Applied Psycholinguistics (2025), **46**, e18, 1–29
doi:10.1017/S0142716425100052

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

The influence of word co-occurrence frequency on predictive processing in first and second languages: A webcam-based eye-tracking study

Haerim Hwang[1] and Sun Hee Park[2]

[1]Department of English, The Chinese University of Hong Kong, Hong Kong, China and [2]Department of Korean Studies, Ewha Womans University, Seoul, Republic of Korea
Corresponding author: Haerim Hwang; Email: haerimhwang@cuhk.edu.hk

(Received 24 July 2024; revised 19 February 2025; accepted 25 April 2025)

Abstract

The questions of whether first language (L1) speakers and second language (L2) learners can both predict what follows based on given linguistic cues and what factors may influence this predictive processing are still underexplored. Prior research has focused on the success or failure of predictions in real-time processing, paying relatively less attention to the speed of prediction. This study addresses these gaps by investigating the role of word co-occurrence frequency and proficiency in L1 and L2 predictive processing, using the Korean classifier system. In a webcam-based visual-world eye-tracking experiment, both L1-Korean speakers and L2-Korean learners showed sound predictive processing, with the frequency of co-occurrence between classifiers and nouns playing a crucial role. Higher co-occurrence frequency expedited predictive processing for L1-Korean speakers and boosted the ability to make online predictions for L2-Korean learners. The study also revealed a proficiency effect, where more advanced L2-Korean learners made predictions regardless of co-occurrence frequency, unlike their less advanced counterparts. Our findings suggest that predictive mechanisms in L1 and L2 operate in a qualitatively similar way. In addition, the use of webcam eye-tracking is expected to create a more inclusive and equitable research environment for (applied) psycholinguistics.

Keywords: classifier; prediction speed; predictive processing; proficiency; webcam-based eye-tracking; word co-occurrence frequency

Introduction

Human language processing is characterized by the gradual integration of cues from multiple sources. During real-time processing, language users make use of linguistic and non-linguistic information to construct meaning and predict what will come next (e.g., Altmann & Mirković, 2009; Levy, 2008; Pickering & Garrod, 2013).

The present study examines predictive processing, which refers to a process that utilizes contextual information to activate or generate expectations about forthcoming linguistic features, such as lexical items, word categories, and syntactic structures, before the actual input is encountered by the language processing system (Kuperberg & Jaeger, 2016). Specifically, we focus on the impact of word co-occurrence frequency in shaping the prediction process in both first language (L1) and second language (L2), while also examining the interacting effect of proficiency in the latter case.

There is a general consensus that L1 speakers are able to use prior linguistic knowledge to predict upcoming elements (e.g., Altmann & Kamide, 1999; Huettig et al., 2010). By contrast, the question of whether L2 learners possess this ability remains controversial, with some studies indicating that they do (e.g., Fang & Wu, 2022; Kim & Grüter, 2021; Mitsugi, 2020) and other studies indicating that they do not (e.g., Grüter et al., 2020; Grüter & Rohde, 2021; Mitsugi & MacWhinney, 2016; Perdomo & Kaan, 2021; Rusk et al., 2020). Toward reconciling these divergent findings, the most up-to-date version of the Reduced Ability to Generate Expectations (RAGE) hypothesis (Grüter & Rohde, 2021; for the original version, see Grüter et al., 2017) postulates that some L2 learners, like some L1 speakers (Federmeier, 2007; Kuperberg & Jaeger, 2016), may not utilize prediction-based processing as it may not be the most optimal strategy for them. For example, certain linguistic information may not be consistently reliable for L2 learners to use in their predictive processing (Kaan & Grüter, 2021). Also, there can be situations where the advantages of making predictions may be offset by the potential expenses, such as the necessity of correcting an erroneous expectation (Kuperberg & Jaeger, 2016).

While the RAGE hypothesis (Grüter & Rohde, 2021) has significantly enhanced our comprehension of L2 predictive processing, the factors determining the degree of optimality in prediction remain largely unknown. In fact, Kaan (2014) has argued that L2 learners and L1 speakers do not differ "in the nature of the predictive mechanisms or in the way these mechanisms are employed, but in what drives these mechanisms" (p. 260). Factors affecting the predictive mechanisms may include individual factors, such as working memory capacity and target language proficiency, as well as linguistic factors, such as lexical frequency and structural complexity. Despite their importance, previous studies on L2 predictive processing have primarily investigated the role of individual factors (e.g., Fang & Wu, 2022; Kim & Grüter, 2021). Furthermore, prior research has concentrated on the success or failure of predictions made during the predictive region in online processing, thus giving comparatively less consideration to the speed of prediction (see also Stone et al., 2021). Speed is a critical element in real-time predictive processing, as it can expose more nuanced timing patterns that are shaped by individual and linguistic factors among successful predictions. Therefore, examining this timing information is essential for shedding light on how prediction mechanisms operate in conjunction with diverse factors.

The aforementioned gaps in previous research serve as the primary motivation for the current study. This study focuses on the role of co-occurrence frequency in predictive processing, building on the idea that frequent usage of lexical items enhances their accessibility and automation, leading to more efficient real-time processing (Gollan et al., 2008). This frequency effect extends beyond single words to phrases and combinations of words (Hopp, 2018). In predictive processing, for example, the phrase *a bottle of* in English can create higher expectations for words like *water* or *wine* compared to *oil*. As such, when a predictive cue frequently co-occurs with a target in real language usage, this high frequency of co-occurrence can strengthen an associative link between the two. This link, possibly stored in memory, enables the cue to more easily and rapidly activate the target with minimal cognitive effort, resulting in more robust and/or rapid predictions. Therefore, investigating co-occurrence frequency in predictive processing should illuminate the underlying mechanisms that drive prediction and the linguistic factors that influence them.

Specifically, this study makes use of the classifier system in Korean. This phenomenon offers an ideal test case for predictive processing and for the role of frequency of word co-occurrence in predictive processing for two reasons. First, classifiers have an association with the nouns they categorize, and thus the presence of a particular classifier can function as a predictive cue for the upcoming noun. For example, the classifier *calwu* is commonly used with nouns for thin objects (e.g., gun, pencil) or those with large volume (e.g., rice); so, when *calwu* appears, it should lead to the prediction that the word for some object of these types will follow. Second, such a classifier-noun association allows for systematic manipulation of co-occurrence frequency through the use of different nouns. For example, in both L1-Korean and L2-Korean corpora, *calwu* co-occurs more frequently with *yenphil* "pencil" than with *chong* "gun".

In addition, the present study also examines how proficiency interacts with co-occurrence frequency in L2 predictive processing. While it is reasonable to expect that L2 learners' predictive processing will converge with that of L1 speakers as their proficiency improves, the existing literature presents a complex and inconsistent picture of the relationship between proficiency and L2 predictive processing (Kaan & Grüter, 2021). A potential explanation for the inconsistent proficiency effects lies in the challenges of both defining what constitutes high proficiency and ensuring that data collection encompasses a sufficient range of proficiency levels. We address this issue by recruiting L2-Korean learners with varying proficiency levels from an institution with a large L2-Korean learner population.

Furthermore, we took an innovative approach by employing a webcam-based visual-world eye-tracking method to explore real-time processing patterns in human participants. We believe that this attempt will expand the possibilities of employing the webcam-based method as a viable alternative for researchers without access to in-lab eye-tracking equipment. Also, the current study goes beyond the examination of prediction success/failure by evaluating the degree of prediction optimality, which can be measured by the speed and/or robustness of prediction. To this end, we employ generalized additive mixed modeling (GAMM) for analysis of our eye-tracking data, which is able to pinpoint the precise moment when statistically significant predictive patterns are observed. As a preview of our findings, the processing patterns of both L1-Korean speakers and L2-Korean learners were influenced by the co-occurrence frequency between classifiers and nouns. Higher co-occurrence frequency led to faster processing for L1 speakers and more robust predictions for L2 learners. Additionally, we observed a proficiency effect, where more advanced learners, similar to L1 speakers, demonstrated faster predictions

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

with higher co-occurrence frequency than with lower co-occurrence frequency, whereas less advanced learners only showed predictive patterns with higher co-occurrence frequency. The methodological approach and findings of this study are expected to reveal more about the underlying mechanisms of L1 and L2 predictive processing.

Classifiers and predictive processing

Classifiers, or numeral quantifiers, indicate the category of the noun they are associated with based on its inherent properties, including shape, material, natural type, and function (Gao & Malt, 2009). In classifier languages, such as Korean and Mandarin, the use of classifiers is mandatory when the noun phrase includes a numeral. Modern Korean has 93 classifiers, according to Hwang et al. (2010); and modern Mandarin has approximately one hundred classifiers, according to Ma (2015).

The complexity of classifiers goes beyond their sheer number. Classifiers are not always associated with homogeneous semantic sets within the same class, and nouns vary in the extent to which they can be considered semantically prototypical members of a class; such nuances make classifiers difficult for language learners to acquire and use accurately. To illustrate, the classifier *tay* in Korean is typically used for machines that have automaticity, such as bikes (Academy of Korean Studies, 2023), but it can also co-occur with pianos. In Mandarin, the classifier *tiáo* is used for nouns that are slender, long, and often flexible, such as ropes (Gao & Malt, 2009), but it can also go along with dogs. This fact indicates that the classifier system can be considered an abstract grammatical attribute of nouns.

There have been a few studies investigating how L1 and L2 users make use of the classifier information to predict forthcoming nouns during real-time sentence processing, all employing the visual-world paradigm. The initial exploration of this topic was conducted by Huettig and his colleagues (2010) with a group of L1-Mandarin speakers. They showed that upon hearing a classifier (e.g., *bǎ*, which is typically used for graspable objects with a handle-like part), the L1-Mandarin speakers sent more looks to the target object associated with that classifier in the visual scene (e.g., *chair*) than to the distractor objects (e.g., *candle*). This finding highlights the predictive role of classifiers in L1-Mandarin processing.

