direction for future research is, therefore, to combine such ecologically grounded paradigms with methods that more directly target linguistic forms and representational levels, enabling stronger tests of how production-based and associative mechanisms contribute to prediction during SI. Second, although the use of coherent narratives enhances ecological validity, it also introduces variability in discourse-level predictability. Future studies should better control text-level predictability, for instance, by computational matching of discourse coherence or through systematically manipulating contextual constraint. Third, the transitional probabilities were derived from a general-purpose corpus (COCA), which may not fully capture the specific lexical associations typical of literary narrative texts like those used in this study. While COCA provides robust estimates of general language patterns, register-specific corpora might yield more sensitive measures of lexical associations in future studies.

Notwithstanding these limitations, this study is an important initial step in examining predictive mechanisms during SI within discourse-rich contexts. Future studies could further explore the range of contexts in which referents within a sentence can be retrieved and how these referents establish a mental model of context (Johnson-Laird, 1983) to support prediction. Additionally, researchers could incorporate various types of visual stimuli (e.g., printed words, contextualised sceneries) to examine the real-time processing of finer-grained linguistic units (e.g., words, phonemes) during SI. Such investigations will not only deepen our understanding of prediction mechanisms during SI but also advance the broader field of bilingual processing by clarifying how interpreters navigate complex linguistic and cognitive demands in real time.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/S1366728926101084>.

Data availability statement. The data that support the findings of this study are openly available in OSF at <https://osf.io/vu45j/>.

Acknowledgements. This project was supported by a Chinese Scholarship Council grant (No 201908060148). We are also grateful to Tobii company for their technical support. We extend our sincere thanks to Prof Yanjing Wu, Dr Xueni Zhang, and the anonymous reviewers for their valuable comments on previous versions of this manuscript.

Competing interests. The authors declare none.

References

- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1).
- Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33(4), 583–609. <https://doi.org/10.1111/j.1551-6709.2009.01022.x>.
- Amos, R. M., & Pickering, M. J. (2020). A theory of prediction in simultaneous interpreting. *Bilingualism: Language and Cognition*, 23(4), 706–715. <https://doi.org/10.1017/S1366728919000671>.
- Amos, R. M., Seeber, K. G., & Pickering, M. J. (2022). Prediction during simultaneous interpreting: Evidence from the visual-world paradigm. *Cognition*, 220, 104987. <https://doi.org/10.1016/j.cognition.2021.104987>.
- Amos, R. M., Seeber, K. G., & Pickering, M. J. (2023). Student interpreters predict meaning while simultaneously interpreting - even before training. *Interpreting*, 25(2), 211–238. <https://doi.org/10.1075/intp.00093.amo>.
- Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474. <https://doi.org/10.1016/j.jml.2007.09.002>.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>.
- Chen, X., & Dong, Y. (2019). Evaluating objective and subjective frequency measures in L2 lexical processing. *Lingua*, 230, 102738. <https://doi.org/10.1016/j.lingua.2019.102738>.
- Chernov, G. V. (1994). Message redundancy and message anticipation in simultaneous interpretation. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 139–155). John Benjamins. <https://doi.org/10.1075/btl.3.13che?locatt=mode:legacy>.
- Chmiel, A. (2021). Effects of simultaneous interpreting experience and training on anticipation, as measured by word-translation latencies. *Interpreting*, 23(1), 18–44. <https://doi.org/10.1075/intp.00048.chm>.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27(1), 3–42. <https://doi.org/10.1017/S0142716406060024>.
- de Groot, A. M. (2011). *Language and cognition in bilinguals and multilinguals*. Psychology Press. <https://doi.org/10.4324/9780203841228>.
- Dell, G. S., & Chang, F. (2013). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1634), 20120394. <https://doi.org/10.1098/rstb.2012.0394>.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121. <https://doi.org/10.1038/nn1504>.
- Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, 20(5), 917–930. <https://doi.org/10.1017/S1366728916000547>.
- Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R library for eye-tracking data analysis*. <https://doi.org/10.32614/CRAN.package.eyetrackingR>; <https://github.com/jwdink/eyetrackingR>.
- Dong, Y., & Cai, R. (2015). Working memory and interpreting: A commentary on theoretical models. In Z. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 63–82). B Multilingual Matters. <https://doi.org/10.21832/9781783093595-008>.
- Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, 35(2), 353–387. <https://doi.org/10.1017/S0272263112000915>.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491–505. <https://doi.org/10.1111/j.1469-8986.2007.00531.x>.
- Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, 42(2), 133–141. <https://doi.org/10.1111/j.1469-8986.2005.00274.x>.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2), 164–203. [https://doi.org/10.1016/S0010-0285\(03\)00005-7](https://doi.org/10.1016/S0010-0285(03)00005-7).
- Ferretti, T. R., McRae, K., & Hatherly, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, 44(4), 516–547. <https://doi.org/10.1006/jmla.2000.2728>.
- Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, 37(5), 617–631. [https://doi.org/10.1016/S0042-6989\(96\)00218-0](https://doi.org/10.1016/S0042-6989(96)00218-0).
- Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, 14(3), 379–399. <https://doi.org/10.1017/S136672891000012X>.
- Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language*, 95, 200–214. <https://doi.org/10.1016/j.jml.2017.04.007>.
- Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 862–877. <https://doi.org/10.1037/0278-7393.31.5.862>.
- Gile, D. (2002). Conference interpreting as a cognitive management problem. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader* (pp. 162–176). Routledge.
- Gile, D. (2009). *Basic concepts and models for interpreter and translator training*. John Benjamins. <https://doi.org/10.1075/btl.8?locatt=mode:legacy>.
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>.
- Hintz, F., Meyer, A. S., & Huettig, F. (2017). Predictors of verb-mediated anticipatory eye movements in the visual world. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(9), 1352–1374. <https://doi.org/10.1037/xlm0000388>.
- Hodzik, E., & Williams, J. N. (2017). Predictive processes during simultaneous interpreting from German into English. *Interpreting*, 19(1), 1–20. <https://doi.org/10.1075/intp.19.1.01hod>.
- Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29(1), 33–56. <https://doi.org/10.1177/0267658312461803>.
- Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, 40(1), 171–199. <https://doi.org/10.1017/S0272263116000437>.
- Huang, H.-W., Meyer, A. M., & Federmeier, K. D. (2012). A “concrete view” of aging: Event related potentials reveal age-related changes in basic integrative processes in language. *Neuropsychologia*, 50(1), 26–35. <https://doi.org/10.1016/j.neuropsychologia.2011.10.018>.
- Huang, Y., & Snedeker, J. (2020). Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition*, 200, 104251. <https://doi.org/10.1016/j.cognition.2020.104251>.
- Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, 1626, 118–135. <https://doi.org/10.1016/j.brainres.2015.02.014>.

