READING ACTIVITY CHARACTERISTICS, L2 READING PROCESSES AND NOTICING

The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions

Jookyoung Jung & Andrea Révész

Abstract

This study examined the extent to which manipulating the characteristics of second language reading activities affects the reading process and noticing of glossed linguistic constructions. Thirty-eight Korean learners of English read two texts under conditions that required more and/or less careful reading. For the condition intended to promote more careful reading, each paragraph of the texts was divided into three or four subparts. For the condition expected to elicit less careful reading, each paragraph was split into two sections. While reading the texts, the participants' eye-movements were recorded. Eleven students were further invited to participate in stimulated recall protocols. The target constructions were English unaccusative verbs and ten pseudowords, which were glossed with Korean translations. The eye-movement and stimulated recall data indicated that, as predicted, the participants processed the texts more carefully and attended to the target verbs more closely when paragraphs were divided into more subparts.

Introduction

Reading is not only an important comprehension skill that most second language (L2) learners strive to develop but also a major source of comprehensible input for L2 acquisition to occur. Hence, L2 reading activities are often designed with the dual aim of promoting development in comprehension ability and fostering the acquisition of L2 knowledge. In order to achieve these goals, it is necessary to explore and identify factors that affect the L2 reading process and the learning that accrues from engaging in reading activities. The characteristics of the reading activity is one variable that is likely to influence the nature of text processing as well as acquisition resulting from reading. While a few studies exist that investigate the influence of types of reading activity on text processing (e.g., Horiba, 2000, 2013; Kaakinen & Hyönä, 2005), little research has explored how the characteristics of reading activities may simultaneously affect the L2 reading process and noticing of L2 features. Given that reading activities are indispensible components of L2 instruction not only as means of developing reading skills but also as carriers of input, it is vital to begin filling this gap in the literature.

Against this background, the aim of this study was twofold. First, we intended to investigate how manipulating the characteristics of a reading activity, while keeping textual input constant, may influence reading processes. In particular, we aimed to explore whether activities designed to elicit more or less careful reading would indeed result in differential reading processes. Second, our goal was to examine how the extent of careful reading required by an activity may affect noticing and hence the possibility of learning L2 lexical and grammatical constructions from glosses. Glosses were operationalised as translations of

linguistic items in the margin of the reading input. In broader terms, we hoped to explore the extent to which reading activity characteristics can facilitate opportunities for learning to read and reading to learn (Han & D'Angelo, 2009), similar to how manipulating the cognitive demands of productive communicative activities have been observed to enhance opportunities for developing new (e.g., Révész, 2009) and practising existing (e.g., Plonsky & Kim, 2016) L2 knowledge and skills. The methodological novelty of our research lay in triangulating eye-tracking with stimulated recall data, a combination which is also just beginning to be utilized in the wider field of L2 research (Brunfaut & McCray, 2015; Smith, 2012).

Literature review

Reading activity characteristics and L2 reading processes

Khalifa and Weir's (2009) cognitive model for reading comprehension was considered as an ideal theoretical basis for the present study, given that this framework views reading as a cognitive process that constantly reacts to the reader's goal, which, in turn, is expected to be influenced by the characteristics of the activity in which the reader is engaged. Khalifa and Weir (2009) presuppose three knowledge sources: the knowledge base, the central core, and metacognitive activity. The knowledge base subsumes the reader's general world knowledge, topic knowledge, text-related knowledge, and linguistic knowledge. The central core entails a hierarchical system of lower-level and higher-level reading processes. The lower-level processes include word recognition, lexical access, syntactic parsing, and establishing propositional meaning, whereas higher-order processes comprise inferencing, building a mental model, and creating a text-level and inter-textual representation. Metacognitive activity is concerned with setting goals, monitoring, and remediation.

Goal setting seems particularly relevant when considering the effects of reading activity characteristics on reading processes. As part of goal setting, the reader determines the type of reading required to perform an activity. This process could lead to *careful* or *expeditious* reading, taking place at either *local* or *global* level. Local reading entails extracting propositions at the clause or sentence level, whereas global comprehension involves understanding the text as a whole. The aim of careful reading is to process all the information in the text to achieve complete comprehension. By contrast, expeditious reading is quick and selective with the goal of identifying specific information, as in skimming or scanning. Thus, the goal-setter enables readers to call upon different reading strategies and skills in response to differential activity demands.

Only a handful of studies exist that have looked into the relationship between reading activity characteristics and L2 reading processes, when textual difficulty was kept constant. In Horiba (2000), the participants verbalized their thoughts while they read either freely as they normally would or for coherence by additionally paying attention to sentence relations. Activity type did not emerge as a predictor of comprehension, operationalised as content recall. However, more think-aloud comments referred to relational (i.e., relating textual information to background knowledge) and integrative processing (i.e., processing relations between sentences) when participants read for coherence. In another study by Horiba (2013), participants were assigned to one of three conditions: reading for understanding new expressions, visualizing situations, and evaluating the author's views. While reading comprehension was not found to vary across reading conditions, a follow-up think-aloud study revealed that lower-level processes were more prevalent when participants read for understanding new expressions, whereas higher-level operations were more frequent when they read for evaluation. These findings, overall, led Horiba to conclude that the effects of

reading conditions may materialize more clearly in reading processes than outcomes. This conclusion, however, needs further verification.

The link between reading activity characteristics and L2 reading processes have received more attention in language testing. A study by Brunfaut and McCray (2015) is of particular relevance here. This study, like the present research, examined the cognitive processes of test-takers by triangulating stimulated recall data with eye-tracking. The participants were 25 test-takers who performed four types of reading assessments (multiple-choice gap-filling, sentence re-ordering, banked gap-filling, and matching headings). Gap-fill items elicited more careful local reading and lower-level processing, while sentence-ordering and matching headings involved proportionately more careful global reading, higher-level processing, and some expeditious reading. As evinced in previous studies investigating L1 (e.g., Kaakinen & Hyönä, 2005) and L2 (e.g., Bax, 2013) reading processes, the authors found that the eye-movement metrics and stimulated recall protocols yielded converging but complimentary findings. The eye movement analyses provided more insights into lower-level processes, whereas the stimulated recalls generated more information about higher-level processes.

To sum up, previous research indicates that L2 reading processes may vary according to reading activity characteristics. So far, however, researchers have primarily focused on the effects of reading activity types; little research has looked into how manipulating the characteristics of the same reading activity type might influence L2 reading. Triangulating eye-tracking with verbal protocol analyses appears a suitable methodological approach to address this research gap.

Glossing and noticing of L2 constructions

The impact of reading activity characteristics on glossing has received even less attention than the link between reading activity manipulations and L2 reading. This also constitutes an important gap in the literature, since glossing (i.e., providing contextualised information about linguistic items in the form of definitions, synonyms or translations) is a way to infuse focus on form into reading instruction. As a means of focus on form, glosses can arguably induce noticing (Leow, 2009), defined by Schmidt (1990) as focal attention directed to linguistic elements accompanied with a low level of awareness. That is, glosses are assumed to have the capacity to trigger both attention and awareness, the two processes associated with noticing (Godfroid, Boers, & Housen, 2013). Input converted into noticed intake (Leow, 2015), in turn, can potentially be rehearsed in working memory and made available for further processing that may eventually lead to L2 development (Leow, 2015; Schmidt, 1990). Thus, reading activities incorporating glosses can create what are considered favorable conditions for encoding new L2 representations by combining a primary focus on meaning with timely opportunities for noticing and processing of L2 form-meaning connections (Long & Robinson, 1998).

Given that noticing, by definition, entails focal attention allocated to linguistic information (Schmidt, 1990), it would appear that glossed linguistic information may be more prone to noticing when a reading activity, as in the present research, necessitates more careful reading. The more intensive and attentive processing of the texts will probably direct learners' attention to the glosses and target linguistic items more frequently. Also, the more accurate understanding needed is expected to lead to more in-depth processing of the targeted constructions. In a similar vein, Robinson (2001) suggested that more cognitively demanding tasks along certain dimensions would lead learners to seek more help from the input, resulting in deeper processing of input made salient through focus on form.

