Language Learning & Technology
ISSN 1094-3501
2023, Volume 27, Issue 1
pp. 1–22

ARTICLE

Using eye-tracking as a tool to develop L2 lexical knowledge

Andrea Révész, *University College London*
Matthew Stainer, *Griffith University*
Jookyoung Jung, *Chinese University of Hong Kong*
Minjin Lee, *Yonsei University*
Marije Michel, *University of Groningen*

Abstract

Eye-tracking is primarily used as a tool to capture attentional processes in second language (L2) research. However, it is feasible to design visual displays that can react to and interact with eye-movements in technology-mediated contexts. We explored whether gaze-contingency can foster L2 development by drawing attention to novel words reactively during reading. In particular, we investigated whether the acquisition of lexis can be facilitated by interactive glosses, that is, making glosses visually salient when triggered by fixations on a target word. We found that interactive, gaze-contingent glosses led to more and longer fixations at target words and glosses but did not lead to superior performance in recognition scores. We observed, however, an interaction between interactivity and form recognition, with more gloss fixations being associated with better performance under the interactive, but with worse outcomes in the non-interactive, condition. We attributed this difference to distinct motivations for viewing glosses in the groups.

Keywords: *Eye-tracking, Gaze-contingency, Glossing, Vocabulary*

Language(s) Learned in This Study: *English pseudowords*

APA Citation: Révész, A., Stainer, M., Jung, J., Lee, M., & Michel, M. (2023). Using eye-tracking as a tool to develop lexical knowledge. *Language Learning & Technology*, *27*(1), 1–22. https://hdl.handle.net/10125/73537

Introduction

Eye-tracking methodology is becoming increasingly popular in second language (L2) research as a tool to capture the cognitive processes in which learners engage while they interact with visual materials (Conklin et al., 2018; Godfroid, 2020), with an increasing number of eye-tracking studies conducted in computer-assisted language learning (CALL) contexts (Michel & Smith, 2017). Given advances in human-computer-interaction research and software engineering, now it is not only possible to record eye-gaze behaviours, but also to design visual information displays that can react to and interact with eye movements (e.g., Duchowski, 2007). This capacity opens up the possibility of using eye-tracking as an interactive technology-mediated educational tool, which provides tailor-made support to learners when they experience difficulty or fail to attend to target stimuli.

This gaze-contingency affordance of eye-tracking holds special promise in fostering L2 development in CALL contexts. In L2 acquisition, it is generally assumed that, to optimise L2 development, learners' attention needs to be drawn to linguistic features (Schmidt, 2001). Researchers have also suggested that a beneficial way to achieve this is to employ reactive focus-on-form interventions, which direct learners' attention to linguistic elements in context where the primary focus is on meaning (Long & Robinson, 1998).

Research on both face-to-face (e.g., Mackey & Goo, 2007) and computer-mediated interaction (e.g., Sauro, 2009) has provided ample evidence that reactive focus-on-form techniques can promote L2 learning. It remains unexplored, however, whether reactive pedagogical interventions can facilitate L2 development during reading activities. Gaze-contingent paradigms may help address this gap through their potential to draw learners' attention to L2 constructions in an adaptive, reactive manner (Slavuj et al., 2017) during digital reading. This is a promising technological affordance, given that, increasingly, technology-mediated written texts constitute a key source of input for L2 learners.

The gaze-contingency paradigm may also facilitate learners' use of help options in CALL environments. Help options, defined as "embedded application resources that assist learners in performing computing operations and/or support language learning" (p. 69), are often ignored by L2 learners (Cárdenas-Claros & Gruba, 2010). In digital reading, glosses constitute one type of language-related help option, but they may be skipped by learners (Warren et al., 2018). Gaze-contingent eye-tracking might have the capacity to draw attention to glosses and other help options that learners may otherwise fail to use.

This study is the first to explore whether gaze-contingent eye-tracking can indeed direct attention to language-related help options in CALL environments and thereby serve as a reactive focus-on-form intervention. Specifically, we aimed to test whether using interactive, gaze-contingent eye-tracking to highlight glosses that learners would otherwise not view can direct learners' attention to L2 vocabulary during digital reading and thus aid development in L2 lexical knowledge. While several meta-analyses have indicated that static, non-reactive glosses can promote lexical development in CALL contexts (Abraham, 2008; Yanagisawa et al., 2020, Yun, 2011), it appears worthwhile to examine whether the positive effects of glossing may be further enhanced through interactive eye-tracking, operationalised here as highlighting glosses when learners do not view them when expected (Warren et al., 2018).

Previous Use of Gaze-contingency in Developmental and Educational Research

The idea of using eye-tracking as an interactive tool is by no means new. Gaze-contingency experiments date back to the 1970s when first language (L1) reading researchers (Rayner, 1975; Reder, 1973), began to use gaze-contingency to investigate perceptual span, the visual area that is viewed with enough resolution to process the text during an eye fixation. It was not until more recently, however, that researchers also started to explore how the gaze-contingency paradigm could be exploited for developmental and educational purposes (Wass et al., 2011).

Among studies employing gaze-contingency, recent work by Lee et al. (2015) is probably the closest to this research. The researchers examined whether the temporal contingency of animated pedagogical agents (i.e., virtual characters designed to assist learning) can promote foreign vocabulary learning. Under the gaze-contingent condition, the agent-initiated eye-contact with the participant, then turned his gaze to a picture depicting a target word. When joint attention was formed (i.e., the participant also fixated on the target picture), the agent uttered the target word. If the participant did not form joint attention, the agent looked at the student and invited them to follow their gaze. Under the control condition, the agent did not interact with the participants' eye-movements. The results yielded an advantage for the gaze-contingent condition.

In our study, similar to Lee et al. (2015), we intended to draw learners' attention to unknown words through interactive eye-tracking and thereby facilitate vocabulary learning. However, rather than using gaze-contingent agents to present words out of context, we attempted to enhance attention to the target words through interactive glosses in context. In the next section, we present a review of previous research on glossing and discuss how the gaze-contingency paradigm may help increase the effectiveness of this technique.

Glossing, Attention, and L2 Development

Glossing is a pedagogical technique, which supplies contextualised information about language items via translations, synonyms, or definitions in paper-based or computer-mediated contexts. In CALL environments, glossing is seen as a type of help option with a potential to assist language learning. Indeed, glosses, as a means of focus on form, appear to have the capacity to trigger noticing (Leow, 2009), that is, to draw focal attention to aspects of linguistic constructions (e.g., a meaning associated with a form) with a low level of awareness (Schmidt, 2001). Once learners have noticed a piece of linguistic information, this may be rehearsed in working memory and followed by further processing, which may lead to L2 learning (Leow, 2015; Schmidt, 2001). Thus, reading activities integrating glosses seem to constitute an ideal platform for encoding new aspects of L2 constructions by combining a principal focus on meaning with opportunities to notice new linguistic information (Long & Robinson, 1998).

