ELSEVIER

Prediction during simultaneous interpreting: Evidence from the visual-world paradigm

Rhona M. Amos [a,*], Kilian G. Seeber [a], Martin J. Pickering [b]

[a] *University of Geneva, Uni Mail, 40, Boulevard du Pont d'Arve, 1205 Genève, Switzerland*
[b] *University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, Scotland, United Kingdom*

A R T I C L E   I N F O

A B S T R A C T

We report the results of an eye-tracking study which used the Visual World Paradigm (VWP) to investigate the time-course of prediction during a simultaneous interpreting task. Twenty-four L1 French professional conference interpreters and twenty-four L1 French professional translators untrained in simultaneous interpretation listened to sentences in English and interpreted them simultaneously into French while looking at a visual scene. Sentences contained a highly predictable word (e.g., The dentist asked the man to open his *mouth* a little wider). The visual scene comprised four objects, one of which depicted either the target object (*mouth; bouche*), an English phonological competitor (*mouse; souris*), a French phonological competitor (*cork; bouchon*), or an unrelated word (*bone; os*). We considered 1) whether interpreters and translators predict upcoming nouns during a simultaneous interpreting task, 2) whether interpreters and translators predict the form of these nouns in English and in French and 3) whether interpreters and translators manifest different predictive behaviour. Our results suggest that both interpreters and translators predict upcoming nouns, but neither group predicts the word-form of these nouns. In addition, we did not find significant differences between patterns of prediction in interpreters and translators. Thus, evidence from the visual-world paradigm shows that prediction takes place in simultaneous interpreting, regardless of training and experience. However, we were unable to establish whether word-form was predicted.

1. Introduction

There is strong evidence that comprehenders often predict what they are about to hear (Huettig, 2015; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). Similarly, most theoretical accounts assume that simultaneous interpreters regularly predict what they are about to hear as they work (Gerver, Longley, Long, & Lambert, 1984; Moser-Mercer, Frauenfelder, Casado, & Künzli, 2000; Setton, 2005). Although several studies have identified instances when interpreters produce a translation (in the target language) before they hear the complete source utterance (Seeber, 2001; Van Besien, 1999; Wilss, 1978), no study has measured predictive processing in trained interpreters and untrained bilinguals online during a simultaneous interpreting task.

In this study, we investigated the time course of prediction in simultaneous interpreting by tracking the eye movements of professional conference interpreters and professional translators untrained in simultaneous interpreting as they looked at a visual scene and simultaneously interpreted English sentences into French. We hypothesized

that, despite the challenging conditions created by the concurrent production task, both interpreters and translators would predict upcoming meaning. However, it was also plausible that neither group would engage in prediction because of the challenging listening conditions, or that simultaneous interpreters might engage in prediction whereas translators might not.

1.1. Prediction in a native language

Much evidence indicates that comprehenders predict semantic, syntactic and phonological aspects of upcoming words in a native language (Pickering & Gambi, 2018). Evidence for semantic prediction comes from Altmann and Kamide (1999), who presented participants with visual scenes showing an agent and four objects, such as a boy sitting on the floor of his room with a cake, a train set, a toy car and a balloon. Participants heard sentences containing a verb which was compatible with either one or all four of the objects serving as its patient, such as "The boy will *eat* the…" or "The boy will *move* the…". When

participants heard the verb "eat" they began looking at the cake before noun onset, but when they heard the verb "move" they did not, indicating that they used information from the verb to predict the meaning of the upcoming noun. Semantic prediction is supported by findings from similar eye-tracking studies using the visual world paradigm (Mani & Huettig, 2012; Rommers, Meyer, Praamstra, & Huettig, 2013) and other studies that do not involve a visual context (Grisoni, McCormick Miller, & Pulvermüller, 2017).

There is evidence that comprehenders predict syntactic and phonological aspects of upcoming words, although this evidence is not fully consistent. In an ERP study, Otten, Nieuwland, and Van Berkum (2007) had participants listen to short texts, half of which ended with a highly predictable noun such as "cross", rather than a less predictable noun, such as "crucifix", in the (Dutch translation of) the sentence "My grandfather and grandmother are very religious. Above the head of their bed hangs a big…". They found a negative deflection on the adjective that agreed in gender with the less predictable noun (starting 300 ms after adjective onset), indicating that participants predicted the syntax of the predictable noun. In another study, in which participants read short texts similar to those used in Otten et al. (2007), Otten and Van Berkum (2008) found a negative ERP effect, but this appeared later relative to adjective onset. In contrast, Wicha, Moreno, and Kutas (2004) found a positive ERP effect on articles that did not agree in gender with a predictable noun in Spanish, and Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005) also found an early positivity on adjectives that did not agree in gender with a predictable noun. However, in a larger-scale study which was a close replication of Van Berkum et al. (2005), Nieuwland, Arkhipova, and Rodríguez-Gómez (2020) failed to replicate these adjective effects, and in fact found weak evidence of a negative effect. It is possible that syntactic predictions just include the gender of the upcoming noun or that people also predict a specific article (see Fleur, Flecken, Rommers, & Nieuwland, 2020). There is also evidence that comprehenders predict the form of an upcoming word (DeLong, Urbach, & Kutas, 2005; Ito, Corley, Pickering, Martin, & Nieuwland, 2016; Ito, Gambi, Pickering, Fuellenbach, & Husband, 2020; Ito, Pickering, & Corley, 2018; Laszlo & Federmeier, 2009), although not all studies have found such effects (Nieuwland et al., 2018).

1.2. Prediction in adverse conditions

The above studies provide evidence, most convincingly from the visual world paradigm, for prediction in a native language in quiet laboratory conditions. But a simultaneous interpreting task means listening in adverse listening conditions. Firstly, during an interpreting task, most interpreters listen in a non-native language, in which they are less proficient than their mother tongue. Secondly, the speech signal that they receive is imperfect, because they produce utterances while listening, which means they must comprehend noisy speech. Thirdly, engaging the production mechanism during comprehension may, in itself, limit prediction. Despite its name, simultaneous interpreting is not fully simultaneous, and production is almost always time-delayed compared to comprehension. This means that the interpreter often produces the same utterance at a lag (something which may be akin to producing an unrelated utterance) and so the production mechanism may not assist in making an appropriate prediction. (On the other hand, when the interpreter produces an utterance whose timing is closely synchronized with the source utterance, the production mechanism might facilitate prediction.) Finally, interpreters comprehend incoming speech under increased cognitive load, because they must remember, reformulate, and produce the incoming message while listening. Therefore, prediction during simultaneous interpreting may be limited or impeded.

Indeed, there is evidence that prediction in L2, prediction in noise, prediction with concurrent (unrelated) production and prediction under cognitive load are all more limited than prediction in L1 in ideal conditions. First, Martin et al. (2013) compared prediction of semantic and

phonological content among native English speakers and late Spanish-English bilinguals as they read highly predictable sentences in English. Following DeLong et al. (2005), sentences contained a more or less predictable noun which was either vowel or consonant initial (e.g., "He was very tired so he sat on *a chair/an armchair.*") Both groups showed a reduced N400 effect on the more predictable noun, but only the native English speakers also showed a reduced N400 effect on the article that preceded the more predictable noun. (Ito, Martin, & Nieuwland, 2017b point out that Martin et al., 2013 use an atypical reference channel, so these results may not generalise well.) More recently, Ito et al. (2018) compared the time-course of prediction in L1 and L2 speakers of English, and found that L2 speakers' predictive eye movements were delayed in comparison to those of L1 speakers, and L2 speakers did not make predictive fixations on phonological competitors. These studies suggest that L2 speakers may not predict word-form as L1 speakers do (see also Ito, Martin, & Nieuwland, 2017a).

Syntactic predictions also may not take place in L2 as they do in L1. For instance, Mitsugi and Macwhinney (2015) showed that intermediate L2 speakers of Japanese did not use case-marking information predictively (although this could have been because of the particularly complex Japanese case-marking system). This in turn suggests that some aspects of predictive processing are not automatic, and so the occurrence of prediction is linked to the time and resources available (Ito & Pickering, 2021).

Semantic prediction, on the other hand, may be less affected by non-nativeness. In a visual-world experiment based on Altmann and Kamide (1999), Dijkgraaf, Hartsuiker, and Duyck (2017) had participants listen to sentences which were either constraining (e.g., *"Mary knits a scarf"*) or not (e.g., *"Mary loses a scarf"*) and look at scenes depicting four objects, of which only one could be knitted, but all could be lost. Participants were one group of Dutch-English bilinguals, who listened in both Dutch and English, and one group of English monolinguals who listened in English. They found that both groups made predictive eye movements, and that the effect of condition (constraining vs. non-constraining) was similar for L1 listening in English and in Dutch, as well as for Dutch participants listening in their L2. This might be because the semantic level is shared between languages, and so participants predict the same meaning.

Now let us consider prediction in noisy conditions. In a Bayesian account, word predictability may have a greater influence on word recognition than bottom-up input when the speech signal is less clear (Norris & McQueen, 2008). There is also evidence that increased attention when listening to noisy speech leads to greater reliance on top-down processing, which may include prediction (Wild et al., 2012). However, listening to noisy speech in an L2 may not affect prediction in the same way. Mayo, Florentine, and Buus (1997) found that even when L2 speakers showed native-like comprehension in quiet conditions, their comprehension was more degraded in noise than that of L1 speakers. Meanwhile, Mattys, Carroll, Li, and Chan (2010) found that L2 speakers relied more on acoustic cues than top-down strategies when listening to noisy speech.

Producing speech that is unrelated to what is being comprehended may also limit prediction. Martin, Branzi, and Bar (2018) had participants read sentences while either engaging in a concurrent task (tongue tapping), listening task (listening to /ta/), or production task (producing /ta/). Participants in the concurrent production condition predicted less than those in the other two groups, suggesting that occupying the production mechanism limits prediction. On the other hand, comprehending and producing closely related utterances at the same time – that is, utterances in different languages but with (ideally) similar meaning – may support, rather than impede, predictive processing. By some accounts, prediction takes place using the production mechanism (see Pickering & Gambi, 2018 for a review), and there is evidence suggesting a link between prediction and production (Adank, 2012; Drake & Corley, 2015; Hintz, Meyer, & Huettig, 2017; Mani & Huettig, 2012). Where comprehenders engage their production mechanism by concurrently

producing the same content, this may trigger predictive processing. The production mechanism might be engaged in this way when comprehension and production are very synchronized.

There is evidence that cognitive load limits prediction. In a study similar to Dijkgraaf et al. (2017), Ito, Corley, and Pickering (2017) had L1 and L2 participants listen to constraining or non-constraining sentences while they either did or did not have to remember five unrelated words. For both groups, the additional load generated by memorizing unrelated words led to delayed predictive eye movements, suggesting that prediction requires cognitive resources in both native and non-native speakers. In addition, Huettig and Janse (2016) found that individual differences in working memory affected predictive processing, with participants with better working memory making more predictive eye movements (although see also Otten & Van Berkum, 2009 who found that predictive processing was similar in participants with both high and low working memory capacity).

