Promoting Grammatical Development Through Textually Enhanced Captions: An Eye-Tracking Study

Minjin Lee & Andrea Révész

<ABSTRACT>

This study launched an investigation into the extent to which textual enhancement in captions can promote learner attention to and subsequent development in second language (L2) grammar. Using eye-tracking, it also intended to extend research on the relationship between attention and L2 learning. A pretest–posttest experimental design was employed, with 3 treatment sessions. Forty-eight Korean learners of L2 English were randomly assigned into a captions group (n = 24) and an enhanced captions group (n = 24). For the enhanced captions group, the components of pronominal anaphoric reference were boldfaced in the treatment task input. Learner attention to anaphora antecedents and personal pronouns was assessed with eye-movement indices, and a written and an oral grammaticality judgment test were used to measure learning gains. Textual enhancement succeeded in directing learner attention to the anaphora antecedents, and led to increased gains in receptive knowledge of pronominal anaphoric reference. However, significant links between attention and L2 development were only observed for the unenhanced captions group. The findings, overall, demonstrate that textually enhanced captioning is a useful pedagogical tool to facilitate development in L2 grammatical knowledge.

Keywords: textual enhancement; captions; eye-tracking; grammatical knowledge; attention

<END ABSTRACT>

In the area of second language acquisition (SLA) research, exposure to comprehensible input is generally regarded as a prerequisite for second language (L2) development to occur. However, L2 learners do not normally process all that is available in the input (Corder, 1967). Only part of the input to which learners have access gets processed and subsequently learned, and attention is considered a principal cognitive mechanism determining what part of the input is selected for further processing (Robinson, 2003; Schmidt, 2001). Therefore, a key question in instructed SLA research and practice is how learner attention can be drawn to linguistic features during L2 learning activities. While this issue has been the subject of much research over the past two decades, L2 researchers have only more recently begun to explore ways in which learner attention can be directed to L2 constructions during activities entailing various modalities, such as audio and pictorial input. This interest is probably due to the fact that multimedia materials (e.g., podcasts, DVDs, and YouTube) are becoming increasingly available and used by L2 learners in both formal and informal L2 settings.

In the context of multimodal activities, a technique that seems to hold special promise is captioning, defined as adding "redundant text that matches spoken audio signals and appears in the same language as the target audio" (Vandergrift, 2007, p. 79). The role of captions in L2 performance and development has attracted much recent research, and, according to a meta-analysis (Montero Perez, Van Den Noortgate, & Desmet, 2013), captions are useful for promoting L2 listening comprehension and vocabulary knowledge. So far, however, little is known about what type of captions (e.g., full, key-word, textually enhanced) can facilitate L2 acquisition most effectively, and how different types of captions may affect development in L2 grammatical knowledge.

To address these gaps in the literature, the aim of this study was to compare the capacity of two types of captions, textually enhanced and non-enhanced, to foster L2 development. The novelty of our research was that we examined the effects of different types of captioning on the acquisition of grammatical knowledge rather than development in lexis or listening comprehension skills. With the help of eye-tracking methodology, we also intended to expand on existing research by assessing the extent to which attention paid to target grammatical constructions is related to L2 learning (e.g., Godfroid, Boers, & Housen, 2013), and begin to explore whether this link may be influenced by type of captioning. In the sections that follow, we describe theoretical and empirical work that provides the background for this research.

<A>INPUT, ATTENTION AND L2 DEVELOPMENT

Given that attention plays a primary role in mediating the process of selecting input for further processing (Robinson, 2003; Schmidt, 2001), researchers have shown a keen interest in determining the extent to which various instructional techniques can attract attention to target L2 constructions. One way to assess the amount of attention drawn to linguistic features in written input is through eye-tracking technology. The assumption underlying eye-tracking is that the location, length and sequence of eye movements are a close reflection of attentional processes, and thereby where and when eyes move can supply information about the nature, order, and timing of cognitive operations while individuals interact with a visual stimulus (Just & Carpenter, 1976). Using eye-tracking, researchers have by now accumulated some evidence that, indeed, those learners who pay more attention to target linguistic constructions display increased development in L2 vocabulary (Godfroid et al., 2013; Montero Perez, Peters, & Desmet, 2015; Pellicer–Sánchez, 2015; Williams & Morris, 2004) and grammatical knowledge (Godfroid & Uggen, 2013; Indrarathne & Kormos, 2017; see, however, Issa, Morgan–Short, Villegas, & Raney, 2015). Nevertheless, the role of attention

in L2 development remains a current concern in L2 research. Thus, one aim of the present study was to contribute to this line of research.