Given that focal attention to glossed information is a prerequisite for it to be further processed, enhancing the salience of glosses (Gass et al., 2017) would probably make it more likely that their potentially positive effects are exploited. Taking the perspective of recent accounts of attention as a multiple system (Chun et al., 2011), making glosses more perceptually salient may help direct learners' external or perceptual attention to glosses. This might increase the probability that glosses also receive internal attention, a condition for further processing, given that internal attention, responsible for executive functioning and cognitive control, is involved in selecting perceptual information for encoding and maintenance in working memory. It would seem especially beneficial to increase the salience of glossed linguistic features reactively when they otherwise receive no external attention. In such cases, learners would receive the needed linguistic information in context when their primary, internal attention is dedicated to processing content. Interactive, eye-gaze contingent glosses seem to have the potential to achieve this type of reactive focus-on-form intervention during reading

Although there is no research available on the impact of interactive, gaze-contingent glosses on attentional allocation and L2 development, considerable research has investigated the extent to which glossing may facilitate vocabulary acquisition. According to Yanagisawa et al.'s (2020) recent meta-analysis, glossing promotes L2 vocabulary learning, with multiple-choice glosses being the most effective gloss type and L1 glosses and L1 plus L2 glosses leading to more learning than L2 only glosses. As expected, Yanagisawa et al. found that gains were more pronounced on recognition than recall delayed post-tests. Interestingly, proficiency and glossing mode (textual, pictorial, auditory) did not influence vocabulary learning. Of relevance to this study are also the results of two meta-analyses focusing on glossing in computer-mediated contexts. Abraham (2008) observed that computer-mediated glosses had a large effect on incidental vocabulary learning, observing, similar to Yanagisawa et al. (2020), a larger effect size for receptive than productive tests. Yun (2011) exclusively focused on hypertext glosses, identifying a significant but not conclusive advantage for multiple as compared to single hypertext glossing.

A few glossing studies, almost all conducted in CALL contexts, also investigated the impact of glosses on attentional allocation. In a pioneering study, Bowles (2004) examined whether paper-and-pen versus computer-delivered translation glosses led to greater awareness and learning in comparison to a no glossing condition. The researcher used think-alouds to assess attentional allocation and meaning recall and recognition tests to measure vocabulary acquisition. Bowles found that participants noticed the target items and recognised their meanings more when they were glossed, but no significant difference emerged between the paper-and-pen and computer-based conditions. In another think-aloud study, Guidi (2009) examined the effects of computer-mediated L1 glossing on attention to and learning of lexical and grammatical constructions. Unlike Bowles, Guidi revealed no difference in awareness and recognition of the target constructions across the glossing and no glossing groups. However, the noticing of both glossed lexical and grammatical items had a positive relationship with recognition scores. Martinez-Fernández (2010), using a paper-based treatment, also examined the noticing and learning of glossed lexical and grammatical items with think-aloud protocols and recognition and written production post-tests. Participants received either L1 translation glosses, L1 translation fill-in-the blank glosses, or no glosses. Like Bowles's study, the gloss conditions led to more reported awareness and learning of lexical constructions, but glossing type did not affect noticing. Glossing, however, had no influence on the noticing and learning of the grammatical feature, as in Guidi's study. To summarise, while there are indications that glossing can enhance noticing and subsequently result in learning, the findings are inconclusive.

Given these conflicting findings, Jung and Révész (2018), in another computer-mediated experiment, investigated whether increasing the salience of glosses through manipulating reading task characteristics can help draw attention to pseudo-lexical items. The researchers combined eye-tracking with stimulated recall data to tap attentional allocation to and awareness of target pseudowords and grammatical constructions and corresponding L1 translation glosses. The study found that an enhanced need to read carefully resulted in greater attention to and awareness of the glossed grammatical feature. However, reading activity characteristics did not influence the relationship between glossing and noticing of the pseudo-lexical items.

Besides investigating the effects of glossing on attentional allocation, an eye-tracking study by Warren et al. (2018) examined the extent to which attention to glosses and targeted linguistic features might predict L2 learning of lexical items in a computer-mediated context. In the experiment, target pseudowords were annotated by pictorial, textual, or multimodal (picture and text) marginal glosses, and form recall and meaning recognition tests measured vocabulary development. The study yielded an advantage for pictorial glosses, but gloss type had no significant influence on the amount of attention dedicated to the pseudowords and glosses. Notably, however, more and longer fixations at the target words led to better form recall and meaning recognition. This result is consistent with previous research, which yielded a positive relationship between fixation counts and/or durations and offline measures of vocabulary uptake (see Godfroid, 2019, for a review). A further finding worth highlighting was that a considerable number of glosses received no fixations (text-only condition: 20%, picture-only: 11.4%; multimodal: 9.3%), thus excluding the possibility that learners benefit from them.

It appears reasonable to assume that interactive, eye-gaze contingent glossing can direct learner attention to glosses which are otherwise ignored (as in Warren et al., 2018) and thereby increase the likelihood that learners benefit from the information they contain. Our aim was to explore this assumption.

Research Questions

1. To what extent does interactive versus non-interactive glossing draw learners' attention to in-text lexical items and their glosses, as reflected in participants' eye-movements?
2. To what extent does interactive versus non-interactive glossing promote the recognition of in-text lexical items, as reflected in participants' performance on a form and a meaning recognition test?
3. To what extent does attention to in-text lexical items and associated glosses relate to the recognition of form and meaning of in-text lexical items? Is this relationship influenced by glossing condition, interactive versus non-interactive?

Methodology

Design

The participants were 45 L2 users of English, who took part in one individual session. They were first administered a background questionnaire, followed by a proficiency test. Then, they were randomly assigned to two groups. During the treatment, both groups read two texts in which pseudo-lexical items were glossed. One group (N = 23) was exposed to interactive glosses (i.e., glosses reacted in response to participants' eye-gaze behaviours), whereas the other group (N = 22) read texts accompanied with traditional non-interactive glosses. While participants were completing the reading activities, their eye-gaze behaviours in both groups were recorded. Immediately after reading each text, participants were administered a short post-reading questionnaire. Finally, they completed an exit questionnaire.

Participants

All 45 participants were L1 Mandarin speakers. Thirty-five were female, and the mean age was 23.36 (SD = 2.81). All participants were university students in the UK, studying for various degrees (education and social science: $n = 32$; medicine, engineering, and science: $n = 7$; humanities: $n = 6$). Their proficiency level was on average in the B2 band according to the Common European Framework of Reference, as determined by their performance on the Use of English and Reading section of a practice Cambridge Advanced English (CAE) test. Out of the maximum 78 points, the participants achieved an average score of 44.13 ($SD = 11.91$; *95% CI* = [40.55, 47.71]). Cronbach's alpha for the CAE scores was .82. The participants had similar demographic characteristics across the groups (see Appendix A). Independent samples t-tests confirmed that there was no significant difference between the interactive and traditional glossing groups in terms of age ($t = .40$, $p = .69$, $d = 0.12$) or proficiency ($t = 1.12$, $p = .28$, $d = 0.33$).

Target Constructions

The target constructions were pseudowords to control for existing lexical knowledge. Five pseudowords were included in both texts taken from Jung and Révész (2018). Each contained seven letters and two syllables (e.g., *stragon*) (see Appendix B for all target words). The pseudowords appeared once, and they were all nouns. They replaced ten original words in the two texts, and were aligned with English morphological and orthographic rules. The plural marker -*s* was retained when the original word was used in the plural.

Texts

The two texts were taken from past TOEFL tests. They were selected because they addressed topics with which participants were unlikely to be familiar. The topic of text one was petroleum resources, and text two was about the Cambrian period. The texts contained 682 and 699 words respectively. The average readability scores, computed using various indices (Automated Readability, Coleman-Liau index, Flesch-Kincaid grade level, Gunning-Fog score, SMOG index), were 11.6 and 13.4 for texts one and two. Based on these values, the texts appeared suitable for the participants. Both texts are available in the IRIS database.