Although no research has examined the impact of manipulating the characteristics of reading activities on learner noticing of glosses, previous research has shown that, overall, glossed texts have positive, though small, effects on L2 development (e.g., Hulstijn, Hollander, & Greidanus, 1996; Watanabe, 1997). A small number of studies (Bowles, 2004; Guidi, 2009; Martinez-Fernández, 2010) also investigated the impact of glossing on the noticing of L2 constructions using think-aloud protocols. Bowles (2004) examined the extent to which computer-based versus paper-and-pen L1-glosses promoted awareness of glossed lexical items as compared to lack of exposure to glosses. The think-aloud comments revealed that the target vocabulary items were noticed more in the glossed conditions, but only negligible difference was observed between the computer and paper-and-pen groups. In Guidi (2009), participants read either unglossed or L1-glossed texts, and glosses were provided to target lexical items and two grammatical constructions. No difference was found in the amount of reported awareness between the gloss and no gloss groups. In Martinez-Fernández (2010), participants were assigned to an L1 translation gloss, L1 translation fill-in-the blank gloss, or no gloss condition. The target features were lexical items and a grammatical feature. As in Bowles (2004), participants reported awareness of more target lexical items in the gloss groups, but noticing was not affected by glossing type. The think-aloud protocols, however, revealed no effects of glossing for the noticing of the grammatical construction, similar to Guidi's findings.

Given the small amount of research available and conflicting findings, more research is needed to explore the extent to which glosses can facilitate noticing during reading. As mentioned above, further research is also warranted to examine how the characteristics of the reading activity may influence this link. Subsequent research would particularly benefit from utilizing verbal protocol data together with eye-tracking. While verbal protocols are suitable for tapping level of awareness (Gass & Mackey, 2017), eye-tracking is presumed to be a

measure of attention (Reichle, Pollatsek, & Rayner, 2006). This combination, therefore, would allow for gauging both the quality (level of awareness) and quantity of attention (Godfroid et al., 2013) triggered by glosses.

Research questions

1. To what extent does manipulating the characteristic of an L2 reading activity, while keeping textual input constant, affect reading processes, as reflected in participants' eye-movements and stimulated recall comments?
2. To what extent does manipulating the characteristic of an L2 reading activity, while keeping textual input constant, affect the noticing of glossed linguistic constructions, as reflected in participants' eye-movements and stimulated recall comments?

The reading activity manipulation in the present study involved creating two versions of a text-ordering activity, which were designed to require more or less careful reading.

Methodology

Design

Thirty-eight L2 users of English participated in the study. They completed two versions of a reading activity (henceforth, Text 1 and Text 2). Two types of target constructions, English unaccusative verbs and pseudo lexical items, were glossed for both Text 1 and Text 2. Following a 2x2 repeated-measures design, participants were exposed to the two texts under a more careful and/or less careful reading condition (see Table 1).

INSERT TABLE 1 HERE

Text order was counterbalanced across participants (see Figure 1). All participants took part in one session. They first completed a background questionnaire. Then, a pretest (a grammaticality judgment test, GJT) was administered, followed by a proficiency test. While participants were carrying out the reading activities, their eye-movements were recorded. Each reading was immediately followed by a short post-reading questionnaire. Eleven participants were further invited to partake in a stimulated recall session right after completing the two reading activities and questionnaires. These students were randomly selected from among the participants who completed both activity versions. This allowed for comparing reading processes and noticing of glossed constructions under the two reading conditions. Finally, participants were administered an exit questionnaire.

INSERT FIGURE 1 HERE

Participants

The 38 participating students were native speakers of Korean, enrolled at a UK university. Thirty-two students were female, and the mean age was 27.84 (*SD* = 4.52). The average length of stay in an English-speaking country was 9.92 months (*SD* = 3.84). The participants had IELTS scores 6.5 (borderline CEFR B2/C1) or higher. To ensure homogeneity of proficiency among participants across the text and activity combinations, all students were administered an adapted version of the *Use of English* section of a practice *Cambridge Proficiency English* (CPE) test. Cronbach's alpha for the CPE scores was .82.

Target constructions

We selected English unaccusativity as the target grammatical construction because this feature poses persistent difficulty for Korean learners (Chung, 2014). Pseudowords were included to control for prior lexical knowledge.

English unaccusative verbs

Intransitive verbs can be classified into unergatives (e.g., Mary *danced*.) and unaccusatives (e.g., The snow *melted*.) (Perlmutter, 1978). An unergative verb assigns an agent role to its subject, where the agent/subject has a deliberate involvement in the event. The subject of an unaccusative verb lacks volitional control and performs a patient role. Researchers (e.g., Zobl, 1989) have found that even high proficiency learners tend to overpassivize unaccusatives (e.g., *My mother was died when I was just a baby* in Zobl, 1989). As Table 2 shows, 15 English unaccusative verbs were identified in the two treatment texts and selected as target constructions.

INSERT TABLE 2 HERE

Pseudowords

Both texts included five pseudowords (see Table 3). They were all nouns and appeared once in the texts. They substituted ten original lexical items, and followed English orthographic and morphological rules. When the original word was in the plural, the plural marker -s was retained. Each pseudoword consisted of seven letters, containing two syllables.

INSERT TABLE 3 HERE

Texts

The two texts were expository passages selected from past TOEFL tests. The texts were chosen based on (a) whether they contained sufficient unaccusative verbs and (b) whether they covered topics likely to be unfamiliar to the participants. Text 1 was about petroleum resources, and the topic of text 2 was the Cambrian period. The length of Text 1 and Text 2 were 682 and 699 words respectively. Average readability, calculated from various indices (Flesch-Kincaid grade level, Gunning-Fog score, Coleman-Liau index, SMOG index, Automated Readability) was 11.6 for Text 1 and 13.4 for Text 2. These values indicated that the texts required at least upper-intermediate proficiency and thus were considered appropriate for the participants, who had at least low advanced proficiency.

Reading activity manipulation

Both texts were divided into five segments. Each segment was presented on one page, following the original TOEFL format. The reading activities involved ordering parts of the segments (henceforth, text-ordering activity) and then answering multiple-choice comprehension questions (henceforth, reading comprehension test). The reading comprehension items were taken from the TOEFL tests, whereas the text-ordering component was added as part of the experimental manipulation.

Under the less careful reading condition, each text segment was split into two subparts, whereas, under the more careful reading condition, the segments were divided into three or four. The participants were asked to determine the correct order of the parts under both reading conditions. The participants were instructed to point and click the capital letter labeling each subpart in the order they considered correct (see Figure 2). We assumed that the version which required the re-ordering of more subparts would require more careful reading at both local and global level (Khalifa & Weir, 2009) due to the decreased clarity and coherence of text structure (Meyer & Ray, 2011).

The comprehension questions asked participants to identify factual information, make inferences, understand rhetorical purpose, recognize vocabulary meaning, simplify/paraphrase a sentence, or select main ideas of the text (Educational Testing Service, 2012). There were nine multiple-choice comprehension items for each text, with one or two questions following each segment. The maximum comprehension score was 10 points for each text. Each item was worth 1 point, except for the last item, for which the total score was 2 points. This item required completing a summary by selecting several responses from a group of multiple-choice options.

Double-spaced Courier font was used to present the texts. Each target item was underlined and a corresponding Korean translation was provided in a marginal gloss. The participants were given 25 minutes for completing the activities. Piloting revealed that this time was sufficient to carry out the activities but put some pressure on participants under the more careful reading condition.

Pretest

To measure the participants' prior knowledge of unaccusativity, an untimed GJT was used as a pretest. The test included 15 grammatical and 15 ungrammatical sentences for the target unaccusative verbs (e.g. *The sun was soon disappeared* vs. *The tension soon disappeared*), and another 15 grammatical and 15 ungrammatical sentences served as distracters. The participants were asked to make binary choices (correct versus incorrect). The maximum score was 30, and the test took approximately 7 minutes. The Cronbach's alpha for the GJT was .62. The relatively low reliability might have been due to the fact that participants showed little knowledge of unaccusativity, as reflected in their close to chance GJT performance on average (see Pretest results below).