<A>CAPTIONING AND L2 DEVELOPMENT

Given the key roles attributed to input and attention in SLA and the increased reliance on multimedia materials in L2 teaching and learning, a key question in SLA research and practice is how learner attention can be drawn to linguistic features during multimodal input-based activities, that is, how focus on form (Long, 2000) can be achieved while learners engage in activities involving multiple modalities. Of the various focus-on-form options that can be integrated into multimedia materials, this study aimed to investigate the effects of typographical enhancement in captions on L2 grammatical development.

As noted earlier, the positive effects of captioning on L2 listening comprehension and vocabulary learning are now well established, with a recent meta-analysis yielding large effect sizes for captioning (Montero Perez et al., 2013). The observed benefits of supplying captions have been explained by the fact that captions can assist learners in segmenting speech into words (Bird & Williams, 2002; Vanderplank, 1988). Access to segmented speech, in turn, is likely to facilitate word recognition (Bird & Williams, 2002; Markham, 1999), which is a key determinant of successful L2 listening (Rost, 2011) and reading comprehension (Grabe, 2012). More successful word recognition also enables learners to identify new lexical items in the input with greater ease, promoting attention to and learning of new vocabulary (Winke, Gass, & Sydorenko, 2010).

It seems reasonable to assume that the availability of captions aids as well in drawing learner attention to grammatical constructions. Having less advanced processing skills, L2 listeners (Rost, 2011) and readers (Grabe, 2012) often need to draw on controlled, conscious rather than automatic processing when decoding aural and written input (Segalowitz, 2003). Due to the availability of captions, the demands on word recognition processes will decrease, which is likely to allow learners to pay more attention to morphosyntactic features and engage in more in-depth processing of grammar. Although the present study does not directly address the impact of providing captions on L2 grammatical development, it begins to explore the effects of different types of captioning on the acquisition of a grammatical feature, pronominal anaphoric reference.

<A>TYPES OF CAPTIONING, ATTENTION, AND L2 DEVELOPMENT

The roles of various types of captions in L2 listening performance and lexical development have already been the subject of a few investigations. As regards listening comprehension, the results of existing research are mixed, with some studies yielding an effect for type of captioning and others finding no difference among various captioning conditions. For example, Montero Perez et al. (2014) revealed no difference in video comprehension scores comparing four groups – no captions, full captions, keyword captions, and full captions with key words highlighted. Guillory (1998) and Yang and Chang (2014), on the other hand, found that caption type had a significant impact on listening scores. Guillory (1998) observed an advantage for key-word over full-text captions when the highlighted words were relevant to the comprehension questions. In Yang and Chang's (2014) study, full captions and keyword only captions proved less effective in promoting listening comprehension than annotated keyword captions, where pictorial symbols were added to increase the salience of reduced forms in the keywords (e.g., a blue dot was added above those letters of the keywords where the corresponding sound was assimilated in the aural text).

The few studies that have looked into the effects of different types of captions on vocabulary acquisition yielded differential gains depending on captioning conditions. In the previously mentioned study by Montero Perez et al. (2014), the researchers used four tests to measure lexical gains: form recognition, meaning recognition, meaning recall, and clip association (testing whether participants could associate words with corresponding video clips). The meaning recall test yielded no significant difference among the four groups, and all participants who watched video clips with captions, regardless of caption type, outperformed the control group in the form recognition and clip association tests. Type of captioning, however, had a significant impact on the meaning recognition test: the groups with keyword captioning and full captioning with highlighted keywords achieved greater gains than the control group.