Reading Activity

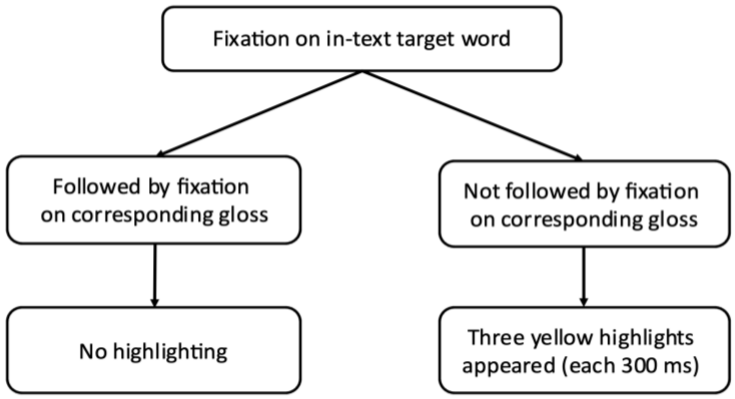
Following the original TOEFL format, the two texts were split into five segments, with each segment presented on a separate page. The participants' task was to answer multiple-choice comprehension questions taken from the TOEFL tests. Responding to the questions involved identifying factual information, making inferences, understanding rhetorical purpose, recognizing vocabulary meaning, simplifying/paraphrasing a sentence, or selecting main ideas of the text (Educational Testing Service, 2012). Thirteen multiple-choice comprehension items accompanied each text, each segment being followed by one to four questions. For each text, the maximum score was 13 points. We used double-spaced Courier font (size: 14) to present the texts. Twenty-five minutes were provided for completing the activities as this time proved sufficient through piloting.

Interactive Eye-tracking Manipulation of Glossing Conditions

Each in-text target pseudoword was underlined and a Mandarin translation was given in a marginal gloss. The translations were provided independently by two L1 Mandarin-speaking applied linguists. The percentage agreement was 100% between them. Participants' eye-movements were recorded under both glossing conditions, but only under the interactive glossing condition did the glosses interact with eye-gaze behaviours. We recorded participants' eye position with a Tobii X2-30 eye tracker, which sampled eye position at 30Hz. A custom script was written by the second author using Matlab Version 2017b. The script first calibrated the eye-tracker using a 9-point calibration grid, and then recorded eye position live. When the eye was stable (within approximately half a degree) for three consecutive samples (around 75 ms), this was marked as a fixation. In the interactive glossing condition, three yellow highlights appeared on the gloss when a fixation was detected on an in-text target word but was not followed by a fixation on the corresponding gloss. Each highlight would last 300 ms before disappearing. When a fixation on the in-text target word was followed by a fixation on the appropriate gloss, no highlighting occurred (see Figure 1 for a schematic representation of the intervention). In this way, reactive focus on form was achieved: learners' attention was directed to the meaning of the pseudowords (available in the L1 glosses) reactively in context when they would have otherwise failed to pay attention to the meaning of targeted lexical items presented in the glosses.

Figure 1

Schematic Representation of the Intervention



Assessment Tasks

The recognition of the pseudo-words was measured with a form and a meaning recognition test. Form and meaning recognition tests were considered suitable to assess the participants' knowledge of the pseudowords as the participants were exposed to each target item only once. Both tests were constructed and administered using the software E-Prime 2.0.

The form recognition test had 20 items, modelled after a test in Leeser (2007, p. 269). Participants were instructed to press "*z* (yes)" or "*m* (no)," according to whether they remembered seeing the lexical items in the texts. Ten items were pseudowords, which had appeared in the texts. The remaining ten items were distractors, which were constructed based on pseudowords in Godfroid et al. (2013). Similar to the target items, the distractors included two syllables and seven letters. The items were presented on a computer screen in randomised order. Participants received 1 point when they responded "yes" for an item which had been included in the texts and when they responded "no" to an item that had not appeared in the readings. Similar, they were allocated zero points when they had pressed "yes" for an item not present in the texts and "no" for an item that had occurred in the readings. Thus, the maximum score was 20 points. The test took approximately three minutes.

The meaning recognition test included 20 multiple-choice items. Participants were asked to select the correct Mandarin translation of a pseudoword from among three options. Ten items were target pseudowords, whereas the other ten were distractors also used in the form recognition test. For the target items, one option included the gloss for the target word, the other two options presented glosses for other target words, and the last option was "*I don't know*" to prevent guessing. The items were presented on a computer screen in a randomised order. Participants received one point for each correct response on the target items, resulting in a total score of 10. Participants took about three minutes to complete the test.

Questionnaires

The background questionnaire elicited demographic information. The post-reading questionnaires included two 7-point Likert-scale items, which measured participants' familiarity with the reading topics. In the exit questionnaire, we asked participants what, in their view, the purpose of the study was and whether they had learnt any new language.

Procedure

After providing informed consent, participants were administered the background questionnaire and the proficiency test. Then, the eye-tracking system was calibrated, followed by the first reading activity and post-reading questionnaire. After a short break, the same procedure was repeated for the second reading activity and post-reading questionnaire. While reading, participants' eye movements were captured with a Tobii X2-30 eye tracker. A forehead and a chin rest were used to keep the viewing distance constant throughout the experiment. Participants sat approximately 600 mm away from the center of the computer screen. The sessions were conducted individually in a quiet room at the first author's institution. The sessions lasted about two hours. Participants were offered a 10-minute break after the proficiency test and between the reading activities.

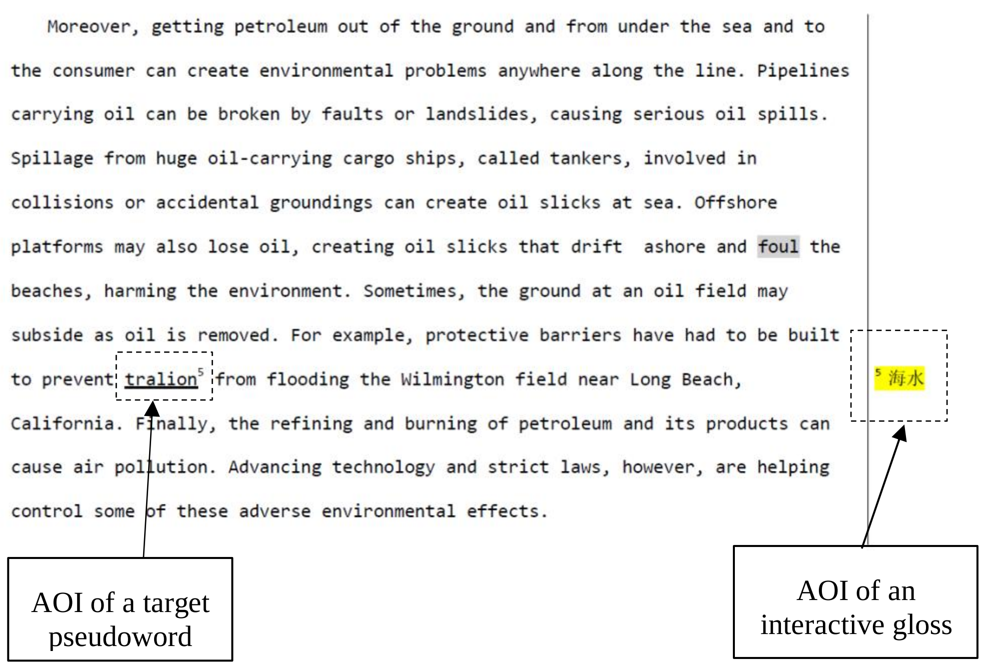
Data Analyses

Eye-movement Data

The eye-tracking data were extracted with Matlab Version 2017b. Two types of areas of interest (AOI) were identified, for the in-text target pseudowords and accompanying glosses (see Figure 2, *the gloss is highlighted in response to the learner's eye-movement*). The AOIs were drawn by hand. For these AOIs, we obtained total number of fixations and total fixation durations. We utilised these measures to gain information about both the frequency and length of fixations at the AOIs. We found strong correlations among these measures (target pseudowords: $r = .82$, $p < .01$; glosses: $r = .75$, $p < .01$). For each index, we expected greater values under the interactive condition as interactive highlighting was anticipated to draw more attention to the glosses and in-text target pseudowords.