Stimulated recall

After completing both reading activities, eleven students were asked to participate in a stimulated recall session prompted by their eye-gaze recordings. It was first explained to the participants in everyday language that the red circles and lines in the recordings indicated their eye-fixations and saccades respectively. They were also instructed to stop the recording at any time they wanted to verbalize what they were thinking while engaged in the original activity. The researchers also interrupted the recordings and prompted the participants to describe their thoughts during the performance of the reading activity on the few occasions when unusual or interesting eye-movements were observed (longer fixations, regressive eye-movements, or re-reading behaviours), but these behaviours were not commented on by the participants. The stimulated recall sessions were video-recorded to capture participants' spatial movements as well. Piloting revealed that participants often pointed at the computer monitor during the protocols (e.g., *I started here, like this (pointing at screen), and it was very difficult*.) The interviews were carried out in Korean.

Questionnaires

The background questionnaire elicited information about the participants' demographics and English language learning experience. The post-reading questionnaire included two Likert-scale items gauging the participants' familiarity with the reading topics. The exit questionnaire asked the participants to provide comments about their experiences during the reading activity. All questionnaires were administered in Korean.

Procedure

Participants' eye-movements during the reading activities (i.e., text-ordering and answering reading comprehension questions) were captured with a mobile Tobii X2-30 eye-

tracker. Participants carried out the activities individually in a quiet room at the researchers' institution while one of the researchers sat at a discrete distance to avoid any disruption caused by her presence. The sessions took approximately two hours for the non-stimulated recall and three hours for the stimulated recall students. To decrease participant fatigue, students were offered a break at several points in the experiment (after completing the pretest, proficiency test, and post-reading questionnaires).

Data Analyses

Eye-movement data

The eye-tracking data were analysed with Tobii Studio 3.0.9 and the R statistical package. To assess the effects of reading condition on L2 reading processes (RQ1), two types of areas of interest (AOIs) were defined: (a) the text and (b) the text and response options combined (see Figure 2). AOIs for the texts were used in extracting indices associated with text reading processes, reflecting participants' comprehension processes. AOIs for the text and response options combined served as the basis for calculating measures of global reading processes, that is, they were presumed to shed light on how participants coped with the activity as a whole. Then, inspired by Brunfaut and McCray (2015), ten indices of text and global processing were calculated based on the eye fixation and saccade data with a series of R-scripts (*http://rpubs.com/GarethMcCray/reading-metrics*). The measures are summarised in Table 4. For each index, we expected greater values under the more careful reading condition, as ordering of the text subparts would likely require more intensive and recursive text processing. The only exception was median forward saccade length. For this measure, we predicted smaller values in the activity requiring more careful reading, as textual processing would probably be more frequently interrupted.

Next, to examine if the reading activity manipulation affected participants' processing of glossed constructions (RQ2), AOIs were defined for each target feature and gloss. While the target areas were identical in pixel size for pseudowords and glosses, those for unaccusative verbs were inevitably dissimilar due to the different verb lengths. This did not confound the results, as both versions of the activity included the same AOIs. Eight eye-tracking measures were extracted for the target constructions and glosses (see Table 4). We hypothesised that, under the more careful reading condition, the values for each measure would be higher, as more thorough text processing would result in more attention to the target words and their glosses.

INSERT FIGURE 2 AND TABLE 4 AROUND HERE

Statistical analyses

Prior to any statistical procedures, we performed a power analysis for all tests using GPower 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007). The sample size was found to be adequate to detect medium effect sizes for all factors of interest with an $\alpha = .05$ and power = .90. The Statistical Package for the Social Sciences (SPSS) 22 was used to compute reliability and descriptive statistics. The rest of the analyses were conducted with version 3.3.0 of the R statistical package, by constructing mixed-effects models using the *lmer* function of the *lme4* package. We first established that, across the two versions of the activity, participants were equivalent in English proficiency and pretest scores, and that topic familiarity had no confounding influence on reading comprehension. It was also confirmed that there were significant differences in perceived mental effort between the two activity versions.

Then, the research questions were addressed, with each mixed effects model including an eye-tracking measure as the dependent variable. In all models, Condition (i.e., reading activity version), Text and their Interaction were the fixed effects (i.e., independent variables), and Subject was the source of random effects. The modelling started by constructing null models that contained only a random intercept for Subject. Next, Condition, Text and the Interaction were added, and it was tested whether their addition improved the fit of the null models. In this step, likelihood ratio tests were conducted using χ^2 statistics. If a significant fixed effect was identified, a maximal random effects structure was produced to examine the magnitude and direction of the fixed effect(s) on the eye-movement measure. Given that maximal random structures can be overly complex with multiple random slope parameters, models often fail to converge. If this was the case, the random effect parameters accounting for the least variance were removed one by one until convergence was achieved. As participants were assigned to the reading conditions following a 2x2 repeated measures design, when Condition emerged as a significant factor in the likelihood ratio tests, multi-level mixed-effects models were developed including a within-subject random slope for Condition. A within-subject random slope for Text was not included in the models, as participants produced only one value, i.e., either an eye-movement index or time taken to complete the activity, for each text.

An absolute *t*-value above 2.0 was the criterion for significance. Effect sizes were computed with the *r.squared GLMM* function from the *MuMln* package. R^2 values above .06, .16 and .36 were considered as small, medium and large, respectively (Plonsky & Oswald, 2014). Collinearity statistics for the fixed effects (Condition and Text) were calculated using the *collin.fnc* function in the *languageR* package. Following Baayen (2008), condition numbers between 0 and 6 were regarded as evidence for no collinearity, around 15 as medium collinearity, and 30 or above as potentially harmful collinearity.

Stimulated recalls

The stimulated recalls were transcribed using the video-transcription software F5, version 2.2. The transcripts were uploaded to NVivo 10.0.3 for qualitative analysis. The first author reviewed the transcripts and identified emergent categories by annotating the data (for the resulting coding scheme, see Table 9). Then, a randomly selected subset of the video-recordings (13.6%) was watched and coded by a second coder to verify coding reliability. Inter-coder agreement was 90 per cent with a kappa of .71, which was acceptable, SE = 1.02, 95% CI [- .98, 3.06]. Next, the comments were further categorized depending on whether they concerned the reading conditions, and frequency counts were calculated for each code under each condition.

Results

Preliminary analyses

English proficiency

The CPE scores were analysed to check if the participants, across reading conditions and texts, were homogeneous in their proficiency (see Table 5 for descriptive statistics). The likelihood ratio tests revealed that the null model was not improved by adding Condition ($\chi^2(1) = .01$, $p = .93$, $R^2 < .01$), Text ($\chi^2(1) = .01$, $p = .99$, $R^2 < .01$), or the Interaction ($\chi^2(1) < .01$, $p = .99$, $R^2 < .01$). In other words, there was no significant difference in proficiency among the participants depending on text or reading condition.

INSERT TABLE 5 HERE

Prior knowledge of unaccusatives

The GJT scores were analysed to test whether the participants had parallel prior knowledge of unaccusatives across the experimental conditions. The maximum score was, again, 30. Under the less careful condition, the mean score of the participants who read Text 1 and Text 2 were 15.74 (n = 19, 95% CI [14.84, 16.64], SD = 4.57) and 15.39 (n = 19, 95% CI [14.49, 16.29], SD = 4.36), respectively. For the version requiring less careful reading, the mean was 14.95 (n = 19, 95% CI [14.05, 15.85], SD = 3.52) for Text 1 and 15.53 (n = 19, 95% CI [14.63, 16.43], SD = 4.02) for Text 2. The likelihood ratio tests constructed with the pretest GJT scores indicated that adding Condition ($\chi^2(1) = .01$, $p = .93$, $R^2 < .01$), Text ($\chi^2(1) = .01$, $p = .99$, $R^2 < .01$), and Interaction ($\chi^2(1) < .01$, $p = .99$, $R^2 < .01$) did not improve the null models. That is, at the pretest, the participants did not differ in their ability to judge the grammaticality of unaccusative sentences across text and reading activity allocation.

