Student interpreters predict meaning while simultaneously interpreting - even before training

Citation for published version:
Amos, RM, Seeber, KG & Pickering, MJ 2023, 'Student interpreters predict meaning while simultaneously interpreting - even before training', *Interpreting*, vol. 25, no. 2, pp. 211-238.
https://doi.org/10.1075/intp.00093.amo

Digital Object Identifier (DOI):
10.1075/intp.00093.amo

Author accepted manuscript to appear in *Interpreting*
https://www.jbe-platform.com/content/journals/1569982x

Student interpreters predict meaning while simultaneously interpreting – even before training

Rhona M. Amos[1], Kilian G. Seeber[1] and Martin J. Pickering[2]
[1] University of Geneva | [2] University of Edinburgh

Abstract
Prediction has long been considered advantageous in simultaneous interpreting, as it may allow interpreters to comprehend more rapidly and focus on their own production. However, evidence of prediction in simultaneous interpreting to date is relatively limited. In addition, it is unclear whether training in simultaneous interpreting influences predictive processing during a simultaneous interpreting task. We report on a longitudinal eye-tracking study which measured the timing and extent of prediction in students before and after two semesters of training in simultaneous interpreting. The students simultaneously interpreted sentences containing a highly predictable word as they viewed a screen containing four pictures, one of which depicted a highly predictable object. They made predictive eye movements to the highly predictable object both before and after their training in simultaneous interpreting. However, we did not find evidence that training influenced the timing or the magnitude of their prediction.

Keywords: simultaneous interpreting, prediction, eye-tracking, longitudinal study

Introduction

Prediction often occurs during language comprehension (for a review, see Pickering & Gambi 2018). It has also long been considered advantageous in simultaneous interpreting (SI) (Chernov 1994; Moser 1978; Seleskovitch 1984). However, despite the potential advantages of predicting during SI, the difficult listening conditions inherent in SI may limit prediction. It has been suggested that only professional interpreters are able to reap the benefits of prediction during SI (Frauenfelder & Schriefers 1997; Moser 1978). We investigate this premise in this longitudinal study, which monitors students' eye movements during an SI task in order to establish whether or not they are able to predict before training and whether training in SI influences their subsequent predictive behaviour.

Prediction in challenging conditions

People predict what they are about to hear, and such prediction supports rapid comprehension (Pickering & Gambi 2018). Altmann and Kamide (1999) had participants listen to sentences containing a verb which was compatible either with one of four objects in a visual scene or with all four of the objects. For example, participants heard "The boy will *eat* the cake" or "The boy will *move* the cake" while viewing a boy next to a cake, a train set, a balloon and a toy car. When the participants heard the constraining verb *eat*, they tended to look at the cake, whereas when they heard the non-constraining verb *move*, they did not look at the cake more than the other objects. Such looks were predictive – that is, they occurred before the participants heard *cake*. This finding shows that listeners make predictions during comprehension. There is now much evidence which demonstrates that people form predictions both while reading (DeLong et al. 2005; Dikker et al. 2010) and while listening (Ito et al. 2020; Kamide et al. 2003; Van Berkum et al. 2005).

However, prediction may be resource intensive and may be impaired when the comprehender lacks resources. In a study based on Altmann and Kamide (1999), Ito, Corley and Pickering (2017) had people listen to their L1 and L2 while they either performed, or did not perform, a concurrent working memory task of remembering unrelated words. In both the L1 and the L2 listeners, predictive eye movements were delayed when the participants carried out the concurrent working memory task. Meanwhile, Huettig and Janse (2016) had participants listen to spoken instructions in Dutch while they viewed four pictures, one of which had a different grammatical gender from the three others – for instance, "Kijk naar *de* afgebeelde *piano*" (look at *the* displayed *piano* [common noun]) where the other objects shown on the screen took the article *het* [neuter noun]. Participants fixated on the target object more than the distractor objects before the onset of the target noun, and the difference in fixations between target and distractor objects was greater for participants with higher working-memory capacity. However, people with low working-memory capacity have also been shown to predict upcoming meaning during language comprehension (Otten & Van Berkum 2009).

Predictive processing may be more limited in L2 than in L1. While several studies demonstrate that prediction also takes place in L2 (Dijkgraaf et al. 2017; Ito et al. 2017, 2018; Martin et al. 2013), there is evidence that prediction in L2 may be slower and less detailed than in L1. For example, Martin et al. (2013) and Ito, Pickering and Corley (2018) found that L1 readers predicted both meaning and form, but L2 readers predicted only meaning. Lew-Williams and Fernald (2010) found that when listening to short sentences containing a gender-marked object and viewing a screen showing the named object and another object of a different gender, native Spanish speakers used the gender cue to look at the named object more quickly whereas non-native speakers did not. Mitsugi and MacWhinney (2015) reported similar findings in a study of Japanese L2 speakers.

It is unclear what underlies this deficit in predictive processing in L2 compared to L1. It is possible that L2 listeners find it more difficult to access grammatical knowledge (McDonald 2006) and to use it to make predictions. Alternatively, L2 comprehension may be more resource-intensive than L1 comprehension, and so listeners use more working-memory resources simply to comprehend L2 and consequently do not have as much working memory available for prediction. But, whatever the cause, it is likely that prediction is particularly limited for people listening in their L2 under a working memory load.

Prediction in interpreting

The challenging conditions outlined in the previous section might limit prediction during SI. However, prediction has been regarded as a strategy used in SI to reduce the constraints on working memory (Kalina 1998; Liontou 2015; Seeber 2001; Setton 2002; Van Besien 1999). If interpreters produce their interpretation with a relatively short time delay (i.e., lag) after hearing the original speaker, they will have to retain information for less time in their short-term memory, and this may reduce their cognitive load (Amos & Pickering 2020). In other words, prediction may help interpreters regulate lag during interpreting to allow for reformulation while not overburdening working memory (Gile 1997).

There has often been an implicit or explicit assumption that only trained and experienced simultaneous interpreters benefit from such prediction (Frauenfelder & Schriefers 1997; Hoffman 1997; M. Liu 2008; Moser 1978; Vandepitte 2001). However, the empirical evidence for this is mixed. Some studies suggest that prediction takes place in SI regardless of training. For instance, Liu, Hintz, Liang and Huettig (2022) found that untrained bilinguals form predictions when listening in their L1 and interpreting simultaneously into their L2 (although this is not the direction in which most simultaneous interpreters work); and Amos, Seeber and Pickering (2022) found that both trained interpreters and untrained bilinguals predict when they interpret from their L2 into their L1. Özkan, Hodzik and Diriker (2022) found a difference between professional and trainee interpreters, with interpreters only appearing to predict – but this was in a listening-only task.

A longitudinal study could provide a more rigorous test of whether training influences predictive processing during SI because it involves the same group of participants before and after training.

Why might training change predictive processing during simultaneous interpreting?

If training in SI does increase predictive processing, it might do so in two ways. On the one hand, students might be instructed to predict (e.g., by paying careful attention to context; see Setton & Dawrant 2016). On the other hand, training might help students to improve domain-specific skills (e.g., the ability to produce translation equivalents quickly) or else lead to cognitive enhancements (e.g., improvements in relevant aspects of working memory), therefore increasing or improving the cognitive resources available for prediction.

Setton and Dawrant (2016) propose exercises to develop prediction techniques in SI, including asking students to complete sentences before hearing their conclusion (see also Liontou 2015). However, it is unclear how and to what extent prediction is taught by interpreter trainers and whether such teaching works. What we do know is that training and experience in SI may lead to an overall improvement in SI performance (M. Liu et al. 2004; Tzou et al. 2012). Tzou, Eslami, Chen and Vaid (2012) found that second-year interpreting students outperformed first-year students when simultaneous interpreting a 15-minute speech, and first-year students outperformed non-interpreters. Meanwhile, Chmiel (2021) found that training increased the speed at which translation equivalents were produced (regardless of context). These overall improvements in performance may mean that more resources are available for predictive processing.