Some factors may mitigate the adverse effects of comprehending in L2 and cognitive load, and thus support prediction during simultaneous interpreting. For instance, highly proficient users of an L2 may predict more similarly to L1 speakers (Kaan, 2014). Chambers and Cooke (2009) had native English speakers listen to sentences in their L2, French, as they looked at four objects, one of which was the referent of a target word (e.g., poule [chicken]), and one of which was the referent of an English interlingual homophone (e.g., pool) in the sentence "Marie va *nourrir/décrire* la poule" [Marie will *feed/describe* the chicken]. When participants heard the constraining verb (feed), they looked at the target object (chicken), and rarely considered the interlingual homophone (pool), showing that they had limited their expectations to the predictable noun before word onset. Proficiency in French and predictive fixations were positively correlated.

Other comparisons of higher and lower proficiency bilinguals have found that higher proficiency bilinguals make predictions based on gender markings (syntax), even when gender markings are not present in their L1, whereas lower proficiency bilinguals do not (Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013; Hopp, 2013). Where L1 and L2 words share grammatical gender, gender-marking based prediction is supported (Dussias et al., 2013; Foucart, Martin, Moreno, & Costa, 2014).

In addition, certain non-language aspects of the simultaneous interpreting setting may also support prediction. Specifically, the referents of utterances are often visually present in a simultaneous interpreting context. For example, during conference interpreting, interpreters almost always see the speaker, and may also view Power-Point presentations or other documents or images while simultaneously interpreting.

1.3. Prediction in simultaneous interpreting

We have reviewed evidence showing that prediction may, but need not always, occur at all linguistic levels (see Pickering & Gambi, 2018), that L2 speakers may predict more slowly than L1 speakers (Ito et al., 2018), that cognitive resources are needed for prediction (Ito, Corley, et al., 2017), and that listening in L2 in noisy conditions may increase reliance on bottom-up processing strategies (Mattys et al., 2010). In addition, there is evidence that concurrent production (of irrelevant speech) may impede prediction (Martin et al., 2018). Although these adverse comprehension conditions may be somewhat mitigated by high L2 proficiency, similarities between L1 and L2, the presence of visual referents, and nearly synchronized concurrent production, the evidence provides reasons to expect that prediction might be impaired in simultaneous interpreting.

In spite of this, most accounts of simultaneous interpreting assume a key role for prediction (Gerver et al., 1984; Moser, 1978; Moser-Mercer et al., 2000; Seeber, 2001; Seleskovitch, 1984; Setton, 2005). It is included as a processing stage in one of the earliest process models of simultaneous interpreting (Moser, 1978). Setton (2005) suggested that

an ability to predict is a prerequisite for success in simultaneous interpreting and Chernov (2004) even proposed that being able to anticipate how a message will develop is what makes simultaneous interpreting possible. Indeed, prediction may allow interpreters to ignore parts of the input and focus entirely on production or memorizing (De Groot, 2011). Prediction has been described as both a *skill* (Moser-Mercer, 2000) and a *strategy* (Seeber, 2001; Setton, 2001; Van Besien, 1999) used by simultaneous interpreters. This implies that interpreters either have or develop (implicitly or explicitly) a special ability to predict during the task of simultaneous interpreting that other groups may not have. In other words, theories from the Interpreting Studies literature posit that trained interpreters alone may use predictive cues during interpretation (Frauenfelder & Schriefers, 1997), and that both training and experience may be necessary to engage (strategically) in prediction during simultaneous interpreting (e.g., Moser, 1978).

Evidence that interpreters predict during interpreting comes from the observation that they sometimes produce an utterance in the target language before hearing it in the source language (Hodzik & Williams, 2017; Seeber, 2001; Van Besien, 1999; Wilss, 1978). However, these studies tend to be based on theories according to which interpreters alone (are able to) predict during a simultaneous interpreting task (Seeber, 2001; Van Besien, 1999; Wilss, 1978), and the question of whether untrained bilinguals predict during a simultaneous interpreting task has not been extensively investigated. Only Hodzik and Williams (2017) considered whether interpreters have a shorter lag compared to untrained bilinguals when interpreting predictable content. They found no significant difference between groups, although note that their interpreters were mainly students of interpreting, rather than trained professionals.

To our knowledge, no study has used the visual-world paradigm to track the time course of prediction in interpreters and bilinguals untrained in interpreting *during* an interpreting task. Use of the visual world paradigm makes it possible for us to study the time-course of prediction while participants are engaged in a concurrent production task, and thus investigate whether, during an interpreting task, prediction is specific to interpreters or generalises to other bilingual populations. In the visual-world paradigm, the experimenter infers prediction from looks to pictures (or objects) before they are mentioned (e.g., Altmann & Kamide, 1999; see Pickering & Gambi, 2018, for discussion). It is true that the visual referents in an interpreting context are unlike such pictures (interpreters may see speakers, as well as their gestures and facial expressions, a presentation, and objects such as name plates and the rostrum rather than four images on a screen). However, a visual context is regularly present in a conference interpreting context (Seeber, 2017) and may aid comprehension, just as the visual scenes presented in the visual-world paradigm may.

A related question is whether interpreters and bilinguals untrained in simultaneous interpreting are able to make word-form predictions during an interpreting task. For example, is it possible to predict the phonological form of an upcoming word as well as its meaning during a simultaneous interpreting task? Such predictions may be particularly advantageous in simultaneous interpreting, as they would allow interpreters to plan their own upcoming utterance with precision (Amos & Pickering, 2020). Word-form prediction would be compatible with an account by which comprehenders use their production mechanism to predict by working through the same stages, in the same order, as during production – from meaning, to syntax, to sound, but without articulating (Pickering & Gambi, 2018). If so, we would expect people to make predictions primarily in the language in which they are comprehending (the source language) during an interpreting task. However, given that both languages are strongly activated in simultaneous interpreting, predictions may also be formed in the target language. Alternatively, the concurrent activation of both languages, and the regular switching of focus from comprehension to production and back between the two languages, may lead to a weaker activation of each language. This might lead to an apparent lack of word-form prediction.

1.4. The current study

We designed a study to test whether simultaneous interpreters and translators untrained in simultaneous interpreting make predictions during simultaneous interpreting, and to shed light on how specific these predictions might be in both groups. We chose to test both simultaneous interpreters and professional translators, so that we could test two groups who are similar in terms of language proficiency and age, and are used to working with two languages at the same time (meaning that both groups are likely able to carry out the task of simultaneous interpreting). Testing both a group of interpreters and a group of translators allows us to see whether prediction during simultaneous interpreting is possible only for trained interpreters, or whether untrained bilinguals also engage in prediction during a simultaneous interpreting task.

Native French-speaking participants listened to highly constraining sentences in English and simultaneously interpreted these sentences into French. Based on Ito et al. (2018), they viewed a visual scene containing three distractors (which were pictures of unrelated objects) and a picture of a critical object, whose name was one of the target word (e.g., *mouth*; bouche), a word phonologically related to the English form of the target word (English phonological competitor; e.g., *mouse*; souris), a word phonologically related to the French form of the target word (French phonological competitor; e.g., *bouchon*; cork), or an unrelated word (e. g., *bone*; os).

The timing of fixations in the visual array is linked to underlying comprehension processes (Tanenhaus, Magnuson, Dahan, & Chambers, 2000). We therefore measured the timing of fixations on the critical objects. If participants predict while simultaneously interpreting, they should fixate target objects more than unrelated objects. Such predictive looks would demonstrate that L2 listeners predict upcoming utterances while simultaneously interpreting. If participants fixate English phonological competitor objects more than unrelated objects, this would demonstrate that L2 listeners pre-activate phonological information in their L2. If they fixate French phonological competitor objects more than unrelated objects, this would demonstrate that listeners engage in cross-linguistic prediction. If the simultaneous interpreter group predicts more or earlier than the translator group, this would demonstrate that interpreters engage in more or earlier prediction during simultaneous interpreting than translators.

2. Material and methods

2.1. Participants

Twenty-five conference interpreters working in Geneva, whose A language was French and whose B or C language was English, participated in the experiment.[1] One participant was excluded from the analysis because they almost never (less than 3% of the time) fixated the depicted objects (on experimental items and filler items). All participants had normal or corrected-to-normal vision and reported no language disorders. Most participants (*n* = 20) were members of the International Association of Conference Interpreters (AIIC), which promotes professional ethics, standards and conditions in conference interpreting, were accredited to international organisations such as the UN, or were both AIIC members and accredited. The four remaining

participants were professional interpreters working in Geneva.

Twenty-five translators working in Geneva, who translated into French from languages including English, participated in the experiment. Three of the participants were trained translators who no longer worked as translators but in related fields (e.g., management). One additional participant was tested but the results were excluded at random because only 24 participants were necessary to match the interpreter condition. All participants had normal or corrected-to-normal vision and reported no language disorders.

We determined sample size from Ito et al. (2018), which used items (sentences and pictures) with the same characteristics and experimental structure as our experiment. There is no comparable work using interpreters (or translators) and so we did not have any data on which to conduct an appropriate a priori power analysis. In addition, the population of French-to-English professional interpreters is extremely limited, especially given that we wished to recruit them from the well-established community in Geneva (a community with which we are highly familiar). For instance, there are currently 87 such interpreters in Geneva who are members of the International Association of Conference Interpreters (AIIC). As recruitment of such busy professionals is challenging, we decided to recruit the same number of participants per group as in Ito et al. (2018).

The groups were matched for factors pertaining to background and language background (see Table 1). All participants completed a language background questionnaire, based on the Leap-Q questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007). Unlike in the original questionnaire, participants were asked to provide language background information only on French and English (rather than on all of their languages), and used a five-point instead of a ten-point scale for the self-proficiency ratings. Participants also provided information about their professional background.

The two groups did not differ in age of acquisition, age of fluency, time living in an area where French was spoken, and time living in an area where English was spoken. However, the translator group had a significantly higher overall current exposure to French (see Table 1). Participants also rated their language proficiency for speaking, reading, and listening, in French and English, on a five-point Likert-type scale (from 1 = "very low" to 5 = "very high"). With the exception of one

---

[1] A language classification system unique to the interpreting profession is used to describe the languages in an interpreter's combination. The A-language is the language in which the interpreter is most proficient, and is a target language into which the interpreter works from any of the languages in his or her combination. The C-language is a source language from which the interpreter works. The B-language is a source language and target language in which the interpreter is perfectly fluent, but nonetheless less proficient than the A-language. (AIIC, 2019)

Table 1

Background information provided by both groups in the language background questionnaire, and a comparison (*t*-test for numerical variables, Wilcoxon test for ordinal variables).