Tsang and Chambers (2011) went a step further by conducting a study on L1-Cantonese speakers to examine how the semantic prototypicality of nouns within their classifier classes influences predictive processing behaviors. Given the possibility that the successful predictive processing patterns observed by Huettig et al. (2010) could be explained by two information sources—a grammatical form-class cue (as in the case of the article-noun association in English) or a semantic cue (as in the case of the association between gender-marked determiners and nouns referring to humans [e.g., *girl*, *boy*] in Spanish)—Tsang and Chambers sought to tease apart these two sources. They achieved this by grouping nouns into four categories based on two criteria: grammatical form-class match/mismatch (G+/G−) and semantic match/mismatch (S+/S−). (However, note that semantic prototypicality is a matter of degree rather than a clear-cut categorization.) In their visual-world eye-tracking experiments, L1-Cantonese speakers showed a greater number of fixations on G+S+ competitor objects, in contrast to unrelated distractors (Experiment 3); and they further exhibited that while G−S+ competitor objects did not elicit significantly more fixations than unrelated distractors, both G+S+ competitor objects and G+S− competitor objects showed competition effects, albeit slightly stronger for the former (Experiment 2). All in all, the findings of Tsang and Chambers (2011) indicated that semantic information did influence predictions, but only when the competitor objects aligned with the grammatical form-class of the target classifier. This suggests that in L1 predictive processing, grammatical form-class serves as the primary cue, and semantic prototypicality serves as the secondary cue.

The presence of both grammatical and semantic cues in the classifier-noun association raises an interesting question regarding the extent to which L2 learners might rely on these cues in their predictive processing. Building on Tsang and Chambers's (2011) study, Grüter et al. (2020) explored this question with 38 L1-Mandarin speakers and 41 L2-Mandarin learners with diverse L1 backgrounds: English ($n = 34$), German ($n = 2$), Spanish ($n = 3$), Dutch ($n = 1$), and Hebrew ($n = 1$). Their L1 group, like the L1 group in Tsang and Chambers's study, used the classifier primarily as a grammatical form-class cue, providing evidence for predictive processing. By contrast, for the L2 group, the G+S− target faced competition from G−S+ nouns that shared semantic features with the classifier they heard but were grammatically incompatible during the predictive region. The finding of semantic competition within the predictive region was taken by the researchers as an indication that L2 learners can engage in predictive processing, albeit of a different type than that of L1 speakers (p. 232), thus supporting the RAGE hypothesis (Grüter et al., 2017; Grüter & Rohde, 2021). Furthermore, they argued that a higher reliance on semantic cues (vs. grammatical cues) may be an L2 learners' strategy to maximize processing efficiency, which is achievable through "different means in different contexts" (p. 232).

While Grüter et al.'s (2020) study offers valuable insights into the role of different cues in predictive processing for L1 speakers and L2 learners, their interpretation of the results leads to a question about the specific contexts and factors that contribute to the efficiency or optimality of predictive processing. Among these factors are individual variables, like working memory capacity and proficiency in the target language, as well as linguistic factors, such as lexical frequency, the latter of which has been underexplored in the context of L2 predictive processing (c.f., Hwang & Kim, 2025). To address this issue, we focus on the influence of word co-occurrence frequency in L1 and L2 prediction, as well as the influence of proficiency in L2 prediction.

Mitsugi (2020) broadened the scope of previous research to Japanese, while also including proficiency as a modulating factor. In her visual-world eye-tracking task, L1-Japanese speakers and L1-English L2-Japanese learners listened to sentences where a numeral classifier (e.g., *-hon* for long and string-like objects) was followed by a noun. The results showed increased fixations toward the target in both L1-Japanese speakers and L2-Japanese learners during the predictive region when the classifier provided information about the noun's identity, compared to when it did not. Interestingly, the study observed no influence of proficiency during the predictive region. It is possible that the use of a self-rating questionnaire to measure proficiency led to this result, as it might not be a reliable or objective measure to accurately capture the nuances of L2 proficiency.

Taken together, prior research on predictive processing involving classifier-noun association has demonstrated that L1 speakers can utilize the grammatical information of classifiers to anticipate upcoming nouns, irrespective of the target language (i.e., Cantonese, Mandarin, and Japanese). When it comes to L2 learners, they also exhibit some evidence of predictions utilizing linguistic cues; however, the two L2 studies discussed in this section showed inconsistent results. Whereas Mitsugi's (2020) research identified robust predictive processing abilities in L2 learners, Grüter et al.'s (2020) study suggested that the specific processing patterns of L2 learners diverge from those of L1 speakers. Although our study does not look at the competition between grammatical and semantic cues, we expand on the existing research by examining L2 predictive processing based on Korean classifiers. Because higher lexical frequency can facilitate online processing, as we will discuss in the following section, this study investigates whether increased word co-occurrence frequency can reduce the cognitive burden on L2 learners, as well as L1 speakers, and facilitate their predictive processing. In doing so, we center our analysis on the speed and robustness of prediction.

Co-occurrence frequency in processing

Language contains a wealth of diverse statistical information, such as frequency and the probability of co-occurrence (Erickson & Thiessen, 2015). Such information can affect how language users encode, process, and retrieve linguistic elements (Ellis, 2002; Wells et al., 2009). Co-occurrence frequency, in particular, has garnered attention in research on both L1 processing (e.g., Brunellière et al., 2017; Jones & Golonka, 2012; Saffran et al., 1996) and L2 processing (e.g., Ellis et al., 2014; Yi et al., 2017), although its role in the context of prediction has not yet been studied.

In the domain of lexical processing, Brunellière et al. (2017) demonstrated that the frequency of co-occurrence between prime-target pairs enhances semantic priming effects in French. In their lexical decision task (Experiment 1), L1-French speakers were presented with three semantic priming contexts: (a) a semantically related context where primes and targets co-occur highly frequently (e.g., *garage–voiture* "garage–car"); (b) a semantically related context where primes and targets co-occur less frequently (e.g., *trafic–voiture* "traffic–car"); and (c) an unrelated context where primes and targets do not have any close semantic relationship (e.g., *ours–voiture* "bear–car"). The results indicated a semantic priming effect in both of the related priming contexts, with a heightened impact observed for word pairs displaying a higher frequency of co-occurrence.

Turning to the domain of phrasal processing, Yi et al. (2017) observed noteworthy impacts of co-occurrence frequency on L1 and L2 speakers' online reading behaviors. L1-Chinese speakers and L2-Chinese learners (with various L1 backgrounds) completed an eye-tracking while-reading task where they read

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

adverbial sequences consisting of two monosyllabic words (e.g., 也-还 *yě-hái* "also-again") that were embedded in sentences. The results showed a facilitative effect of the co-occurrence frequency of two words (i.e., phrasal frequency in this case) on reading times for both L1 and L2 groups. That is, the greater frequency of Chinese sequences led to faster reading, as indicated by shorter first fixation duration and total reading time.

In sum, psycholinguistic research has demonstrated that individuals exhibit sensitivity to co-occurrence frequency information in language processing, irrespective of whether they are L1 or L2 speakers. However, previous studies have primarily investigated its effects within either lexical or phrasal domains (c.f., Kim et al., 2020), and its impact on predictive processing has not been explored as yet. To address this gap, we systematically manipulate this factor in our audio stimuli based on the co-occurrence frequency between classifiers and nouns.

The present study

This study is guided by the following research questions (RQs):

RQ 1. Can L1-Korean speakers and L2-Korean learners predict the upcoming noun based on the linguistic cue of a classifier?

RQ 2. Do L1-Korean speakers and L2-Korean learners show evidence of classifier-noun co-occurrence frequency effects in their predictive processing? If so, when and how?

RQ 3. Does proficiency play a role in L2-Korean learners' predictive processing? If so, when and how?

RQ 1 examines whether L1 speakers and L2 learners can generate expectations by making use of classifiers in Korean. We predict that L1 speakers will show predictive processing patterns, in line with extensive previous research on L1 prediction. Regarding L2 learners, their processing patterns are less predictable. They may show robust predictive processing patterns (e.g., Mitsugi, 2020) or exhibit delayed prediction (e.g., Fuchs, 2022). Alternatively, they may not be able to use classifiers to make predictions.

Regarding RQ 2, if L1 speakers and L2 learners can make predictions based on classifiers, our prediction is that both groups will exhibit a lexical co-occurrence frequency effect in their predictive processing, potentially influencing the speed or robustness of their predictions.

As for RQ 3, we do not have a clear prediction. One possibility is that proficiency influences L2-Korean learners' predictive processing such that more advanced learners engage in prediction, regardless of co-occurrence frequency, or display faster and more robust prediction compared to less advanced learners when co-occurrence frequency is higher and/or lower. Alternatively, there may be no proficiency effects whatsoever, with all learners showing an indication of engaging in prediction, or with neither advanced nor less advanced learners showing any indication of engaging in prediction.

Method

Participants

This study involved the participation of 76 individuals, comprising 22 L1-Korean speakers and 54 L2-Korean learners. All participants were recruited from the second author's institution in the Republic of Korea, which has a large population of L2-Korean learners enrolled in its Korean language courses. At the time of testing, all L2 participants were registered in classes at an upper intermediate or advanced level. These participants were chosen for their accessibility and the potential for variation in their proficiency levels. Data from nine L2 learners had to be excluded: two of them encountered technical difficulties; five did not complete a proficiency task; and two had Cantonese as their L1, which differed from the majority of our L2 participants. Additionally, we excluded data from one L1-Korean speaker and 15 L2-Korean learners who failed to meet the criterion of 80% accuracy (e.g., Baker & Love, 2022) in the fill-in-the-blank knowledge task that served as a screening task in this study (see Section "Materials").

The data cleaning process resulted in a sample of 21 L1-Korean speakers and 30 L2-Korean learners, whose L1 was Chinese. The fact that Chinese employs classifiers, as in Korean, is expected to facilitate our L2 participants' prediction process based on classifiers. However, there are also specific differences in classifier usage between these two languages (see Section "Materials"), and this can reveal an interesting picture of how these differences shape the patterns of L2 prediction based on classifiers (see Section "Discussion").

The L2 group's average age at which they began learning Korean was 19.90 years; they had an average of 3.82 years of immersion experience in Korea; and their C-test proficiency scores averaged 32.40 out of 50 (see Section "Procedure"). Table 1 summarizes the background information of both participant groups. Prior to the commencement of the study, administrative and ethical clearances were obtained at the first author's institution, and informed consent was obtained from each participant.