- Huetting, F., & Pickering, M. J. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in Cognitive Sciences*, 23(6), 464–475. <https://doi.org/10.1016/j.tics.2019.03.008>.
- Huetting, 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. <https://doi.org/10.1016/j.actpsy.2010.11.003>.
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, 10(4), 785–813. <http://doi.org/10.3758/BF03196544>.
- Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(4), 635–652. <https://doi.org/10.1037/xlm0000315>.
- Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, 55(7), 3461–3493. <https://doi.org/10.3758/s13428-022-01969-3>.
- Ito, A., & Pickering, M. J. (2021). Chapter 2. Automaticity and prediction in non-native language comprehension. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 25–46). John Benjamins. <https://doi.org/10.1075/bpa.12.02ito>.
- Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English. *Journal of Memory and Language*, 98, 1–11. <https://doi.org/10.1016/j.jml.2017.09.002>.
- Johnson-Laird, P. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge University Press.
- Jörg, U. (1995). Bridging the gap: Verb anticipation in German-english simultaneous interpreting. In M. Snell-Hornby, Z. Jettmarová, & K. Kai (Eds.), *Translation as intercultural communication: Selected papers from the EST congress, Prague 1995* (pp. 217–228). John Benjamins. <https://doi.org/10.1075/btl.20.22jor>.
- Kaan, E., & Grüter, T. (2021). Prediction in second language processing and learning: Advances and directions. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 1–24). John Benjamins. <https://doi.org/10.1075/bpa.12.01kaa>.
- Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156. [https://doi.org/10.1016/S0749-596X\(03\)00023-8](https://doi.org/10.1016/S0749-596X(03)00023-8).
- Karaca, F., Brouwer, S., Unsworth, S., & Huetting, F. (2021). Prediction in bilingual children: The missing piece of the puzzle. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 115–138). John Benjamins. <https://doi.org/10.1075/bpa.12.06kar>.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, 30(3), 481–529. https://doi.org/10.1207/s15516709cog0000_65.
- Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, 57(4), 519–543. <https://doi.org/10.1016/j.jml.2007.01.003>.
- Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, 95(1), 95–127. <https://doi.org/10.1016/j.cognition.2004.03.002>.
- Kukona, A., Cho, P., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 326–347. <https://doi.org/10.1037/a0034903>.
- 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. <https://doi.org/10.1016/j.cognition.2010.12.002>.
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146(4), 23–49. <https://doi.org/10.1016/j.brainres.2006.12.063>.
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59. <https://doi.org/10.1080/23273798.2015.1102299>.
- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63(4), 447–464. <https://doi.org/10.1016/j.jml.2010.07.003>.
- Liu, Y., Hintz, F., Liang, J., & Huetting, F. (2022). Prediction in challenging situations: Most bilinguals can predict upcoming semantically-related words in their L1 source language when interpreting. *Bilingualism: Language and Cognition*, 25(5), 801–815. <https://doi.org/10.1017/S1366728922000232>.
- Lozano-Argüelles, C., & Sagarra, N. (2022). Interpreting experience enhances the use of lexical stress and syllabic structure to predict L2 word endings. *Applied Psycholinguistics*, 42(5), 1135–1157. <https://doi.org/10.1017/S0142716421000217>.
- Lozano-Argüelles, C., Sagarra, N., & Casillas, J. V. (2020). Slowly but surely: Interpreting facilitates L2 morphological anticipation based on suprasegmental and segmental information. *Bilingualism: Language and Cognition*, 23(4), 752–762. <https://doi.org/10.1017/S1366728919000634>.
- Lozano-Argüelles, C., Sagarra, N., & Casillas, J. V. (2023). Interpreting experience and working memory effects on L1 and L2 morphological prediction. *Frontiers in Language Sciences*, 1, 1065014. <https://doi.org/10.3389/flang.2022.1065014>.
- Mani, N., & Huetting, F. (2012). Prediction during language processing is a piece of cake - but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 843–847. <https://doi.org/10.1037/a0029284>.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>.
- McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55(3), 381–401. <https://doi.org/10.1016/j.jml.2006.06.006>.
- McDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, 43, 1735–1751. [http://doi.org/10.1016/S0042-6989\(03\)00237-2](http://doi.org/10.1016/S0042-6989(03)00237-2).
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475–494. <https://doi.org/10.1016/j.jml.2007.11.006>.
- Mitsugi, S. (2017). Incremental comprehension of Japanese passives: Evidence from the visual-world paradigm. *Applied Psycholinguistics*, 38(4), 953–983. <http://doi.org/10.1017/S0142716416000515>.
- Mitsugi, S., & MacWhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition*, 19(1), 19–35. <https://doi.org/10.1017/S1366728914000881>.
- Moser-Mercer, B., Künzli, A., & Korac, M. (1998). Prolonged turns in interpreting: Effects on quality, physiological and psychological stress (pilot study). *Interpreting*, 3(1), 47–64. <https://doi.org/10.1075/intp.3.1.03mos>.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods*, 36(3), 402–407. <https://doi.org/10.3758/BF03195588>.
- Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, 8, 89. <https://doi.org/10.1186/1471-2202-8-89>.
- Otten, M., & van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, 45(6), 464–496. <https://doi.org/10.1080/01638530802356463>.
- Özkan, D., Hodzik, E., & Diriker, E. (2022). Simultaneous interpreting experience enhances the use of case markers for prediction in Turkish. *Interpreting*, 25(2), 186–210. <https://doi.org/10.1075/intp.00085.ozk>.
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144(10), 1002–1044. <https://doi.org/10.1037/bul0000158>.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347. <https://doi.org/10.1017/s0140525x12001495>.