	Interpreters (*n* = 24)	Translators (n = 24)	*Comparison*
Background			
Age *(yrs)*	45.17 ± 10.98	40.75 ± 12.49	
Languages spoken	4.96 ± 1.40	4.08 ± 1.02	*
Time in profession *(yrs)*	16.92 ± 10.2	12.96 ± 10.05 (*n21*)	
French language			
Age *(yrs)* of Acquisition	2.33 ± 2.57	1.42 ± 2.47	
Age *(yrs)* of Fluency	4.21 ± 2.55	3.27 ± 2.27	
Time *(yrs)* living in FR area	35.89 ± 13.58 (*n20*)	34.81 ± 14.47	
Current exposure %	47.92 ± 14.36	62.83 ± 15.02	***
English language			
Age *(yrs)* of Acquisition	8.71 ± 4.53	10.12 ± 3.29	
Age *(yrs)* of Fluency	15.42 ± 6.88	17.50 ± 6.26	
Time *(yrs)* living in EN area	3.18 ± 3.80 (*n23*)	6.17 ± 10.36	
Current exposure %	27.21 ± 9.04	21.38 ± 12.52	
Self-rated English proficiency			
Speaking	4.33 ± 0.76	4.08 ± 0.88	
Reading	4.88 ± 0.34	4.79 ± 0.41	
Listening	4.88 ± 0.34	4.33 ± 0.87	**

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

translator, all participants rated their French proficiency as 5 for all areas of ability. There were no between-group differences in self-rated reading or writing proficiency in English. However, interpreters rated their listening ability in English as higher than translators (see Table 1). It is likely that this difference in perceived listening proficiency is linked to the different professional profiles of the two groups, given that listening in English is a key component of the interpreting process.

2.2. Stimuli

Experimental stimuli consisted of 32 English sentences, each paired with a visual array containing three distractor objects and one of four critical objects (see Appendix 1). Critical objects appeared in each of the four quadrants equally frequently following a Latin-square design. The experimental sentences each contained a highly predictable word (e.g., *mouth,* in "The dentist asked the man to open his *mouth* a little wider.") at varied positions in the sentence (range = 5th – 20th word, M = 10.8, SD = 3.18), but never sentence finally. The sentences consisted of a mean of 15.4 words (range = 10–21, SD = 2.85). Stimuli sentences were based on Ito et al. (2018) and Block and Baldwin (2010), or else were designed by the authors.

There were an additional 32 filler sentences. These sentences were designed to not be constraining for any particular word. These sentences were paired with the same visual scenes as the experimental sentences, but the quadrants in which the objects appeared were varied. Filler sentences mentioned distractor objects 75% of the time, so together with the experimental sentences, which mentioned a critical object present 25% of the time (i.e., in the target condition), the sentences mentioned one of the objects in the visual scene 50% of the time.

The sentences were recorded at a sampling rate of 48 kHz in a sound-proof recording studio by a male native Southern British English speaker. The speaker read the experimental sentences at a rate of 2.01 syllables per second (SD = 0.27). The mean sentence duration was 9.87 s (SD = 1.63). The mean onset time of the critical word was 6.28 s (SD = 1.94) (see Table 2).

The predictability of the target words was assessed using a cloze probability test (online, via LimeSurvey). First, 20 native speakers of French who were proficient in English read 40 English sentences that we judged could be completed with a predictable word, and completed each sentence with the first word that came to mind. Then, 14 different native speakers of French were asked to complete the same English sentences with a word in French. This was important in ensuring that there was one clearly predictable word in both languages. We were left with 27 sentences that were high cloze for both the English and the French word. We then constructed five further sentences. Twelve participants completed these sentences with a word in English, and 10 participants with a word in French. The final mean cloze probabilities were 90.7% (SD = 11.9, range = 60–100%) for the English word, and 91.7% (SD = 10.9, range = 57–100%) for the French word (see Table 2).

Each of the visual scenes contained four objects: a critical object and three distractor objects. In the target condition, the critical object corresponded to the predictable word (e.g., *mouth* [French: *bouche*]). In the English competitor condition, the English name of the critical object phonologically overlapped at onset with the predictable word (e.g.,

mouse [*souris*]). In the French competitor condition, the French name of the critical object phonologically overlapped at onset with the French translation of the predictable word (e.g., *cork* [*bouchon*]). The mean number of phonemes shared between the predictable words and English competitor words was 2.2 (SD 0.55) out of a mean of 3.6 phonemes (61%). The mean number of phonemes shared between French translations of predictable words and French competitor words was 2.1 (SD 0.53) out of a mean of 4.2 phonemes (50%).[2] English and French names of the predictable objects were unrelated to each other phonologically. The translation of the English phonological competitor was unrelated to the French translation of the predictable word, and the translation of the French phonological competitor was unrelated to the English translation of the predictable word. In the unrelated condition, the name of the critical object did not have phonological onset overlap with either the English or French name of the object.

We conducted an online picture naming test to assess naming agreement for the depicted objects in English and in French. French native speakers who were proficient in English and who did not participate in either the eye-tracking experiment or the cloze probability test looked at pictures of objects and gave the first name that came to mind when they looked at the picture. We did not discriminate between correct and incorrect spellings of the same word (e.g., stappler/stapler) or related words sharing the same meaning and phonological onset (e.g., mike/microphone). In selecting the pictures used for the phonological competitors, we took into account naming agreement ratings only in the relevant language. Some of the items were changed and re-tested and each of the objects in the final stimuli set was named in French and in English by at least 12 participants. Naming agreement for experimental objects was 88.4% in English (SD: 12.7, range 50% - 100%) and 94.4% in French (SD: 9.6, range 50% - 100%) (see Table 3 for a breakdown of these results).

The study comprised 32 experimental and 32 filler sentences (a total of 64 items). There were 32 arrays containing four images. Each array was shown twice, once with an experimental item and once with a filler item. Each experimental list contained two half lists, each made up of the 32 visual arrays paired with 16 experimental and 16 filler sentences. Visual arrays paired with experimental items in one half-list were paired with fillers in the other half-list, and vice versa. Experimental images were counterbalanced in the full lists, resulting in 4 different sets of items, and 8 experimental lists in total. Some of the experimental sentences included a word that was also a critical object in another sentence. In these cases, the sentence in which the word corresponded to a critical object was always played before the sentence simply mentioning the name of the object.

Table 2
Properties of experimental sentences.

Cloze rating	Sentence duration (words)	Sentence duration (seconds)	Syllables per second	Critical word onset (seconds)
English: 90.7 ± 11.9 (60–100) French: 91.7 ± 10.9 (57–100)	15.4 ± 2.85	9.87 ± 1.63	2.01 ± 0.27	6.28 ± 1.94

Table 3
Mean picture naming agreement.

Object type	Naming agreement % (English)	Naming agreement % (French)
Target	91.1 ± 10.1 (58.3–100)	96.1 ± 7.6 (75–100)
EN competitor	87.8 ± 13.6 (58.3–100)	95.3 ± 12.5 (41.7–100)*
FR competitor	76.8 ± 21.7 (41.7–100)*	92.2 ± 12.5 (50–100)
Unrelated	86.2 ± 14.0 (50–100)	95.1 ± 7.9 (50–100)
Distractor	88.3 ± 16.6 (33.3–100)	94.7 ± 10.4 (50–100)

* These values are provided for information only. They are not included in the overall mean calculation.

---

[2] For three of the French competitor words, only one phoneme overlapped with the target word, but the words were closely related in orthography (aile/aimant, lit/lion and nuage/nuque). This was also the case for one English competitor word (bee/beard).

2.3. Procedure

Before the experiment began, participants 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 then started with a picture familiarization task. Participants saw all 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 English and French name. At the same time, participants heard the English and French name for the object through their headphones. After that, they were asked to name each object using the words that had been provided. The order of language presentation was counterbalanced so that half of the participants heard and saw the English word followed by the French word, and the other half saw the reverse. Participants were instructed to look at and listen to the names given to the objects, so that they could name the objects using the same words later.

Objects were considered as correctly named if both the English and French word were correctly repeated by the participant. Incorrectly named objects (3.1% Interpreters, 4.4% Translators) were repeated, and the experimenter prompted participants who did not provide the correct name for the object on second viewing.

In the eye-tracking experiment, participants were seated in front of a computer screen, at a distance of approximately 60 cm, in the experimental laboratory (LaborInt) of the Interpreting Department of the University of Geneva. The computer was set up inside a portable ISO4043-compliant interpretation booth. The participant's dominant eye[3] was tracked on an SR Research EyeLink® 1000 remote desktop-mounted eye-tracker. Participants were asked to listen to, and simultaneously interpret, the sentences into French, and subsequently to judge whether the sentence had mentioned any of the objects shown on the display. After the instructions, the eye-tracker was calibrated using the nine-point calibration grid. Pictures were presented on a viewing monitor at a resolution of 1024 × 768 pixels. Each trial started with a drift correction, followed by a 500 ms blank screen. The visual scene was presented just over 1000 ms before onset of the predictable word in experimental trials. On filler trials, the presentation was just over 1000 ms before the onset of a word that referred to a distractor, or else at an arbitrary mid-sentence point if the sentence did not mention anything in the scene. Mean preview time for the experimental items was 1053 ms (SD 21 ms) for the Interpreter group and 1029 ms (SD 18 ms) for the Translator group. For the Filler items, mean preview time was 1047 ms (SD 18 ms) for the interpreter group and 1025 ms (SD 11 ms) for the translator group (with the difference being due to a change in display computer across groups). The picture stayed on the screen until offset of the spoken sentence. A blank screen then appeared for 4000 ms, after which audio recording of the interpretation stopped. After this, the following question appeared: "Did the sentence mention any of the pictures?". Participants gave their answer using a keyboard, pressing 1 for "Yes" and 2 for "No", and the next trial started.

The experiment started with four practice trials, after which participants were given a chance to ask questions. The experimenter also checked whether participants were interpreting the trial sentences simultaneously, and, if necessary, reminded participants that they should interpret simultaneously. The eye-tracker was then recalibrated before participants began the experiment. No feedback was given during the experiment. The experimenter monitored the eye-tracking display and recalibrated if necessary. The session lasted about 50 min.

3. Results

3.1. Comprehension question accuracy

The mean accuracy for comprehension questions in the experimental trials was 97.3% (SD 2.9%) for the Interpreter group and 97.5% (SD 3.6%) for the Translator group. Incorrectly answered trials were excluded from the eye-tracking analysis.