Topic familiarity

To assess whether topic knowledge affected participants' comprehension, topic familiarity was measured using post-reading Likert scale items (Item 1: *I thought this topic of the reading was familiar.*; Item 2: *I had some background knowledge about the reading topic.*). The maximum value for each item was 7.00. The responses to the items correlated significantly, Text 1: $r(38) = .80$, $p < .01$, Text 2: $r(38) = .87$, $p < .01$, suggesting that the items assessed overlapping constructs. For participants who performed the version designed to elicit less careful reading, the mean value was 2.62 for Text 1 (n = 19, 95% CI [1.93, 3.31], SD = 1.54) and 1.87 for Text 2 (n = 19, 95% CI [1.50, 2.24], SD = .83), whereas the mean under the more careful reading condition was 2.05 for Text 1 (n = 19, 95% CI [1.60, 2.50], SD = 1.01) and 2.23 for Text 2 (n = 19, 95% CI [1.60, 2.86], SD = 1.39). Likelihood ratio tests found that adding Familiarity and its interactions with the fixed effects did not improve the null model, Familiarity ($\chi^2(1) = 1.56$, $p = .21$, $R^2 < .01$), Condition*Familiarity ($\chi^2(1)$

= .89, p = .34, R^2 < .01), Text*Familiarity ($\chi^2(1) = 1.47$, p = .22, R^2 < .01), and Familiarity*Condition *Text ($\chi^2(4) = 5.37$, p = .25, R^2 = .01). That is, topic familiarity did not affect reading comprehension across the two activity versions or texts.

Text-ordering performance and reading comprehension scores

Table 6 presents the descriptive statistics for participants' text-ordering activity performance and comprehension test scores. Likelihood ratio tests revealed that both Condition ($\chi^2(1) = 74.60$, p < .01, R^2 = .19) and Text ($\chi^2(1) = 12.56$, p < .01, R^2 = .03) had a significant influence on text-ordering performance, but not their Interaction ($\chi^2(1) = .23$, p = .63, R^2 = .02). The effect size was medium for Condition and small for Text. According to the maximal-structure models, text-ordering performance was significantly better for the less careful reading conditions (Estimate = .44, t = 8.85) and for Text 2 (Estimate = .20, t = 3.79). Another set of likelihood ratio tests indicated that Condition ($\chi^2(1) = 1.47$, p = .22, R^2 = .02), Text ($\chi^2(1)$ < .01, p = .94, R^2 = .02), or Interaction ($\chi^2(1) = .65$, p = .42, R^2 < .01) did not have an influence on the comprehension scores.

INSERT TABLE 6 HERE

Time taken for activity completion

The mean time under the more careful reading condition was 835.45 (n = 38, 95% CI [818.19, 819.47], SD = 203.93) and 757.43 (n = 38, 95% CI [818.19, 819.47], SD = 203.93) under the less careful reading condition. A likelihood ratio test yielded a significant difference between the time taken to complete the two versions of the activity ($\chi^2(1) = 9.67$, p < .01, R^2 = 06). The summary of a maximal-structure model demonstrated that it took

significantly longer to complete the versions intended to elicit more careful reading (Estimate = 102.52, t = 3.30).

Next, we analysed whether time on activity differed by text version. Participants on average took 774.04 seconds to complete the activities based on Text 1 (n = 38, 95% CI [773.40, 774.68], SD = 192.44) and 818.83 seconds based on Text 2 (n = 38, 95% CI [818.19, 819.47], SD = 203.93). A likelihood ratio test revealed that there was no significant difference between these times ($\chi^2(1) = 2.79$, $p = .09$, $R^2 = .01$).

Eye-movement data

Reading activity characteristics, eye-gaze behaviours, and reading processes

Table 7 presents the descriptive statistics for the eye-movement measures computed to investigate reading processes (RQ1). From a series of likelihood ratio tests, Condition emerged as a significant predictor for the following measures: number of fixations for texts and responses combined ($\chi^2(1) = 13.39$, $p < .01$, $R^2 = .12$), number of fixations for texts only ($\chi^2(1) = 29.46$, $p < .01$, $R^2 = .27$), sum of fixation durations for texts only ($\chi^2(1) = 16.91$, $p < .01$, $R^2 = .17$), number of forward saccades ($\chi^2(1) = 23.25$, $p < .01$, $R^2 = .21$), and number of regressions ($\chi^2(1) = 18.26$, $p < .01$, $R^2 = .16$). The interaction of Condition and Text also improved the null models for the sum of fixation durations for texts and responses combined ($\chi^2(1) = 13.24$, $p < .01$, $R^2 = .22$) and proportion of regressions ($\chi^2(1) = 4.17$, $p = .04$, $R^2 = .03$). No significant effects were observed for the rest of the measures.

INSERT TABLE 7 HERE

As shown in Table 8, post hoc multi-level mixed effects models confirmed that the versions designed to promote more careful reading generated more eye fixations for the

activity as a whole (text and responses combined) and the text segments only. Participants also fixated longer on the texts, and made significantly more forward saccades and regressive eye-movements under the more careful reading condition. That is, as intended, participants appeared to engage in more careful and recursive reading, as manifested in the longer overall eye-gaze duration and increased number of fixations on the texts, and more forward saccades and regressions. Turning to the interactions, the more careful reading condition led to an increase in the sum of fixation durations on both text and response sections in Text 2 but decreased the index in Text 1. Also, the condition created to elicit more careful reading resulted in greater proportion of regressions in Text 1, but lower in Text 2. The R^2 values for these relationships ranged from .12 to .22, indicating small to medium effect sizes. The only exception was a very small effect size (R^2 = .03) for the Interaction on the proportion of regressions.

INSERT TABLE 8 HERE

Reading activity characteristics, eye-gaze behaviours, and noticing

Table 9 presents the descriptive statistics for the eye-movement measures calculated to examine the effects of reading activity manipulation and text version on the noticing of glosses and glossed constructions. The likelihood ratio tests indicated that both Condition and Text significantly improved the model fit for the number of fixations (Condition: $\chi^2(1) = 20.44$, $p < .01$, $R^2 = .23$, Text: $\chi^2(1) = 7.88$, $p < .01$, $R^2 = .09$) and sum of fixation durations (Condition: $\chi^2(1) = 12.50$, $p < .01$, $R^2 = .15$, Text: $\chi^2(1) = 10.89$, $p < .01$, $R^2 = .12$) on the target unaccusative verbs. Interaction between Condition and Text, however, did not improve model fit (number of fixations: $\chi^2(1) = .97$, $p = .32$, $R^2 = .34$, sum fixation durations: $\chi^2(1) < .01$, $p = .93$, $R^2 = .28$). No significant effects were found for the rest of the indices.

INSERT TABLE 9 HERE

As Table 10 illustrates, post-hoc multi-level mixed-effects models confirmed Condition and Text as significant predictors of the number and sum of fixations on the target unaccusative verbs. The participants fixated significantly more often and longer on the target verbs when performing the versions constructed to promote more careful reading and when working on Text 2. The effect sizes were medium for both number of fixations (R^2 = .34) and sum of fixations (R^2 =.28).