Figure 2

An Example Page with AOIs



Statistical Analyses

We used the Statistical Package for the Social Sciences (SPSS) 25 to compute reliability and descriptive statistics and to run independent samples *t*-tests to compare the participants' topic familiarity with the texts across groups. To check for potential differences in the two groups' reading comprehension scores and to address the research questions, we constructed a series of mixed-effects regression models using the *lmer* and *glmer* functions of the *lme4* package in the R statistical package. In each analysis, the non-interactive group served as the reference category. An alpha level of .05 was set as the criterion for significance. To measure effect sizes, Cohen's *d* values were computed for the independent-samples *t*-tests and marginal R^2 values were obtained for the mixed effects models using the *MuMln* package. Cohen's *d* values above 0.40, 0.70, and 1.00 were considered as small, medium and large, respectively.

Results

Preliminary Analyses

Topic Familiarity

We found no significant difference in participants' familiarity with the text topics, when we compared participants' responses on the post-reading Likert scale items with a series of independent samples *t*-tests. The descriptive statistics for the questionnaire items and the results of the *t*-tests are presented in Appendix C and Appendix D respectively.

Reading Comprehension

The mean comprehension scores per item ($n = 13$) for the interactive and the non-interactive groups were 0.73 ($SD = 0.45$, 95% CI [0.69, 0.77]) and 0.67 ($SD = 0.47$, 95% CI [0.65, 0.73]) respectively. A binomial mixed effects regressions analysis revealed that interactivity had no significant influence on reading comprehension scores ($z = 1.56$, $p = .12$, see Appendix E for full results).

Eye-tracking Data

To ensure that our eye movement processing led to data that was in line with previous research, we examined participants' average fixation durations. Rayner's (1998) influential review on eye tracking identified an average fixation duration of 225 ms during silent reading. Cop, Drieghe and Duyck (2015) showed that reading in an L2 results in increases in fixation durations of around 20 ms on average for L2 users with upper-intermediate proficiency. In our processed data, the mean fixation duration was 245.77 ms ($SD = 82.02$), reflecting what is expected based on previous research.

Exit Questionnaire Data

Only four participants, in the interactive group, thought that they had learnt some new vocabulary. No participant identified the purpose of the study correctly; the majority thought that the goal was to test reading skills.

RQ1: The Effects of Interactivity on Learner Attention

The descriptive statistics for fixation counts and total fixation durations for the in-text target pseudowords and glosses are given in Table 1. To address the effects of interactivity on attentional allocation, we first ran a series of linear mixed-effect regression models. Group was the fixed effect in each model, and the random effects were participant and item, the latter referring to the in-text target pseudowords or glosses. In addition to the by-item and by-participant intercepts, we also tried adding a random slope for condition by item, but the resulting maximal models led to singular fit. In each model, the dependent variable was the number of fixations or total duration of fixations on the in-text target words or glosses. As shown in Table 2, the analyses found a significant difference between the interactive and non-interactive groups for each eye-tracking measure, explaining 2% to 5% of the variation. These results mean that glossing through interactive eye-tracking was more successful in drawing attention to the in-text target words and accompanying glosses than non-interactive, traditional glossing.

Table 1

Descriptive Statistics for Fixation Counts and Durations for Target Pseudowords and Glosses

		Target pseudowords				Glosses			
				95% CI				95% CI	
Measure	**Group**	Mean	SD	Lower	Upper	Mean	SD	Lower	Upper
Fixation Count	Interactive	2.24	0.91	1.28	3.20	1.28	0.52	0.75	1.80
	Non-interactive	1.48	1.14	0.51	2.44	0.78	0.64	0.24	1.31
Fixation Duration	Interactive	634.24	286.54	390.05	878.43	393.48	229.46	235.24	551.72
	Non-interactive	310.68	228.19	65.64	555.73	177.50	159.10	18.45	336.55

Table 2

Results for the Models Examining the Effect of Interactivity on Fixation Counts and Durations on Target Pseudowords and Glosses

		Fixed effects						Random effects	
Measure	**Predictors**	Est	SE	t	p	R^2m	R^2c	Factor	SD
Target pseudowords									
Fixation Count	Intercept	1.48	0.45	3.24	<.001	.02	.41	Participant	0.85
	Interactivity	0.77	0.31	2.50	.016			Item	1.39
Fixation Duration	Intercept	310.68	115.50	2.69	.016	.05	.32	Participant	185.9
	Interactivity	323.56	77.44	4.18	<.001			Item	351.1
Glosses									
Fixation Count	Intercept	0.78	0.25	3.08	.007	.03	.42	Participant	0.48
	Interactivity	0.50	0.17	2.89	.006			Item	0.76
Fixation Duration	Intercept	177.50	75.66	2.35	.031	.04	.24	Participant	135.8
	Interactivity	215.98	59.12	3.65	<.001			Item	217.4

Note. measure ~ 1 + interactivity + (1|participant) + (1|item)

To get a more nuanced picture of attentional allocation, we also examined how often participants fixated on glosses after they had fixated on the corresponding in-text target lexical item. In particular, we investigated how many times participants fixated on a gloss in the window up to eight fixations following a fixation at the corresponding in-text target word. Our rationale for considering eight subsequent fixations was that, if participants had already planned subsequent saccadic movements, then any effect of interactivity may be evident across a more extended timescale. Out of the eight subsequent fixations, participants fixated at glosses 39% of the time under the interactive condition (M = 0.39, SD = 0.04), whereas the proportion of looks at the glosses was only 23% for the non-interactive condition (M = 0.23, SD = 0.01).

Next, we ran a mixed-model ANOVA to explore whether there were significant differences in the number of gloss fixations among the eight fixations following a fixation at the in-text target word. The dependent variable was the proportion of times that participants looked at a gloss after looking at the in-text target word. One of the independent variables was the interactivity condition (between-subjects factor). The other independent variable was defined as the number of fixations examined post fixating at the in-text target word (within-subjects factor). For example, for value 4, we inspected how many times the participant fixated at the gloss within the four fixations (F1 to F4) they made following a fixation at the in-text target word. The results of this ANOVA are shown in Table 3, with marginal means plotted in Figure 3. In sum, we found significant main effects for both interactivity and fixation number, with a significant interaction.

Table 3

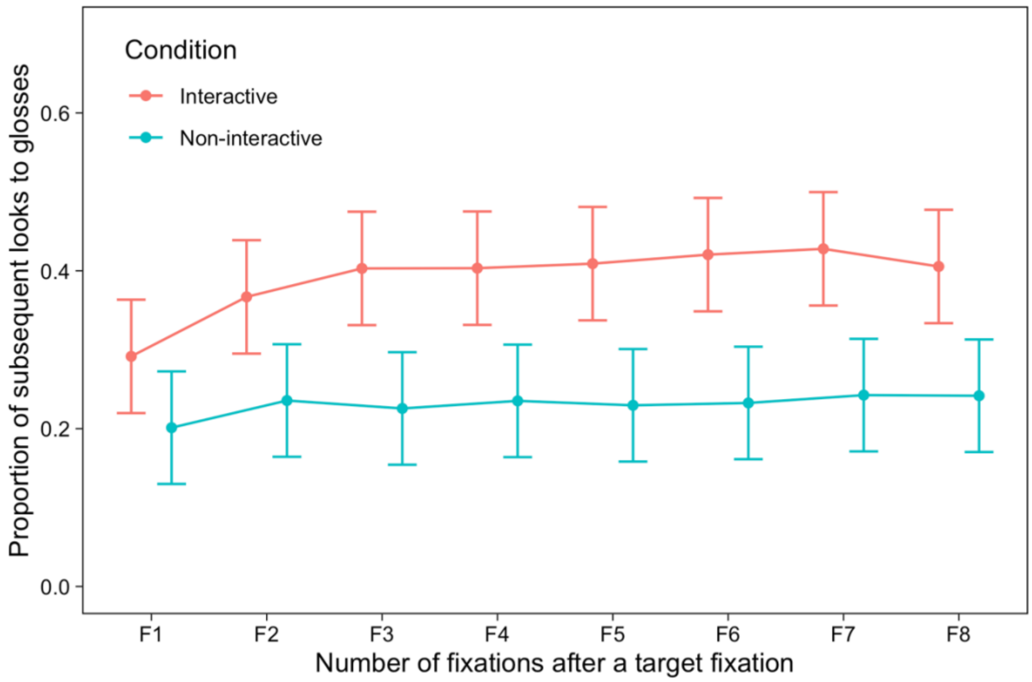
Results of Mixed-design ANOVA Exploring Proportions of Fixations to Glosses in the Fixations Following a Target Word Fixation