3.2. Eye-tracking data analyses

We analysed data from the two groups separately using a linear mixed effects model with the lme4 package (Bates, Mächler, Bolker, & Walker, 2015), using the optimx optimiser (Nash, 2014) in R Studio Version 1.4.1717 (RStudio Team, 2021). Following (Ito et al., 2018), proportions of fixations on target, English competitor, French competitor, and unrelated objects were calculated separately, using the Eye-Link's DataViewer, for 50 ms bins.[4] Blinks and fixations outside the computer screen were included in calculation of the proportion of fixations. However, bins containing only blinks or fixations outside the computer screen were then excluded from the analysis. We explored the time-course of effects by running the model for each bin from 1000 ms before target word onset to 1000 ms after onset, in order to consider prediction and bottom-up activation of phonology. The model evaluated the arcsine-transformed fixation proportions on critical objects as predicted by condition for each bin. The unrelated condition was used as a reference group, or baseline condition, using the relevel function in RStudio, so that we could test the effects of each critical condition relative to the unrelated baseline condition (target vs. unrelated, English competitor vs. unrelated, and French competitor vs. unrelated). The model included random intercepts and (de-correlated) random slopes for participants and items (Barr, 2008a). We used the bobyqa method within the optimx optimizer, which allowed all models to converge. However, as the models returned singular fit warnings, we also carried out a Bayesian linear mixed model analysis to check our findings (again using the optimx optimiser, this time with the nlminb method). Where the Bayesian model returned different results for any time bin, this is noted in the footnotes. As in Ito et al. (2018), and similarly to Borovsky, Elman, and Fernald (2012), we base our conclusions on periods over which a minimum of three consecutive bins are significantly different between a critical condition and the baseline condition. We consider that any significant divergence begins at the start time of the first of these consecutive bins, and that the difference between a critical and the baseline condition is significant when the *t*-value has absolute values exceeding 2 (Baayen, Davidson, & Bates, 2008).

We also carried out a cluster-based permutation analysis. This allowed us to compensate for some of the disadvantages of a bin-by-bin analysis by allowing us to identify time ranges over which an effect was statistically reliable, correct for multiple comparisons and avoid the element of arbitrariness in the choice of length of time bins (Barr, Jackson, & Phillips, 2014). Since the bin-by-bin analyses for the experimental items did not reveal significant differences between English and French competitor and unrelated objects, we considered fixations on Target and Unrelated objects only in the cluster-based permutation analyses. We ran the cluster-based permutation analyses for the filler items for the conditions identified as significant in our bin-by-bin linear mixed model analyses. We used the "clusterperm" package in RStudio to run by-subject and by-item ANOVAs for each time bin for the period from −1000 ms before until 1000 ms after word onset. We

---

[3] Dominance was assessed using a sighting ocular dominance test.

[4] 50 ms bins were chosen as this was the time period used in Ito et al., 2018 on which our experiment is based. Although there is no clear consensus on how long time bins should be (Pyykkönen-Klauck & Crocker, 2016), we consider that 50 ms provides a balance between showing how effects develop over time and the number of statistical models required to analyse these effects.

then detected clusters of time bins by applying a clustering threshold using the detect_clusters_by_effect function. We then ran permutation tests by subject and by item to calculate a Monte Carlo *p*-value which evaluated the probability of such clusters occurring by chance.

3.2.1. Linear-mixed model for experimental items by group

As shown in Fig. 1, interpreters looked at the target object consistently more than the unrelated object from −600 ms before, until 1000 ms after, predictable word onset. This result was supported by the results of a cluster-based permutation analysis, which detected a cluster for the period of −600 ms to 1000 ms in both the by-item and the by-subject analyses. Using a Monte Carlo estimate, we calculated that the probability of this cluster occurring by chance was minimal for both the by-item and by-subject analysis ($p < .001$). Thus, interpreters showed predictive looks to the target object, and persisted in these looks. Although they tended to look more at the English competitor object than the unrelated object from 550 ms to 900 ms after word onset, this difference was not statistically significant. Interpreters therefore predicted predictable words, but we did not find evidence that they predictively activated their phonology.

Translators also fixated the target object significantly more than the unrelated object from −400 ms onwards (except in one time bin at −150 ms) but they did not predictively activate the English phonological information of the predictable word[5] (see Fig. 2). We carried out the same cluster-based permutation analysis as for the interpreter group. Based on our by-subject analysis we identified a cluster running from −450 ms before word onset until 1000 ms after word onset with a probability of chance occurrence of $p < .001$. The by-item analysis returned slightly different results, with a first cluster whose probability of occurring by chance was minimal starting only at −50 ms before word onset until 600 ms after word onset (Monte Carlo *p*-value of $<.001$) and continuing again from 750 ms until 1000 ms after word onset (Monte Carlo p-value of $<.001$). Thus, our analyses showed that translators engaged in prediction. Translators did not fixate either the English or the French phonological competitor more than the unrelated object at any point in the analysis window. In sum, translators (like interpreters) predicted predictable words while carrying out a simultaneous interpreting task, but there was no evidence of word-form prediction.

3.2.2. Linear mixed model for filler items by group

We then analysed the filler items to examine whether interpreters fixated the target, English competitor, or the French competitor object more than the unrelated object even when the predictable word was not mentioned in the sentence. Filler sentences mentioned one of the distractors from the visual stimuli set 75% of the time. Where filler sentences did not mention any object present on the screen (25%), we chose an arbitrary mid-point in the sentence and considered the 1000 ms before and after this point. We used the same linear mixed effects model as for the experimental analysis. As shown in Fig. 3, in the interpreter group, there were no additional fixations on critical objects at any point before onset of the distractor word. There were significantly more fixations on the French phonological competitor object than the unrelated object from 850 ms to 1000 ms after word onset.[6] We ran a cluster-based permutation analysis to further investigate this result. Both the by-subject and by-item analyses detected a cluster of bins at 850 to 1000 ms (Monte Carlo *p* value: by subject $p < .05$, by item $p < .01$). The by-subject analysis also detected a cluster during which fixation

proportions on the French phonological competitor were less than those on the unrelated object. This cluster ran from 300 to 450 ms after word onset and its probability of occurring by chance was $p < .05$. These results do not follow the same pattern as the results from the analysis of the experimental items, which show no significant increase or decrease of fixations on the French phonological competitor object compared to the unrelated object.

We carried out the same analysis on the filler items for the translator group. Translators looked at the target object more than at the unrelated object from −350 ms to −100 ms, at the English competitor object more than at the unrelated object from −300 ms to 0 ms and at the French competitor object more than at the unrelated object from −250 ms to 0 ms in the predictive time window.[7] We ran a cluster-based permutation analysis to consider the fixations on each of the competitor objects compared to the unrelated object. We found a similar pattern. For the French competitor object, we found a cluster in the by-item analysis from −250 ms to −50 ms with a Monte Carlo *p*-value of $<.001$, and in the by-subject analysis, we found a cluster from −350 ms to −50 ms with a Monte Carlo p-value of $<.01$. For the English competitor object, we found a cluster in the by-item analysis from −300 to 50 ms with a Monte Carlo p-value of $<.001$, and in the by-subject analysis, we found a cluster from −300 ms to 0 ms with a Monte Carlo p-value of $<.001$. For the Target object, we found a cluster in the by-item analysis from −350 ms to −150 ms with a Monte Carlo p-value of $<.001$, and in the by-subject analysis, we found a cluster from −400 ms to −200 ms with a Monte Carlo p-value of $<.01$. As shown in Fig. 4, these results appear to be due to a decrease in fixations on the unrelated object. They do not follow the same pattern as the results from the experimental sentences.

3.2.3. Between-group analysis

To compare results from the interpreter and translator groups on the experimental items, we ran our linear mixed model on both groups together and specified an interaction by group. We did not find any interaction between group and condition before the onset of the predictable word. After onset, we found a significant interaction between fixations on the target vs. unrelated object and group for three consecutive bins from 750 ms until 900 ms after word onset.[8] There was also an interaction between the fixations on the English competitor vs. unrelated object and group for two bins from 750 ms to 850 ms after word onset (although this did not pass our three-bin threshold, and only one bin was significant when we ran the Bayesian model). Although the by-subject cluster-based permutation analysis detected an interaction between group and fixations on the target compared to the unrelated object at two time points, at −600 ms and from 800 to 850 ms, these clusters may have occurred by chance (Monte Carlo *p*-value at −600 ms: $p = .842$, Monte Carlo p-value from 800 to 850 ms: $p = .295$). We detected no clusters when we ran the corresponding by-item analysis.

Finally, we calculated the average arcsine transformed fixation proportions over the three time bins from −600 ms to −400 ms, during which time our by-group analyses had indicated interpreters had begun predicting but the translators had not, and ran our linear-mixed model over this time period. We found that the interaction between the fixations on the target condition compared to the unrelated condition, and the professional group (interpreter/translator) approached, but did not reach, significance ($t = −1.725$).

We thus find that both interpreters and translators predict during an interpreting task, and that neither group predicts the word-form of a predictable word (see linear mixed models). We did not find significant

---

[5] A Bayesian linear mixed model with random slopes and intercepts found a significant difference starting only at −350 ms and a lack of significant difference between target and unrelated object from −200 to −50 ms and at 50 ms. The results for the other bins were the same.

[6] The Bayesian analysis did not return a significant difference in fixations over three consecutive bins. Only the bins at 850 ms and 950 ms were significant.

[7] In the Bayesian analysis, for the target vs. unrelated object, the bin at −150 was not significant; for the English competitor vs. unrelated object, the bin at −300 ms was not significant, and for the French competitor vs. unrelated object, the bins at −250 and −200 ms were not significant.

[8] Using the Bayesian model, the difference was significant from 750 ms to 850 ms only.
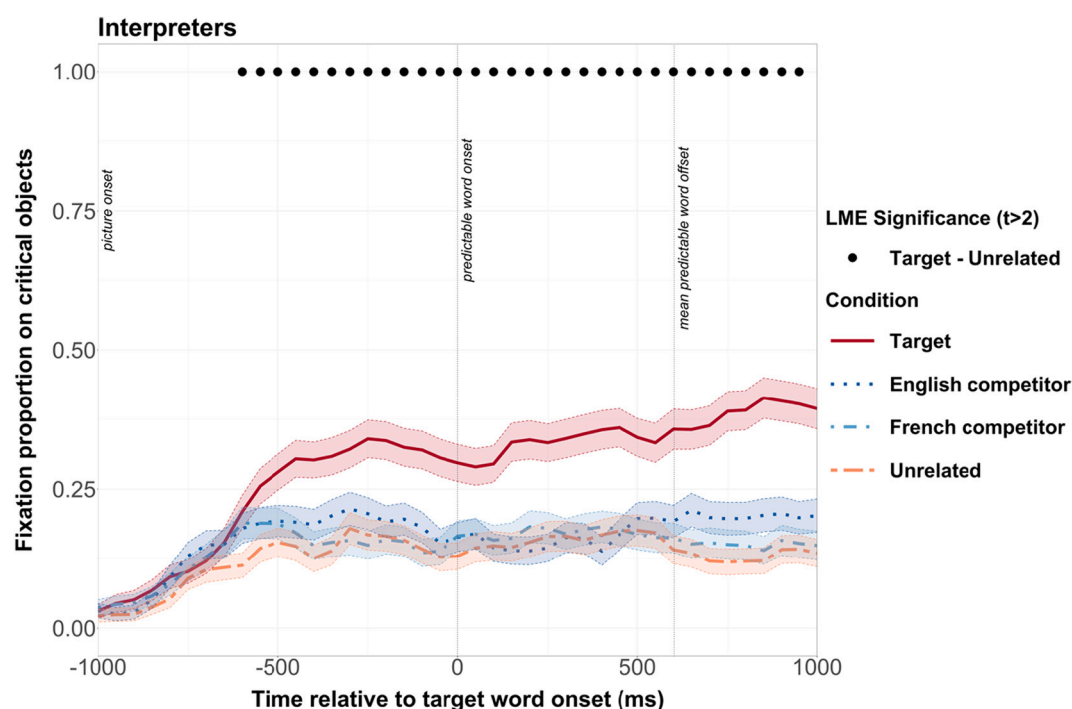
Fig. 1. Eye-tracking results for the interpreter group. Graph showing fixation proportions on target, English competitor, French competitor, and unrelated objects for the Interpreter group for the experimental sentences. Time 0 ms shows target word onset. Mean target offset was at 602 ms. Picture onset was just before −1000ms. The black dots along the top of the graph show 50 ms bins in which the difference between target and unrelated conditions was significant ($|t| > 2$).