INSERT TABLE 10 HERE

Stimulated recall protocols

Eight meta-codes emerged from the stimulated recall comments: *high difficulty, low difficulty, ability to concentrate on activity, comprehension, word-level cue, discourse-level cue, noticing target unaccusative verbs,* and *noticing target pseudowords*. Each meta-code was broken down into sub-codes (see Table 11 for examples). More annotations were counted for the more careful *(n* = 374) than the less careful reading conditions (*n* = 230) overall and for most individual codes. When participants described their performance under the more careful reading condition, they more frequently reported experiencing difficulty and feeling unconfident. Reference to certain reading strategies, such as careful reading, skimming, and searching for hints, were also more frequent among comments on the version of the activity, which was designed to elicit more careful reading. Likewise, participants reported relying on linguistic cues more often when describing their thoughts under the more careful reading condition: at the word level, the participants remembered utilising keywords,

pronouns and transitional words more frequently; at the discourse level, there were more comments indicating that the participants struggled to order the text segments. They also reported with greater frequency that they focused on the first and the final sentence of each text segment and coherence between sentences. Lastly, although few comments concerned the target items and the glosses, there were more comments referring to the unaccusative verbs and their glosses when participants recalled their thoughts during the activities created to facilitate more careful reading.

For a few codes, however, recalls for the less careful reading condition yielded more comments. For example, the participants remembered rereading the texts more frequently during this activity version. In addition, they reported focusing more on articles, first mention of words, and sentence connectives. Finally, among the few comments generated about the glosses, participants referred to noticing of glosses for pseudowords more often when recalling their performance on the activity constructed to promote less careful reading.

INSERT TABLE 11 HERE

Discussion

The present study investigated the effects of manipulating the characteristics of reading activities on L2 reading processes and noticing of glossed linguistic constructions. The reading activity involved ordering jumbled texts, in which paragraphs were split into two (less careful reading condition) versus three to four (more careful reading condition) segments.

Reading activity characteristics and L2 reading processes

Our first research question asked the extent to which manipulating the characteristics of L2 reading activities affected reading processes, as reflected in participants' eye-movements and stimulated recall comments. We hypothesised that the need to re-order more subparts would prompt more careful reading at both the local and global level (Khalifa & Weir, 2009), given the more intensive and attentive reading required to identify inter-sentential relations (Meyer & Ray, 2011). As predicted, both the eye-movement and stimulated recall data confirmed that the version designed to elicit more careful reading, as intended, generated more thorough and intensive text processing. When performing the versions constructed to elicit less careful reading, the participants fixated more frequently on the activity (text and response options combined) and more often and longer on the texts. Participants also produced a larger number of forward saccades and regressive eye-movements, indicating that they engaged in more attentive and recursive text processing. The effect sizes ranged from small to medium. The stimulated recall comments revealed that, under the more careful reading condition, participants more frequently employed certain reading strategies, such as skimming, careful reading and searching for hints. They also recalled more extensive use of lexical and discourse cues. That is, the eye-tracking and stimulated recall data seem to converge and confirm that, when carrying out the activity versions created to promote more careful reading, participants indeed processed the texts more carefully and intensively.

It is also important to point out, however, that for some of the eye-movement measures, no significant difference (median fixation duration, median forward saccade length, median regression length, and proportion of regressions) or an interaction effect (sum of fixation durations for text and response options combined and proportion of regressive movements) was found between the two reading conditions, contrary to our expectations. A possible explanation may lie in that, although the two versions of the activity led to a

differential amount or quantity of processing, they did not prompt reading processes that were qualitatively different in nature. Unlike frequency and sum of duration measures, medians of fixation, saccade and regression lengths are likely to capture qualitative differences in reading. For example, longer saccade lengths are probably more associated with global reading, since global reading (e.g., reading for gist) necessitates less detailed comprehension (Brunfaut & McCray, 2015). On the other hand, shorter saccades are more likely to reflect engagement in local reading (e.g., reading for detail), requiring more thorough text processing. The same reason might explain that no considerable activity effects emerged for regressive movements (although a significant interaction was identified for this index, the effect size was small). Proportion of regressive movements is also likely to vary when readers engage in qualitatively different processes. A gap-fill activity with a given set of words, for instance, would likely involve more regressive movements than the text-ordering activity here, as readers would probably revisit the list of words on a number of occasions while working on the activity. The stimulated recall data also suggest that our reading activity manipulation had primarily quantitative effects on reading processes: participants recalled using certain strategies with greater frequency under the more careful reading condition, but rarely mentioned qualitatively different strategies.

Let us now attempt to explain the interaction identified for sum of fixation durations for text and responses combined, that is, why participants would fixate shorter overall under the more careful reading condition for Text 1, while the pattern was in the expected direction for Text 2 with longer overall fixation duration in the more careful reading condition. A possible clue lies in that participants achieved considerably lower mean scores on the text-ordering activity under the more careful reading condition for Text 1 (Mean =1.00) than Text 2 (Mean =1.95), suggesting that increasing the need to engage in careful reading resulted in proportionately greater demands for Text 1. This might have left less attention for answering

the comprehension questions, which, in turn, might have led to shorter fixations on the Text 1 comprehension questions (but not the text itself). This account is consistent with the fact that, as discussed above, sum of fixation durations for text only were, just as for Text 2, higher for the more careful reading condition for Text 1. Another possible explanation is that the more careful reading condition for Text 1, which appeared to be even more demanding than that for Text 2, encouraged participants to engage in the text more thoroughly and repeatedly, resulting in quicker completion of the comprehension items.

Finally, it is worth noting that the eye-gaze data yielded largely overlapping patterns for Text 1 and Text 2, only one measure yielded an interaction between text version and activity version. This suggests that the effects of the reading activity manipulation observed here might generalize to other academic expository texts. Clearly, future research is needed to confirm this and to test whether our results would transfer to other genres.

Reading activity manipulation and noticing glossed linguistic constructions

Our second research question investigated the extent to which manipulating the characteristics of reading activities may influence the noticing of glossed linguistic constructions, as reflected in participants' eye-movements and stimulated recall comments. We expected that participants would notice glossed information to a greater degree under the more careful reading conditions, as the more intensive and attentive text processing required would direct attention to the glosses and target linguistic items more frequently. Also, the more precise understanding needed would promote more in-depth processing of the target form-meaning relationships. As hypothesised, the eye-movement indices revealed that the target unaccusative verbs received considerably more attention in the more careful reading condition, evidenced in the significantly greater number of fixations and longer fixation durations on the target verbs. The effect size for these relationships was large. Similarly, the

stimulated recall data demonstrated that all of the comments related to the target verbs (n=3) concerned the more careful reading condition. It should be pointed out, however, that the number of comments on the target verbs was relatively small. Considering that the stimulated recall comments indicate a low level of awareness (Schmidt, 1990) or depth of processing (Leow, 2015), the findings appear to imply that the impact of the reading activity manipulation on learners' awareness appeared to be only marginal. Taken together, the more careful reading condition was more likely to trigger attention to and awareness of the unaccusative verbs, but its impact seemed more pronounced on attention than awareness.

Interestingly, however, the reading activity manipulation did not affect the overall amount of attention paid to the glosses associated with the target verbs. That is, the increased need to engage in careful reading, according to the eye-movement data, did not encourage learners to check the glosses with greater frequency or process them longer. In fact, verb glosses were often ignored; the average number of fixations to all verb glosses was below 4 for both texts regardless of activity version although Text 1 and Text 2 included 8 and 7 target verbs respectively. The participants also made few stimulated recall comments about the verb glosses, suggesting that they might have rarely been the focus of attention. This was probably because the unaccusative construction is of low communicative value. Therefore, if participants were familiar with the meaning of the root verb, they might have disregarded the glosses as the grammatical information in them was not essential to comprehension.

It was also contrary to our expectations that activity version had no impact on the noticing of pseudowords and their glosses, as indicated by a lack of a significant difference in the number and sum of eye-fixations at pseudowords and their glosses across the reading conditions. One reason for this may be that the processing of pseudowords was less essential to the completion of the activity than that of the unaccusatives. If the pseudowords had been

selected based on task-essentialness, the difference in reading conditions might have affected the extent to which they were attended to and processed.

Finally, like for RQ1, the eye-movement data generated similar patterns for Text 1 and Text 2, implying that the results might generalise to other academic expository texts.