Hervais-Adelman, Moser-Mercer, Murray and Golestani (2017) found that approximately 14 months of interpreting training led to increases in cortical thickness in areas of the brain associated with speech-sound processing, executive control of attentional resources and working memory (with decreases in cortical thickness in a control group). They suggest that these findings may reflect the beneficial effects of SI training on executive and attentional skills (see also Van de Putte et al., 2018). Babcock, Capizzi, Arbula and Vallesi (2017) tested short-term memory (measured using simple span tasks), working memory (measured through complex span tasks), alerting and orienting networks and task switching in interpreting students, translation students and non-language students before and after approximately two years of university-level education in their respective disciplines. After this education, all the groups showed improved scores in tasks measuring task-switching ability, but the interpreting trainees alone improved their short-term memory. Babcock and Vallesi (2017) also found that professional interpreters performed better than multilingual controls on the letter-span task. However, in Babcock et al. (2017), different group sizes could have affected the results of the comparison between interpreting and translation students, and in Babcock and Vallesi (2017), the results were not corrected for multiple comparisons. Chmiel (2016) found higher L2 reading-span scores in trainees after training than before it (although this might simply reflect an improvement of L2 reading skills after training). Meanwhile, Lozano-Argüelles, Sagarra and Casillas (2020), in a listening-only task, found that interpreters anticipated word endings based on segmental and supra-segmental cues at a faster rate than non-interpreters (see also Lozano-Argüelles & Sagarra 2021); and Fan, Collart and Chan (2022), also in a listening-only task, found subtle differences in brain-response patterns to predictable content. Finally, Dong and Liu (2016) found that interpreting training led to an advantage in switching and working memory updating compared to other types of language training.

In sum, while the findings of some of these studies should be read with the caveats previously expressed, there is some evidence that training in SI might change the availability of cognitive resources during an SI task and that such resources may be used for prediction.

The current study

We designed a longitudinal study to investigate two main questions. First, do interpreting students predict upcoming meaning when interpreting from English into their A (i.e., native) language before training? Secondly, does the timing or extent of prediction during interpreting change after training? Testing the same group of participants before and after SI training provides a stringent test of whether such training affects prediction. Our study also investigated a potential link between prediction, production and training during SI. Do interpreting trainees start interpreting predictable utterances more quickly than unpredictable sentences, and does training affect this? Is there a link between the time at which students start interpreting sentences and the time at which they start making predictive eye movements?

We hypothesized that interpreting students would predict upcoming meaning before undergoing training in interpreting but that the timing or extent of prediction would change after training. With regard to the link between prediction and production, we expected the students to begin their interpretation of predictable sentences more quickly after training than before it. Finally, in line with Amos et al. (2022), we hypothesized that when students began interpreting earlier in relation to the predictable word, they would also make more predictive eye movements and make them sooner.

Methods

Participants

Twenty-two students admitted to the Master's in Conference Interpreting at the University of Geneva participated in the study. Two cohorts of participants were recruited over two years (so that we could recruit sufficient participants). One further participant was tested but did not complete the course of studies and so that student's data was excluded. All the participants had passed entrance exams which were designed to ensure that they had sufficient language proficiency and an aptitude for conference interpreting. In addition, all the participants completed a language background questionnaire (see Table 1) based on the Leap-Q questionnaire (Marian et al. 2007) and provided information about their professional background.

The Master's in Conference Interpreting comprises three semesters of 14 weeks, with consecutive interpreting taught over three semesters and SI added in the second and third semesters. Students receive at least 10 hours per week of formal teaching in SI, participate in practice groups for an additional 6–8 hours a week and also practice independently. Students' performance is assessed using continuous assessment (including grades during the second semester) and a final exam. Participants in the first cohort (class of 2020) received two semesters of onsite training in SI and those in the second cohort (class of 2021) received two semesters of blended training (online and onsite) because of the COVID-19 pandemic.

All the participants interpreted from English into their native language (A language), which was French (5), German (5), Spanish (5), Italian (4) or Russian (3). Three participants were studying English as a B language, meaning that they also interpreted simultaneously (and consecutively) from their native language into English. Two further participants interpreted into English in the consecutive mode only. The other 17 participants studied English as a C language, meaning that they interpreted from English into their A language, but not from their native language into English. Four participants had had some form of prior training in SI (either via a short course or as an optional module).

Table 1: Information obtained from the language background questionnaire

	Student interpreters (***n*** = 22)
Background	
Age (*years*)	**27.27** (*SD*: 5.40, range: 21–40)
Languages spoken	**4.73** (*SD*: 1.35, range: 3–8)
Highest level of education	
Bachelor's	**50%**
Master's	**50%**
Self-reported English-language measures	
Age (*years*) of acquisition	**8.32** (*SD*: 4.19, range: 0–15)
Age (*years*) of fluency	**14.50 (***SD*: 5.47, range: 3–22)
Age (*years*) reading proficiency acquired	**13.91** (*SD*: 4.39, range: 6–22)
Age (*years*) written proficiency acquired	**14.95** (*SD*: 4.35, range: 6–25)
Self-reported exposure to English	
Years living in an English-speaking area	**1.90** (*SD*: 5.13, range: 0–24)
Years spent in an EN school/work environment	**3.28** (*SD*: 5.05, range: 0–18)
Years living in an EN family environment	**1.98** (*SD*: 6.23, range: 0–24)
Current exposure %	**21.50** (*SD*: 9.49, range: 10–50)
Self-rated English proficiency (on a scale of 1–5)	
Speaking	**4.05** (*SD*: 0.72)
Reading	**4.64** (*SD*: 0.49)
Listening	**4.64** (*SD*: 0.49)

Stimuli

The stimuli were the same as those used in Amos et al. (2022). There were 32 predictable sentences, each containing a highly predictable word (e.g., *camel,* in "*The traveller went to the desert because he wanted to ride a camel and go exploring*.") at various positions in the sentence but never in the sentence-final position; there were also 32 unpredictable sentences (e.g., "*He could hear what they were saying because the door had been left open*."). These sentences were designed not to be constraining for any word. The sentences were divided into two sets so that half the participants received Stimulus Set 1 before training and Stimulus Set 2 after training and the other half received the sets in reverse.

Each predictable sentence was paired with a visual array of three distractor objects and one of two critical objects. In the target condition, the name of the critical object corresponded to the predictable word (e.g., *camel*). In the unrelated condition, the name of the critical object (e.g., *barrel*) was semantically and phonologically unrelated to the predictable word. Critical objects appeared in each of the four quadrants equally frequently following a Latin square design. If a participant fixated on the picture of the predictable noun more than the picture of an unrelated object before hearing the predictable word, this would imply that they predicted the upcoming word.

The same set of objects was also paired with an unpredictable sentence. Unpredictable sentences mentioned a distractor object 50% of the time (and did not mention an object on the screen 50% of the time). This meant that, together with the predictable sentences, which also mentioned a critical object 50% of the time (i.e., in the target condition), sentences mentioned one of the objects in the visual array 50% of the time. We used the unpredictable sentences to check for visual bias towards critical objects. If a participant did not fixate on the picture of the predictable noun more than the picture of the unrelated object when they were presented with an unpredictable sentence, this would allow us to exclude visual bias as a reason for fixations on the predictable noun in the predictable sentences.

The target and unrelated visual conditions were counterbalanced in each set, resulting in two main lists. Each main list comprised two half-lists, each containing 16 visual arrays paired with eight predictable and eight unpredictable sentences. The visual arrays paired with predictable sentences in one half-list were paired with unpredictable sentences in the other half-list. For each main list, half the participants saw one half-list first and the other half saw the other half-list first, so that the order of presentation of the images with predictable and unpredictable sentences was counterbalanced. Therefore, there were four lists for each stimulus set.