Fig. 2. Eye-tracking results for the translator group. Graph showing fixation proportions on target, English competitor, French competitor and unrelated objects for the Translator group. Time 0 ms shows target word onset. Mean target offset was at 602ms. Picture onset was just before −1000 ms. The black dots along the top of the graph show 50ms bins in which the difference between target and unrelated conditions was significant ($|t| > 2$).
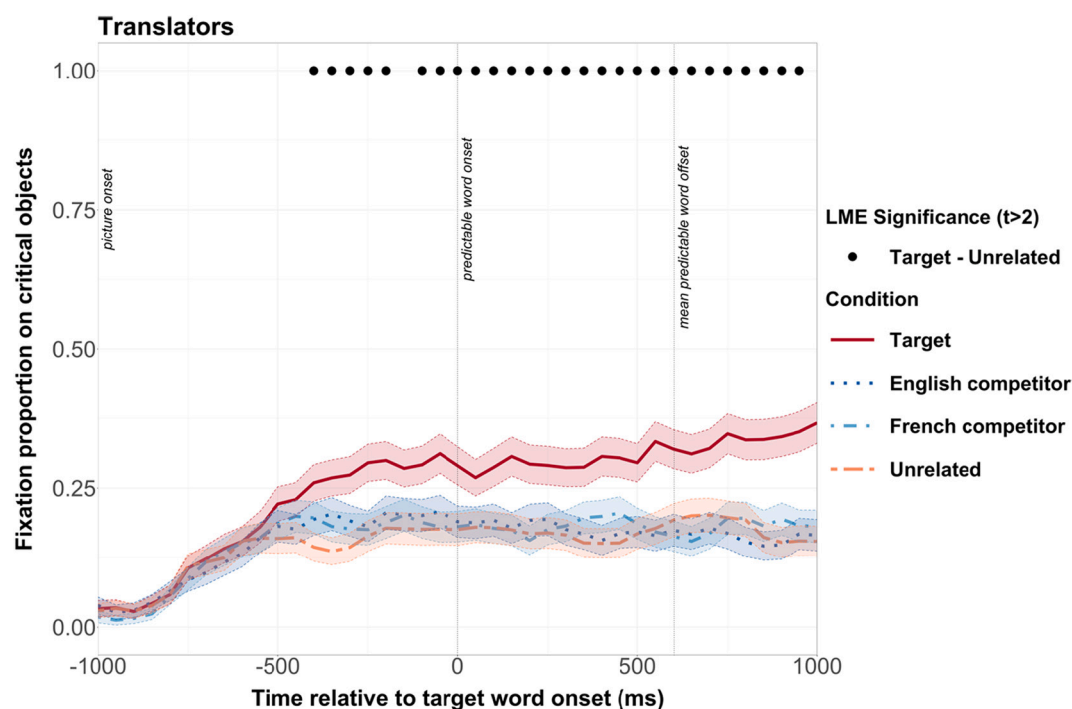
differences between predictive fixations in the two groups over consecutive time bins. We found a significant difference between fixation proportions on target and unrelated objects that depended on profession after the onset of the predictable word, from 750 ms to 900 ms.

3.2.4. Relation between eye movements and lag in simultaneous interpretation

We also analysed the eye-tracking data in light of our recordings of participants' interpretations. Based on our audio recordings, we considered when participants began interpreting, whether they
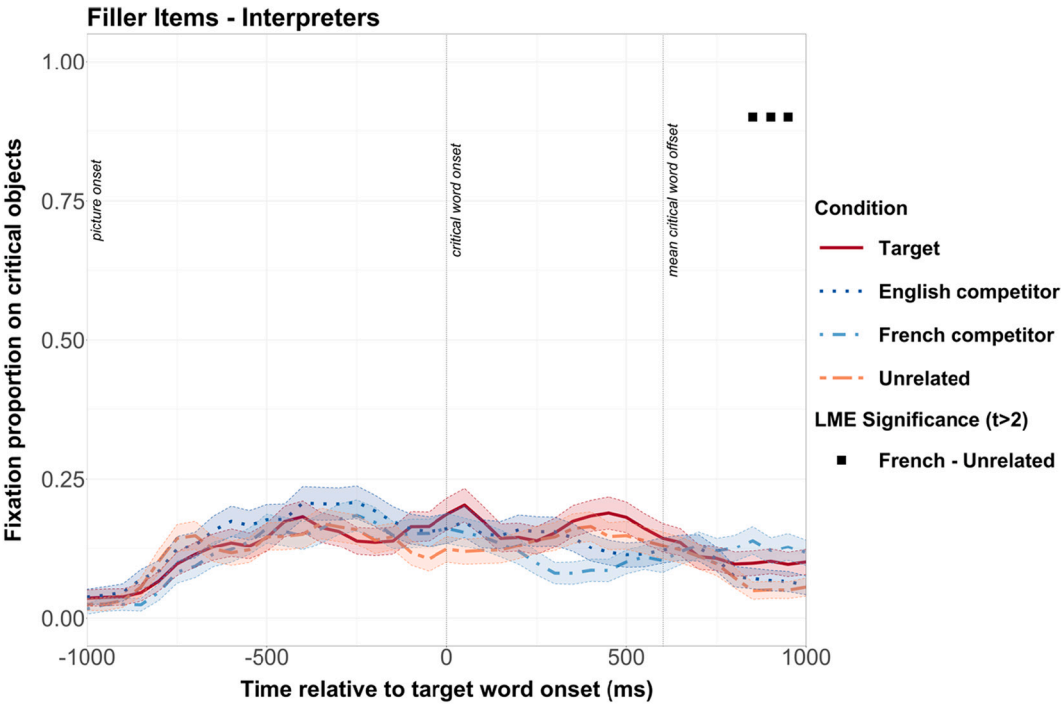
Filler Items - Interpreters

Fig. 3. Graph showing fixation proportions on target, English competitor, French competitor, and unrelated objects for the Interpreter group for the filler sentences. Time 0 shows the time at which the distractor word was mentioned, or the arbitrary mid-point in the sentence. The filled squares along the top show significant differences between the French competitor object and the unrelated object (t| > 2).
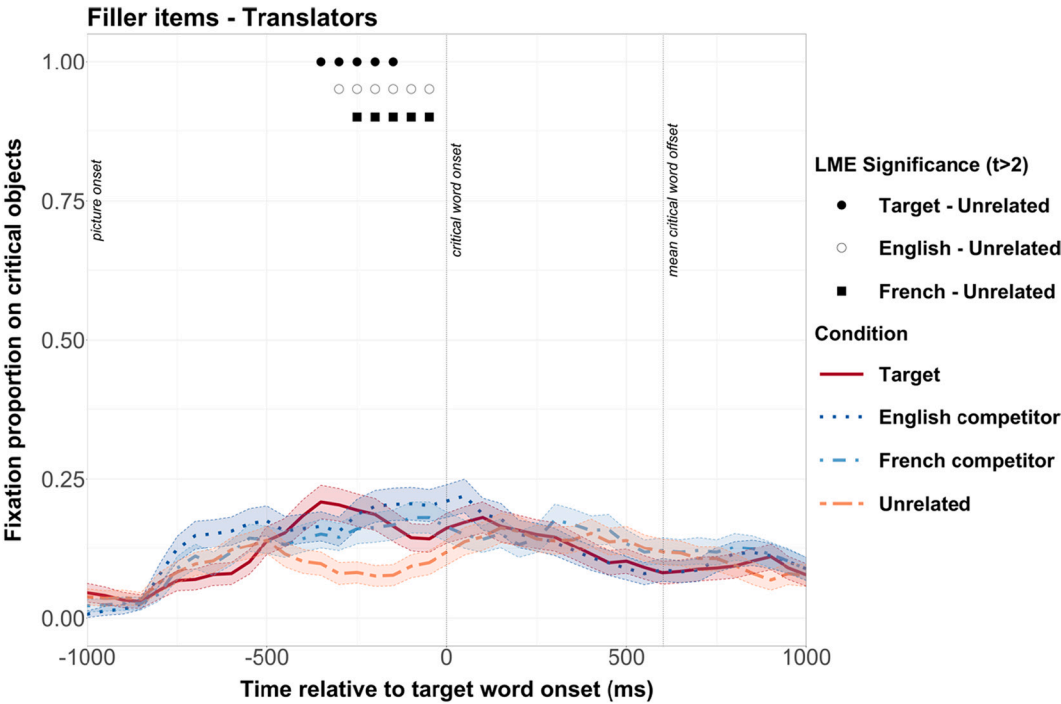
Filler items - Translators

Fig. 4. Graph showing fixation proportions on target, English competitor, French competitor and unrelated objects for the Translator group for the filler sentences. Time 0 shows the time at which the distractor word was mentioned, or the arbitrary mid-point in the sentence.

completed their interpretation within 4000 ms of the end of the spoken sentence, and whether they used the equivalent noun in French to translate the predictable word in English.

We excluded one of the translator participants from this part of the analysis because the audio recording had failed. We also excluded four trials from the analysis because four participants in the Translator group

failed to provide an interpretation for one item. We again excluded items for which the comprehension question was answered incorrectly.

For the items for which participants answered the comprehension question correctly, 85.7% (SD: 14.3) were interpreted within the allocated time of 4000 ms after the end of the sentence (85.8% of items for interpreters (SD: 16.4%) and 85.6% for translators (SD: 14.7%)).