Materials

Visual-world eye-tracking task

The main focus of this study is a webcam-based visual-world eye-tracking task, which aims at investigating the real-time predictive processing patterns in L1-Korean speakers and L2-Korean learners. The task was designed using a Latin square design, which involved manipulating the co-occurrence frequency factor (see below) and yielded two separate lists. In each list, there were 12 critical sentences (along with 48 fillers), wherein each of the following six classifiers was used twice: *kwon* (for e.g., books), *pel* (for e.g., clothes, bowls, and spoons), *calwu* (for e.g., slender objects), *tay* (for e.g., machines, instruments, and transportations), *cang* (for e.g., thin and flat objects), and *songi* (for e.g., objects that form a bunch). All the classifiers used in this study were grammatical classifiers, but they varied in terms of their semantic informativity or transparency in relation to the nouns they were paired with. Half of them—*kwon*, *cang*, and *songi*—can be considered more semantically informative, as they typically occur with semantically similar objects.

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

Table 1. Background information of participants

	Age at testing	Age of L2 onset	Length of residence in Korea (in years)	Proficiency score (out of 50)
L1-Korean speakers ($n = 21$)	26.43 ($SD = 3.72$; Range = 22–35)	NA	NA	43.86 ($SD = 2.87$; Range = 39–49)
L2-Korean learners ($n = 30$)	27.43 ($SD = 4.55$; Range = 22–41)	19.90 ($SD = 4.46$; Range = 6–34)	3.82 ($SD = 3.67$; Range = 0–18)	32.40 ($SD = 8.97$; Range = 12–45)

The other half—*pel*, *calwu*, and *tay*—are considered less semantically informative, as they can co-occur with a wider variety of nouns.

The phonetic structures of our Korean classifiers exhibit both similarities and differences with their Chinese counterparts; and some Korean classifiers have etymological connections to Chinese, as outlined in Appendix A in Supplemental Online Material. Regarding phonetic similarity, we identified pairs of the target Korean classifiers and their corresponding Chinese classifiers as phonetically similar if they shared (a) the same onset and nucleus (cf. cohort priming) or (b) the same nucleus and coda (cf. rhyme priming). As for etymological connections, we categorized pairs of the target Korean classifiers and their corresponding Chinese classifiers as connected if the former originated from the latter. Of the 24 items examined in the two experimental lists, three showed phonetic similarity between the Korean classifier *cang* and the Chinese one *zhāng*. On the other hand, eight items revealed etymological connections between the Korean classifiers *kwon* and *tay* and the Chinese ones *běn* and *tái*, respectively. Although the (dis)similarities between L1 and L2 are not the primary focus of this study, we revisit this issue in Section "Discussion."

Importantly, the task had two conditions that were based on the frequency of occurrence of the specific classifiers with the specific nouns in the critical sentences: a higher co-occurrence frequency condition ("Higher condition"; e.g., (1)), and a lower co-occurrence frequency condition ("Lower condition"; e.g., (2)). For a comprehensive list of the critical sentences, see Appendix B in Supplemental Online Material.

(1) Higher co-occurrence frequency condition
Onul Yengswu-ka han calwu, Cinhuy-ka han calwu-uy yenphil-ul
today Yengswu-NOM one CL Cinhuy-NOM one CL-GEN pencil-ACC
chac-ass-e-yo.
find-PST-DEC-POL
"Yengswu (found) a (pencil) and Cinhuy found a pencil today."

(2) Lower co-occurrence frequency condition
Onul Yengswu-ka han calwu, Cinhuy-ka han calwu-uy chong-ul
today Yengswu-NOM one CL Cinhuy-NOM one CL-GEN gun-ACC
chac-ass-e-yo.
find-PST-DEC-POL
"Yengswu (found) a (gun) and Cinhuy found a gun today."

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

The difference between the two conditions in the classifier-noun co-occurrence frequency was checked using both L1-Korean and L2-Korean corpora. We counted instances of the target noun co-occurring with its specific target classifier in the corpora. We did not consider instances of the target noun co-occurring with the general classifier *kay*. For the L1-Korean corpus, we utilized the NIKL (National Institute of Korean Language) corpus, including both written and spoken data collected in 2018 (National Institute of Korean Language, 2021), which contains 18,313,158 eojeols (i.e., basic spacing units in Korean, each of which carries lexical and/or morphosyntactic information) and 42,256,866 morphemes in total; and then, we conducted a natural language processing-based analysis to automatically extract the classifier-noun pairs tested in this study, using the "KoNLPy" package (Park, 2014) in Python (Python Software Foundation, 2022). For the L2-Korean corpus, we used the Korean Learners' Corpus Search Engine (National Institute of Korean Language, 2023), including both written and spoken data, which consists of 4,013,233 eojeols and 6,230,590 morphemes; and then, we manually searched for the target classifier-noun pairs on the website. In both corpora, we found a significant difference between the two conditions in terms of the log-transformed frequency of co-occurrence between classifiers and nouns, with the higher condition showing higher frequency than the lower condition (L1-Korean corpus: $t(11) = 3.551$; $p = .005$; Cohen's $d = 1.509$; L2-Korean corpus: $t(11) = 5.955$; $p < .001$; Cohen's $d = 2.497$).

All critical sentences comprised a total of nine eojeols. The words chosen for the experiment were carefully evaluated by the second author, an expert in Korean language education with extensive experience teaching the same learner population involved in this study; she confirmed that the words were within the participants' ability to understand. To expand the predictive region for L2 learners, who often show delayed processing (e.g., Felser & Cunnings, 2012; Hopp, 2017), the sentences repeated the classifier twice before the target noun, as presented in (1)–(2), making use of the fact that Korean permits the coordination of two conjuncts without a verb and the head of its object noun phrase (i.e., target noun) in the first conjunct. A male L1-Korean speaker recorded all experimental sentences using Audacity, ensuring natural prosody. The predictive region (underscored in (1)–(2)), spanning from the offset of the first classifier to the onset of the target noun, had a duration of 2,537 ms on average. This long duration of the prediction region enabled us to compare prediction speed across the experimental conditions.

Accompanying the sentences were visual scenes, as illustrated in Figure 1. These scenes comprised clipart images depicting two objects. The two objects on each visual scene were designated as areas of interest (AOIs) for further analysis. Specifically, in the case of (1), the target was *yenphil* "pencil", while its competitor was *othopai* "motorcycle". The placement of these objects was counterbalanced for both the Higher and Lower conditions, resulting in the creation of four distinct lists.

We also ensured that the word frequency of both targets and competitors was controlled independently of their co-occurrence frequency. The targets and competitors showed no differences in their log-transformed frequency in both the NIKL corpus and the Korean Learners' Corpus Search Engine (all ps $> .1$).

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

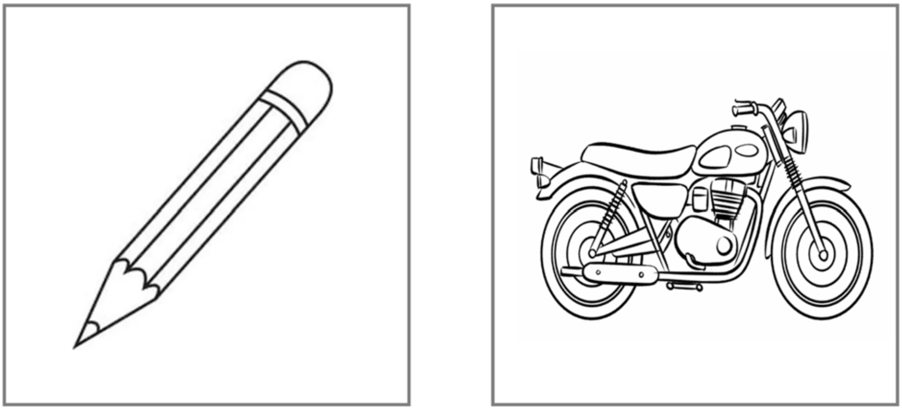

Figure 1. Example visual scene in the visual-world eye-tracking task.

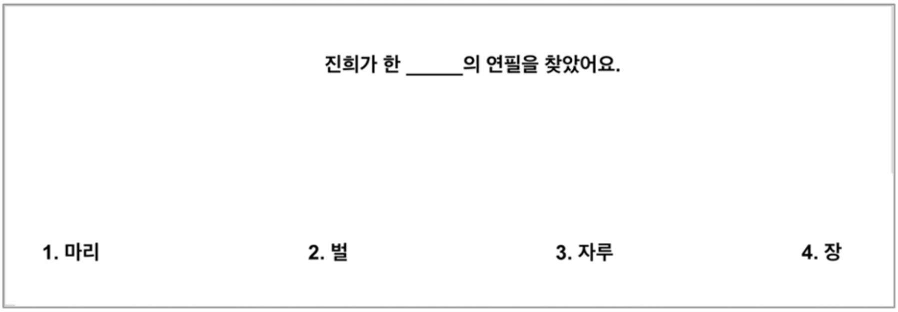

Figure 2. Example item in the fill-in-the-blank task.

Fill-in-the-blank knowledge task

To assess participants' knowledge of the associations of grammatical classifiers and nouns used in this study, a fill-in-the-blank task was administered. The task employed essentially identical sentences as the eye-tracking task, but was simplified by excluding the first conjunct. In this task, participants were asked to choose the correct classifier from a set of four multiple-choice options to complete the blank. The task was multiple-choice to keep participants from using a general classifier, *kay*, which is generally applicable to a variety of objects. Figure 2 shows a sample item of this task, where the sentence in (1) without the first conjunct is given in the Korean writing system, Hangul, and the answer is #3, *calwu*. Appendix C in Supplemental Online Material provides a full list of the critical items included in the fill-in-the-blank task.

Procedure

All participants completed a webcam-based eye-tracking task, a language background questionnaire, a fill-in-the-blank task, and a C-test, in that order. All participants performed these tasks at a place of their convenience, with our

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

research assistants providing instructions and support via Zoom. The first task was a webcam-based eye-tracking task designed and implemented on GazeRecorder (https://gazerecorder.com/). To create this task, we first made a video presenting all our stimuli for each of the four lists and uploaded it to the web. We then put the links to the four videos on GazeRecorder. This type of online eye-tracking task, while not widely researched, has shown promise in cognitive science and psycholinguistics (e.g., Semmelmann & Weigelt, 2018; Vos et al., 2022; Yang & Krajbich, 2021), particularly for visual-world paradigms (see Section "Discussion").