In another study, Montero Perez et al. (2015) further investigated the usefulness of different types of captions in promoting vocabulary knowledge, this time not only examining the impact of captioning on vocabulary learning but also on the allocation of attention to target lexis. Type of captioning was operationalised as access to either keyword captions or full captions. Another independent variable in the study was the presence versus absence of test announcement, resulting in an incidental (no announcement) and an intentional (announcement) condition. The participants were assigned to four treatment groups: full captioned video and incidental, full captioned video and intentional, keyword captioned video and incidental, and keyword captioned video and intentional groups. Development in vocabulary knowledge was gauged with the same type of assessments as in Montero Perez et al. (2014): form recognition, meaning recognition, meaning recall, and clip association. Attention allocation to the target words was measured with three eye-tracking indices: gaze duration (i.e., sum of fixations before leaving the target word area), which captured initial processing (Rayner, 1998); second pass reading time (i.e., time spent rereading the target word area), reflecting re-analysis of information; and total fixation duration. The results revealed an advantage for keyword captioning in terms of gaze duration and performance on the form recognition test. The keyword-captions group also showed greater second pass reading times and total fixation duration under the intentional condition. However, significant links between attentional allocation and learning gains were only attested for the full-captions groups on the form recognition test: total fixation time and second pass reading time led to better scores when learners were made aware of the forthcoming test, whereas longer gaze durations were associated with greater gains when the vocabulary test was not announced. Interestingly, in the absence of test announcement, those who had higher second pass reading times displayed lower form recognition scores in the full captions group.

The two studies by Montero Perez and colleagues, overall, suggest that increasing the visual salience of target areas in captions has the capacity to capture learner attention to target lexis and generate vocabulary gains. Notably, however, only when captions were not visually enhanced did Montero Perez et al. (2015) find significant links between attention and L2 development in vocabulary knowledge. The researchers interpreted this finding as suggesting that more attention allocated to the target constructions evidenced in longer gazes may not indicate more elaborate processing. Clearly, more research is needed to explore these relationships. To extend this line of research, the present study examined the extent to which textual enhancement in captions benefits attention to and learning of grammatical constructions, and whether it moderates the relationship between attention and grammatical development.

<A>TEXTUAL ENHANCEMENT, ATTENTION, AND L2 GRAMMATICAL DEVELOPMENT

While the effects of textually enhanced captions on attention to and learning of grammar have not been studied to date, a few eye-tracking studies have already examined potential links between textual enhancement, attention, and L2 grammatical development in the context of unimodal language learning activities. In these studies, written input was enhanced via typographical cues such as the use of boldfacing (Indrarathne & Kormos, 2017; Simard & Foucambert, 2013), underlining (Simard & Foucambert, 2013; Winke, 2013), and different colours (Issa et al., 2015; Loewen & Inceoglu, 2016; Winke, 2013), in an attempt to promote the visual salience of the targeted grammatical constructions and thereby direct learner attention to the enhanced linguistic features. Textual enhancement led to increased attention, as captured in eye-tracking indices, in only two of the studies (Simard & Foucambert, 2013; Winke, 2013); and none of the studies reported benefits for the enhanced group in terms of learning. Indrarathne and Kormos (2017), however, found a positive relationship between amount of attention and gain scores for the enhanced group.

According to Leow and Martin (2017), these findings suggest that, when the readers' primary goal is to extract meaning from written input, textual enhancement alone does not have the capacity to trigger sufficient depth of processing for learning to occur. Leow and Martin also observe, however, that developmental benefits for textual enhancement were more likely to be found in studies when textual enhancement was used together with other attention-getting tools (e.g., feedback, explicit instruction), probably due to the higher level of processing achieved through the combination of techniques. The present study put this observation to a new test by investigating the impact of textual enhancement on grammatical development while keeping captioning, another pedagogical tool, constant.

<A>RESEARCH QUESTIONS

Against this background, we formed the following research questions with regard to our target linguistic construction, the use of pronominal anaphoric reference:

RQ1. To what extent does textually enhanced versus unenhanced captioning draw learners' attention to pronominal anaphoric reference?
RQ2. To what extent does textually enhanced versus unenhanced captioning affect L2 development in the knowledge of pronominal anaphoric reference, as measured by a written and an oral grammaticality judgment test (GJT)?
RQ3. To what extent is attention to pronominal anaphoric reference in captions related to L2 development? Is this relationship influenced by textual enhancement?