Participants interpreted 97.1% (SD: 5.9%) of items simultaneously (defined as beginning interpretation before the end of the sentence): 97.7% (SD: 5.6%) of items for interpreters and 96.4% (SD: 6.6%) for translators. On average, participants began interpreting −2519 ms before onset of the target word (SD: 2733 ms) (interpreters at −2305 ms (SD: 2847 ms), translators at −2745 ms (SD: 2847 ms)). Mean onset of the interpretation was at 3782 ms (SD: 1791 ms) after the beginning of the sentence (3982 ms (SD: 1671 ms) for interpreters and 3570 ms (SD: 1888 ms) for translators). Although it may appear surprising that translators began interpreting slightly earlier than interpreters, this may be because translators understood the instruction to "interpret simultaneously" as meaning to begin their interpretation as soon as possible, whereas interpreters perhaps first waited for a meaningful unit before beginning their interpretation.

When participants predict more quickly, they may prepare their own utterance more quickly thanks to this prediction. And the reverse: preparing and producing their utterance more quickly may support predictive processing. We carried out an exploratory analysis to see if there was indeed a relationship between prediction and lag. We used the same linear-mixed model to investigate the time at which participants fixated the critical objects in relation to the time, relative to the target word, at which they began their interpretation (relative onset time) and whether or not they completed their interpretation. We split the data on the mean, with all items during which participants began interpreting at least −2519 ms before the target word and completed their interpretation before audio recording stopped at 4000 ms after the end of the sentence in one group, and all items during which participants began interpreting less than −2519 ms before the target word in the second group.

As shown in Fig. 5, when interpretation began at an earlier relative onset time, and participants completed their interpretation within 4000 ms of the end of the English sentence, the proportion of fixations on the target vs. the unrelated object was significantly higher ($|t| > 2$) from −550 ms until the end of the analysed time window at 1000 ms after word onset. In addition, there were significantly more fixations on the English competitor object compared to the unrelated object ($|t| > 2$) from −400 to −250 ms before word onset.[9]

As shown in Fig. 6, when interpretation began at a later relative onset time, the proportion of fixations on the target vs. the unrelated object was significantly higher

($|t| > 2$) from −50 ms before word onset until before word onset until 1000 ms after word onset.[10] There were no significant differences in the proportion of fixations on the English or French competitor object compared to the unrelated object at any point of the analysed time window.

Thus, on trials in which interpretation began earlier relative to the onset of the target word, predictive eye movements also began numerically earlier. In addition, for trials in which the relative onset time of interpretation was earlier, we found evidence suggesting phonological prediction. We further explored these findings by running a linear mixed model that included an interaction term for relative onset time (measured in seconds and centred) and condition. We included random intercepts and decorrelated random slopes for item and participant. We found an interaction of relative onset time and condition (target vs. unrelated) ($|t| > 2$) from −200 ms to 0 ms and an interaction of relative onset time and condition (English competitor vs. unrelated) ($|t| > 2$)

from −200 ms to 50 ms.[11] There was also an interaction between relative onset time and condition (French competitor vs. unrelated) ($|t| > 2$) from −150 ms to 0 ms. Thus, in the time period from −200 ms before word onset until word onset itself, the divergence between the proportion of fixations on the target versus the unrelated object depended on the relative onset time, with greater divergence when relative onset time was earlier. In the time period from −200 ms before word onset until 50 ms after word onset, the divergence between the proportion of fixations on the English competitor versus unrelated object depended on the relative onset time, with greater divergence when relative onset time was earlier. From −150 ms before word onset until word onset, the divergence between the proportion of fixations on the French competitor object versus the unrelated object depended on the relative onset time.

Raw data and scripts for these analyses are available on Open Science Framework at: https://osf.io/3sjfd/.

4. Discussion

We investigated the time-course of prediction during a simultaneous interpreting task in professional simultaneous interpreters and professional translators untrained in interpreting. L1 French interpreters and translators listened to English sentences containing a highly predictable word and simultaneously interpreted these sentences into French. The results showed that both professional simultaneous interpreters and professional translators predict upcoming information while simultaneously interpreting. Simultaneous interpreters and translators both made predictions. There were no significant differences in predictive patterns between interpreters and translators. In an exploratory analysis, we found that on trials where participants had a shorter lag, they predicted to a greater extent than on trials where they lagged further behind the original.

4.1. Evidence for prediction during simultaneous interpreting

Our findings show that both professional simultaneous interpreters and professional translators, untrained in simultaneous interpreting, make predictive eye movements during a simultaneous interpreting task. This evidence supports theories of simultaneous interpreting which assume that prediction takes place during simultaneous interpreting (e. g., Chernov, 2004; Seleskovitch, 1984). It also demonstrates that training is not necessary for prediction to take place during simultaneous interpreting.

We found robust prediction in L2 listeners who are completing an additional task. This finding contrasts with some recent findings and extends others. For example, Ito, Corley and Pickering (2017) found that L2 listeners did not make predictive eye movements before target word onset when they had an additional cognitive load, and Dijkgraaf et al. (2017) found semantic prediction in L2 listeners when there was no additional task. Our study shows that highly proficient bilinguals make predictive eye movements even in the combination of challenging conditions present during a simultaneous interpreting task (noise, concurrent production and cognitive load). In other words, challenging listening conditions in L2 do not necessarily prevent prediction.

Now let us compare our experiment with Experiments 1 and 2 of Ito et al. (2018). The professional simultaneous interpreters (L2 speakers of English) in our study made predictive eye movements earlier than the L2 speakers in Experiment 2 of Ito et al.'s, 2018 study (a significant difference over at least three time bins first emerged only at −350 ms, similar to our translator group, who began predictive fixations at −400 ms). In fact, the predictive fixations on target objects made by

---

[9] The Bayesian linear mixed model showed a significant difference between the target and unrelated conditions emerged at 500 ms (with two bins not returning significance at 50 and 100 ms) and the significant difference between the English competitor object and the unrelated object first emerged at −350 ms.

[10] The Bayesian linear mixed model showed a significant difference between the target and unrelated conditions emerged at word onset and lasted until 1000 ms after word onset (with one bin not significant at 250 ms).

[11] The Bayesian model returned significant interactions at the same time bins, with the exception that the interaction between target vs. unrelated conditions and onset time in seconds began one bin later, at −150 ms.
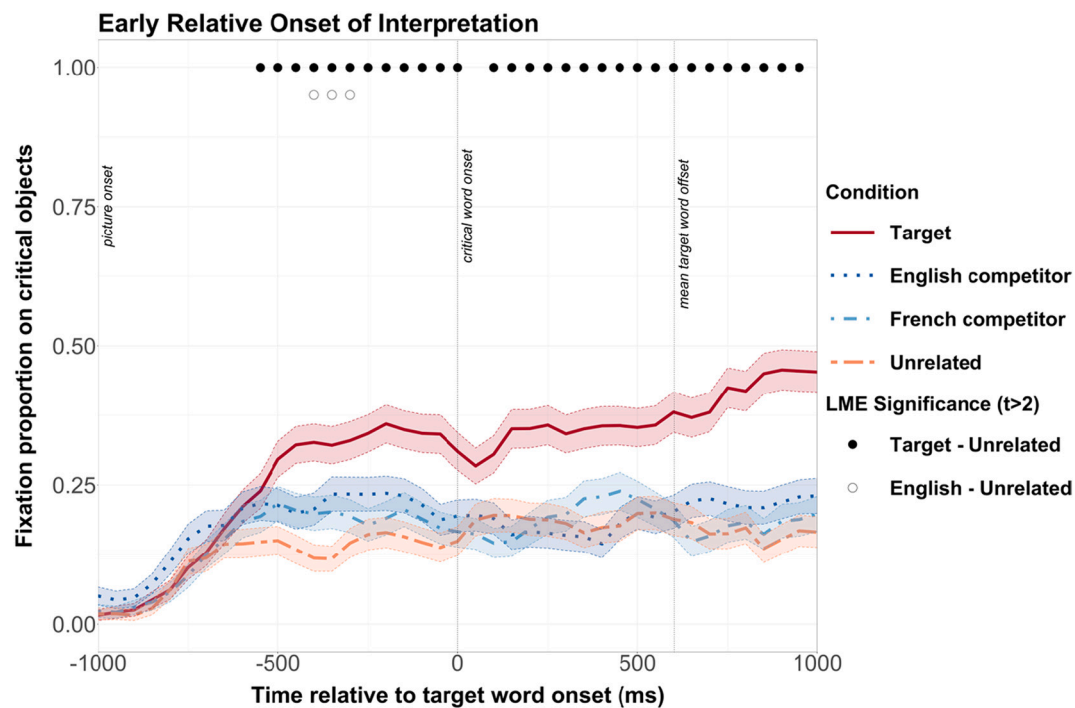
Fig. 5. Graph showing trials in which participants began interpreting at least −2519 ms before the onset of the target word. Time 0 ms shows target word onset. Mean target offset was at 602 ms. Picture onset was just before −1000ms. The black dots along the top of the graph show 50ms bins in which the difference between target and unrelated conditions was significant (|t| > 2). The open dots show 50 ms bins in which the difference between the English competitor and unrelated objects was significant (|t| > 2).
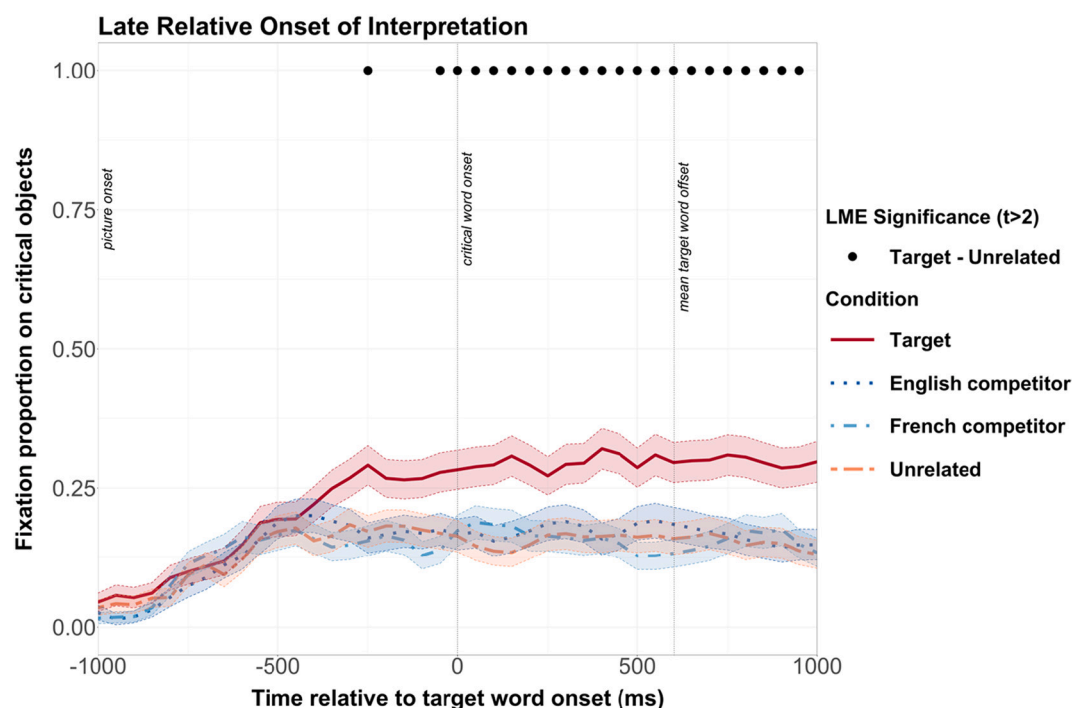

Fig. 6. Graph showing trials in which participants began interpreting at −2519 ms or less before the onset of the target word, or after onset of the target word. Time 0 ms shows target word onset. Mean target offset was at 602 ms. Picture onset was just before −1000 ms. The black dots along the top of the graph show 50 ms bins in which when the difference between target and unrelated conditions was significant (|t| > 2).

simultaneous interpreters in our study had a remarkably similar timing to the predictive fixations made by L1 speakers in Experiment 1 of Ito et al. (2018), with predictive fixations in both of these experiments beginning at −600 ms before onset of a predictable word. However,

unlike the L1 speakers in Ito et al. (2018), neither interpreters nor translators fixated on English competitor objects significantly more than on unrelated objects during the predictive time window. The results suggest that during a simultaneous interpreting task, people may predict

predictable words, but may not pre-activate their phonological form, even when they are trained in interpreting.