Participants were instructed to sit in front of their computer screens, maintaining a comfortable viewing distance. Their eye movements while observing the visual scenes were read using their personal webcams and then stored on the GazeRecorder server. The task began with an initial phase in which a 16-point array is displayed on the screen for the purpose of eye calibration. Participants then received the following instruction in Korean: "In this task, you will be asked to view pictures presented on the computer screen while simultaneously listening to Korean sentences. You will begin with a practice session involving three sentences." Following the practice session, the participants were given the next instruction in Korean: "The practice session is now complete. You will now proceed to the main session." Each trial began with the display of a fixation cross at the center of the screen, lasting for 1,000 ms. Next, the recorded sentences and visual scenes were presented in a predetermined order that was pseudo-randomized. Throughout the task, eye movements were monitored and recorded online, and recalibration was conducted whenever necessary. This task took around 10–15 minutes to complete.

As the second task, participants were requested to complete a questionnaire (Appendix D in Supplemental Online Material) to supply information regarding their language background, which took about five minutes. They then completed a fill-in-the-blank knowledge task (10–15 minutes). The order of sentences in this task was pseudo-randomized for each participant. The questionnaire and fill-in-the-blank task were created and presented using PCIbex Farm (Zehr & Schwarz, 2018).

The last task was a C-test, which took 10–15 minutes. This task was conducted to assess the Korean proficiency of our participants, particularly the L2 learners. In this task, participants were asked to fill in the missing letters for 50 items spread across two passages (see Lee-Ellis, 2009, Appendix A, Passages 1 and 4). Note that we selected only two passages from the original five passages in Lee-Ellis's C-test, drawing from our pilot study, which demonstrated the ability of this abbreviated C-test to generate a range of proficiency scores. The total time required to complete all the tasks was approximately 50 minutes.

Analysis

Fill-in-the-blank knowledge task

The fill-in-the-blank knowledge task served as a screening test to ensure that our participants had robust offline knowledge of classifier and noun associations. We tested whether there was any significant difference between the L1-Korean speakers and the L2-Korean learners in their offline knowledge of classifier-noun associations by constructing a logistic mixed-effects regression model on the accuracy in the

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

fill-in-the-blank knowledge task with the maximal structure. This model included *Group* (L1 vs. L2) and *Co-occurrence frequency* (Higher vs. Lower) as fixed effects and *Item* as a random effect. The analysis was conducted in R (R Core Team, 2024) using the "lme4" package (Bates et al., 2015).

Visual-world eye-tracking task

The data collected from the eye-tracking task consisted of the screen pixel coordinates representing the approximate gaze location. To account for variations in participant screen sizes and resolutions, GazeRecorder standardizes the pixel locations on the screen. Specifically, these locations represent the positions of the estimated gaze as a proportion relative to the participants' screen width and height, with the pixel coordinate at the center of the screen defined as (0.5, 0.5), irrespective of screen resolution.

As an initial step prior to the main analysis, we undertook a data cleaning process. Initially, we excluded data collected during instructional periods (9.87% of the entire data). Furthermore, we eliminated data points with incorrect time stamps, where the subsequent data points had earlier time points (0.02% of the entire data), likely due to technical issues, as well as duplicated data points and those during intervals between stimuli (17.63% of the entire data).

For our main analysis, we isolated the critical items, which accounted for 17.17% of the entire dataset. Next, we aggregated the data into 50-ms bins and then converted fixation proportions for each 50-ms time bin into empirical logits ("elogits" hereafter). This transformation was carried out to address a concern related to the bounded nature of the data (e.g., Barr, 2008; Mahr et al., 2015). To facilitate easier data interpretation, on the other hand, we reset the onset of the target noun (i.e., the offset of the predictive region) to 0 ms.

To analyze the eye-tracking data within the predictive region, we used two statistical analysis methods. Our primary method was applying a GAMM to elogit-transformed fixation proportions, which allowed us to address the timing issue related to prediction speed. This recently developed approach is able to consider non-linear effects over time and pinpoint specific time periods where the differences between conditions reach statistical robustness (Baayen et al., 2017; Rusk et al., 2020; Wieling, 2018; Wood, 2006).

For RQ 1 and RQ 2, we created the hyper-factor for analysis, following the conventions established by previous work (e.g., Rusk et al., 2020; Wieling, 2018), by crossing all the factors we had—i.e., *Group* (L1 vs. L2), *Co-occurrence frequency* (Higher vs. Lower), and *Fixation area* (Target vs. Competitor)—resulting in eight levels: (a) L1-Higher-Target (reference level), (b) L1-Higher-Competitor, (c) L1-Lower-Target, (d) L1-Lower-Competitor, (e) L2-Higher-Target, (f) L2-Higher-Competitor, (g) L2-Lower-Target, and (h) L2-Lower-Competitor. For RQ 3, our analysis was limited to the L2 data. We first divided the L2 learners into two proficiency groups based on the median score (i.e., 35): More advanced ($n = 15$) and Less advanced ($n = 15$). Next, we created the hyper-factor for analysis by crossing the three factors—i.e., *Proficiency group* (More advanced vs. Less advanced), *Co-occurrence frequency* (Higher vs. Lower), and *Fixation area* (Target vs.

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

Competitor)—resulting in eight levels: (a) More advanced-Higher-Target (reference level), (b) More advanced-Higher-Competitor, (c) More advanced-Lower-Target, (d) More advanced-Lower-Competitor, (e) Less advanced-Higher-Target, (f) Less advanced-Higher-Competitor, (g) Less advanced-Lower-Target, and (h) Less advanced-Lower-Competitor.

In each GAMM we constructed, the hyper-factor was included as both a parametric predictor and a smooth term to, respectively, account for potential differences in intercepts among its levels and capture any time-related changes. The model also had random effects for *Participant* and *Item*, similar to conventional mixed-effects regression models. The final model discussed in the "Results" section is the adjusted model that accounted for autocorrelation within the time-series data by incorporating a *rho* value, which is reported in the results table. Given our emphasis on temporal processing patterns, we focus on smooth terms in this paper, which capture the dynamic relationships between each condition and the elogit-transformed fixation pattern over time. (We do not discuss the parametric coefficients, which represent the elogit-transformed fixation proportions for each condition compared to the reference condition, aggregated across the predictive region.) However, the statistical significance of a smooth term only indicates that an effect is present, and it does not describe the specific temporal dynamics of processing. Hence, we generated difference plots based on the model results to identify the exact time periods where the differences between the conditions in comparison achieved statistical significance.

However, GAMMs do not directly reveal interactions between factors being tested, such as *Group/Proficiency group* and *Co-occurrence frequency* in our case. To address this issue, we supplemented our analysis with a growth curve analysis of target advantage scores, which were obtained by calculating the difference between the elogit-transformed fixation proportions of the target and those of the competitor. The present study effectively combined the benefits of both methods, confirming the interaction effect of *Group/Proficiency group* and *Co-occurrence frequency* using a growth curve analysis and pinpointing the temporal source of this interaction effect using a GAMM.

Our growth curve model for RQ 1 and RQ 2 included *Group* (L1-Korean vs. L2-Korean) and *Co-occurrence frequency* (Higher vs. Lower) as fixed effects, whereas our growth curve model for RQ 3 included *Proficiency group* (More advanced vs. Less advanced) and *Co-occurrence frequency* (Higher vs. Lower) as fixed effects. Both models accounted for *Participant* and *Item* variability through random effects. We incorporated second-order orthogonal polynomials to capture time-dependent changes in our data (e.g., Beyersmann et al., 2021). The linear time term (i.e., the first-order polynomial component of the growth curve) indicates fixation changes in slope, showing "whether fixations rose or fell faster in a given window" (Henry et al., 2022, p. 409), and the quadratic time term (i.e., the second-order polynomial component of the growth curve) represents acceleration of those fixation changes or curvature differences.

All the analyses were carried out in R (R Core Team, 2024), employing the "itsadug" package (van Rij et al., 2022) for the GAMM and the "lme4" package (Bates et al., 2015) for the growth curve analysis.

Table 2. Output from the logistic mixed-effects regression model for the fill-in-the-blank data

	β	*SE*	*z*	*p*
(Intercept)	4.348	0.968	4.490	< .001
Group	−2.317	1.812	−1.279	.201
Co-occurrence frequency	−1.494	1.833	−0.815	.415
Group × Co-occurrence frequency	0.809	3.662	0.221	.825

Notes: $R^2{}_m = 0.226$; $R^2{}_c = 0.597$.

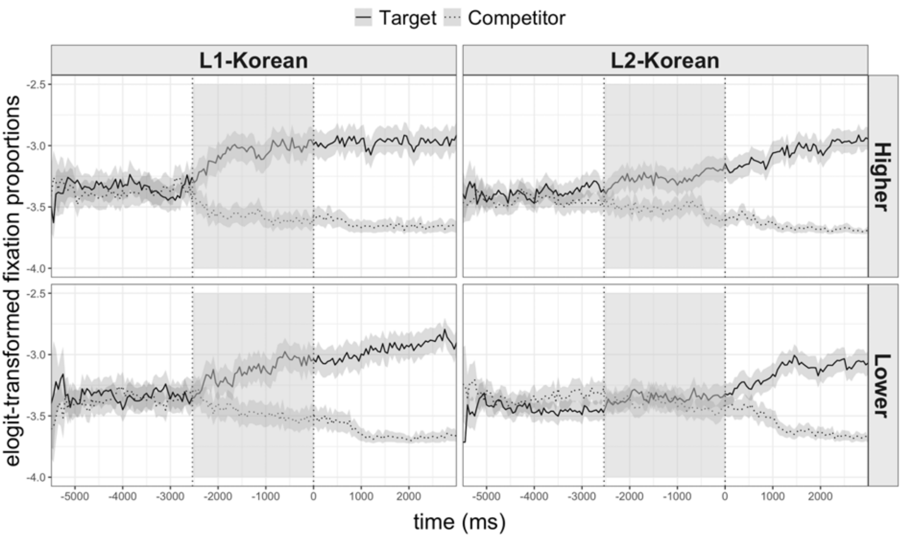

Figure 3. Elogit-transformed proportion of fixations to the target versus the competitor by group and co-occurrence frequency.
Notes: The shaded area around the lines indicates 95% confidence intervals. The highlighted section between −2,537 ms (first classifier offset) and 0 ms (noun onset) indicates the predictive region.