- RStudio Team.** (2024). *RStudio: Integrated development environment for R (version 2024.04.2+764) [computer software]*. RStudio, PBC <https://posit.co/downloads/>
- Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K.** (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. <https://doi.org/10.1016/j.jml.2013.11.002>.
- Seeber, K. G.** (2001). Intonation and anticipation in simultaneous interpreting. *Cahiers de Linguistique Française*, *23*(4), 61–97.
- 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. <https://doi.org/10.1111/1467-9280.00151>.
- Sweller, J.** (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (pp. 37–76). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>.
- Timarová, Š., Dragsted, B., & Hansen, I. G.** (2011). Time lag in translation and interpreting: A methodological exploration. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 121–146). John Benjamins. <https://doi.org/10.1075/btl.94.10tim>.
- 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), 443–467. <https://doi.org/10.1037/0278-7393.31.3.443>.
- Wicha, N. Y., Moreno, E. M., & Kutas, M.** (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. <https://doi.org/10.1162/0898929041920487>.
- Wilss, W.** (1978). Syntactic anticipation in German-English simultaneous interpreting. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication* (pp. 343–352). Springer. https://doi.org/10.1007/978-1-4615-9077-4_30.
- Wlotko, E. W., & Federmeier, K. D.** (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, *49*(6), 770–785. <https://doi.org/10.1111/j.1469-8986.2012.01366.x>.
- Zirnsteina, M., van Hell, J. G., & Kroll, J. F.** (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition*, *176*, 87–106. <https://doi.org/10.1016/j.cognition.2018.03.001>.