All the sentences were recorded by a male native speaker of southern British English, who read the predictable sentences at a rate of 2.01 syllables per second (*SD* = 0.27) and the unpredictable sentences at a rate of 1.86 syllables per second (*SD* = 0.25). The mean sentence duration was 9.87 s (*SD* = 1.63) for the predictable sentences and 9.04 s (*SD* = 1.85) for the unpredictable sentences. The mean onset time of the predictable word was at 6.28 s (S*D* = 1.94) for the predictable sentences. As mentioned above, 50% of the unpredictable sentences contained a word corresponding to one of the images. The critical timepoint in the unpredictable sentences was 4.06 s (*SD* = 2.06); it was selected based on when the depicted word was mentioned or selected randomly.

Before the experiment began, the predictability of each predictable sentence was assessed by groups of at least 14 L2 speakers of English (whose L1 was French). The mean cloze probabilities were 89.1% for Stimulus Set 1 and 92.3% for Stimulus Set 2. The mean position of the critical word in the predictable sentences was 9.25 s (*SD* = 1.41) in Stimulus Set 1 and 10.4 s (*SD* = 3.44) in Stimulus Set 2. The properties of the two sets of predictable sentences are summarised in Table 2. Welsh's two-sample *t*-tests were run for all the stimuli characteristics and no differences were found between the predictable sentences in the two sets (see Table 2).

Table 2. Properties of predictable sentences

	Stimulus Set 1	**Stimulus Set 2**	**Welsh's *t*-test result**
Mean length (s)	10.01 ± 1.450	9.743 ± 2.148	$|t| = 0.45$, $p = 0.66$

Mean predictable word onset (s)	6.468 ± 1.837	6.086 ± 2.026	|t| = 0.55, *p* = 0.58
Mean predictable word offset (s)	7.084 ± 1.827	6.664 ± 2.073	|t| = 0.61, *p* = 0.55
Mean cloze value (%)	89.1 ± 12.8	92.3 ± 11.1	|t| = 0.77, *p* = 0.45

An online picture-naming test was run before the experiment to assess naming agreement in English for all critical and distractor objects. A different group of at least 12 L2 speakers of English (who were L1 speakers of French) provided a name for each picture. Table 3 shows the average picture-naming agreement ratings for the critical objects in both stimulus sets. A two-sample *t*-test showed that there was no difference in the picture-naming agreement for target versus unrelated objects in Set 1 ($t(29.7) = 0.375$, $p = 0.71$) or in Set 2 ($t(21.7) = 1.88$, $p = 0.074$); nor was there any difference between Sets 1 and 2 in picture-naming agreement in the target condition ($t(25.2) = 0.00017$, $p = 1$) or the unrelated condition ($t(26.9) = 1.39$, $p = 0.18$).

Table 3. Picture-naming agreement ratings (%) for the target and unrelated objects in both stimulus sets

Condition	**Stimulus Set 1**	**Stimulus Set 2**
Target	91.2 ± 12.3	91.2 ± 7.7
Unrelated	89.6 ± 11.2	82.8 ± 16.0

Procedure

The participants were tested at two different timepoints: before and after training in SI. They participated in the pre-training experiment in the first two weeks of the second semester, when they started SI training. Post-training, they participated in the experiment in the second-last week of term before the university holidays and their final exams. They were assigned either to a group which received Stimulus Set 1 before training and Stimulus Set 2 after training or to a group which received Stimulus Set 2 before training and Stimulus Set 1 after training. The participants' A language was considered when assigning them to a group so that different A languages were represented to the same extent in both groups (as far as possible).

The participants first read and signed an informed consent form approved by the Ethics Committee of the Faculty of Translation and Interpreting at the University of Geneva. The experiment started with a picture familiarization task: the participants saw all the objects appearing in the experiment in an automatically generated randomized order. The objects were shown on the screen one at a time above a caption showing their name (in English). The participants were then instructed to look at the objects and listen to their names so that they could name the objects later. After that, they were asked to name each object using the English name that had been provided. Objects were considered to have been named correctly if the participants repeated the name that they had seen. Incorrectly named objects (2.34% before training; 1.07% after training) were shown again and the experimenter prompted the participants when they did not name the object accurately on the second viewing.

In the eye-tracking experiment, the participants were seated in front of an SR Research Eyelink® 1000 desktop-mounted eye-tracker in remote mode which recorded gaze at a 500 Hz sample rate. It was set up inside a portable interpreting booth in the experimental laboratory (LaborInt) of the Interpreting Department at the University of Geneva. The participants' dominant eye was tracked (with dominance assessed using an ocular sighting test). They were instructed to listen to and simultaneously interpret the sentences into their A language and then to judge whether the sentence had mentioned any of the objects shown on the display. This comprehension question allowed us to check whether participants had understood the critical word. After the instructions were administered, the eye-tracker was calibrated using the nine-point calibration grid. The eye-tracker was then recalibrated before the participants began the experiment.

Figure 1 shows the trial sequence. Each trial started with the presentation of a blank screen for 500 ms, the screen remaining blank as the sentence began. The visual array appeared 1,000 ms before the onset of the critical word and remained visible until sentence offset. The participants then viewed a blank screen for 4,000 ms before viewing a screen with the question: “Did the sentence mention any of the pictures?” The participants could answer by pressing 1 for “Yes” or 2 for “No”.

The experiment started with two practice trials, after which the participants were given an opportunity to ask questions. During the practice trials, the experimenter checked whether the participants were interpreting the sentences and whether they were doing so simultaneously (defined as beginning their interpretation before the end of the sentence) and, if necessary, reminded the participants to interpret simultaneously. The session lasted about 30 minutes.

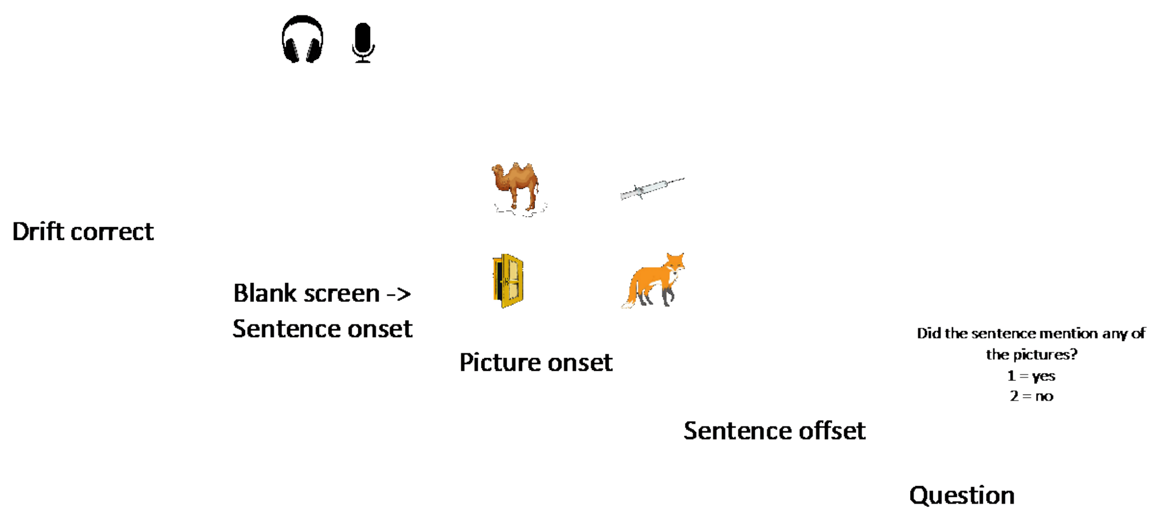

Figure 1. Trial sequence for the experimental trial: “The traveller went to the desert because he wanted to ride a camel and go exploring.” The same pictures were paired with the unpredictable sentence: “He could hear what they were saying because the door had been left open.”

Analyses

Following Ito et al. (2018), the proportions of time spent fixating on the target and unrelated objects were calculated using EyeLink’s DataViewer for 50 ms bins. Blinks and fixations outside the computer screen were included in the calculation of the proportion of fixations. However, bins containing only blinks or fixations outside the computer screen were excluded from the analysis. We also excluded incorrectly answered trials. Before training, the mean accuracy for comprehension questions in the predictable sentences was 96.9% (*SD* = 4.0%) and 95.2% (*SD* = 7.5%) in the unpredictable sentences and after training it was 99.1% (*SD* = 2.5%) for predictable and 98.6% (*SD* = 2.7%) for unpredictable sentences.