	Sum of Squares	df	Mean Square	F	p	η^2_p
Within Subjects Effects						
Fixation	0.20	7	0.03	6.41	< .01	.15
Fixation * Interactivity	0.07	7	0.01	2.33	.03	.06
Residual	1.15	252	<0.01			
Between Subjects Effects						
Interactivity	1.95	1	1.95	12.1	<.01	.25
Residual	5.80	36	0.16			

To tease apart the significant interaction, we examined whether there were differences at each post-target-fixation applying a Bonferroni-adjusted alpha criterion (.006). There was no significant difference between the interactivity conditions when considering the first (p = .079) or the first two (p = .012) fixations following a fixation at the in-text target word. However, there were differences when we examined the proportions of fixations at the gloss following three to eight fixations after fixation at the in-text target (all p values < .002). This demonstrates that the positive effect of the interactive highlighting was not immediate but developed across multiple fixations.

Figure 3

Marginal Mean Plot (with 95% CI) for the Proportion of Subsequent Looks to Glosses across the Eight Fixations after Fixating at the Target Word



RQ2: The Effects of Interactivity on Form and Meaning Recognition

Table 4 presents the descriptive statistics for the form and meaning recognition total scores. To investigate the effects of interactivity on learners' recognition scores, we ran two separate linear mixed-effect regression models. The fixed effect was group in each model, and the random effects were participant and item. This time item referred to the test items in the form or meaning recognition tests. In addition, we also added a random slope for condition by item. The resulting maximal model led to a better-fitting model for the meaning but not the form recognition results. Participants' form or meaning recognition responses served as the dependent variable. Table 5 shows that in neither analysis did a significant difference emerge between the interactive and non-interactive groups' performance on the form or meaning recognition tests. These results indicate that interactive glossing did not lead to significantly better recognition of the form and meaning associated with glossed pseudo-lexical items.

Table 4

Descriptive Statistics for the Form and Meaning Recognition Tests by Group

Test	Group	N	Mean	SD	95% CI	
					Lower	Upper
Form recognition	Interactive	23	13.96	4.41	13.05	14.86
	Non-interactive	22	12.73	2.81	11.48	13.98
Meaning recognition	Interactive	23	5.47	1.59	4.79	6.17
	Non-interactive	22	5.04	1.73	4.28	5.81

Note. The total score for the form and meaning recognition tests were 20 and 10 points respectively.

Table 5

Results for the Models Examining the Effect of Interactivity on Form and Meaning Recognition Scores

		Fixed effects						Random effects	
Test	Predictors	Est	SE	z	p	R^2m	R^2c	Factor	SD
Form Recognition	Intercept	0.91	0.15	6.12	<.001	<.01	<.01	Participant	0.03
	Interactivity	-0.33	0.20	-1.63	.103			Item	<0.01
Meaning recognition	Intercept	0.19	0.14	1.34	.180	<.01	.02	Participant	0.174
	Interactivity	-0.17	0.24	-0.74	.460			Item (Intercept)	0.138
								Item (Interactivity)	0.414

Note. Form recognition: measure ~ 1 + interactivity + (1|participant) + (1|item)
Meaning recognition: measure ~ 1 + interactivity + (1|participant) + (1+interactivity|item)

RQ3: Interactivity, Attention, and Form and Meaning Recognition

Finally, we ran a series of binomial linear mixed-effects regressions to examine the extent to which attention to in-text pseudowords and their associated glosses predicted the recognition of these words and whether these relationships were influenced by glossing condition. Each model had three fixed effects: interactivity, count/duration of fixations at in-text pseudowords and glosses, and the interaction between interactivity and fixation count/duration. As in the previous analyses, participant and item served as random effects. In addition, we also added a random slope for condition by item when this improved model fit. The dependent variables were participants' form or meaning recognition scores. Our predictors of interest were the main effect for the eye-gaze measures and the interaction between interactivity and the eye-movement indices.

None of the analyses found a significant main effect for fixation count or duration (see Appendix F for all model results). As shown in Table 6, however, a significant interaction effect emerged from the analysis examining the effects of interactivity on the relationship between frequency fixations at glosses and participants' form recognition scores. As shown in Figure 4, under the interactive condition, more fixations at glosses led to better form recognition scores, whereas, under the non-interactive condition, more fixations at glosses resulted in lower form recognition scores. In other words, the more participants looked at interactive glosses made salient through highlighting, the more likely they were to recognize the form of the corresponding in-text pseudowords. In contrast, the more they looked at static glosses, the less likely they were to recall the glossed in-text pseudoword forms.

Table 6

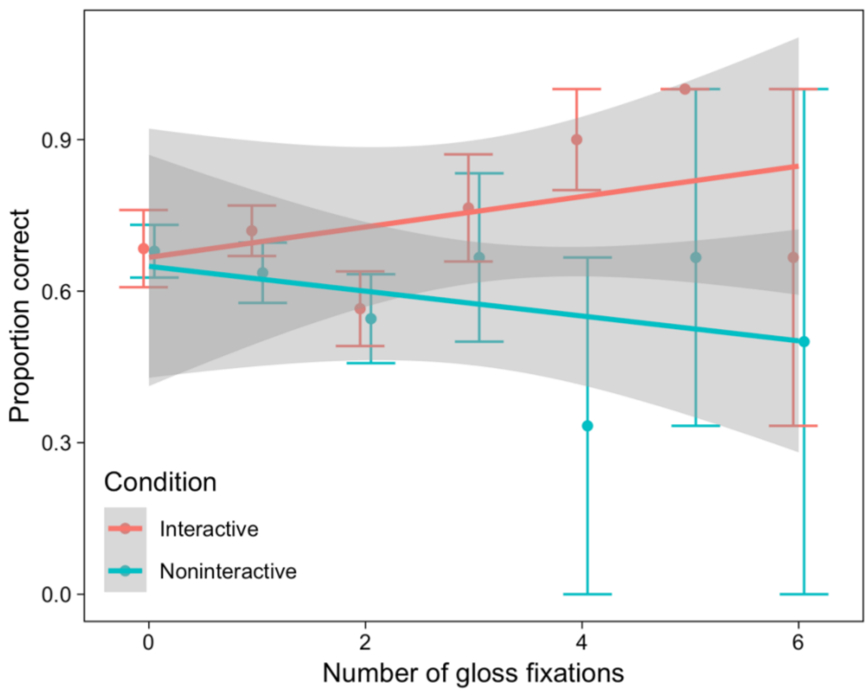
Results for Model Examining the Effects of Interactivity on the Relationship between Number of Fixations at Glosses and Form Recognition

		Fixed effects						Random effects	
Dependent variable	**Predictors**	Est	SE	z	p	R^2m	R^2c	Factor	SD
Glosses									
Form Recognition	Intercept	0.74	0.20	3.68	<.001			Participant	0.067
	Interactivity	-0.11	0.32	-0.33	.745	<.01	<-.01	Item (Intercept)	0.102
	FC	-0.19	0.12	-1.58	.115	<.01	<.01	Item (Interactivity)	0.293
	Interactivity*FC	0.32	0.16	2.00	.046	.01	.01		

Note. form/meaning recognition response ~ 1 + interactivity * fixation count/duration + (1|participant) + (1+interactivity|item)

Figure 4

Interaction Effect Between Number of Fixations at Glosses and Form Recognition



Finally, we considered whether fixating at the corresponding gloss after the in-text target lexical item promotes the recognition of the lexical item. In particular, we examined whether fixations at the in-text target words were followed or not followed by a fixation on the gloss in the subsequent eight fixations (coded as 1 and 0, respectively), and then used this value along with the interactivity condition as a predictor in a binomial mixed effects regression analysis. The dependent variable in the regression analysis was participants' binomial score on items (correct/incorrect) in the form and meaning recognition tests. As shown in Table 7, we found that, regardless of interactivity, those who viewed the gloss within the eight fixations following a fixation at the target word were significantly more likely to identify the meaning of that word on the test.