Implications

On the theoretical front, the stimulated recall results corroborated hypotheses deduced from Khalifa and Weir's (2009) model of L2 reading, suggesting that, depending on reading activity characteristics, readers can call upon differential strategies and skills to achieve their goals. By theorising the relationships between reader goals, metacognitive activities, and reading processes, Khalifa and Weir's model would appear as a useful theoretical starting point for future work exploring links between reading activity characteristics and L2 reading. We also found some evidence for our hypothesis that increasing the need to engage in careful reading would generate greater attention and awareness of glossed linguistic constructions.

At the methodological level, we confirmed that combining eye-tracking with stimulated recall is a useful way to tap reading processes, enabling the investigation of both lower (e.g., saccades) and higher-order (e.g., strategies) reading operations. Triangulating eye-tracking and stimulated recall data is new to glossing research, and we found it helpful to assess both qualitative and quantitative aspects of attentional allocation (Godfroid et al., 2013), and thereby gain information about the effects of reading activity characteristics not only on attention or awareness but both processes.

Finally, the findings of this study suggest that increasing the need to carry out careful reading, although having limited impact on awareness, may result in more attention to target L2 features, as shown in the case of the target unaccusative verbs. That said, a potential pedagogical implication is that manipulating reading activities in such a way that they elicit

more careful reading can promote more attention to textually enhanced grammatical constructions, which may otherwise remain unattended. For example, a manipulation that induces learners to evaluate text coherence or structural organization to a greater extent would probably encourage more attentive text processing.

Limitations and future research

One limitation of this research lies in the use of the stimulated recall methodology. As stimulated recall involves *a posteriori* recollection of cognitive processes, it is possible that only a subset of the conscious processes during performance was reported. Another issue concerns the selection of target lexical items. We used two criteria: single occurrence in the text and being a noun. Additional factors that would ideally be considered in future research include concreteness, inferrability, and position in sentence. It would also be worthwhile to explore whether the activity manipulations would have a stronger impact on noticing if the processing of lexical items were made essential to the successful completion of the activity. In this study, the target lexical items appeared to have low functional load. A further weakness originates from the within-subjects design. Although this allowed controlling for individual differences to a greater extent, it made it impossible to investigate the combined effects of reading activity characteristics and glossing on L2 development. Adopting a between-subject pretest-posttest design could address this limitation. The study would also have benefited from using a more high-precision eye-tracker, especially as regards the noticing measures. Compared to the reading indices, the AOIs for noticing were relatively small, thus more prone to error. Finally, this study utilized a single activity type, one type of reading activity manipulation, an academic expository text, and low-advanced/advanced Korean L2 readers. Future research should examine whether our results would extend to different activity types, activity manipulations, genres, proficiency levels, and L1 speakers.

Conclusion

This study aimed to launch a new line of research into how manipulating the characteristics of L2 reading activities may affect L2 reading processes and the noticing of glossed constructions while reading. In general, we hoped to help fill a gap in instructed SLA research on how the characteristics of reading activities may influence opportunities to learn to read and read to learn (Han & D'Angelo, 2009). A methodological innovation of our research was the triangulation of eye-gaze with stimulated recall data. Reflecting a prediction we derived from Khalifa and Weir's (2009) reading model, we observed that manipulating a text-ordering activity resulted in more careful reading. As hypothesised, we also found that an increased need to engage in careful reading led to greater attention to and awareness of a glossed grammatical construction, with attention appearing to be more affected by the manipulation of reading activity characteristics than awareness. In contrast to what we expected, however, no relationship emerged between our manipulation and noticing of glossed pseudowords.

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing, 30*(4), 441-465.

Bowles, M. (2004). L2 glossing: To CALL or not CALL. *Hispania, 87*(3), 541-552.

Brunfaut, T., & McCray, G. (2015). Looking into test-takers' cognitive processes whilst completing reading tasks: a mixed-method eye-tracking and stimulated recall study. *ARAGs Research Reports Online,* AR/2015/001. London, UK: British Council.

Chung, T. (2014). Multiple factors in the L2 acquisition of English unaccusative verbs. *IRAL, 52*(1), 59-87.

Educational Testing Services (2012). *Official TOEFL iBT Tests*. McGraw-Hill: New York.

Faul, F., Erdfelder, E., Lang, A. –G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavioral Research Methods, 39,* 175-191.

Gass, S. M., & Mackey, A. (2017). *Stimulated recall methodology in second language research*. New York: 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.

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.

Han, Z. & D'Angelo, A. (2009). Balancing between comprehension and acquisition: Proposing a dual approach. In Z. Han & N.J Anderson (Eds.), *Second language reading research and instruction: Crossing the boundaries* (pp. 173-191). Ann Arbor, MI: The University of Michigan Press.

Horiba, Y. (2000). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes, 29,* 223-267.

Horiba, Y. (2013). Task-induced strategic processing in L2 text comprehension. *Reading in a Foreign Language, 25*(2), 98-125.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal, 80*(3), 327-339.

Kaakinen, J. K., & Hyönä, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eyetracking, and recall. Discourse Processes, 40, 239–257. doi:10.1207/s15326950dp4003_4.

Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language learning.* Cambridge: Cambridge University Press.

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). Ann Arbor, MI: The University of Michigan Press.

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. New York, NY: 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). New York: Cambridge 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 dissertation, Georgetown University, Washington, D.C.

McCray, G. (2016). *Analysing Tobii Studio extracted eye-tracking data in reading: Extracting reading measures from Tobii Studio eye-tracking data in R*. Retrieved from http://rpubs.com/GarethMcCray/reading-metrics.

Meyer, B. J. F., & Ray, M. N. (2011). Structure strategy interventions: Increasing reading comprehension of expository text. *International Electronic Journal of Elementary Education*, *4*(1), 127-152.

Perlmutter, D. M. (1978). Impersonal and the unaccusative hypothesis. In *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society* (pp. 157-190). Berkeley, CA: University of California, Berkeley Linguistics Society.

Plonsky, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics, 36*, 73-97.

Plonsky, L., & Oswald, F. L. (2015). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7*, 4 –22.

Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition, 31*, 437-470.

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition, 38*(4), 703-737.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*, 27–57.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Smith, B. (2012). Eye tracking as a measure of noticing: a study of explicit recasts in SCMC. *Language Learning & Technology, 16*(3), 53-81.

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition, 19*, 289-307.

Zobl, H. (1989). Canonical typological structures and ergativity in English L2 acquisition. In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language* (pp. 203–221). Cambridge University Press.

Figure 1. Procedure of the study

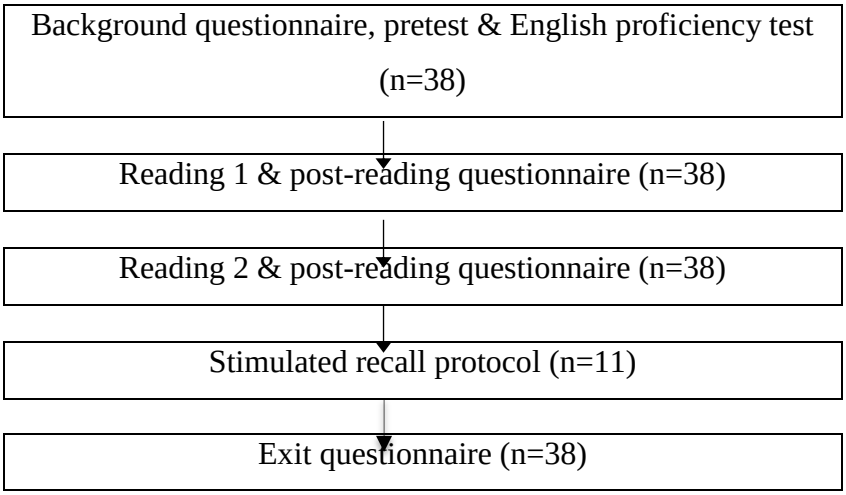

Figure 2. An example page with AOIs (enclosed in rectangles)