We then considered whether prediction took place during SI. In this analysis, our dependent variable was the proportion of arcsine-transformed fixations and our independent variable was condition (Target vs Unrelated). We used dummy coding and set the Unrelated condition as a baseline, to which we compared the Target condition. We considered our pre- and post-training data separately. As in Amos et al. (2022), we ran a linear-mixed model analysis for each bin from 1,000 ms before target word onset to 1,000 ms after the onset, using the lme4 package (Bates et al. 2015). Our model contained random intercepts and decorrelated random slopes by item and by participant (Barr 2008) using the bobyqa method in the optimx optimiser, which allowed all the models to converge. The model used was lmer(arcsine_fixations ~ condition + (condition || participant) + (condition ||item)). However, since there were singular fit warnings for the models for multiple bins, we checked our results using a Bayesian linear mixed model. Here we used the default priors and the optimx optimiser, this time with the nlminb method. The model evaluated the arcsine-transformed fixation proportions on critical objects, as predicted by the condition for each bin. The model was blmer (arcsine_fixations ~condition + (condition || participant) + (condition || item). Following Ito et al. (2018) and Borovsky, Elman and Fernald (2012), we based our conclusions on periods where consecutive time bins showed a significant difference in order to account for familywise error. We then ran a series of linear-mixed models using the same parameters and the same dependent variable to check (1) whether people fixated on predictable objects even when listening to the unpredictable sentences (to check for visual bias towards predictable objects), (2) whether there was a difference in prediction patterns before and after training and (3) whether the extent of predictive processing was related to the time at which the participants began their interpretation relative to the predictable word.

We also explored whether there was a difference in the onset of the interpretation between the predictable and the unpredictable sentences and whether this difference changed during training. As predictable and unpredictable sentences were spoken at different rates, we also included speech rate as an independent variable that could influence the interpretation onset time. Here our dependent variable was the onset time of interpretation. We ran a linear mixed model with random slopes and decorrelated random intercepts in which the onset time of interpretation was our dependent variable and in which the independent variables were type of sentence (predictable vs unpredictable), stage (pre- vs post-training) and speech rate (measured as syllables per second). The model formula was lmer(onset time ~ TrainingStage * SentenceType * SpeechRate + (TrainingStage * SentenceType * SpeechRate || participant) + (TrainingStage * SentenceType * SpeechRate || item)).

Finally, we ran a series of correlational analyses to investigate whether the participants' extent of prediction was linked to their grades in SI from English into their A language (cor.test(ExtentofPrediction, FinalENGrade)).

All our models consistently used the maximum random effects structure supported by the data (Barr et al. 2013). The full R script used to analyse the data is available at <https://osf.io/htj3r/>.

Results

Predictive eye movements before and after training

Figure 2 shows eye-tracking data for the participants before and after training in SI. Before training, they looked at the target object more than the unrelated object from –500 ms before predictable word onset and until the end of the time window analysed (1,000 ms after predictable word onset). However, one bin was not significant at –350 ms before word onset and another was not significant at 350 ms after word onset (the bin at –500 ms was not significant in the Bayesian model). Therefore, the participants predictively looked at the target object and persisted in these predictive looks from –350 ms before predictable word onset.

After training, they began looking significantly more at the target object than at the unrelated object at –350 ms before the onset of the predictable word and continued to do so until the end of the time window analysed (1,000 ms after predictable word onset). The participants therefore looked predictively at the target object and persisted in these predictive looks from –350 ms before predictable word onset. Therefore, both before and after training they predicted the semantic content of the predictable words while engaged in an SI task.

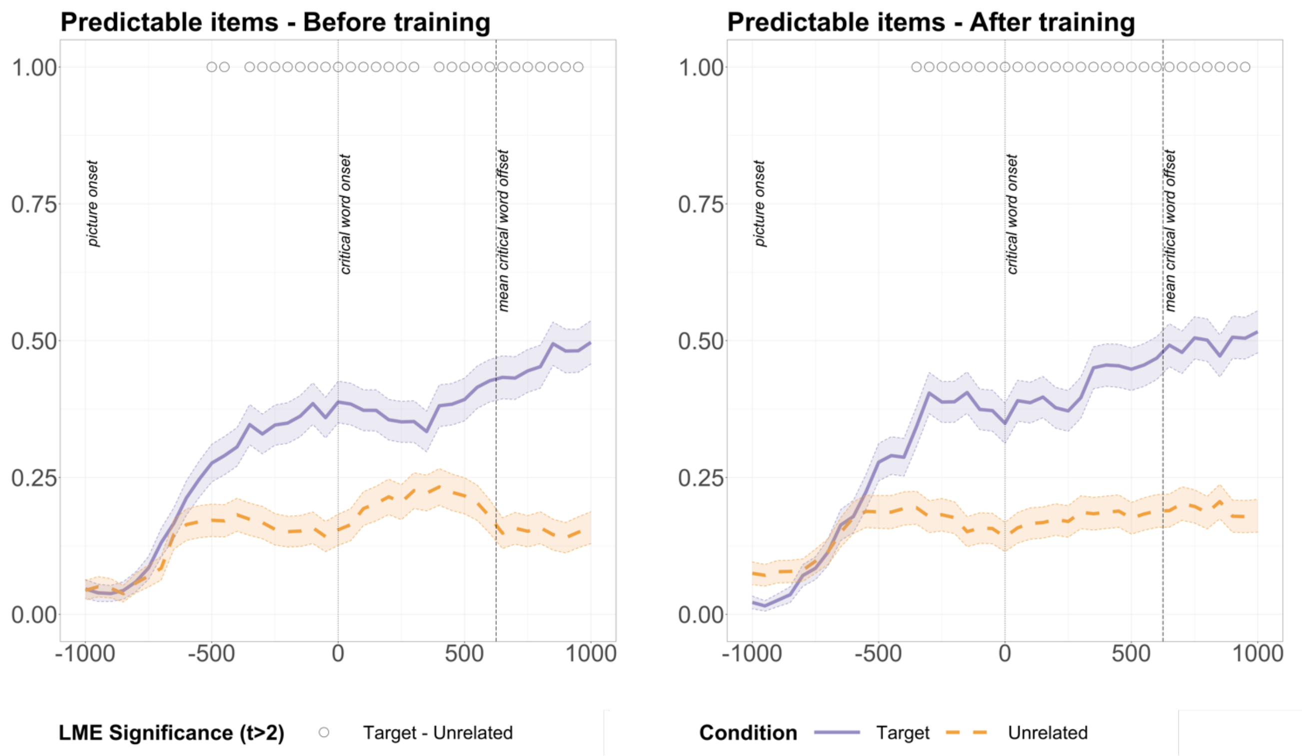

Figure 2. Eye-tracking results for the participants before (left) and after (right) training in SI ($n = 22$). The graph shows fixation proportions on target and unrelated objects for the experimental sentences. Time 0 ms shows predictable word onset. The mean predictable word offset was 602 ms; picture onset was at –1,000 ms. The open dots along the top of the graph show 50 ms bins in which the difference between target and unrelated conditions was significant (|t| >2). The thick transparent lines are error bars representing standard error.

Analysis of unpredictable sentences

We then checked whether our results could have been the consequence of visual bias. To do so, we analysed the unpredictable items to examine whether the participants fixated on the target object more than the unrelated object. We were interested in ascertaining whether they did so even when a predictable word was not mentioned in the sentence and the sentence either mentioned one of the distractors from the visual array or did not mention any of the objects present in the visual display. First we analysed the pre-training data. As shown in Figure 3, we did not find evidence of a sustained divergence in the fixation proportions between the target and the unrelated objects, and there was no indication that the target image was fixated more than the unrelated image at any point. Therefore, in the unpredictable items, there was no indication that the target images attracted more attention than the unrelated images and thus no indication that the results of our main analyses of the predictable sentences could reflect visual bias. We carried out the same analysis on the unpredictable items after training in SI. We did not find any differences between the fixation proportions on the target and the unrelated objects and therefore there was no indication that the findings for the predictable sentences could reflect visual bias.