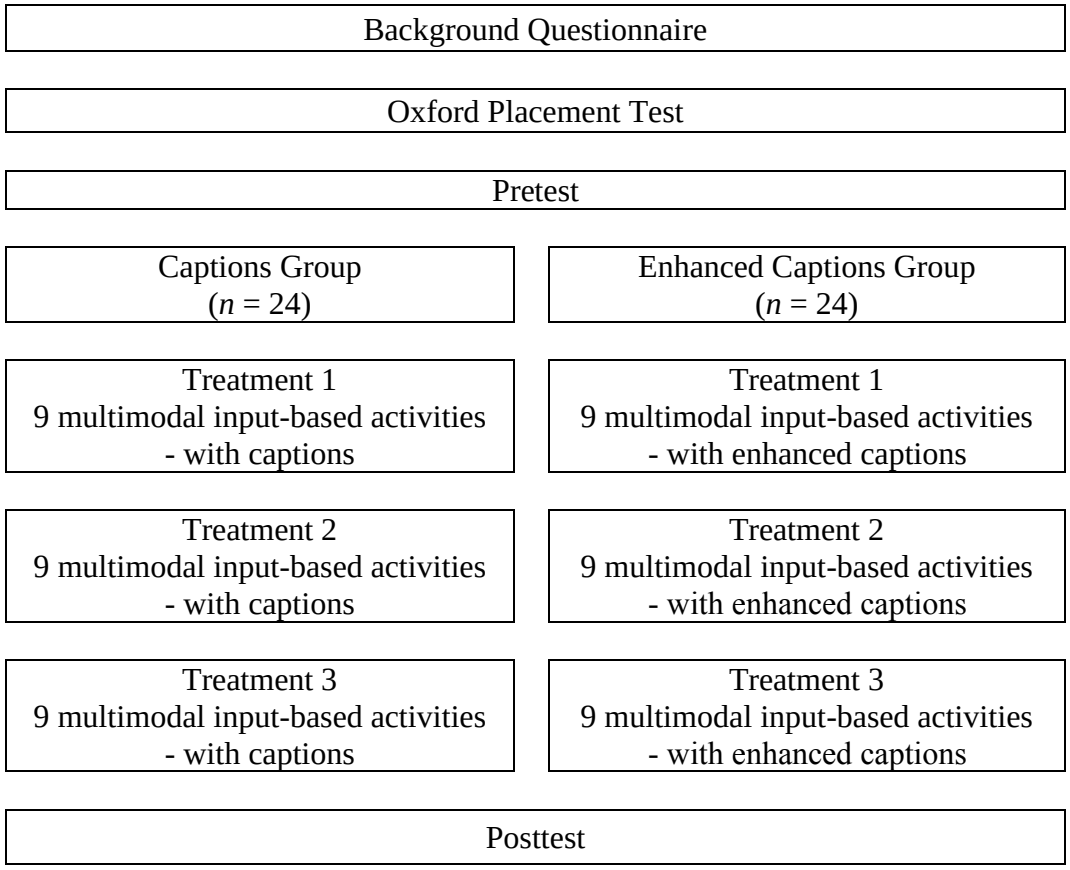
<A>METHOD

<B>Design

As shown in Figure 1, the study employed a pretest–posttest experimental design. The participants were 48 English as a Foreign Language (EFL) students, who were randomly assigned to two comparison groups: a captions group (n = 24) and an enhanced captions group (n = 24). First, all participants were administered a background questionnaire, the Oxford Placement Test (OPT), and a pretest. Then, participants engaged in 3 treatment sessions, each involving the completion of 9 multimodal input-based activities. The format of the multimodal input-based activities was the same for the two groups. The groups, however, differed as to whether they completed activities with regular captions or captions with textually enhanced input. While participants worked on the treatment tasks, their eye

movements were recorded using a Tobii X2-30 mobile eye-tracker. Finally, a posttest was administered. The pretest and the posttest each included a written and an oral grammaticality judgment task (GJT).