Results

Fill-in-the-blank knowledge task

As shown in Table 2, our logistic mixed-effects regression analysis found no significant effects of *Group*, *Co-occurrence frequency*, or their interaction (all *ps* > 0.1). Both groups achieved high scores on the offline fill-in-the-blank knowledge task. The L1 group's mean score was 96.03 (*SD* = 0.20), and the L2 group's mean score was 92.45 (*SD* = 0.26). This result indicates that both groups had firm offline knowledge of classifier-noun associations.

Visual-world eye-tracking task

Figure 3 displays a considerable group difference in the elogit-transformed proportions of fixations on the target and competitor images. During the predictive

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

Table 3. Output from the growth curve analysis model for the eye-tracking data including group and co-occurrence frequency

	β	*SE*	*t*	*p*
(Intercept)	0.255	0.063	4.034	< .001
Linear term	**0.516**	**0.072**	**7.170**	**< .001**
Quadratic term	−0.019	0.132	−0.142	.887
Group	**−0.277**	**0.068**	**−4.080**	**< .001**
Linear term	**−0.310**	**0.085**	**−3.631**	**< .001**
Quadratic term	−0.161	0.099	−1.626	.104
Co-occurrence frequency	**−0.165**	**0.012**	**−14.002**	**< .001**
Linear term	−0.090	0.173	−0.521	.603
Quadratic term	**−0.369**	**0.180**	**−2.047**	**.041**
Group × Co-occurrence frequency	0.020	0.022	0.900	.368
Linear term	**0.430**	**0.154**	**2.794**	**.005**
Quadratic term	0.160	0.204	0.780	.435

Notes: $R^2_m = 0.042$; $R^2_c = 0.159$.

region, L1-Korean speakers consistently exhibited a greater tendency to fixate on the Target image compared to the Competitor image, irrespective of the *Co-occurrence frequency* condition. By contrast, L2-Korean learners exhibited a greater tendency to fixate on the Target image over the Competitor image only when the co-occurrence frequency of classifiers and nouns was higher.

The results of our growth curve analysis for RQ 1 and RQ 2 are summarized in Table 3. The analysis revealed a significant effect of the linear time term ($p < .001$), indicating overall time-related changes in fixations during the predictive region. We also found significant effects of *Co-occurrence frequency* on both the intercept ($p < .001$) and quadratic time term ($p = .041$), as well as significant effects of *Group* on the intercept ($p < .001$) and linear time term ($p < .001$). Notably, there was a significant interaction between *Co-occurrence frequency* and *Group* on the linear time term ($p = .005$), suggesting that the L1 and L2 groups exhibited distinct fixation patterns depending on *Co-occurrence frequency*, particularly when considering time-related changes.

Now, Table 4 provides a summary of the results obtained from our GAMM analysis for RQ 1 and RQ 2. The analysis revealed significant time-related effects in the data, as indicated by the p-values associated with all by-participant and by-item smooth terms ($ps < .001$), as well as the two smooth terms for L1-Lower-Target ($p = .026$) and L1-Lower-Competitor ($p = .029$).

Understanding the results of the GAMM analysis requires a visual examination of the model's output (e.g., Rusk et al., 2020). Figure 4 provides a graphical representation of the effects of *Group* and *Co-occurrence frequency* on the time-related changes of elogit-transformed fixation proportions within the predictive region. Importantly, this figure presents vertical dashed lines where the error bar

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

Table 4. Output from the generalized additive mixed model for the eye-tracking data including group and co-occurrence frequency

A. Parametric coefficients	*β*	*SE*	*t*	*p*
(Intercept)	−3.066	0.059	−51.953	< .001
L1-Higher-Competitor	**−0.506**	**0.068**	**−7.412**	**< .001**
L1-Lower-Target	−0.068	0.094	−0.724	.469
L1-Lower-Competitor	**−0.403**	**0.083**	**−4.874**	**< .001**
L2-Higher-Target	**−0.220**	**0.083**	**−2.651**	**.008**
L2-Higher-Competitor	**−0.453**	**0.067**	**−6.755**	**< .001**
L2-Lower-Target	**−0.288**	**0.077**	**−3.772**	**< .001**
L2-Lower-Competitor	**−0.403**	**0.083**	**−4.874**	**< .001**
B. Smooth terms	*edf*	Ref.*df*	*F*	*p*
s(timebin): L1-Higher-Target	2.436	2.640	2.925	.077
s(timebin): L1-Higher-Competitor	1.817	2.004	1.843	.161
s(timebin): L1-Lower-Target	**1.001**	**1.001**	**4.956**	**.026**
s(timebin): L1-Lower-Competitor	**1.965**	**2.197**	**3.653**	**.029**
s(timebin): L2-Higher-Target	1.001	1.001	2.329	.127
s(timebin): L2-Higher-Competitor	1.044	1.054	1.391	.245
s(timebin): L2-Lower-Target	1.000	1.001	0.475	.491
s(timebin): L2-Lower-Competitor	1.001	1.001	0.463	.496
s(timebin, *participant*): L1-Higher-Target	**114.911**	**189.000**	**9.462**	**< .001**
s(timebin, *participant*): L1-Higher-Competitor	**81.690**	**189.000**	**2.401**	**< .001**
s(timebin, *participant*): L1-Lower-Target	**110.589**	**189.000**	**8.327**	**< .001**
s(timebin, *participant*): L1-Lower-Competitor	**76.834**	**189.000**	**3.440**	**< .001**
s(timebin, *participant*): L2-Higher-Target	**168.345**	**269.000**	**8.783**	**< .001**
s(timebin, *participant*): L2-Higher-Competitor	**146.940**	**269.000**	**3.244**	**< .001**
s(timebin, *participant*): L2-Lower-Target	**110.589**	**189.000**	**8.327**	**< .001**
s(timebin, *participant*): L2-Lower-Competitor	**160.279**	**269.000**	**5.121**	**< .001**
s(timebin, *item*): L1-Higher-Target	**64.050**	**108.000**	**2.665**	**< .001**
s(timebin, *item*): L1-Higher-Competitor	**46.424**	**108.000**	**2.143**	**< .001**
s(timebin, *item*): L1-Lower-Target	**65.507**	**108.000**	**6.369**	**< .001**
s(timebin, *item*): L1-Lower-Competitor	**42.504**	**108.000**	**4.747**	**< .001**
s(timebin, *item*): L2-Higher-Target	**60.006**	**107.000**	**4.560**	**< .001**
s(timebin, *item*): L2-Higher-Competitor	**43.721**	**107.000**	**1.985**	**< .001**
s(timebin, *item*): L2-Lower-Target	**54.448**	**107.000**	**4.929**	**< .001**
s(timebin, *item*): L2-Lower-Competitor	**65.286**	**107.000**	**3.621**	**< .001**

Note: *Rho* value to correct for autocorrelation within the model: −0.394.

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

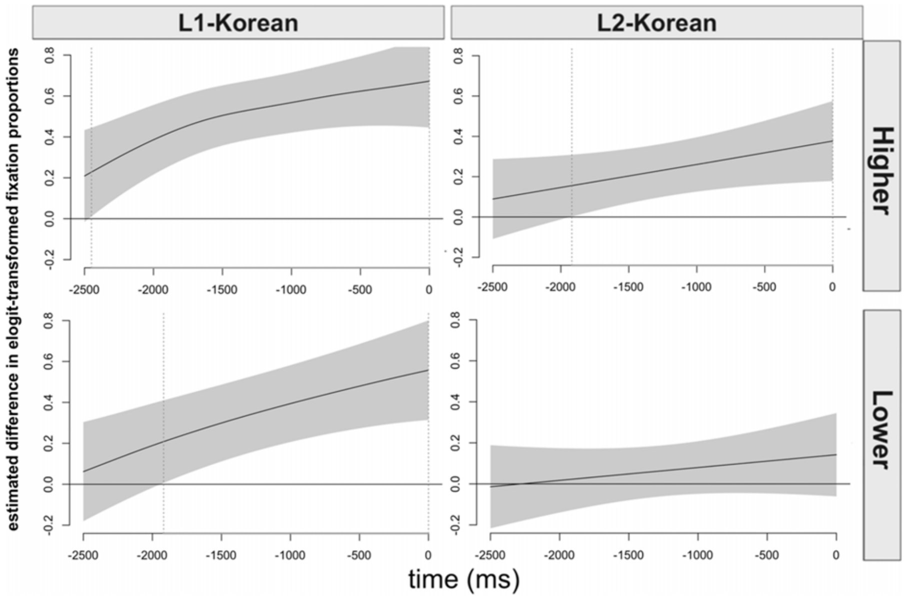

Figure 4. Difference between the elogit-transformed proportions of fixations to the target and the competitor over time during the predictive region by group and co-occurrence frequency.
Notes. A value greater than zero on the y-axis indicates that the Target received more fixations than the Competitor, whereas a value less than zero suggests the opposite pattern, with the Competitor receiving more fixations than the Target. The vertical dashed lines represent the time period where significant differences between the conditions under comparison emerge.

shading does not cover the x-axis to indicate significant differences between the conditions in comparison. Inspecting this figure, therefore, enables us to identify significant differences between the looks to the Target and Competitor by *Group* and *Co-occurrence frequency* and to determine the exact time when such differences become apparent. Our pair-wise comparisons between the Target and Competitor in the L1-Korean group showed a significant difference in both the Higher condition (from −2449.49 ms) and the Lower condition (from −1944.44 ms), with the former reaching statistical significance earlier. By contrast, the L2 group's elogit-transformed fixation proportions on the Target and Competitor were statistically different only in the Higher condition, with a latency of −1944.44 ms, which was 505.05 ms later when compared to the L1 group.

As shown in Table 5, the growth curve analysis for RQ 3 showed a significant effect of the linear time term ($p < .001$). It also revealed a significant effect of *Proficiency group* on both its intercept ($p < .001$) and quadratic time term ($p = .003$), and a significant effect of *Co-occurrence frequency* on both its intercept ($p < .001$) and quadratic time term ($p = .043$). There was also a significant interaction between *Proficiency group* and *Co-occurrence frequency* across the intercept ($p = .024$), linear time term ($p = .001$), and quadratic time term ($p = .029$).

Further, the GAMM analysis for RQ 3 showed significant time-related effects in the data, as shown in Table 6: All by-participant and by-item smooth terms were statistically significant (ps $< .001$).