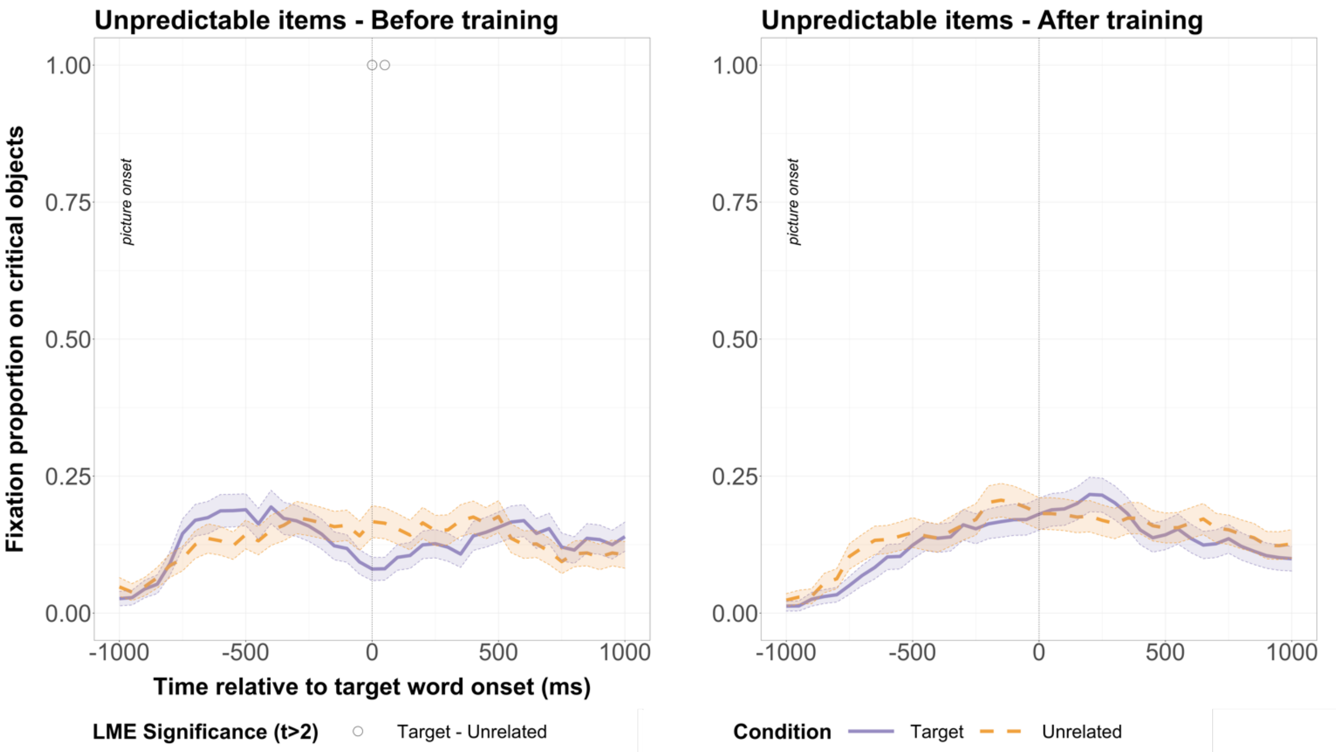

Figure 3. Eye-tracking results for students before (left) and after (right) training in SI ($n = 22$). The graph shows the fixation proportions on the target and the unrelated objects for the unpredictable sentences. The picture onset was at –1,000 ms. In half of the sentences, an object on the screen was mentioned at 0 ms and for the other sentences 0 ms represents a random point in the sentence. The thick transparent lines are error bars representing standard error.

Effects of training on prediction

Both before and after training, predictive eye movements begin at roughly the same time (see Figure 2). However, this finding does not indicate whether or not the extent of prediction (as reflected in the size of the difference between fixations on the target and the unrelated objects at different time points) was similar before and after training. To determine whether this was the case, we ran linear-mixed models over all the time bins from –1,000 to 1,000 after the predictable word in the full dataset and included an interaction term for training (before vs after). There was no interaction between training and predictive fixations on the target versus the unrelated objects, which indicated that predictive fixations did not depend on training. There was an interaction of training and fixation proportions on the target versus the unrelated object on an isolated bin at +350 ms after word onset, but this does not constitute reliable evidence of a difference in fixation proportions across the conditions as a whole. Therefore, there is no reliable evidence of differences in fixation proportions to the target and the unrelated objects either before or after the onset of the predictable word.

Relationship between fixation proportions and relative speech onset

We then considered the time lag between the participant speech onset and the onset of the predictable word in the predictable sentence and whether there was a relationship between this time lag and the fixations made on the predictable object. We identified the speech onset time using Praat (Boersma & Weenink, 2002) and also used a voice-onset detection script to identify the time at which the intensity of the interpretation stream exceeded 50 dB (sound pressure level). We then checked manually whether the results of this onset detection matched the start of the speech onset and adjusted all the onset times manually when the interpretation onset was later because it had started with a hesitation (e.g., *emm, err*) or when the onset time was earlier because the intensity threshold was not met by the first sound (e.g., unvoiced fricative). We adjusted 7.81% of all of the onset times manually. No interpreting was provided for two trials (0.14%). We then determined the time lag by subtracting the interpretation onset time from the onset time of the predictable word.

We ran a linear mixed model that included an interaction term for time lag (measured in seconds and centred), stage (before and after training) and condition for each time bin. We included decorrelated random slopes for time lag, stage and condition within participant and decorrelated random slopes for time lag and condition within item (because over the two different timepoints all the items were shown and therefore the items were balanced over timepoints). The model formula was lmer(arcsine_fixations ~condition*time lag*TrainingStage + (condition*time lag*TrainingStage || participant) + (condition*time lag || item)). We found no consistent evidence of a three-way interaction between time lag, stage (pre- vs post-training) and condition (target vs unrelated). The models for only three bins – after word onset, at +450 ms, +600 ms and +650 ms – showed a significant three-way interaction. There was no interaction at any time between the time lag and the stage (before or after studies). There was an interaction between condition and stage at only one timepoint (+350 ms). However, there was a two-way interaction between time lag and condition for three bins from –500 ms to –350 ms and another interaction over two bins from –200 ms to –100 ms.

In view of this, we dropped the three-way interaction term from our model and considered whether, regardless of stage, time lag was linked to predictive fixations on target versus unrelated objects (model: lmer(arcsine_fixations ~condition*time lag+ (condition*time lag||participant) + (condition*time lag||item)). We found evidence of a difference in fixation proportions on the target and the unrelated objects that depended on the time lag. The interaction between time lag and condition was significant in the predictive time window at –350 ms and from –200 ms to 100 ms. The interaction was also significant after the predictive time window from +250 ms to +450 ms, from +500 ms to +600 ms and from +750 ms until +850 ms. Therefore, there was no evidence of a sustained interaction between condition (target vs unrelated) and time lag in the predictive window. However, after the onset of the predictable word, there was an interaction between the difference in fixations on the target versus the unrelated object and the time lag of the interpretation. The direction of the interaction indicates that when interpreting began later relative to the predictable word, there were more fixations on the target compared to those on the unrelated object.

Effects of training on speech onset

We then considered whether there was a difference in interpretation onset time for the predictable and the unpredictable sentences that depended on training. We first trimmed the speech-onset time data. After that we considered the pre-training and the post-training data separately and removed all speech onsets which fell 2.5 *SD*s above the by-participant mean to reduce the rightward skew of our data (pre-training: 23 responses/3.28%; post-training: 17 responses/2.41%). None of the speech onsets fell 2.5 *SD*s below the by-participant mean.