<INSERT FIGURE 1 ABOUT HERE>
FIGURE 1
The Experimental Design



<B>Participants

All the 48 participants were Korean undergraduate students studying English as a foreign language in Seoul, South Korea. There were 31 female and 17 male students, with an age range of 19 to 25 (M = 22.53, SD = 1.89). The students had received at least 9 years of English instruction prior to the study, given that English is taught from grade three of elementary school throughout high school in South Korea. At the time of the study, students took 2 to 3 English classes a week, each class lasting approximately 40 to 45 minutes (the target construction was not the focus of instruction during the study period). The classes mainly focused on developing reading skills, along with instruction targeting vocabulary and grammar. Only students who had not lived abroad were selected.

The participants' proficiency level fell into the B1–B2 bands according to the Common European Framework for Reference (CEFR), as determined by the Oxford Placement Test. The captions and enhanced captions groups achieved comparable scores on both the listening (captions group: M = 73.88, SD = 4.49; enhanced captions group: M = 74.21, SD = 5.07) and grammar (captions group: M = 64.08, SD = 6.91; enhanced captions group: M = 63.38, SD =

7.44) sections of the OPT. Independent samples *t*-tests confirmed that there was no significant difference between the two groups on either the listening, $t(46) = -.24$, $p = .81$, $d = .07$, or grammar, $t(46) = .34$, $p = .73$, $d = .10$, component of the OPT.

<B>Target Linguistic Construction

Anaphoric reference was chosen as the target linguistic construction for the present study. The term anaphoric reference or anaphora describes the "relation between two linguistic elements, in which the interpretation of one (called an anaphor) is in some way determined by the interpretation of the other (called an antecedent)" (Huang, 2005, p. 231). Among the various categories of anaphoric reference, we selected one type of pronominal anaphora, the use of the third-person pronouns (he, she, and they), as the target linguistic construction for the current research.

The processing of pronominal anaphoric reference requires integrating information from two sources, a referential antecedent (e.g., Mark) and a pronoun (e.g., he). The integration of information online from multiple sources has been shown to cause difficulty for L2 users (e.g., Roberts, Gullberg, & Indefrey, 2008), thus the processing of pronominal anaphoric reference is also likely to pose a challenge. Processing pronominal anaphoric reference is expected to be especially demanding for Korean learners of English, given that the two languages have different anaphoric systems. Unlike in English where pronominal anaphora is the most widely used anaphoric reference type, in Korean zero anaphora, full-noun anaphora, and demonstrative plus noun anaphora are more common. In addition, the use of third-person pronouns in Korean is a very recent phenomenon; third-person pronouns are less frequently used in Korean than in English (An, 2008).

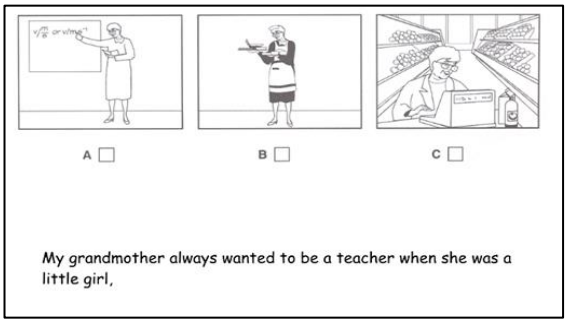
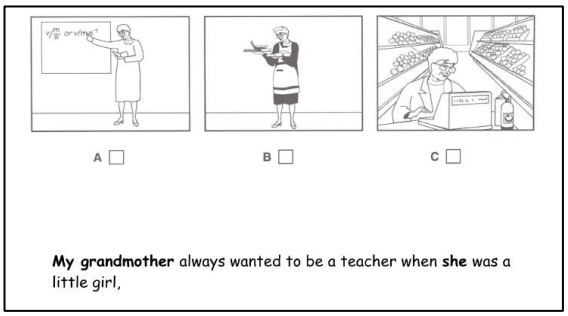
<B>Treatment Task

The multi-modal input-based activity that participants completed during the 3 treatment sessions were adapted from items included in Listening Part 1 of a series of Practice Cambridge Preliminary English Tests (PET). This test is designed for CEFR B1 level students, so it was considered appropriate for the participants. We also piloted the materials with learners from comparable backgrounds, and the piloting process provided further confirmation that the level was suitable for the participants.