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

Table 5. Output from the growth curve analysis model for the eye-tracking data including proficiency group and co-occurrence frequency

	β	*SE*	*t*	*p*
(Intercept)	0.146	0.052	2.814	.013
Linear term	**0.339**	**0.095**	**3.577**	**< .001**
Quadratic term	0.159	0.166	0.957	.340
Proficiency group	**0.192**	**0.046**	**4.148**	**< .001**
Linear term	0.008	0.110	0.072	.943
Quadratic term	**0.384**	**0.132**	**2.924**	**.003**
Co-occurrence frequency	**−0.181**	**0.016**	**−11.176**	**< .001**
Linear term	-0.164	0.223	-0.735	.463
Quadratic term	**−0.475**	**0.234**	**−2.032**	**.043**
Proficiency group × Co-occurrence frequency	**−0.066**	**0.029**	**−2.251**	**.024**
Linear term	**−0.653**	**0.203**	**−3.211**	**.001**
Quadratic term	**−0.598**	**0.274**	**−2.180**	**.029**

Notes: ${R^2}_m = 0.025$; ${R^2}_c = 0.080$.

Table 6. Output from the generalized additive mixed model for the eye-tracking data including proficiency group and co-occurrence frequency

A. Parametric coefficients	β	*SE*	*t*	*p*
(Intercept)	−3.205	0.074	−43.160	< .001
More advanced-Higher-Competitor	**−0.321**	**0.087**	**−3.701**	**< .001**
More advanced-Lower-Target	−0.078	0.105	−0.744	.457
More advanced-Lower-Competitor	**−0.219**	**0.087**	**−2.513**	**.012**
Less advanced-Higher-Target	−0.150	0.104	−1.437	.151
Less advanced-Higher-Competitor	**−0.325**	**0.087**	**−3.725**	**< .001**
Less advanced-Lower-Target	**−0.237**	**0.090**	**−2.636**	**.008**
Less advanced-Lower-Competitor	**−0.201**	**0.096**	**−2.085**	**.037**
B. Smooth terms	*edf*	Ref.*df*	*F*	*p*
s(timebin): More advanced-Higher-Target	1.000	1.000	0.272	.602
s(timebin): More advanced-Higher-Competitor	1.001	1.001	1.130	.288
s(timebin): More advanced-Lower-Target	1.001	1.001	0.433	.510
s(timebin): More advanced-Lower-Competitor	1.001	1.001	1.474	.225
s(timebin): Less advanced-Higher-Target	1.000	1.001	3.742	.053
s(timebin): Less advanced-Higher-Competitor	1.000	1.000	0.450	.503

(*Continued*)

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

Table 6. (*Continued*)

A. Parametric coefficients	*β*	*SE*	*t*	*p*
s(timebin): Less advanced-Lower-Target	1.000	1.000	0.006	.940
s(timebin): Less advanced-Lower-Competitor	1.000	1.000	0.003	.956
s(timebin, *participant*): More advanced-Higher-Target	**82.900**	**135.000**	**8.528**	**< .001**
s(timebin, *participant*): More advanced-Higher-Competitor	**77.894**	**134.000**	**3.781**	**< .001**
s(timebin, *participant*): More advanced-Lower-Target	**79.419**	**134.000**	**8.679**	**< .001**
s(timebin, *participant*): More advanced-Lower-Competitor	**85.631**	**134.000**	**5.335**	**< .001**
s(timebin, *participant*): Less advanced-Higher-Target	**79.114**	**134.000**	**7.615**	**< .001**
s(timebin, *participant*): Less advanced-Higher-Competitor	**65.819**	**134.000**	**2.697**	**< .001**
s(timebin, *participant*): Less advanced-Lower-Target	**66.521**	**134.000**	**3.296**	**< .001**
s(timebin, *participant*): Less advanced-Lower-Competitor	**66.908**	**134.000**	**4.987**	**< .001**
s(timebin, *item*): More advanced-Higher-Target	**61.994**	**108.000**	**4.602**	**< .001**
s(timebin, *item*): More advanced-Higher-Competitor	**42.927**	**107.000**	**1.696**	**< .001**
s(timebin, *item*): More advanced-Lower-Target	**58.291**	**107.000**	**4.518**	**< .001**
s(timebin, *item*): More advanced-Lower-Competitor	**54.657**	**107.000**	**2.169**	**< .001**
s(timebin, *item*): Less advanced-Higher-Target	**43.249**	**107.000**	**2.716**	**< .001**
s(timebin, *item*): Less advanced-Higher-Competitor	**49.150**	**107.000**	**2.055**	**< .001**
s(timebin, *item*): Less advanced-Lower-Target	**53.250**	**107.000**	**3.354**	**< .001**
s(timebin, *item*): Less advanced-Lower-Competitor	**60.760**	**107.000**	**3.071**	**< .001**

Note: *Rho* value to correct for autocorrelation within the model: –0.366.

Crucially, our visual inspection of Figure 5 revealed that less advanced L2 learners showed evidence of prediction only in the Higher condition at –1262.63 ms, consistent with our findings on the entire L2 dataset. However, more advanced L2 learners exhibited patterns similar to those of L1-Korean speakers, displaying evidence of prediction in both the Higher condition (at –2222.22 ms) and the Lower condition (at –883.83 ms), albeit with a delayed onset in the latter condition.

Discussion

Regarding RQ 1, the study showed that L1-Korean speakers looked at the image of the target noun significantly more often than they looked at the image of the competitor upon hearing the classifier and before hearing the target noun. A similar pattern was observed among L2-Korean learners when the classifiers and nouns had a higher frequency of co-occurrence. These results indicate that both groups make

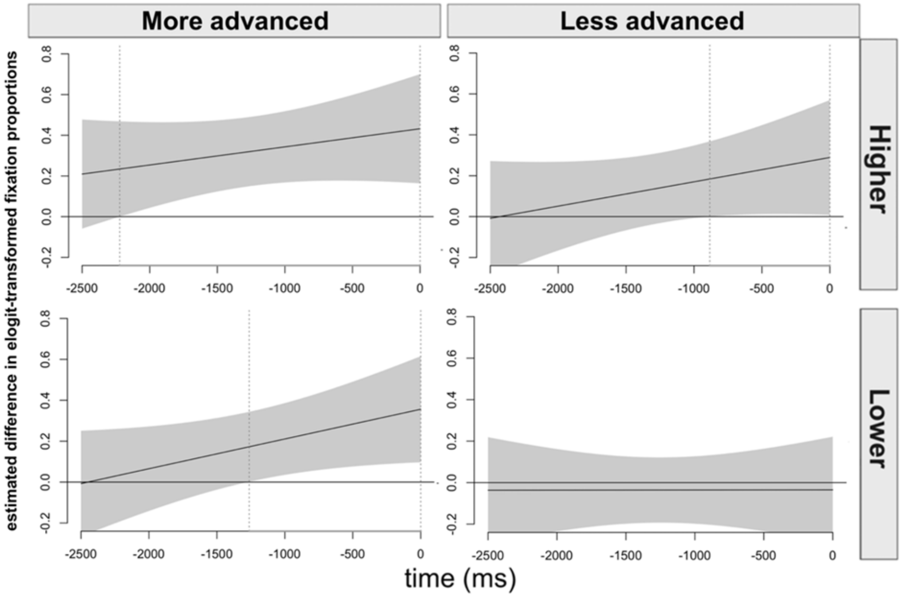

Figure 5. Difference between the elogit-transformed proportions of fixations to the target and the competitor over time during the predictive region by proficiency group and co-occurrence frequency. *Notes*: A value greater than zero on the y-axis indicates that the Target received more fixations than the Competitor, whereas a value less than zero suggests the opposite pattern, with the Competitor receiving more fixations than the Target. The vertical dashed lines represent the time period where significant differences between the conditions under comparison emerge.

use of grammatical classifiers to generate expectations for the upcoming nouns. Notably, the result from the L1 group essentially echoed the findings of earlier studies on the same phenomenon in Cantonese (Tsang & Chambers, 2011) and Mandarin (Grüter et al., 2020; Huettig et al., 2010). It also corresponds with most past studies where L1 speakers have exhibited robust predictive processing patterns regardless of the target languages or phenomena (Altmann & Mirković, 2009; Levy, 2008; Pickering & Garrod, 2013).

The result from the L2 learner group in this study is consistent with Mitsugi's (2020) findings but inconsistent with Grüter et al.'s (2020). The two studies done by us and Mitsugi demonstrated L2 learners' ability to predict the upcoming noun during the predictive region (at least in the Higher condition or in the case of more advanced learners in our study). On the other hand, Grüter et al.'s study's main finding was that, unlike L1 speakers, L2 learners' predictive processing was primarily driven by semantic cues rather than grammatical ones. Although this finding provides novel insights into semantic competition in L2 predictive processing, what deserves our focus for comparison is how Grüter et al.'s L2 learner participants performed on the G−S− competitor, which is unrelated to the target classifier, as in the case of the competitor in our study. The L2 learners of Grüter et al.'s study, unlike those of our study, did not seem to show any notable differences between the target object and the unrelated G−S− competitor during the predictive region (p. 228, Figure 2; p. 230, Figure 3).

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

The inconsistency between our study and Mitsugi's (2020), on the one hand, and Grüter et al.'s (2020) study, on the other, can be attributed to L2 learners' offline grammatical knowledge. All three studies employed a fill-in-the-blank task to assess this knowledge, but with a minor difference: Whereas our study presented a full sentence that needed to be completed with a classifier (see Figure 2), Mitsugi and Grüter et al. provided a noun phrase with a numeral and head noun (e.g., 一 ______ 鱼 "a/one ____ fish"). Our research only included participants who had solid knowledge of classifiers, as evidenced by their high average accuracy score of 92.45% on the fill-in-the-blank task, with scores ranging from 80% to 100%. The L2 participants in Mitsugi's study achieved an average score of 86.50% on a similar fill-in-the-blank task, ranging from 68.75% to 100%. In stark contrast, Grüter et al.'s L2 participants struggled to reach even chance level, with an average score of 46.3% and a wide range of scores from 8% to 92%. Although the researchers excluded incorrect items in their eye-tracking data analysis, it is questionable whether the L2 learners who showed such low accuracy in the offline task had acquired the necessary grammar knowledge to be employed in real-time processing. That is, the lack of fixation differences between the target object and the unrelated G−S− competitor in their study is likely due to the participants' limited knowledge of the target grammar, which was not available to inform their online processing. In this context, it is plausible to posit that the more solid offline knowledge present in the L2 group of our study and Mitsugi's study enabled them to effectively incorporate grammatical information into their predictive processing. However, the fact that the fill-in-the-blank tasks used in these studies were not identical in terms of the length and complexity of the stimuli compromises the comparability of the results.