4.2. No clear evidence that simultaneous interpreters predict earlier or more than translators

Although both simultaneous interpreters and translators predicted during our simultaneous interpreting task, we found more robust predictive patterns in the interpreter group. However, we found no significant between-group differences in patterns of prediction. This suggests that training and experience do not affect the time course of prediction in challenging conditions, such as present in simultaneous interpreting. Thus our results do not lend support to theories of interpreting that posit that prediction in simultaneous interpreting is a skill or strategy.

4.3. Prediction of word-form during simultaneous interpreting

We did not find any evidence of prediction of word form in English by either the translator or the interpreter group during the simultaneous interpreting task, nor was there evidence of word-form activation after word onset (see Section 4.4). This contrasts with Ito et al. (2018) who found phonological prediction in the absence of an interpreting task (see also Kukona, 2020). Therefore, it may be that, as in Martin et al. (2018), simultaneous interpreting engaged participants' production mechanisms in producing a translation of what they had already heard, and so they could not use those mechanisms to the same extent to make word-form predictions. The load associated with the simultaneous interpreting task may also have limited eye movements that were linked to linguistic activation (for example, see Prasad & Mishra, 2020). Alternatively, it may be that even highly proficient L2 speakers do not predictively activate word form during comprehension.

Further, despite engaging in a simultaneous interpreting task into French, professional translators or interpreters showed no evidence of activating the French form of the target word. This finding is similar to Ito et al. (2018), who did not find that Japanese-English bilinguals participants activated the phonology of the Japanese translation of the English target word.

4.4. Lack of cross-linguistic activation even after word onset

We did not find any evidence of word-form activation after word onset. Neither the English nor the French competitor object attracted significantly more looks than the unrelated object even once the phonologically related word had been produced. One explanation for this is that, despite the target-absent design, participants narrowed their expectations before the onset of the predictable word and therefore did not look at the competitor at or after word onset (Barr, 2008b; Dahan & Tanenhaus, 2004; Weber & Crocker, 2012). The linguistic input may thus have led to a lack of fixations on what were less relevant objects (see Huettig & Altmann, 2005, 2007 for a discussion).

This lack of relevance of the competitor object in the predictable context may have been compounded by the fact that participants were switching between two languages, meaning that at any given point in time, each language may have been activated to a greater or lesser extent. For instance, once participants had comprehended the target word in English, they may have focused more on producing its translation in French. However, they may still have been uttering another, unrelated word in French, in the period immediately following onset of the target word. In addition, although comprehension and image presentation were time-locked and were uniform across participants, participants' production was neither time-locked to image presentation, nor did all participants produce the same sentence constituents at the same time. This may explain the apparent lack of French phonological activation. One or more of these factors may have led to the lack of cross-linguistic activation after the onset of the target word.

4.5. Exploratory findings suggest that lag and prediction are linked

In an exploratory analysis, we found that participants began to make predictive eye movements earlier when production was more closely synchronized with comprehension (i.e., when interpretation began earlier relative to the target word) than when production and comprehension were less synchronized. In addition, when production and comprehension were more synchronized, participants appeared to activate the phonological form of the upcoming word predictively.

Following a prediction-by-production account, it may be that greater synchronicity between production and comprehension leads to greater prediction. In other words, if participants' own utterances are more synchronized with the utterances that they comprehend, predictive processing might be supported; whereas if participants' own utterances are less synchronized with what they comprehend, this situation might be more akin to engaging the production mechanism in an unrelated task. Another possibility is that greater prediction leads to greater synchronicity of production and comprehension – that is, participants produce their own utterance more quickly as their comprehension has been more rapid due to greater prediction.

5. Conclusions

We reported an experiment that investigated whether professional interpreters and professional translators make predictive eye movements during a simultaneous interpreting task. We found that both professional interpreters and translators untrained in interpreting predicted upcoming language, but neither group pre-activated phonological information associated with a target word. There were no significant differences in predictive patterns between groups. Thus, high-proficiency bilinguals routinely engage in prediction during simultaneous interpreting, even though it is a complex and difficult task, and this prediction appears to be independent of training and experience.

Funding

Declaration of Competing Interest

None.

Appendix 1. Experimental sentences and critical visual objects

List of experimental sentences used in the experiment, with cloze values provided for the English and French word. Object names are provided in French and English, and listed as the target object, the English competitor object, the French competitor object and the Unrelated object. One of the objects named was shown in the display alongside three distractor objects.

Item	Sentence (English cloze %, French cloze %)	Bilingual object name* (English cloze %, French cloze %)
		mouth/bouche (83.3, 91.7)
		mouse/souris (100, 100)
1	In order to have a closer look, the dentist asked the man to open his <u>mouth</u> a little wider. (100,100)	cork/bouchon (75, 83.3)
		bone/os (100, 100)
		stairs/escalier (100, 100)
2	In an emergency, we cannot use a lift; instead, we need to use the <u>stairs</u> for our safety. (100, 100)	stapler/agrafeuse (83.3, 100)
		wipers/essuie-glace (41.7, 100)
		calculator/calculatrice (91.7, 100)
		rainbow/arc-en-ciel (100, 100)
3	If the sun comes out during a heavy shower, you can sometimes see a <u>rainbow</u> in the sky. (100, 100)	radio/radio (100, 100)
		tree/arbre (91.7, 91.7)
		goat/chèvre (91.7, 91.7)
		cloud/nuage (100, 100)
4	The tourists expected rain when the sun went behind the <u>cloud</u>, but the weather got better later. (95, 100)	clown/clown (100, 91.7)
		neck/nuque (91.7, 100)
		train/train (83.3, 100)
5	The man didn't know the time because he forgot to wear the <u>watch</u> that he usually wears. (100, 100)	watch/montre (91.7, 100)
		washing machine/machine-à-laver
		hot-air balloon/montgolfière
		tray/plateau (58.3, 100)
6	Bob proposed and gave her a <u>ring</u> that had cost him half his monthly wage. (85, 93)	ring/bague (91.7, 91.7)
		ribbon/ruban (66.7, 100)
		chopstick/baguette (41.7, 100)
		letter/letter (100, 83.3)
7	The man was gathering honey, when he was stung by a <u>bee</u> and gave a cry. (95, 100)	bee/abeille (91.7, 100)
		beard/barbe (100, 100)
		apricot/abricot (66.7, 100)
		helmet/casque (83.3, 100)
8	People can easily go to the island on foot since the government built a <u>bridge</u> last year. (90, 100)	bridge/pont (100, 100)
		brick/brique (75, 100)
		pump/pompe (83.3, 100)
		meat/viande (83.3, 100)
9	The traveller went to the desert because he wanted to ride a <u>camel</u> and go exploring. (90, 86)	camel/chameau (91.7, 100)
		camera/camera (91.7, 100)
		hat/chapeau (100, 100)
		barrel/tonneau (50, 91.7)
10	The woman found the room was too hot and humid, so to get some fresh air, she opened the <u>window</u> completely. (100, 100)	window/fenêtre (83.3, 100)
		windmill/eolienne (75, 100)
		fennel/fenouil (58.3, 83.3)
		globe/globe (66.7, 91.7)
		wing/aile (100, 100)
11	The bird cannot fly because it injured its <u>wing</u> when it had a fight with another bird. (100, 93)	witch/sorcière (91.7, 100)
		magnet/aimant (66.7, 100)
		flag/drapeau (83.3, 100)
		car/voiture (83.3, 100)
12	Amber went to the dealership to purchase a new <u>car</u> the very next day. (80, 79)	castle/château (100, 100)
		veil/voile (75, 91.7)
		bear/ours (83.3, 100)
		shield/bouclier (83.3, 100)
13	To protect against an enemy's bullet or arrows, soldiers used to carry a <u>shield</u> all the time. (70, 57)	sheep/mouton (91.7, 100)
		candle/bougie (100, 91.7)
		onion/oignon (100, 75)
		pencil/crayon (91.7, 100)
14	Before he began to draw, he sharpened his <u>pencil</u> and got out some paper. (67, 92)	penguin/pingouin 91.7, 58.3)
		cream/crème (83.3, 91.7)
		cherry/cerise (100, 100)
		desk/bureau (100, 83.3)
15	In order to study, Karen sat down at her <u>desk</u> and opened her book. (90, 100)	dentist/dentist (58.3, 41.7)
		bubble/bulle (83.3, 83.3)
		chick/poussin (75, 100)
16	In the night sky it is easier to see all the <u>stars</u> and the moon. (100, 100)	star/étoile (91.7, 100)
		stamp/timbre (75, 100)
		label/etiquette (41.7, 100)
		key/clé (100, 100)
		fish/poisson (100, 100)
17	To make sushi, the chef went to the market to buy some <u>fish</u> early in the morning. (60, 93)	finger/doigt (100, 100)
		pear/poire (100, 100)
		moon/lune (100, 83.3)
		coat/manteaux (58.3, 100)
18	During winter, it's best to put on your heaviest <u>coat</u> and a hat. (70, 86)	comb/peigne (75, 100)
		mango/mangue (100, 83.3)
		pineapple/ananas (75, 100)
		shelf/étagère (91.7, 75)
19	The maid dusted the books on the <u>shelf</u> every week. (90, 79)	shell/coquillage (58.3, 100)
		pond/étang (75, 75)

(*continued*)