We also considered whether any difference in the interpretation onset time for the predictable and the unpredictable sentences was dependent on training. Because there was a difference in the rate at which the predictable and the unpredictable sentences were spoken, we also checked whether any speech rate interacted with sentence predictability and training to influence the onset time of the interpretation. The model was lmer(Onset ~ Sentence Type * Stage * Speech Rate + (Sentence Type * Stage * Speech Rate || participant) + (Sentence Type * Stage * Speech Rate || item)). Prior to training, the mean interpretation onset time was 3,076 ms for the predictable and 3,081 ms for the unpredictable sentences. After training, the mean interpretation onset time was 2,812 ms for the predictable and 2,872 ms the unpredictable sentences (see Figure 4).

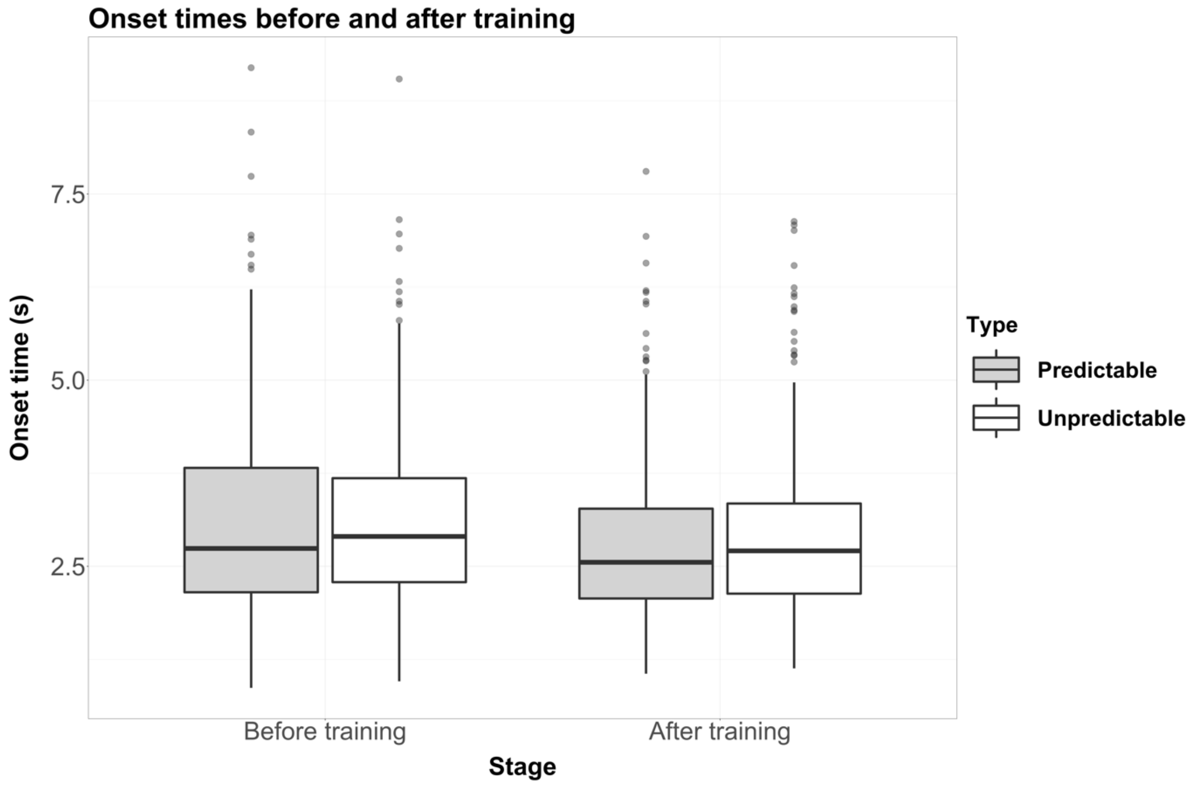

Figure 4: Onset times of the interpretation in predictable and unpredictable sentences before and after training in SI

We found no interaction (|t| = 1.025) between the type of sentence (predictable vs unpredictable), the stage (before and after interpreting) and the speech rate. This suggested that training did not have a differential effect on the onset time of the predictable and the unpredictable sentences when the speech rate was taken into account. However, we did find a main effect of speech rate, which showed that when the speech rate was faster, the participants began interpretation significantly earlier.

Table 4. Results of the linear mixed model for the effect of stage of training, sentence predictability and speech rate on the interpretation onset time

	Interpretation onset time			
Predictors	*Estimates*	*CI*	*t*	*p*
(Intercept)	6.43	4.71–8.15	**7.341**	**<0.001**
Stage	–0.67	–1.62–0.29	–1.367	0.172

Type [Predictable]	–1.60	–4.07–0.87	–1.272	0.204
Speech rate	–1.68	–2.57–0.79	**–3.719**	**<0.001**
Stage * type [Predictable]	–0.77	–2.03–0.48	–1.209	0.227
Stage * Speech rate	0.24	–0.25–0.73	0.969	0.333
Type [Predictable] * Speech rate	0.97	–0.31–2.24	1.488	0.137
Stage * type [Predictable] * Speech rate	0.34	–0.31 0.98	1.025	0.305
$N_{participant}$		22		
N_{item}		64		
Observations		1,366		

Extent of prediction and student performance

Because we did not find a difference between prediction during interpreting before and after training, we considered whether prediction was instead linked to student ability. We therefore investigated whether the extent of prediction was linked to final exam grades (on first attempt) for SI from English into the participants' A language. We computed the difference in the mean arcsine-transformed fixation proportions between the target and the unrelated conditions in the window, during which we found evidence of prediction. Assuming that it takes 200 ms to programme eye movements (Saslow, 1967), we included time bins from –350 ms before word onset until 200 ms after word onset because this was the predictive time period during which we found evidence of prediction before and after training in consecutive time bins. We used this measure as a proxy for the extent of prediction.

We then explored the relationship between the participants' extent of prediction before and after training and their final grades for SI for English into their A language by computing their correlations (the pass grade being 4 on a scale of 0–6, with 6 being the highest grade). We did not find a significant correlation between students' grades and their extent of prediction before and after training (before training: r (20) = 0.33, p = 0.14; after training: r (20) = 0.30, p = 0.17), nor was there a correlation between the extent of prediction before and after training (r (20) = 0.18, p = 0.42).