Table 7

Results for the Models Examining the Effects of Interactivity on the Relationship between Successive Gloss Fixations (GF) and Form and Meaning Recognition Scores

		Fixed effects						Random effects	
Dependent variable	**Predictors**	Est	SE	z	p	R^2m	R^2c	Factor	SD
Form Recognition	Intercept	0.63	0.30	2.11	.04			Participant	<.01
	Interactivity	0.02	0.39	0.04	.97	<.01	<-.01	Item	<.01
	Gloss Fixation	0. 25	0.39	0.63	.53	<-.01	<-.01		
	Interactivity*GF	-0.58	0.56	-1.04	.30	.01	.01		
Meaning	Intercept	-0.37	0.29	-1.28	.20			Participant	<.01
Recognition	Interactivity	0.04	0.38	2.08	.92	<.01	<.01	Item	<.01
	Gloss Fixation	0.77	0.37	0.11	.04	.03	.02		
	Interactivity*GF	-0.22	0.54	-0.41	.68	<.01	<.01		

Note. form/meaning recognition response ~ 1 + interactivity * fixation count/duration + (1|participant) + (1|item)

Discussion

Our first research question asked the extent to which interactive, gaze-contingent glossing can draw attention to lexical items, as reflected in eye-movements. We found that the interactive glossing condition, overall, generated higher fixation counts and longer fixation durations for both the in-text target pseudowords and the accompanying glosses. More detailed analyses revealed that participants in the interactive glossing condition were more likely to fixate at the corresponding gloss in the 3rd to 8th fixations after fixating at the in-text target word, indicating a slightly delayed advantage for highlighting. Although the effect sizes were relatively small, these results confirm that interactive, gaze-contingent glosses were more successful in directing attention to the in-text target lexical items and their glosses than non-interactive, static glosses. A likely explanation is that highlighting glosses dynamically, when they had otherwise been ignored, made the glosses salient (Gass et al., 2017) to learners and thus succeeded in drawing their external, perceptual attention to the glosses and the related in-text lexical items. These results are promising given that previous studies examining the effects of gloss types on attentional allocation have yielded null findings (Bowles, 2004; Martinez- Fernández, 2010; Warren et al., 2018). If future studies confirm these results, it would appear that highlighting glosses interactively, reacting to students' reading behaviors using eye-gaze contingency, has greater capacity to attract learners' external attention to glosses and in-text target words than static, non-gaze-contingent glosses. It is important to acknowledge, however, for further processing to occur, external attention needs to be coupled with internal attention. Our next research questions were concerned with exploring whether this increased external attention helped engage internal learning mechanism and thereby promote development.

In particular, our second research question addressed the extent to which interactive, gaze-contingent glosses can promote the recognition of lexical items, as indicated by participants' performance on a form and a meaning recognition test. Although the interactive glossing condition proved more effective in directing learners' external attention to glosses and in-text target pseudowords, it did not lead to significantly better performance on the form and meaning recognition tests. One way to account for the lack of significant effects might be that participants had modest exposure to the in-text target lexical items and glosses. In the interactive condition, the average fixation number for both the in-text target words and glosses were relatively low (lexical items: 2.24, glosses: 1.28). Probably, repeated occurrence of the target lexical items and associated glosses (Pellicer-Sánchez, 2016; Warren et al., 2018) would have instigated more external attention to them, thereby increasing the probability of internal attention to and subsequent recognition of the in-text target pseudowords. The absence of repeated incidence of the target words might have been further exacerbated by the fact that the processing of in-text target words and their meaning was not always essential to answering the comprehension questions, making it less likely that internal attention is generated to them.

Our final research question asked the extent to which attention to glossed lexical items relates to the recognition of their form and meaning, and whether these relationships are influenced by glossing condition. The analyses yielded a significant interaction between number of gloss fixations and interactivity for the form recognition test. When participants were exposed to non-interactive glosses, more fixations at the glosses were associated with lower form recognition scores. In other words, those non-interactive participants who looked at the glosses more often were less likely to recognize the form of the in-text target words. On the other hand, when participants received interactive, eye-gaze contingent glosses, a greater number of fixations at the glosses predicted better form recognition scores. That is, interactive participants who viewed the glosses with greater frequency were more likely to recognize in-text target word forms.

One possible explanation might be that, under the non-interactive condition, the glosses were more probable to attract those participants' attention who struggled with text comprehension. These participants might have visited the glosses hoping that the information they provide would help them understand the texts and answer the comprehension questions. If so, after they have processed the meaning of the in-text target words provided in the glosses, they might have seen little value in engaging further processing of the actual in-text target pseudowords, as only word meaning entailed in the glosses was potentially useful for facilitating comprehension. In contrast, under the interactive condition, the eye-gaze contingent highlighting might have drawn those learners' attention to the glosses who would have otherwise ignored them, that is, these learners did not visit the glosses in the hope of gaining extra assistance with text comprehension. Therefore, for them, interactive glossing might have functioned as true reactive focus on form, directing their attention to linguistic information that they would have otherwise failed to notice (Long & Robinson, 1998). Hence, the participants in the interactive glossing group were probably more prone to process the related in-text target words in greater depth, realizing that their attention was drawn to new linguistic information.

This interpretation can also be related to a distinction between early and late reading processes. Early reading processes entail word recognition and lexical access reflected in early eye-tracking measures (e.g., first fixation duration), whereas late reading processes involve text re-analysis and lexical integration manifest in late eye-tracking indices (e.g., total reading time) (Conklin et al., 2018). It is possible that participants in the non-interactive group only engaged in early processing of the in-text target words, given that the glosses drew their attention away from the target lexis and prevented them from engaging in deeper processing of their form. On the other hand, interactive glossing might have interrupted the initial processing of the pseudowords but, in doing so, might have attracted learners' focal attention to their form, resulting in stronger retention.

Two questions arise, however, following this reasoning. First, if participants in the interactive condition had indeed engaged in deeper processing of the in-text target vocabulary, it might appear surprising that higher form recognition scores were not also associated with higher fixations at the in-text target words. A lack of a significant increase in fixations on the in-text target words might be explained by a non-direct link between the number of eye-fixations and differences in depth of processing (e.g., Leow, 2015), that is, more eye-fixations do not necessarily indicate deeper processing. A second seemingly intriguing result is that we found no significant link between the eye-gaze indices and the meaning recognition scores for the interactive group. A likely explanation for this is that recognizing the meaning of target lexis requires deeper processing than recognising their form, and a single exposure to the pseudowords was probably insufficient to engage learners in adequately deep processing for meaning recognition to occur. Additionally, a greater depth of processing might have been hampered by the fact that the target lexis, as discussed above, was not always relevant to answering the comprehension questions.