Item	Sentence (English cloze %, French cloze %)	Bilingual object name* (English cloze %, French cloze %)
		pig/cochon (91.7, 91.7)
		book/livre (100. 91.7)
20	The student went to the library to read a <u>book</u> but in the end he ended up chatting with his friends. (90, 100)	bull/taureau (100, 100)
		unicorn/licorne (91.7, 100)
		dice/dé (91.7, 100)
21	Joan fed her baby some warm <u>milk</u> and then put him to bed. (70, 71)	milk/lait (100, 100)
		mirror/miroir (100, 91.7)
		leash/laisse (58.3, 50)
22	She went to the beauty parlour to perm her <u>hair</u> in preparation for the party. (90, 86)	salt/sel (100, 83.3)
		hair/cheveux (100, 83.3)
		helicopter/hélicoptère (100, 100)
23	One day, the caterpillar will turn into a beautiful <u>butterfly</u> and fly away. (100, 100)	chimney/cheminée (100, 100)
		bamboo/bambou (91.7, 100)
		butterfly/papillon (100, 100)
		button/bouton (91.7, 100)
24	Catherine carried her computer in a shoulder <u>bag</u> until she found it was giving her back problems. (100, 79)	umbrella/parapluie (91.7, 100)
		giraffe/giraffe (91.7, 100)
		bag/sac (66.7, 75)
		bat/chauve-souris (100, 100)
25	After every meal it's good to brush your <u>teeth</u> or else chew gum. (100, 100)	salad/salade (50, 66.7)
		kiwi/kiwi (100, 100)
		teeth/dents (91.7, 100)
		teapot/théière (100, 100)
26	Dad carved the turkey with a <u>knife</u> for Christmas dinner. (90, 93)	dancer/danseuse (41.7, 66.7)
		horse/cheval (100, 100)
		knife/couteau (83.3, 100)
		knight/chevalier (66.7, 91.7)
27	He loosened the tie around his <u>neck</u> and immediately felt better. 100, 100)	pillow/coussin (58.3, 91.7)
		rabbit/lapin (83.3, 91.7)
		neck/cou (91.7, 100)
		nest/nid (91.7, 100)
28	A flat tyre forced Katy to pull up at the side of the <u>road</u> and call for assistance. (100, 93)	crown/couronne (91.7, 100)
		handcuff/menottes (58.3, 100)
		road/route (100, 100)
		rope/corde (100, 100)
29	The referee blew his <u>whistle</u> to signal the end of the match. (100, 92)	lipstick/rouge-à-lèvres (100, 100)
		glasses/lunettes (83.3, 100)
		whistle/sifflet (91.7, 100)
		whisk/fouet (66.7, 91.7)
30	He scraped the cold food from his <u>plate</u> into the bin. (80, 71)	saw/scie (100, 100)
		pen/stylo (83.3, 100)
		plate/assiette (91.7, 83.3)
		plant/plante (91.7, 91.7)
		vacuum cleaner/aspirateur (58.3, 100)
31	John was very tired so he decided to go straight to <u>bed</u> and sleep. (100, 100)	grape/raisin (91.7, 83.3)
		bed/lit (100, 100)
		belt/ceinture (91.7, 100)
		lion/lion (100, 100)
		glove/gant (100, 100)
		ladder/echelle (91.7, 100)
32	To reach the roof, the workman climbed up the <u>ladder</u> that was against the wall. (100, 92)	laptop/ordinateur (83.3, 100)
		scarf/écharpe (100, 100)
		bench/banc (66.7, 100)

* Object name in English and in French based on the object depicted.

Appendix 2. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104987.

References

Adank, P. (2012). Design choices in imaging speech comprehension: An activation likelihood estimation (ALE) meta-analysis. *Neuroimage, 63*(3), 1601–1613. https://doi.org/10.1016/j.neuroimage.2012.07.027

AIIC. (2019). The AIIC A-B-C. Retrieved from: https://aiic.org/site/world/about/profession/abc.

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. https://doi.org/10.1017/S1366728919000671

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Barr, D. J. (2008a). 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. (2008b). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition, 109*(1), 18–40. https://doi.org/10.1016/j.cognition.2008.07.005

Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology: General, 143*(1), 404–413. https://doi.org/10.1037/a0031813

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1).

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods, 42*(3), 665–670. https://doi.org/10.3758/brm.42.3.665

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

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 1029–1040. https://doi.org/10.1037/a0015901

Chernov, G. V. (2004). *Inference and anticipation in simultaneous interpreting: A probability-prediction model*. Philadelphia: John Benjamins Publishing.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30* (2), 498–513. https://doi.org/10.1037/0278-7393.30.2.498

De Groot, A. M. B. (2011). *Language and cognition in bilinguals and multilinguals: An introduction*. New York, NY: Psychology Press.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117–1121. 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

Drake, E., & Corley, M. (2015). Articulatory imaging implicates prediction during spoken language comprehension. *Memory & Cognition, 43*(8), 1136–1147. https://doi.org/10.3758/s13421-015-0530-6

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

Fleur, D. S., Flecken, M., Rommers, J., & Nieuwland, M. S. (2020). Definitely saw it coming? The dual nature of the pre-nominal prediction effect. *Cognition, 204*. Article 104335.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1461–1469. https://doi.org/10.1037/a0036756

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

Gerver, D., Longley, P., Long, J., & Lambert, S. (1984). Selecting trainee conference interpreters: A preliminary study. *Journal of Occupational Psychology, 57*(1), 17–31.

Grisoni, L., McCormick Miller, T., & Pulvermüller, F. (2017). Neural correlates of semantic prediction and resolution in sentence processing. *The Journal of Neuroscience.*. https://doi.org/10.1523/jneurosci.2800-16.2017

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

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

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96*(1), B23–B32. https://doi.org/10.1016/j.cognition.2004.10.003

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15*(8), 985–1018. https://doi.org/10.1080/13506280601130875

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience, 31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

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. https://doi.org/10.1017/S1366728917000050

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86*, 157–171. https://doi.org/10.1016/j.jml.2015.10.007

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*. https://doi.org/10.1016/j.neuropsychologia.2019.107291

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017a). 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., Martin, A. E., & Nieuwland, M. S. (2017b). Why the A/AN prediction effect may be hard to replicate: A rebuttal to Delong, Urbach, and Kutas (2017). *Language, Cognition and Neuroscience, 32*(8), 974–983. https://doi.org/10.1080/23273798.2017.1323112

Ito, A., & Pickering, M. (2021). Automaticity and prediction in non-native language comprehension. In T. Grüter, & E. Kaan (Eds.), *Prediction in second-language processing and learning* (pp. 26–46). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.12.

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

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

Kukona, A. (2020). Lexical constraints on the prediction of upcoming words: Insights from the visual world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(11), 2153–2162. https://doi.org/10.1037/xlm0000935

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

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language, 61*(3), 326–338. https://doi.org/10.1016/j.jml.2009.06.004

Mani, N., & Huettig, 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

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. https://doi.org/10.1044/1092-4388(2007/067)

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is production: The missing link between language production and comprehension. *Scientific Reports, 8*(1), 1079. https://doi.org/10.1038/s41598-018-19499-4

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

Mattys, S. L., Carroll, L. M., Li, C. K. W., & Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication, 52*(11), 887–899. https://doi.org/10.1016/j.specom.2010.01.005

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research, 40* (3), 686–693. https://doi.org/10.1044/jslhr.4003.686

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. https://doi.org/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* (pp. 353–368). New York: Plenum Press.

Moser-Mercer, B. (2000). Simultaneous interpreting: Cognitive potential and limitations. *Interpreting, 5*(2), 83–94. https://doi.org/10.1075/intp.5.2.03mos

Moser-Mercer, B., Frauenfelder, U. H., Casado, B., & Künzli, A. (2000). Searching to define expertise in interpreting. In B. E. Dimitrova, & K. Hyltenstam (Eds.)*, Vol. 40. Language processing and simultaneous interpreting* (pp. 107–133). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software, 60*(2), 14. https://doi.org/10.18637/jss.v060.i02

Nieuwland, M. S., Arkhipova, Y., & Rodríguez-Gómez, P. (2020). Anticipating words during spoken discourse comprehension: A large-scale, pre-registered replication study using brain potentials. *Cortex, 133*, 1–36. https://doi.org/10.1016/j.cortex.2020.09.007

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife, 7*, Article e33468. https://doi.org/10.7554/eLife.33468

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357–395. https://doi.org/10.1037/0033-295X.115.2.357

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. A. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes, 45*(6), 464–496. https://doi.org/10.1080/01638530802356463

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

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

Prasad, S., & Mishra, R. K. (2020). Concurrent verbal working memory load constrains cross-linguistic translation activation: A visual world eye-tracking study on Hindi–English bilinguals. *Bilingualism: Language and Cognition*, 1–30. https://doi.org/10.1017/S1366728920000024

Pyykkönen-Klauck, P., & Crocker, M. W. (2016). Attention and eye movement metrics in visual world eye tracking. In P. Knoeferle, P. Pyykkönen-Klauck, & M. W. Crocker (Eds.), *Visually situated language comprehension* (pp. 67–82). Amsterdam/Philadelphia: John Benjamins Publishing Company. doi:10.1075.aicr.93.03pyk.

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia, 51*(3), 437–447. https://doi.org/10.1016/j.neuropsychologia.2012.12.002

RStudio Team. (2021). *RStudio: Integrated development environment for R*. PBC, Boston, MA: RStudio. URL http://www.rstudio.com/.

Seeber, K. G. (2001). Intonation and anticipation in simultaneous interpreting. *Cahiers de Linguistique Française, 23*, 61–97. https://clf.unige.ch/files/5514/4102/7732/04 -Seeber_nclf23.pdf.

Seeber, K. G. (2017). Multimodal processing in simultaneous interpreting. In J. W. Schwieter, & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 461–475). https://doi.org/10.1002/9781119241485.ch25

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

Setton, R. (2001). Deconstructing SI: A contribution to the debate on component processes. *The Interpreter's Newsletter, 11*, 1–26. URI: http://hdl.handle.net /10077/2443.

Setton, R. (2005). Pointing to contexts: A relevance-theoretic approach to assessing quality and difficulty in interpreting. In H. V. Dam, J. Engberg, & H. Gerzymisch-Arbogast (Eds.), *Knowledge systems and translation* (pp. 275–312). Berlin/New York: Walter De Gruyter.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*(6), 557–580. https://doi.org/10.1023/a:1026464108329

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

Van Besien, F. (1999). Anticipation in simultaneous interpretation. *Meta, 44*(2), 250–259. https://doi.org/10.7202/004532ar

Weber, A., & Crocker, M. W. (2012). On the nature of semantic constraints on lexical access. *Journal of Psycholinguistic Research, 41*(3), 195–214. https://doi.org/10.1007/s10936-011-9184-0

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

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *The Journal of Neuroscience, 32*(40), 14010–14021. https://doi.org/10.1523/jneurosci.1528-12.2012

Wilss, W. (1978). Syntatic anticipation in German-English simultaneous interpreting. In D. Gerver, & H. W. Sinaiko (Eds.), *Language interpretation and communication* (pp. 343–352). New York: Plenum Press.