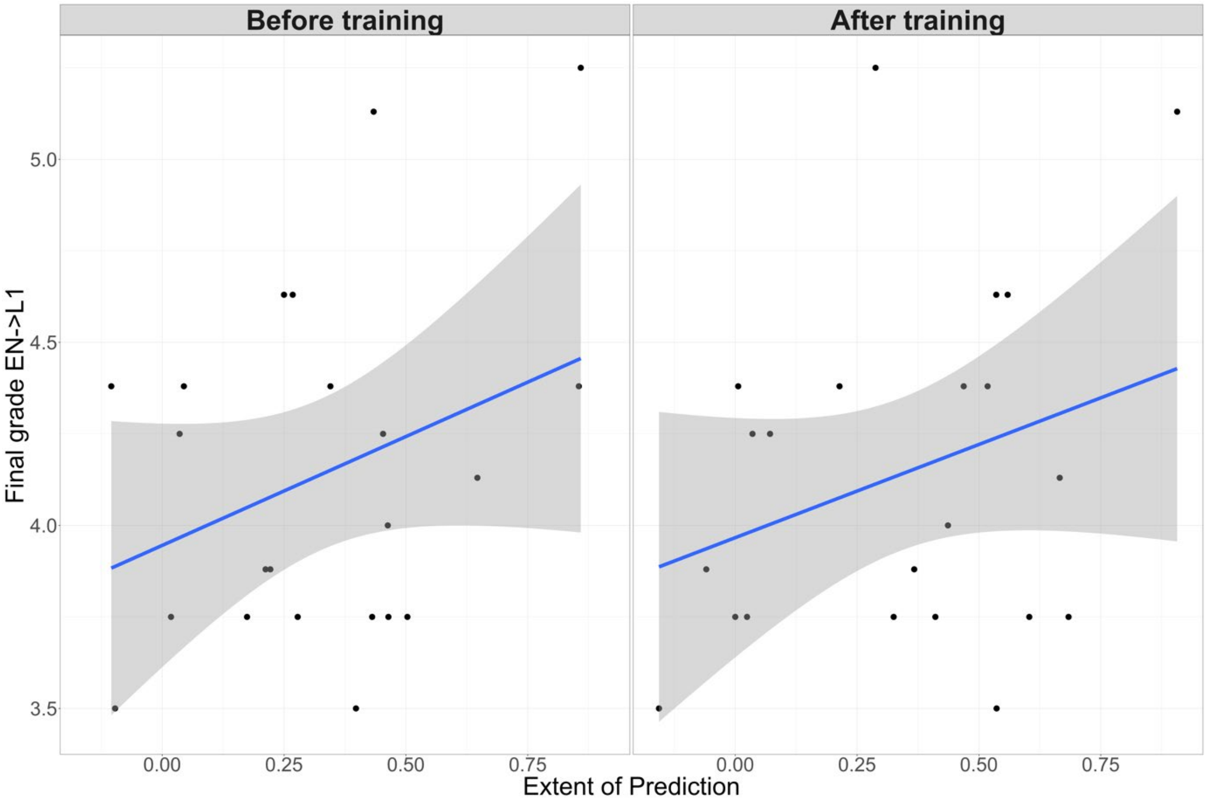

Figure 5. Graph comparing the extent of participants' prediction before and after training and the grade they achieved in their final exam from English into their A language after completing two semesters of training.

Discussion

Our study shows that students have a strong tendency to predict upcoming words, both before and after training in SI. This supports the hypothesis that prediction of meaning takes place even in the challenging listening conditions present during an SI task. However, we found no significant differences in the patterns of prediction after training in SI; this suggests that training in SI does not influence predictive processing during an SI task. Semantic prediction therefore appears to take place even when cognitive resources are limited and does not appear to be influenced by SI training. This complements the finding in Amos et al.'s (2022) between-group study by demonstrating that patterns of prediction are similar for the same group of participants at two different timepoints: before and after training. But it does not support accounts which suggest that only experienced interpreters make use of contextual effects to predict (e.g., Frauenfelder & Schriefers 1997; M. Liu 2008; Vandepitte 2001).

The participants in the present study predicted both before and after training, and we found no evidence that training affected the extent of their prediction. This is in line with Chmiel's (2021) study, which found that training did not affect prediction as measured by the influence of context in word-translation latencies. We did, however, find that the speech rate at which sentences were spoken had a significant effect on the onset time of the interpretation, with the onset time being earlier for sentences spoken at a faster rate. This could be because the participants began their interpretation sooner so as to keep up better with the speaker, or else because the participants comprehended more elements of the sentence earlier and for this reason could begin their interpretation (or both). Whereas the effect of speech rate on lag was not a focus of this study, a future study could explore whether lag is consistently shorter when the input speech rate is faster.

In an exploratory analysis, we found evidence of a link between the time lag between the onset of the interpretation and the onset of the critical word, and predictive fixations on the target compared to the unrelated object. Unlike in Amos et al. (2022), this effect was apparent only after the onset of the predictable word. This difference in fixations on the target compared to the unrelated object was greater when the participants began their interpretation later. It may be that people fixate proportionately more on the predictable object because they are concentrating harder in order to remember it for when they will have to produce it later in their own utterance. Alternatively, the participants might be focusing on the image in order to anticipate language production, as they might do when interpreting between language pairs with mismatching syntactic structures (Setton & Dawrant 2016).

There are alternative explanations for the lack of difference in predictive behaviour before and after training. In this study, ten of the 22 students failed their interpreting exam from English on the first attempt. If students had not fully mastered the task of interpreting at the second timepoint in our study, we might wonder whether, with more training or experience, they might predict to a greater extent. This could explain the discrepancy between our findings and those of Özkan et al. (2022), who compared experienced interpreters with student interpreters. However, we also did not find a correlation between the students' final grades and the extent of their prediction, so we have no evidence to suggest a link between level of expertise (as measured by grades) and predictive processing.

The lack of an effect of training might be linked to the stimuli used in our study, which were sentences judged as highly predictable in their own right. The interpreting literature tends to refer to *knowledge* that interpreters use in order to make predictions in certain situations rather than an ability to form predictions in general (Moser 1978). The sentences we used were designed to be highly predictable in English on their own (without further context). In other words, training in SI could improve prediction in trainee interpreters by providing them with knowledge rather than by training predictive processing either directly (by training prediction) or indirectly (through cognitive enhancement). Using stimuli that provide a fuller context could test this hypothesis. Another point is that the stimulus sentences were highly constraining (cloze probability of around 90%), so that predicting these sentences may have been easy for students, even before training, leaving little room for progress after training. A future study could consider whether students predict to a greater extent after training when listening to sentences that are moderately, rather than extremely, predictable. The listening conditions could also be made more difficult – for example, by adding a noisy background or using accented speech.

To conclude, we found evidence that prediction takes place during SI in both trained and untrained interpreters, but we did not find evidence to suggest that semantic prediction during SI changes over time due to training or that it is linked to performance (as measured by grades). However, a number of questions have emerged that could be the focus of future studies.

Funding information

Rhona M. Amos was funded by a Swiss National Science Foundation (SNF) doc.mobility fellowship, with project number P1GEP1_188165.

References

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

Amos, R. M. & Pickering, M. J. (2020). A theory of prediction in simultaneous interpreting. *Bilingualism: Language and Cognition 23* (4), 706–715. doi: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

Babcock, L., Capizzi, M., Arbula, S. & Vallesi, A. (2017). Short-term memory improvement after simultaneous interpretation training. *Journal of Cognitive Enhancement 1* (3), 254–267. doi:10.1007/s41465-017-0011-x

Babcock, L. & Vallesi, A. (2017). Are simultaneous interpreters expert bilinguals, unique bilinguals, or both? *Bilingualism: Language and Cognition 20* (2), 403–417. doi:10.1017/S1366728915000735

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. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67* (1).

Boersma, P. & Weenink, D. (2002). Praat: Doing phonetics by computer (Version 6.2.08). http://www.praat.org.

Borovsky, A., Elman, J. L. & Fernald, A. (2012). Knowing a lot for one’s age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology 112* (4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

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*. Amsterdam: John Benjamins, 139–153.

Chmiel, A. (2016). In search of the working memory advantage in conference interpreting – training, experience and task effects. *International Journal of Bilingualism 22* (3), 371–384. doi:10.1177/1367006916681082

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

DeLong, K. A., Urbach, T. P. & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience 8*: 1117. doi: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. doi:10.1017/S1366728916000547

Dikker, S., Rabagliati, H., Farmer, T. A. & Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science 21* (5), 629–634. doi:10.1177/0956797610367751

Dong, Y. & Liu, Y. (2016). Classes in translating and interpreting produce differential gains in switching and updating. *Frontiers in Psychology 7*: 1297. doi:10.3389/fpsyg.2016.01297

Fan, D. C., Collart, A. & Chan, S-H. (2022). When two languages are competing: An ERP study of sentence processing in expert and novice interpreters. *Interpreting 24* (1), 1–37. https://doi.org/10.1075/intp.00069.fan

Frauenfelder, U. & Schriefers, H. (1997). A psycholinguistic perspective on simultaneous interpretation. *Interpreting 2* (1/2), 55–89. doi:10.1075/intp.2.1-2.03fra

Gile, D. (1997). Conference interpreting as a cognitive management problem. In F. Pöchhacker & M. Shlesinger (Eds), *The interpreting studies reader*. London: Routledge, 162–178.