The items in this test provide test-takers with three pictures (A, B, and C), a question, and an audio-recording; and the test-takers' task is to listen to the recording and answer the question by choosing the correct picture based on the information provided in the recording. For the purposes of the study, items containing third person pronominal anaphora reference were selected. Then, the software Camtasia 8.0 was used to add captions to the audio-recordings of the PET original items, with the captions appearing below the pictures on the computer screen. The captions were synchronised to the audio recordings. For the enhanced captions group, the target linguistic constructions (antecedent and pronoun) were additionally enhanced and presented in boldface (see Figure 2), as in Simard and Foucambert (2013) and Indrarathne and Kormos (2017). A total of 27 multimodal input-based activities were developed for the 3 treatment sessions, with 9 activities in each set (see Appendix A for an example set with captions). The duration of the activities was approximately 30 seconds to 1 minute. Cronbach α, calculated based on the response options, was found to be acceptable ($\alpha = .68$).

<INSERT FIGURE 2 ABOUT HERE>
FIGURE 2

Example Captions by Experimental Group

Captions Group	
Enhanced Captions Group	

<B>Grammaticality Judgment Tests

A written GJT and an oral GJT were used to assess the participants' pretest–posttest gains in the use of third-person pronominal anaphoric reference. In this way, we were able to determine whether the modality of the outcome measures mediated the relationship between textual input enhancement and development in the use of the target construction. The written GJT and the oral GJT had the same format; they only differed in modality. Before making a grammaticality judgment, the written GJT asked participants to read a given item on a computer screen, whereas the oral GJT involved participants in listening to the item. The GJTs were developed following guidelines offered by Keating and Jegerski (2015).

Thirty-two target items, including an antecedent and a third personal pronoun anaphora, were constructed. The items were identical in terms of sentence structure, number of syllables of words in the same position, and number of syllables in the items. Care was taken to include only high-frequency words by selecting words from the New General Service List (Brezina & Gablasova, 2015). Each item was formulated in four versions. Two versions were grammatical, including (a) a singular antecedent with a singular pronoun or (b) a plural antecedent followed by a plural pronoun. The other two versions were ungrammatical, one with (c) a singular antecedent and plural pronoun and another with (d) a plural antecedent and singular pronoun. An example for each version is given below:

a. My sister saw the market on the street. She bought some cookies. (grammatical, singular antecedent – singular pronoun)
b. My sisters saw the market on the street. They bought some cookies. (grammatical, plural antecedent – plural pronoun)

c. *My sister saw the market on the street. They bought some cookies. (ungrammatical, singular antecedent – plural pronoun)
d. *My sisters saw the market on the street. She bought some cookies. (ungrammatical, plural antecedent – singular pronoun)

In addition to the target items, 48 distractors were developed using three constructions: (a) the passive (16 items), (b) verb + gerund or verb + to infinitive (16 items), and (c) participial adjectives ending in either -ed or -ing (16 items). The purpose of including these constructions, which do not have direct equivalents in Korean, was to distract learners' attention from the target construction and thereby prevent them from easily identifying the focus of the assessment (Keating & Jegerski, 2015). Similar to the target items, 4 versions were developed for each distractor. All distractors had a similar syntactic structure and were designed to be of equal length. They were also designed to be of similar length to the target items. We created a total of 128 target sentences (32 items, 4 versions each) and 192 distractors (48 items, 4 versions each), which were then distributed into four sets of 80 items, counterbalanced across the 4 sets as follows:

Set A: 1a, 2b, 3c, 4d, etc.
Set B: 1b, 2c, 3d, 4a, etc.
Set C: 1c, 2d, 3a, 4b, etc.
Set D: 1d, 2a, 3b, 4c, etc.

Thus, each set contained 32 target sentences and 48 distractors.