Note that our study and Grüter et al.'s (2020) study also differ with respect to the L1 backgrounds of the L2 participants. In our study, the L2 learners' L1 was Chinese, which presents classifiers. In Grüter et al.'s (2020) study, however, the L2 learners had varying L1s, including English, German, Spanish, Dutch, and Hebrew, none of which have classifiers. Given the attested effects of L1 in shaping L2 processing (e.g., Foucart & Frenck-Mestre, 2011), it also appears likely that the presence of classifiers in our L2 group's L1 contributed to their predictive processing abilities. However, given that Mitsugi (2020) also showed successful predictive processing with their L1-English L2-Japanese learners, whose L1 lacks classifiers, casts doubt on this possibility.

Furthermore, it is important to acknowledge that the three studies being compared also exhibit several other dissimilarities. These include variations in the target language (Chinese, Japanese, Korean), the specific eye-tracking system used (webcam-based eye-tracking, SMI RED250 eye-tracker, Eyelink 1000), and the analysis methodology employed (linear mixed-effects regression, GAMM). The age of onset and length of residence may also have had an impact on the mixed results, although this information is not fully reported in all studies, except for Grüter et al.'s (2020) study, which reported an age of onset of 13 or older in their L2 learners. Therefore, future research should systematically manipulate and account for these factors to unveil the specific factors that led to the inconsistent results from the studies under comparison.

Regarding our RQ 2, both L1-Korean speakers and L2-Korean learners displayed an effect of the frequency of co-occurrence between classifiers and nouns in predictive processing, whereby a higher frequency expedited the predictions in the case of L1-Korean speakers and boosted the ability to make online predictions in the case of L2-Korean learners. The beneficial influence of co-occurrence frequency on processing has been observed in both L1 (e.g., Brunellière et al., 2017; Jones & Golonka, 2012; Saffran et al., 1996) and L2 contexts (e.g., Ellis et al., 2014; Yi et al., 2017). However, this effect has primarily been discussed in relation to lexical or phrasal processing (e.g., Brunellière et al., 2017; Yi et al., 2017). The current study, which includes both L1 and L2 groups, expands the scope of previous research on the effect of lexical co-occurrence frequency to include predictive processing.

As for RQ 3, the current study found that more advanced L2-Korean learners exhibited predictive processing patterns similar to those of L1-Korean speakers, although with a slight delay in processing. Less advanced learners showed evidence of predictive processing only when the classifier-noun co-occurrence frequency was higher, but more advanced learners demonstrated evidence of predictive processing even when the classifier-noun co-occurrence frequency was lower. These results lead us to hypothesize that L2 learners can possess what we call a "Sound Ability to Generate Expectations" (SAGE; Hwang & Kim, 2025), similar to that of L1 speakers. By "sound" ability, we imply that there is no fundamental difference in the quality of predictive processing mechanisms between L2 learners and L1 speakers. This hypothesis contrasts with the RAGE hypothesis (Grüter et al., 2017) on predictive processing as well as the Shallow Structure Hypothesis (Clahsen & Felser, 2006, 2018) on general sentence processing, which posits that L2 learners process sentences in a different manner compared to L1 speakers, owing to difficulties in making use of their grammatical knowledge in real-time processing.

Meanwhile, the predictive processing ability of L2 learners can be masked or their processing patterns can differ quantitatively from L1 speakers, showing delayed processing, which might be (mis)interpreted as a sign of reduced abilities. In our view, these processing patterns are attributable to various linguistic factors, such as low lexical frequency and high structural complexity, as well as individual factors, such as low proficiency in the target language and limited working memory capacity (see also Kaan, 2014). Many such factors can reveal, enhance, or impede the ability, efficiency, and optimality, which may be represented by the speed of prediction. In this sense, the SAGE hypothesis is in line with the utility account of prediction in L2 (Kaan & Grüter, 2021), which suggests that L2 learners are more likely to make predictions when the processing demands are not overly taxing.

Taking advantage of the data available to us, we conducted a further analysis of our L2 data, factoring in etymological connections between our learners' L1 Chinese and L2 Korean, which may arguably facilitate predictive processing. (Due to the limited number of tested classifiers showing L1-L2 phonetic similarities [$k = 3$], we failed to construct a growth curve model in R.) Specifically, we distinguished between etymologically connected and unconnected pairs based on whether our Korean classifiers originated from the Chinese counterparts (for our annotation, see Appendix A). We then built a growth curve model, following the same procedure outlined in the "Analysis" section, with *Etymological connection* and *Co-occurrence frequency* as fixed effects and *Participant* and *Item* as random effects. This model revealed both main and interaction effects related to *Etymological connection* on the intercept, linear time term, and quadratic time term (all ps $< .001$). To further

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

examine these interaction effects, we analyzed difference plots derived from a GAMM. The results showed that L2 learners demonstrated predictive behavior when the frequency of classifier-noun co-occurrence was higher, but not when it was lower. Notably, within the Higher condition, the presence of etymological connections between L1 and L2 classifiers accelerated predictive processing, resulting in earlier evidence of prediction when the target classifiers were etymologically connected to the corresponding ones in Chinese (–2247.47 ms vs. –858.59 ms).

Regarding our fill-in-the-blank task results, one reviewer suggested that they may provide counter-evidence to the prediction-by-production theory (Pickering & Gambi, 2018), which maintains that language users make predictions most effectively using their production system. In the present study, both L1-Korean speakers and L2-Korean learners exhibited co-occurrence frequency effects in the eye-tracking task, but they did not demonstrate these effects in the fill-in-the-blank task, where they accurately filled in sentences with target words irrespective of co-occurrence frequency. While this contrast between production and prediction patterns is interesting, it is important to acknowledge the inherent differences between our two tasks. The fill-in-the-blank task involved classifier production, but the eye-tracking task involved noun prediction based on classifiers. Furthermore, the tasks employed distinct measurements, with one focusing on accuracy and the other on fixations over time. To gain a deeper understanding of the production-prediction relationship, one should develop production and prediction tasks that are more comparable in terms of their design and measurement approaches in further research.

Finally, the results derived from this study support the webcam-based eye-tracking system as a practical and cost-effective alternative to in-lab eye-tracking systems, particularly in the context of visual-world experiments. Recently, psycholinguistics research has begun using this method on a variety of platforms (e.g., PCIbex: Schwarz & Zehr, 2021; WebGazer.js: Papoutsaki, 2015) to test their validity against in-lab eye-tracking equipment, although GazeRecorder that we used in this study has not been explored in this regard. Using WebGazer.js, Vos et al. (2022) successfully replicated their findings on the predictive processing of verb aspect from the SMI Red500 eye-tracker. Based on this result, they proposed the webcam-based method as a promising approach for examining the fine-grained temporal aspects of predictive processing in the visual-world paradigm, while pointing out some limitations. One limitation is a 50-ms delay in the onset of the tested effect, potentially due to browser or webcam-based program processing speed; and another limitation pertains to the calibration process, which many of their participants failed to complete (see also Slim & Hartsuiker, 2023).

The current study's L1 group also showed signs of predictive processing around 50 ms after the predictive region began (see Figure 4). This raises the possibility that the time-related limitation may also apply to the GazeRecorder system. However, we must exercise caution, as we have not directly compared our GazeRecorder results with those obtained using in-lab equipment. Regarding calibration, we did not encounter any issues with GazeRecorder, suggesting that it may be more robust than WebGazer.js. Taken together, future research utilizing the visual-world eye-tracking paradigm can consider this webcam-based eye-tracking system as an

https://doi.org/10.1017/S0142716425100052 Published online by Cambridge University Press

alternative when an in-lab eye-tracking device is not accessible (e.g., Semmelmann & Weigelt, 2018; Vos et al., 2022; Yang & Krajbich, 2021).

Limitations and future directions

The current study has arguably a small sample size, with 21 L1 participants and 30 L2 participants. This limitation likely reduces the statistical power of our analysis, underscoring the need for future research to expand the participant pool and enhance the generalizability of the results. Moreover, all our L2 participants' L1 was Chinese, a language that features classifiers. To examine the impact of L1 on predictive processing based on classifiers, we plan to recruit participants whose L1 does not have a classifier system.

Regarding our stimuli, there was variation in semantic informativity across our classifiers. Despite the fact that various classifiers differ in nature and that their semantic informativity or transparency falls along a continuum, this issue can be addressed in future studies. One potential solution is to control semantic informativity by carefully choosing nouns that do not exhibit semantically transparent associations with the classifiers that they belong to.

Furthermore, the repetition of the same classifier twice, which aimed to lengthen the predictive region, may be viewed as a potential issue. Although we recognize that this design choice is not optimal and may not mirror real-life language use, our findings do not seem to suggest that this repetition resulted in artificial facilitation of predictive processing or awareness of the study's purpose among our L2 participants as they did not show evidence for prediction in the Lower condition (despite their intact offline knowledge of classifier-noun associations for the same condition). Future research can address this issue by adding an adjective before the target noun instead of an additional classifier, as shown in (3), to create a longer predictive region.

(3)	Cinhuy-ka	han calwu-uy	phalan	yenphil-ul	chac-ass-e-yo.
	Cinhuy-NOM	one CL-GEN	blue	pencil-ACC	find-PST-DEC-POL
	"Cinhuy found a blue pencil."				

Finally, we opted to treat co-occurrence frequency as a categorical variable in a factorial design, which allowed our GAMM analysis to pinpoint the exact time period where statistical differences between conditions under comparison arise. Nevertheless, treating co-occurrence frequency as a continuous variable can offer deeper insights into its effects on prediction, and we suggest that future research consider this approach.

Replication package. The complete (1) study materials, (2) analysis code, and (3) data are available at https://osf.io/9gu6w/.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0142716425100052

Acknowledgments. We would like to express our gratitude to Sanghyun Han for his help with the development of the audio materials employed in the eye-tracking experiment.