Hervais-Adelman, A., Moser-Mercer, B., Murray, M. M. & Golestani, N. (2017). Cortical thickness increases after simultaneous interpretation training. *Neuropsychologia 98*, 212–219. https://doi.org/10.1016/j.neuropsychologia.2017.01.008

Hoffman, R. R. (1997). The cognitive psychology of expertise and the domain of interpreting. *Interpreting 2* (1/2), 189–230. https://doi.org/10.1075/intp.2.1-2.08hof

Ito, A., Corley, M. & Pickering, M. J. (2017). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition 21* (2), 251–264. doi:10.1017/S1366728917000050

Ito, A., Gambi, C., Pickering, M., Fuellenbach, K. & Husband, E. (2020). Prediction of phonological and gender information: An event-related potential study in Italian. *Neuropsychologia 136*: 107291. doi:10.1016/j.neuropsychologia.2019.107291

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

Kalina, S. (1998). *Strategische Prozesse beim Dolmetschen. Theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen*. Tübingen: Gunter Narr.

Kamide, Y., Altmann, G. T. M. & 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. doi:10.1016/S0749-596X(03)00023-8

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. doi:10.1016/j.jml.2010.07.003

Liontou, K. (2015). Anticipation. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies*. London/New York: Routledge, 15–17.

Liu, M. (2008). How do experts interpret? Implications from research in Interpreting Studies and cognitive science. In G. Hansen, A. Chesterman & H. Gerzymisch-Arbogast

(Eds.), *Efforts and models in interpreting and translation research: A tribute to Daniel Gile*. Amsterdam: John Benjamins. 159–177.

Liu, M., Schallert, D. L. & Carroll, P. J. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting 6* (1), 19–42. doi:10.1075/intp.6.1.04liu

Liu, Y., Hintz, F., Liang, J. & Huettig, 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. doi:10.1017/S1366728922000232

Lozano-Argüelles, C. & Sagarra, N. (2021). Interpreting experience enhances the use of lexical stress and syllabic structure to predict L2 word endings. *Applied Psycholinguistics 42* (5), 1135–1157. doi: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. doi:10.1017/S1366728919000634

Marian, V., Blumenfeld, H. K. & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research 50* (4), 940–967. doi:10.1044/1092-4388(2007/067)

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A. & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language 69* (4), 574–588. https://dx.doi.org/10.1016/j.jml.2013.08.001

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

Mitsugi, S. & Macwhinney, B. (2015). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition 19* (1), 19–35. doi:10.1017/S1366728914000881

Moser, B. (1978). Simultaneous interpretation: A hypothetical model and its practical application. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication*. New York: Plenum Press, 353–368.

Otten, M. & Van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research 1291*, 92–101. https://doi.org/10.1016/j.brainres.2009.07.042

Özkan, D., Hodzik, E. & Diriker, E. (2022). Simultaneous interpreting experience enhances the use of case markers for prediction in Turkish. *Interpreting* 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. doi:10.1037/bul0000158

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America 57* (8), 1030–1033. doi:10.1364/JOSA.57.001030

Seeber, K. G. (2001). Intonation and anticipation in simultaneous interpreting. *Cahiers de Linguistique Française 23*, 61–97.

Seleskovitch, D. (1984). Les anticipations de la compréhension. In D. Seleskovitch & M. Lederer (Eds.), *Interpréter pour traduire*. France: Didier Erudition, 273–283.

Setton, R. (2002). Deconstructing SI: A contribution to the debate on component processes. *The Interpreters' Newsletter 11*, 1–26.

Setton, R. & Dawrant, A. (2016). *Conference interpreting: A trainer's guide*. Amsterdam: John Benjamins.

Tzou, Y.-Z., Eslami, Z. R., Chen, H.-C. & Vaid, J. (2012). Effect of language proficiency and degree of formal training in simultaneous interpreting on working memory and interpreting performance: Evidence from Mandarin–English speakers. *International Journal of Bilingualism 16* (2), 213–227. doi:10.1177/1367006911403197

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. doi:10.1037/0278-7393.31.3.443

Van Besien, F. (1999). Anticipation in simultaneous interpretation. *Meta 44* (2), 250–259.

Van de Putte, E., De Baene, W., García-Pentón, L., Woumans, E., Dijkgraaf, A. & Duyck, W. (2018). Anatomical and functional changes in the brain after simultaneous interpreting training: A longitudinal study. *Cortex 99*, 243–257. https://doi.org/10.1016/j.cortex.2017.11.024

Vandepitte, S. (2001). Anticipation in conference interpreting: A cognitive process. *Revista Alicantina de Estudios Ingleses/Alicante Journal of English Studies 14*, 323–335. doi:10.14198/raei.2001.14.18

Appendix

Stimulus Set No. 1

	Sentence + cloze value (%)	**Object name + naming agreement (%)**
1	In order to have a closer look, the dentist asked the man to open his mouth a little wider. (100)	mouth (83.3)
		bone (100)
2	In an emergency, we cannot use a lift; instead, we need to use the stairs for our safety. (100)	stairs (100)
		calculator (91.7)
3	If the sun comes out during a heavy shower, you can sometimes see a rainbow in the sky. (100)	rainbow (100)
		goat (91.7)
4	The tourists expected rain when the sun went behind the cloud, but the weather got better later. (95)	cloud (91.7)
		train (83.3)
5	The man didn't know the time because he forgot to wear the watch that he usually wears. (100)	watch (91.7)
		tray (58.3)
6	Bob proposed and gave her a ring that had cost him half his monthly wage. (85)	ring (91.7)
		letter (100)
7	The man was gathering honey, when he was stung by a bee and gave a cry. (95)	bee (91.7)
		helmet (83.3)
8	People can easily go to the island on foot since the government built a bridge last year. (90)	bridge (100)
		meat (83.3)
9	To make sushi, the chef went to the market to buy some fish early in the morning. (60)	fish (100)
		moon (100)
10	During winter, it's best to put on your heaviest coat and a hat. (70)	coat (58.3)
		pineapple (75)
11	The maid dusted the books on the shelf every week. (90)	shelf (91.7)
		pig (91.7)
12	He scraped the cold food from his plate into the bin. (80)	plate (91.7)
		grape (91.7)
13		milk (100)

	Joan fed her baby some warm milk and then put him to bed. (70)	salt (100)
14	She went to the beauty parlour to perm her hair in preparation for the party. (90)	hair (100)
		bamboo (91.7)
15	One day, the caterpillar will turn into a beautiful butterfly and fly away. (100)	butterfly (100)
		giraffe (91.7)
16	Catherine carried her computer in a shoulder bag until she found it was giving her back problems. (100)	bag (66.7)
		kiwi (100)

Stimulus Set No. 2

1	The traveller went to the desert because he wanted to ride a camel and go exploring. (90)	camel (91.7)
		barrel (50)
2	The woman found the room was too hot and humid, so to get some fresh air, she opened the window completely. (100)	window (83.3)
		globe (66.7)
3	The bird cannot fly because it injured its wing when it had a fight with another bird. (100)	wing (100)
		flag (83.3)
4	Amber went to the dealership to purchase a new car the very next day. (80)	car (83.3)
		bear (83.3)
5	To protect against an enemy's bullet or arrows, soldiers used to carry a shield all the time.(70)	shield (83.3)
		onion (100)
6	Before he began to draw, he sharpened his pencil and got out some paper. (67)	pencil (91.7)
		cherry (100)
7	In order to study, Karen sat down at her desk and opened her book. (90)	desk (100)
		chick (75)
8	In the night sky it is easier to see all the stars and the moon. (100)	star (91.7)
		key (100)
9	After every meal it’s good to brush your teeth or else chew gum. (100)	teeth (91.7)
		horse (100)
10	Dad carved the turkey with a knife for Christmas dinner. (90)	knife (83.3)
		rabbit (83.3)
11	He loosened the tie around his neck and immediately felt better. (100)	neck (91.7)
		handcuff (58.3)
12	A flat tyre forced Katy to pull up at the side of the road and call for assistance. (100)	road (100)
		glasses (83.3)
13	The referee blew his whistle to signal the end of the match. (100)	whistle (91.7)
		pen (83.3)
14	The student went to the library to read a book but in the end he ended up chatting with his friends. (90)	book (100)
		dice (91.7)
15	John was very tired so he decided to go straight to bed and sleep. (100)	bed (100)
		glove (100)
16	To reach the roof, the workman climbed up the ladder that was against the wall. (100)	ladder (91.7)
		bench (66.7)