Both the written GJT and the oral GJT were administered using E-Prime 2.0, which allowed for recording reaction times (RTs). In each test version, all items (target and distractors) were randomly presented. First, participants were required to press "z" or "m" to indicate whether they judged an item to be grammatical or ungrammatical, respectively. Next, they were asked to provide confidence ratings and source attributions regarding their answers (the data generated are not included in the present study). Fixation crosses were used to indicate transition between items. The reliability of each version of the GJT was evaluated using Cronbach's alpha. All four versions were found to be highly reliable (set A: α = .82; set B: α = .77; set C: α = .84; and set D: α = .77).

The four sets of GJTs (A, B, C, and D) were counterbalanced across modality and testing sessions within participants using a Latin Square design. The instructions were provided in Korean, the participants' native language. Before the GJTs were administered, participants completed a set of practice items in order to familiarise themselves with the format of the written and oral GJTs.

<B>Data Collection Procedure

For each participant, the experiment took place on three days over the span of one week. On day one, the background questionnaire, the OPT, and the pretest were administered. On day two, the participants took part in the first and second treatment sessions. On day three, they participated in the third treatment session, followed by the posttest (see Figure 1). The OPT lasted approximately 40 minutes: 10 minutes were allocated to the listening section and 30 minutes to the grammar section. Each GJT took about 20 to 25 minutes. The duration of each set of 9 treatment tasks was between 9 and 10 minutes. All 3 sessions were individual, and were administered by the first author.

A Tobii X2-30 mobile eye tracker with a temporal resolution of 30 Hz was used to record the participants' eye movements while performing the treatment tasks. This remote

eye-tracking system was mounted onto a laptop computer with a 15-inch screen. The participants were seated approximately 60 cm from the computer screen. The eye-tracking system was calibrated before each set of treatment tasks, using 9 points for calibrating each eye. The experiment was presented with Tobii Studio 3.3.0 software (Tobii Technology, 2014).

<B>Scoring and Data Analysis

The OPT comprised 200 questions: 100 questions in the listening section and 100 questions in the grammar section. Following the OPT scoring guidelines, one point was given for each correct answer, resulting in a maximum score of 100 for each section. Turning to the scoring of the written and the oral GJT, each correct response was awarded one point, thus the total score for each GJT (excluding distractors) was 32 points. Participants' reaction time (RT: the time between the appearance of a sentence on the computer screen and the participant's response) was measured in milliseconds (Jiang, 2012). For each participant, we calculated mean RTs and SDs for the correctly judged sentences only, and potential outliers were identified using the resulting means and SDs (Jiang, 2012). RTs that differed from a participant's mean by more than two standard deviations were considered as outliers (Jiang, 2012), and were trimmed to two standard deviations above or below the mean.

In the treatment tasks, participants were awarded one point for every correct picture choice, resulting in a total score of 27 (9 points per set).

When analysing the eye-movement data, first the quality of the recordings was inspected using the gaze samples measure provided by Tobii Studio 3.3.0. This index is expressed as a percentage, and is calculated "by dividing the number of eye tracking samples that were correctly identified by the number of attempts" (Tobii Studio 3.3.0. User Manual, 2014, p. 39). A value of 100% would mean that the movement of both eyes was found during the full recording (a highly unlikely outcome as participants blink, etc.). For this study, 63.7% was set as a cut-off point for acceptable level of recording quality. This value was calculated by subtracting one standard deviation (18.2%) from the mean percentage (81.9%) of time participants spent viewing the screen.

Drawing on previous eye-tracking research (e.g., Conklin & Pellicer–Sánchez, 2016; Winke, 2013), four measurements – first pass reading duration, second pass reading duration, total fixation duration, and number of visits – were used to examine the amount of attention paid to the two areas of interest: the antecedents and associated personal pronouns. First pass reading time is the sum of all fixation durations during the first visit to the area of interest. This measure is regarded as an index of initial processing. Second pass reading duration is defined as the sum of fixation durations when the eyes return to an area of interest after the first visit. In other words, second pass reading duration captures rereading in the area of interest, which is associated with re-analysis of the input. Total fixation duration is the sum of all fixation durations, that is, the sum of first and second pass reading times. Finally, a visit refers to all the fixations made within an area of interest from the time a participant's eyes first enter that area of interest and until they leave.