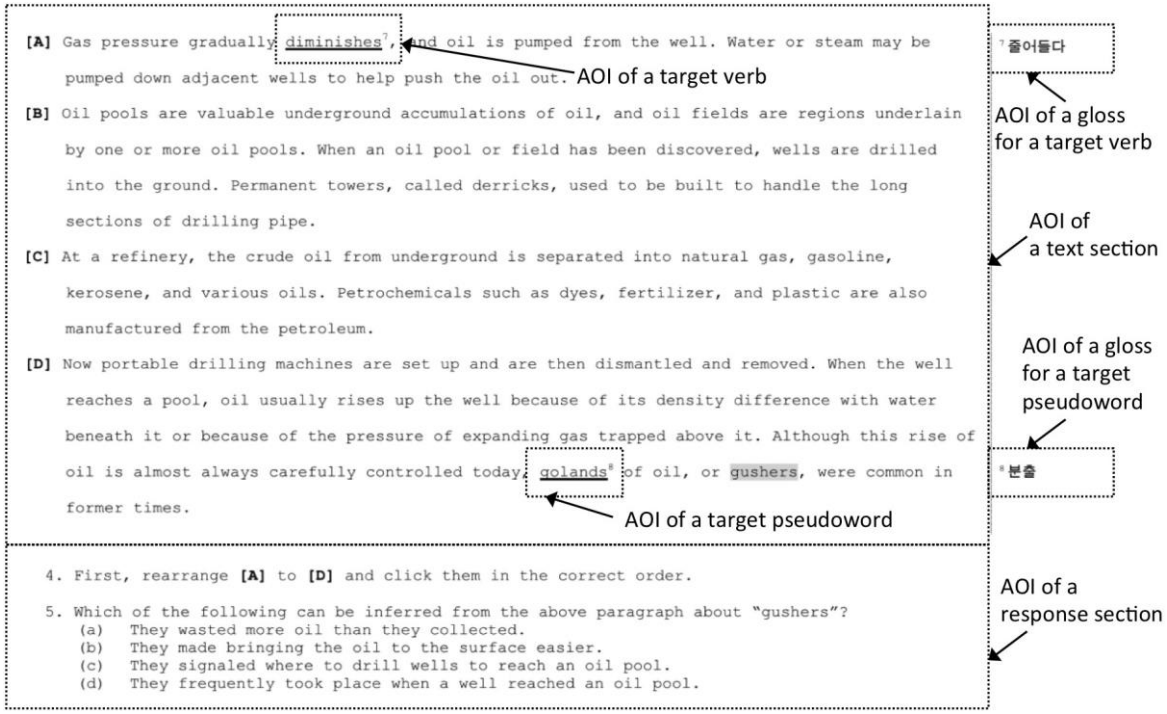
[A] Gas pressure gradually diminishes[7], and oil is pumped from the well. Water or steam may be pumped down adjacent wells to help push the oil out.
[B] Oil pools are valuable underground accumulations of oil, and oil fields are regions underlain by one or more oil pools. When an oil pool or field has been discovered, wells are drilled into the ground. Permanent towers, called derricks, used to be built to handle the long sections of drilling pipe.
[C] At a refinery, the crude oil from underground is separated into natural gas, gasoline, kerosene, and various oils. Petrochemicals such as dyes, fertilizer, and plastic are also manufactured from the petroleum.
[D] Now portable drilling machines are set up and are then dismantled and removed. When the well reaches a pool, oil usually rises up the well because of its density difference with water beneath it or because of the pressure of expanding gas trapped above it. Although this rise of oil is almost always carefully controlled today, golands[8] of oil, or gushers, were common in former times.
AOI of a target verb
AOI of a target pseudoword
[7] 줄어들다
[8] 분출
AOI of a gloss for a target verb
AOI of a text section
AOI of a gloss for a target pseudoword
4. First, rearrange [A] to [D] and click them in the correct order.
5. Which of the following can be inferred from the above paragraph about "gushers"?
(a) They wasted more oil than they collected.
(b) They made bringing the oil to the surface easier.
(c) They signaled where to drill wells to reach an oil pool.
(d) They frequently took place when a well reached an oil pool.
AOI of a response section

Table 1. Group assignment

A	**B**
Less careful reading, Text 1	Less careful reading, Text 2
More careful reading, Text 2	More careful reading, Text 1
C	**D**
More careful reading, Text 1	More careful reading, Text 2
Less careful reading, Text 2	Less careful reading, Text 1

Table 2. Target unaccusative verbs

Text 1	Text 2
subside	date to
ascend	originate
accumulate	consist of
cease	persist
diminish	evolve
drift	disappear
collect	emerge
settle	

Table 3. 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

Table 4. Eye-movement measures

Focus	Measure
Reading processes	
Global processing	Number of fixations on texts and responses combined
	Sum of fixation durations on texts and responses combined (ms)
Text reading	Number of fixations on texts
	Sum of fixation durations for texts (ms)
	Median fixation duration on texts (ms)
	Number of forward saccades
	Median forward saccade length (px)
	Number of regressions
	Median regression length (px)
	Proportion of regressive movements
Noticing	
Verbs	Number of fixations
	Sum of fixation durations (ms)
Verb glosses	Number of fixations
	Sum of fixation durations (ms)
Pseudowords	Number of fixations
	Sum of fixation durations (ms)
Pseudoword glosses	Number of fixations
	Sum of fixation durations (ms)

Table 5. Descriptive statistics for proficiency test

	Less careful reading (*n* = 19)		More careful reading (*n* = 19)	
	Mean	*SD*	Mean	*SD*
Text 1	11.40 [10.50, 12.30]	5.05	9.22 [8.32, 10.12]	4.81
Text 2	9.53 [8.63, 10.30]	4.65	11.00 [10.10, 11.90]	5.22
Total	10.40 [8.81, 11.99]	5.01	10.10 [8.48, 11.72]	5.11

Maximum score = 30. Each student's score is included twice due to the design.

Table 6. Descriptive statistics for text-ordering and reading comprehension scores

	Less careful reading (*n* = 19)		More careful reading (*n* = 19)	
	Mean	*SD*	Mean	*SD*
Text-ordering				
Text 1	3.11 [2.21, 4.01]	1.10	1.00 [0.10, 1.90]	1.11
Text 2	3.89 [2.99, 4.79]	1.05	1.95 [1.05, 2.85]	1.08
Total	3.50 [2.86, 4.14]	1.13	1.47 [0.83, 2.11]	1.18
Reading comprehension				
Text 1	5.47 [4.57, 6.37]	1.78	5.58 [4.68, 6.48]	1.84
Text 2	5.00 [4.10, 5.90]	2.24	6.42 [5.52, 7.32]	1.74
Total	5.24 [4.60, 5.88]	2.00	6.00 [5.36, 6.64]	1.82

The maximum score: text-ordering = 5.00; reading comprehension test = 10.00.

Table 7. Descriptive statistics for eye-movement measures of reading processes

	Global processing		Text-reading							
	Number of fixations	Sum of fixation durations	Number of fixations	Sum of fixation duration	Median fixation duration	Number of forward saccades	Median forward saccade length	Number of reg-ressions	Median reg-ression length	Proportion of reg-ressions
Less careful reading										
Text 1										
Mean	2836.74	739.13	1570.84	426.50	221.11	1024.53	96.42	403.53	-164.50	0.28
SD	564.66	176.81	387.51	108.56	28.82	265.82	9.47	127.02	42.14	0.04
95% CI Low	2589.65	659.50	1397.66	375.40	208.26	912.34	92.44	349.01	-182.22	0.26
95% CI Up	3102.04	817.25	1753.23	477.32	234.36	1149.32	100.82	463.99	-146.59	0.30
Text 2										
Mean	2893.58	775.73	1457.68	405.42	225.21	920.11	95.55	400.58	-163.37	0.30
SD	602.12	202.67	317.13	115.38	34.12	192.94	10.37	129.73	47.31	0.05
95% CI Low	2641.97	695.76	1317.84	359.33	210.76	829.82	91.14	345.35	-183.58	0.28
95% CI Up	3181.36	870.48	1603.83	463.83	240.55	1000.88	100.00	465.37	-143.11	0.32
More careful reading										
Text 1										
Mean	3120.95	597.09	1894.21	500.29	216.53	1152.74	97.71	501.74	-154.58	0.30
SD	588.73	151.41	379.76	121.05	40.40	271.31	14.42	154.87	49.24	0.05
95% CI Low	2874.44	532.49	1723.48	442.90	197.58	1036.15	91.77	435.69	-176.66	0.28
95% CI Up	3376.99	661.88	2049.78	549.46	232.89	1276.35	104.29	569.47	-132.88	0.32
Text 2										
Mean	3487.79	861.94	2112.37	542.36	211.32	1317.32	96.05	536.26	-167.39	0.29
SD	694.27	201.21	529.07	146.73	36.09	332.42	12.10	182.00	49.60	0.04
95% CI Low	3169.48	768.09	1878.34	469.44	195.26	1155.04	90.54	463.08	-191.52	0.27
95% CI Up	3788.66	942.47	2355.19	602.22	227.26	1457.02	101.85	622.14	-147.18	0.31