Limitations and Future Research Directions

Before drawing our conclusions, it is important to acknowledge the limitations of this research and consider how future studies could explore the usefulness of the gaze-contingency paradigm in L2 research. One limitation has to do with the sampling rate of the eye-tracker utilized. Although we screened the data with great care, replicating the study with a higher-precision eye-tracker would decrease the possibility of error and thereby increase the reliability and validity of findings. A more sophisticated eye-tracker would also help increase ecological validity by allowing free head movement without compromising data accuracy. Another limitation concerns the low number of the target pseudowords. In a replication study, it would be interesting to examine whether increasing the occurrence of the pseudowords would lead, in the gaze-contingent group, to better performance on the recognition tests. Future research would also benefit from investigating how making the in-text target pseudowords more relevant to comprehension would interact with gaze-contingent glossing. It would additionally be worthwhile to examine whether enhancing the salience of interactive glossing might lead to greater attention and, hence, learning. Increased salience might be achieved, for example, through highlighting glosses and target words simultaneously. Further, it would be valuable to triangulate the data with verbal report comments and interview data to examine learners' attentional processes directly and tap their perceptions about the usefulness of eye-gaze contingency. This would allow for less speculative interpretations of the results than in the present study. In follow-up experiments, a larger sample size would also help increase the power of the study, and a within-subject design would assist in controlling for the potential effect of individual differences.

Another exciting avenue for future research would be to explore alternative applications of the gaze-contingency paradigm to facilitate the use of various help options in CALL environments (Cárdenas-Claros & Gruba, 2010). For example, gaze-contingency could be used to direct learner attention to L2 constructions through textual enhancement. While students read, targeted L2 constructions could be enhanced reactively whenever learners fixate on them. Comparing the effectiveness of various textual enhancement techniques (e.g., highlighting, bold-facing) in gaze-contingent applications could generate important pedagogical implications. Future research could also explore whether eye-gaze contingency could be utilized to promote the use of online dictionaries and hyperlinks that learners might otherwise neglect to access. Interactive eye-tracking might help draw attention to these help options in an adaptive manner, at moments when learners would benefit most from their assistance (Slavuj et al., 2017).

Conclusion

The purpose of this study was to initiate research into the use of the gaze-contingency paradigm as a means to achieve reactive focus on form in the context of L2 digital reading. Specifically, we intended to assess whether interactive, gaze-contingent eye-tracking can draw learner attention to glosses that would otherwise be ignored and thereby direct learners' attention to the corresponding L2 lexical constructions and promote their retention. Our results revealed that interactive glosses led to higher fixation counts and longer fixation durations at both in-text target words and glosses than non-interactive glosses, indicating that gaze-contingent glossing, as expected, was more successful than static glossing in drawing attention to target lexis. However, interactive glossing did not result in superior performance on the form and meaning recognition tests, probably due to lack of repeated occurrence of the target items (Pellicer-Sánchez, 2016; Warren et al., 2018). Finally, we found a significant interaction between interactivity and number of gloss fixations for the form recognition results. Under the interactive, eye-gaze contingent condition, greater fixation count was associated with better performance on the form recognition test, whereas, under the non-interactive condition, more fixations at the glosses were linked to poorer recognition scores. We attributed this difference to a distinction in the two groups' motivation for viewing glosses, with gaze-contingent glossing functioning as reactive focus on form and non-interactive glossing providing a scaffold to support reading comprehension.

Continued research into the usefulness of gaze-contingency in facilitating attention and L2 learning seems essential. If future studies confirm that this technology can help achieve reactive focus on form when processing written input, this technique might help decrease entrenchment and fossilisation by drawing learners' attention to linguistic features that may go unnoticed. Further research is also warranted into whether the eye-gaze contingency paradigm has the potential to increase the use of CALL help options more generally and thereby facilitate L2 development. If further studies confirm the usefulness of interactive eye-tracking for L2 learning in CALL applications, manufacturers could be encouraged to integrate an eye-tracker into devices such as digital readers, tablets, and computers to enable the use of eye-gaze contingent CALL applications.

References

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*, 199–226. https://doi.org/10.1080/09588220802090246

Bowles, M. (2004). L2 glossing: To CALL or not to CALL. *Hispania*, *87*, 541–552. https://doi.org/10.2307/20063060

Cárdenas-Claros, M. S., & Gruba, P. A. (2010). Help options in CALL: A systematic review. *CALICO Journal*, *27*, 69–90. https://doi.org/10.11139/cj.27.1.69-90

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *An introduction to eye-tracking: A guide for applied linguistics research*. Cambridge University Press. https://doi.org/10.1017/9781108233279

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PloS one*, *10*, e0134008 https://doi.org/10.1371/journal.pone.0134008

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. Springer

E-Prime (Version 2.0) [Computer software and manual]. Psychology Software Tools Inc.

Gass, S. M., Spinner, P., & Behney, J. (2017). Salience in second language acquisition and related fields. In S.M. Gass, P. Spinner, & L. Behney (Eds.), *Salience in second language acquisition* (pp. 1–18). Routledge

Godfroid, A. (2019). Investigating instructed second language acquisition using L2 learners' eye-tracking data. In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 44–57). Routledge. https://doi.org/10.4324/9781315165080

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge

Godfroid, A., Housen, A., & Boers, F (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*, 483–517. https://doi.org/10.1017/S0272263113000119

Guidi, C. (2009). *Glossing for meaning and glossing for form. A computerized study of the effects of* *glossing and type of linguistic item on reading comprehension, noticing, and L2 learning*. Unpublished dissertation, Georgetown University, Washington, D.C.

IBM Corp. (2017). IBM SPSS Statistics for Windows (Version 25.0) [Computer software]. IBM Corp.

Jung, J., & Révész, A. (2018). The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions. *Studies in Second Language Acquisition*, *40*, 755–780. https://doi.org/10.1017/S0272263118000165

Lee, H., Kanakogi, Y., & Hiraki, K. (2015) Building a responsive teacher: how temporal contingency of gaze interaction influences word learning with virtual tutors. *Royal Society Open Science*, *2*, 140361. https://doi.org/10.1098/rsos.140361

Leeser, M. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: topic familiarity and working memory. *Language Learning, 57*, 229–270. https://doi.org/10.1111/j.1467-9922.2007.00408.x

Leow, R. P. (2009). Modifying the L2 reading text for improved comprehension and acquisition: Does it work? In Z.-H. Han & N. J. Anderson (Eds.), *Second language reading research and instruction: Crossing the boundaries* (pp. 83–100). The University of Michigan Press

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. J. Doughty & J. Williams (Eds.), *Focus on form in second language acquisition* (pp. 15–41). Cambridge University Press

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 407–453). Oxford University Press

Martinez-Fernández, A. M. (2010). *Experiences of remembering and knowing in SLA, L2 development, and text comprehension: A study of levels of awareness, type of glossing, and type of linguistic item* [Unpublished doctoral dissertation]. Georgetown University

The Math Works, Inc. (2017). *MATLAB* (Version 2017a) [Computer software]. https://www.mathworks.com/

Michel, M. C., & Smith, B. (2017). Eye-tracking research in computer-mediated language learning. In S. Thorne & S. May (Eds.), *Language, education and technology* (pp. 1–12). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-02237-6

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition, 38*, 97–130. https://doi.org/10.1017/S0272263115000224

R Core Team (2020). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7*, 65–81. https://doi.org/10.1016/0010-0285(75)90005-5

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Reder, S. M. (1973). On-line monitoring of eye-position signals in contingent and noncontingent paradigms. *Behavior Research Methods & Instrumentation*, *5*, 218–228. https://doi.org/10.3758/BF03200168

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning* & *Technology*, *13*, 96–120. http://dx.doi.org/10125/44170

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Slavuj, V., Meštrović, A., &. Kovačić B. (2017) Adaptivity in educational systems for language learning: a review. *Computer Assisted Language Learning, 30*, 64–90. https://doi.org/10.1080/09588221.2016.1242502

Warren, P., Boers, F., Grimshaw, G., & Siyanova-Chanturia, A. (2018). The effect of gloss type on learners' intake of new words during reading: Evidence from eye-tracking. *Studies in Second Language Acquisition, 40*, 883–906. https://doi.org/10.1017/S0272263118000177