Funding statement. This research received no specific grant from any funding agency.

Competing interests. The authors declare none.

References

Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, **33**, 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x

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

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, **94**, 206–234. https://doi.org/10.1016/j.jml.2016.11.006

Baker, C., & Love, T. (2022). It's about time! Time as a parameter for lexical and syntactic processing: an eye-tracking-while-listening investigation. *Language, Cognition and Neuroscience*, **37**, 42–62. https://doi.org/10.1080/23273798.2021.1941147

Barr, D. J. (2008). Analyzing 'visual world'eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, **59**, 457–474. https://doi.org/10.1016/j.jml.2007.09.002

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

Beyersmann, E., Montani, V., Ziegler, J. C., Grainger, J., & Stoianov, I. P. (2021). The dynamics of reading complex words: evidence from steady-state visual evoked potentials. *Scientific Reports*, **11**, 15919. https://doi.org/10.1038/s41598-021-95292-0

Brunellière, A., Perre, L., Tran, T., & Bonnotte, I. (2017). Co-occurrence frequency evaluated with large language corpora boosts semantic priming effects. *Quarterly Journal of Experimental Psychology*, **70**, 1922–1934. https://doi.org/10.1080/17470218.2016.1215479

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, **27**, 3–42. https://doi.org/10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, **40**, 693–706. https://doi.org/10.1017/S0272263117000250

Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, **24**, 143–88. https://doi.org/10.1017/S0272263102002024

Ellis, N. C., O'Donnell, M.B., & Romer, U. (2014). The processing of verb–argument constructions is sensitive to form, function, frequency, contingency, and prototypicality. *Cognitive Linguistics*, **25**, 55–98. https://doi.org/10.1515/cog-2013-0031

Erickson, L. C., & Thiessen, E. D. (2015) Statistical learning of language: theory, validity, and pre- dictions of a statistical learning account of language acquisition. *Developmental Review*, **37**, 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Fang, S., & Wu, Z. (2022). Syntactic prediction in L2 learners: evidence from English disjunction processing. *International Review of Applied Linguistics in Language Teaching*, **62**, 429–456. https://doi.org/10.1515/iral-2021-0223

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology*, **44**, 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Felser, C., & Cunnings, I. (2012). Processing reflexives in a second language: the timing of structural and discourse-level constraints. *Applied Psycholinguistics*, **33**, 571–603. https://doi.org/10.1017/S0142716411000488

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, **14**, 379–399. https://doi.org/10.1017/S136672891000012X

Fuchs, Z. (2022). Facilitative use of grammatical gender in Heritage Spanish. *Linguistic Approaches to Bilingualism*, **12**, 845–871. https://doi.org/10.1075/lab.20024.fuc

Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes*, **24**, 1124–1179. https://doi.org/10.1080/01690960802018323

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, **58**, 787–814. https://doi.org/10.1016/j.jml.2007.07.001

Grüter, T., Lau, E., & Ling, W. (2020). How classifiers facilitate predictive processing in L1 and L2 Chinese: the role of semantic and grammatical cues. *Language, Cognition and Neuroscience*, **35**, 221–234. https://doi.org/10.1080/23273798.2019.1648840

Grüter, T., & Rohde, H. (2021). Limits on expectation-based processing: use of grammatical aspect for co-reference in L2. *Applied Psycholinguistics*, **42**, 51–75. https://doi.org/10.1017/S0142716420000582

Grüter, T., Rohde, H., & Schafer, A. J. (2017). Coreference and discourse coherence in L2: the roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*, **7**, 199–229. https://doi.org/10.1075/lab.15011.gru

Henry, N., Jackson, C. N., & Hopp, H. (2022). Cue coalitions and additivity in predictive processing: the interaction between case and prosody in L2 German. *Second Language Research*, **38**, 397–422. https://doi.org/10.1177/0267658320963151

Hopp, H. (2017). Individual differences in L2 parsing and lexical representations. *Bilingualism: Language and Cognition*, **20**, 689–690. https://doi.org/10.1017/S1366728916000821

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, **17**, 5–27.

Huettig, F., Chen, J., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behavior, but only during linguistic processing. *Journal of Cognition and Culture*, **10**, 39–58. https://doi.org/10.1163/156853710X497167

Hwang, H., & Kim, K. (2025). Effects of lexical frequency in predictive processing: higher frequency boosts L1 speed and facilitates L2 prediction. *Language Learning.* Advance online publication. https://doi.org/10.1111/lang.12718

Hwang, S., Kwon, H-C., & Yoon, A. (2010). KorLexClas 1.5: a lexical semantic network for Korean Numeral Classifier. *Journal of KIISE: Software and Applications*, **37**, 60–73.

Jones, L. L., & Golonka, S. (2012). Different influences on lexical priming for integrative, thematic, and taxonomic relations. *Frontiers in Human Neuroscience*, **6**, 1–17. https://doi.org/10.3389/fnhum.2012.00205

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

Kaan, E., & Grüter, T. (2021). Prediction in second language processing and learning: advances and directions. In E. Kaan & T. Grüter (Eds.), *Prediction in second-language processing and learning* (pp. 1–24). John Benjamins.

Kim, H., & Grüter, T. (2021). Predictive processing of implicit causality in a second language: a visual-world eye-tracking study. *Studies in Second Language Acquisition*, **43**, 133–154. https://doi.org/10.1017/S0272263120000443

Kim, H., Shin, G-H., & Hwang, H. (2020). Integration of verbal and constructional information in the second language processing of English dative constructions. *Studies in Second Language Acquisition*, **42**, 825–847. http://doi.org/10.1017/S0272263119000743

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

Lee-Ellis, S. (2009). The development and validation of a Korean C-Test using Rasch analysis. *Language Testing*, **26**, 245–274.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, **106**, 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Ma, A. (2015). *Hanyu geti liangci de chansheng yu fazhan [The Emergence and Development of Chinese Individual Classifiers]*. China Social Sciences Press.

Mahr, T., McMillan, B. T., Saffran, J. R., Weismer, S. E., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, **142**, 345–350. https://doi.org/10.1016/j.cognition.2015.05.009

Mitsugi, S. (2020). Generating predictions based on semantic categories in a second language: a case of numeral classifiers in Japanese. *International Review of Applied Linguistics*, **58**, 323–349. https://doi.org/10.1515/iral-2017-0118

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

National Institute of Korean Language. (2021). NIKL Spoken Corpus (v.1.2). Retrieved from https://corpus.korean.go.kr

National Institute of Korean Language. (2023). Korean Learners' Corpus Search Engine. Retrieved from https://kcorpus.korean.go.kr/

Papoutsaki, A. (2015). Scalable webcam eye tracking by learning from user interactions. In *Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems* (pp. 219–222). Association for Computing Machinery. https://doi.org/10.1145/2702613.2702627

Park, E. (2014). KoNLPy: Korean natural language processing in Python. In E. L. Park & S. Cho (Eds.), *Proceedings of the 26th annual conference on human & cognitive language technology* (pp. 133–136). Human & Language Technology.

Perdomo, M., & Kaan, E. (2021). Prosodic cues in second-language speech processing: a visual world eye-tracking study. *Second Language Research*, **37**, 349–375. https://doi.org/10.1177/0267658319879196

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

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

Python Software Foundation. (2022). Python language reference (Version 3.8.16). Retrieved from http://www.python.org

R Core Team. (2024). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Available from https://www.R-project.org/

Rusk, B. V., Paradis, J., & Järvikivi, J. (2020). Comprehension of English plural-singular marking by Mandarin-L1, early L2-immersion learners. *Applied Psycholinguistics*, **41**, 547–577. https://doi.org/10.1017/S0142716420000089

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Schwarz, F., & Zehr, J. (2021). Tutorial: introduction to PCIbex–an open-science platform for online experiments: design, data-collection and code-sharing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, **43**. Retrieved from https://escholarship.org/uc/item/1ng1q4c6

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: a first look. *Behavior Research Methods*, **50**, 451–465. https://doi.org/10.3758/s13428-017-0913-7

Slim, M. S., & Hartsuiker, R. J. (2023). Moving visual world experiments online? A web-based replication of Dijkgraaf, Hartsuiker, and Duyck (2017) using PCIbex and WebGazer. js. *Behavior Research Methods*, **55**, 3786–3804. https://doi.org/10.3758/s13428-022-01989-z

Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analyses of visual world data: applications to bilingual research. *Bilingualism: Language and Cognition*, **24**, 833–841. https://doi.org/10.1017/S1366728920000607

The Academy of Korean Studies. (2023, October 7). Encyclopedia of Korean Culture. Retrieved from https://encykorea.aks.ac.kr/Article/E0078797

Tsang, C., & Chambers, C. (2011). Appearances aren't everything: shape classifiers and referential processing in Cantonese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **37**, 1065–1080. https://doi.org/10.1037/a0023601

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, D. (2022). itsadug: interpreting time series and autocorrelated data using GAMMs. R package version 2.4.1.

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the Visual World Paradigm. *Glossa Psycholinguistics*, **1**, 1–37. https://doi.org/10.5070/G6011131

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & McDonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cognitive Psychology*, **58**, 250–271. https://doi.org/10.1016/j.cogpsych.2008.08.002

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, **70**, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Wood, C. (2006). Metrical stress sensitivity in young children and its relationship to phonological awareness and reading. *Journal of Research in Reading*, **29**, 270–287. https://doi.org/10.1111/j.1467-9817.2006.00308.x

Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, **16**, 1485–1505. https://doi.org/10.1017/S1930297500008512

Yi, W., Lu, S., & Ma, G. (2017). Frequency, contingency and online processing of multiword sequences: an eye-tracking study. *Second Language Research*, **33**, 519–549. https://doi.org/10.1177/0267658317708009

Zehr, J., & Schwarz, F. (2018). Returning to non-entailed presuppositions again. In U. Sauerland & S. Solt (Eds.), *Proceedings of Sinn und Bedeutung 22* (pp. 463–480). Leibniz-Zentrum Allgemeine Sprachwissenschaft (-ZAS-). https://doi.org/10.21248/zaspil.61.2018.507

Cite this article: Hwang, H. & Park, S. H. (2025). The influence of word co-occurrence frequency on predictive processing in first and second languages: A webcam-based eye-tracking study. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716425100052