Table 8. Summary of mixed-effects models for eye-movement measures of reading processes

	Fixed effects			Random effects		Effect size
				by subject	by condition: subject	
	Estimate	*SE*	*t*	*SD*	*SD*	R^2
Number of fixations on texts and responses						.12
Intercept	3084.76	87.96	35.07°	449.10	.00	
Condition	463.57	119.99	3.86°	–	–	
Number of fixations on texts						.17
Intercept	1758.78	58.37	30.13°		.00	
Condition	496.96	81.37	6.11°	–	–	
Sum of fixation durations on texts						.14
Intercept	468642	17110	27.39°	83718	.00	
Condition	109063	24851	4.34°	–	–	
Number of forward saccades						.21
Intercept	1103.67	39.44	27.98°	205.00	.00	
Condition	279.29	52.09	5.36°	–	–	
Number of regressions						.16
Intercept	460.53	21.52	21.41°	114.13	.00	
Condition	126.73	27.20	4.66°	–	–	
Sum of fixation durations on texts and responses						.14
Intercept	743.47	28.00	26.55°	151.80	–	
Text	150.72	26.68	5.65°	–	–	
Sum of fixation durations on texts and responses						.22
Intercept	743.47	27.29	27.24°	151.70	–	
Condition*Text	150.72	62.44	3.76°	–	–	
Proportion of regressive movements						.03
Intercept	.29	.01	41.95°	.04	–	
Condition*Text	-.03	.01	-2.01°	–	–	

Significance: °| *t* | > 2.0.

Table 9. Descriptive statistics for eye-movement measures of noticing

	Verb		Verb gloss		Pseudoword		Pseudoword gloss	
	Number of fixations	Sum of fixation durations	Number of fixations	Sum of fixation durations	Number of fixations	Sum of fixation durations	Number of fixations	Sum of fixation durations
Less careful reading								
Text 1								
Mean	30.68	7.94	3.42	0.89	18.58	4.97	2.26	0.49
SD	12.21	2.85	2.61	0.75	8.44	2.36	2.64	0.63
95% CI Low	25.58	6.66	2.32	0.56	15.11	4.02	1.11	0.23
95% CI Up	36.31	9.18	4.53	1.21	22.58	6.06	3.42	0.78
Text 2								
Mean	38.05	11.45	3.37	0.87	20.26	5.25	2.79	0.62
SD	8.26	3.34	1.80	0.78	7.89	2.06	2.42	0.53
95% CI Low	34.58	10.04	2.58	0.58	17.00	4.41	1.74	0.39
95% CI Up	41.58	12.96	4.16	1.28	24.05	6.28	3.89	0.86
More careful reading								
Text 1								
Mean	43.95	11.79	3.63	0.70	20.89	5.40	2.63	0.51
SD	9.66	3.56	3.08	0.64	9.39	2.63	2.24	0.53
95% CI Low	39.90	10.24	2.37	0.45	17.11	4.31	1.69	0.29
95% CI Up	47.84	13.34	5.00	0.98	25.11	6.55	3.68	0.77
Text 2								
Mean	57.26	15.51	3.58	0.82	19.11	4.89	2.32	0.57
SD	21.82	6.93	3.78	1.16	8.11	2.56	3.23	0.95
95% CI Low	47.90	12.61	2.16	0.38	15.84	3.95	1.05	0.21
95% CI Up	67.47	18.54	5.31	1.34	22.74	6.01	3.84	1.02

Table 10. Summary of mixed-effects models for eye-movement measures of noticing

	Fixed effects			Random effects		Effect size
				by subject	by condition:subject	
	Estimate	*SE*	*t*	*SD*	*SD*	R^2
Number of fixations on target verb						.34
Intercept	42.49	1.77	24.01°	6.45	.00	
Condition	16.60	3.13	5.31°	–	–	
Text	10.34	2.86	3.62°	–	–	
Condition*Text	6.07	6.32	.96	–	–	
Sum of fixation duration on target verb						.28
Intercept	11.68	.57	20.58°	1.99	1.15	
Condition	4.11	1.01	4.08°	–	–	
Text	3.62	.90	4.04°	–	–	
Condition*Text	.18	1.97	.09	–	–	

Significance: °| *t* | > 2.0.

Table 11. Code frequency for stimulated recall comments (*n* = 604)

Meta-code/code	More careful reading (*n* = 374)	Less careful reading (*n* = 230)	Example
High difficulty	*57*	*13*	
Difficulty (High)	43	*9*	*It wasn't easy at all.*
Unconfident completion	14	*4*	*I wasn't sure about my text ordering.*
Low difficulty	*7*	*11*	
Difficulty (Low)	7	*8*	*It wasn't that difficult.*
Confident completion	0	*3*	*I was thinking that I understood the content well.*
Ability to concentrate on activity	*13*	*11*	
Concentration (Low)	11	*9*	*I could not concentrate well on the activity in the beginning.*
Concentration (High)	2	*2*	*I could concentrate better on the activity this time.*
Comprehension	*122*	*88*	
Overall comprehension	24	*25*	*I could not understand (A) when I first read it.*
Re-reading	20	*25*	*I tried to read this again.*
Careful reading	24	*15*	*I thought (B) came first, so I had to understand (B) perfectly before reading (A).*
Skimming	*22*	*12*	*I didn't read carefully, because I just wanted to see the overall structure.*
Searching for hints	26	*8*	*I was trying to find something that connects these text segments.*
Refer to previous passage	6	*3*	*I was thinking about the content of the previous passage.*
Word-level cues	*84*	*42*	
Keyword	40	*14*	*I thought "soft-bodied animal" was the keyword here.*
Signal word	18	*8*	*I assumed "finally" must indicate the last part of the text.*
Pronoun	14	*3*	*It wasn't the first, because it follows "these".*
Second mention	8	*5*	*I saw some repeated words. Repeated words were useful when deciding on order.*
First mention	2	*6*	*This was the first time "drilling" was mentioned.*
Article	2	*6*	*For instance, "a" became "the".*
Discourse-level cues	*78*	*56*	
Logical flow	24	*23*	*(B) gave a general statement, while (A) gave a concrete example.*

Wrestle to order segments	33	*11*	*I was debating about the order between these two segments.*
First sentence	17	*11*	*I thought focusing on the first sentence would be enough to decide on the order.*
Sentence connection	3	*9*	*I was checking if (A)-final and (B)-front, or (B)-final and (A)-front were connected.*
Final sentence	1	*2*	*If the sentences were connected, I thought there must be a clue in the final sentence.*
Noticing – Target unaccusative verbs	*9*	*4*	
Noticing glosses	6	*4*	*I could notice the glosses naturally, as they were in Korean.*
Noticing target verbs	3	*0*	*I thought "diminish" might be an important word here.*
Noticing – Target pseudowords	*4*	*5*	
Noticing glosses	2	*5*	*The gloss for "golands" helped me to learn that it had a different meaning from "gusher".*
Noticing target words	2	*0*	*It was my first time seeing this word.*