Wass, S., Porayska-Pomsta, K., & Johnson, M. H. (2011). Training attentional control in infancy. *Current Biology, 21*, 1543–1547. https://doi.org/10.1016/j.cub.2011.08.004

Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How do different forms of glossing contribute to L2 vocabulary learning from reading?: A meta-regression analysis. *Studies in Second Language Acquisition, 42*, 411–438. https://doi.org/10.1017/S0272263119000688

Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning, 24*, 39–58. https://doi.org/10.1080/09588221.2010.523285

Appendix A. Background Information

Group	Gender	Mean Age SD	Course	Mean CAE Score SD 95% CI
Interactive glossing	female: 19 male: 4	23. 52 2.41	education, social science: n = 17 medicine, engineering, science: n = 3 humanities: n = 3	42.22 12.39 [38.86, 47.57]
Non-interactive glossing	female: 16 male: 6	23.18 3.23	education, social science: n = 15 medicine, engineering, science: n = 4 humanities: n = 3	46.14 11.34 [41.11, 51.16]

Appendix B. Target Pseudowords

Text 1		Text2	
Pseudowords	**Original words**	**Pseudowords**	**Original words**
Stragon	Bottom	Cabrons	Changes
Golands	Spouts	Fration	Absence
Phosens	Discoveries	Zenters	Clues
Klenear	Surface	Morbits	Descendants
Tralion	Seawater	Stovons	Conditions

Appendix C. Descriptive Statistics for Post-Reading Questionnaire Items

Text/Item	Group	N	Mean	SD	95% CI	
					Lower	Upper
Text 1, Item 1	Interactive	23	3.65	2.06	2.76	4.54
	Non-interactive	22	3.31	1.73	2.55	4.08
Text 1, Item 2	Interactive	23	3.52	1.83	2.73	4.31
	Non-interactive	22	3.18	1.56	2.49	3.87
Text 2, Item 1	Interactive	23	3.30	1.43	2.69	3.92
	Non-interactive	22	3.27	1.45	2.63	3.92
Text 2, Item 2	Interactive	23	3.04	1.30	2.48	3.60
	Non-interactive	22	2.86	1.42	2.23	3.50

Appendix D. Results of Independent-samples T-tests for Post-Reading Questionnaire Items

Text/Item	t	p	D
Text 1, Item 1	.59	.56	.18
Text 1, Item 2	.67	.51	.20
Text 2, Item 1	.07	.94	.02
Text 2, Item 2	.44	.66	.13

Appendix E. Results for the Model Examining the Effect of Interactivity on Reading Comprehension Scores

	Fixed effects					Random effects	
Predictors	Est	SE	Z	p	R^2	Factor	SD
Intercept	1.01	0.29	3.52	<.01		Participant	0.70
Interactivity	0.41	0.26	1.56	.12	.01	Item	1.07

Appendix F. Results for the Models Examining the Effects of Interactivity on the Relationship between Attention and Form and Meaning Recognition

		Fixed effects						Random effects	
Dependent variable	**Predictors**	Est	SE	Z	p	R^2m	R^2c	Factor	SD
Target pseudowords									
Form	Intercept	0.55	0.19	2.88	<.01			Participant	0.09
Recognition	Interactivity	0.44	0.30	1.47	.14	<.01	<-.01	Item (Intercept)	0.03
	FC	<-0.01	0.06	-0.07	.94	<.01	<.01	Item (Interactivity)	0.17
	Interactivity*FC	-0.04	0.08	-0.51	.61	<.01	<.01		
Meaning	Intercept	-0.03	0.21	-0.16	.88			Participant	<0.01
Recognition	Interactivity	0.51	0.33	1.53	.13	<.01	<-.01	Item (Intercept)	0.32
	FC	0.02	0.06	0.29	.77	<.01	.01	Item (Interactivity)	0.50
	Interactivity*FC	-0.13	0.08	-1.55	.12	<.01	<.01		
Form	Intercept	0.49	0.18	2.64	<.01			Participant	0.12
Recognition	Interactivity	0.35	0.28	1.24	.22	<.01	<-.01	Item (Intercept)	0.04
	FD	<0.01	<0.01	0.49	.62	<.01	<-.01	Item (Interactivity)	0.16
	Interactivity*FD	<-0.01	<0.01	-0.35	.73	<.01	<.01		
Meaning	Intercept	<-0.01	0.21	-0.17	.87			Participant	0.12
Recognition	Interactivity	0.03	0.31	1.08	.28	.01	<-.01	Item (Intercept)	0.33
	FD	<0.01	<0.01	0.34	.73	<.01	<.01	Item (Interactivity)	0.50
	Interactivity*FD	<-0.01	<0.01	-0.91	.37	.01	.01		
Glosses									
Form	Intercept	0.74	0.20	3.68	<.01			Participant	0.07
Recognition	Interactivity	-0.11	0.32	-0.33	.75	.01	<-.01	Item (Intercept)	0.10
	FC	-0.19	0.12	-1.58	.12	<.01	<.01	Item (Interactivity)	0.29
	Interactivity*FC	0.32	0.16	2.00	.05	.01	.01		
Meaning	Intercept	-0.09	0.22	-0.39	.70			Participant	0.16
Recognition	Interactivity	0.28	0.33	0.84	.40	<.01	.02	Item (Intercept)	0.33
	FC	0.09	0.12	0.71	.48	<.01	<.01	Item (Interactivity)	0.49
	Interactivity*FC	-0.11	0.15	-0.73	.50	<.01	<.01		
Form	Intercept	0.47	0.18	2.63	<.01			Participant	0.10
Recognition	Interactivity	0.21	0.27	0.77	.44	<.01	<-.01	Item (Intercept)	0.05
	FD	<0.01	<0.01	0.70	.49	.01	.01	Item (Interactivity)	0.18
	Interactivity*FD	<0.01	<0.01	0.14	.89	<.01	<.01		
Meaning	Intercept	-0.09	0.17	-0.54	.59			Participant	0.12
Recognition	Interactivity	0.31	0.24	1.29	.20	<.01	.02	Item	<0.01
	FD	<0.01	<0.01	1.04	.30	<.01	<.01		
	Interactivity*FD	<-0.01	<0.01	-1.28	.20	<.01	<.01		

Note. form/meaning recognition response ~ 1 + interactivity * fixation count/duration + (1|participant) + (1|item)/(1+interactivity| item)

About the Authors

Andrea Révész is Professor of Second Language Acquisition at University College London. Her research interests lie in second language acquisition (SLA) and instruction, with particular emphases on the roles of tasks, interaction, input, and individual differences in SLA and the cognitive processes involved in second language performance and learning.

E-mail: a.revesz@ucl.ac.uk

Matt Stainer is Lecturer in the School of Applied Psychology at Griffith University in Australia. He completed a PhD in Psychology in the Active Vision Lab at the University of Dundee. His work specialises in how we use eye movements to extract information from our environment.

E-mail: m.stainer@griffith.edu.au

Jookyoung Jung is Assistant Professor in the Department of English at the Chinese University of Hong Kong. Her research interests include task-based language teaching, second language reading and writing, technology-mediated L2 learning, and the role of individual differences in second language acquisition.

E-mail: jookyoungjung@cuhk.edu.hk

Minjin Lee received her PhD in Applied Linguistics from University College London and is a lecturer at Yonsei University, South Korea. Her research interests include second language acquisition and task-based language teaching, with particular emphasis on understanding cognitive processes that underlie second language learning.

E-mail: mlee.minjin@gmail.com

Marije Michel (PhD Applied Linguistics, University of Amsterdam) is Professor of Language Learning at Groningen University in the Netherlands. Her research and teaching focus on second language acquisition and processing with specific attention to task-based language pedagogy, digitally-mediated interaction, and writing in a second language.

E-mail: m.c.michel@rug.nl