<B>Statistical Analyses

We used SPSS 22.0 for statistical analyses. As a preliminary measure, the Kolmogorov–Smirnov test (Appendix B) was carried out to examine whether the data were normally distributed. The test confirmed that, for both groups, the distributions for each measure was not significantly different from normal. A series of independent samples *t*-tests was run to compare the eye-gaze behaviours of the captions and enhanced captions groups during the

treatment sessions (RQ1). Next, mixed-model analysis of variance (ANOVA) tests were carried out to examine the effects of textual input enhancement on the pretest–posttest GJT gains (RQs 2). The relationships between the eye-tracking measures and GJT gain scores were established with Pearson correlational analyses (RQ3). An alpha level of $p < .05$ was set for all tests. Effect size estimates were obtained by calculating Cohen's *d* for the independent-samples *t*-tests, and eta-squared (η^2) and partial eta-squared (ηp^2) values for the mixed-model ANOVAs (Norouzian & Plonsky, 2017). The eta-squared values (η^2) were computed using the sums of squares for within- and between-subjects variables combined. Following Plonsky and Oswald (2014), *d* values of .40, .70, and 1.00, η^2 values of .06, .16, and .36, and *r* values of .25, .40 and .60 were considered as small, medium, and large, respectively.

<A>RESULTS

<B>Effects of Textual Enhancement on Treatment Task Performance

The descriptive statistics for the task completion scores achieved by the participants on the 27 multimodal treatment activities are presented in Table 1. An independent samples *t*-test revealed no significant difference between the captions and enhanced captions groups on the participants' treatment task performance, $t(46) = -.58$, $p = .56$, $d = .17$.

<INSERT TABLE 1 ABOUT HERE>
TABLE 1
Task Completion Scores on Treatment Task by Group

	Captions Group (n = 24)				Enhanced Captions Group (n = 24)			
			95% CI				95% CI	
	M	*SD*	Lower	Upper	*M*	*SD*	Lower	Upper
Task completion score	22.54	2.75	21.38	23.70	23.08	3.62	21.55	24.61

Note. The maximum score was 27 points.

<B>Effects of Textual Enhancement on Learner Attention to Target Constructions (RQ1)

To assess the effectiveness of textual enhancement in drawing learners' attention to the target construction, the participants' fixation durations and total number of visits to both areas of interest – antecedents and respective personal pronouns – were compared.

<C>Fixation Duration. Table 2 provides the descriptive statistics by group for the three indices of fixation duration, that is, first pass reading duration, second pass reading duration, and total fixation duration. The means indicate the sum of fixation durations for the 27 treatment tasks combined.

For the antecedents, a series of independent samples *t*-tests found no significant difference in first pass reading duration between the captions and enhanced captions groups, $t(46) = -1.75$, $p = .09$, $d = .51$, but revealed that the two groups significantly differed in terms of second pass reading duration, $t(46) = -2.33$, $p = .02$, $d = .67$, and total fixation duration, $t(46) = -2.41$, $p = .02$, $d = .70$. The effect size for second pass reading time and total fixation duration was in the small range. This means that the enhanced captions group spent significantly longer time rereading the antecedent and reading the antecedent overall than the captions group.

Another series of independent samples *t*-tests revealed that, for the pronouns, none of the fixation duration measures differed significantly across the two groups, first pass reading

duration: $t(46) = -1.80$, $p = .08$, $d = .51$; second pass reading duration: $t(46) = -.20$, $p = .84$, $d = .05$; total fixation duration: $t(46) = -1.45$, $p = .15$, $d = .41$.

<INSERT TABLE 2 ABOUT HERE>

TABLE 2
Fixation Durations on Areas of Interest by Group

	Captions Group (n = 24)				Enhanced Captions Group (n = 24)			
			95% CI				95% CI	
	M	*SD*	Lower	Upper	*M*	*SD*	Lower	Upper
Antecedent								
First Pass (ms)	175	74	144	207	223	110	177	270
Second Pass (ms)	191	98	150	233	270	134	214	327
Total Time (ms)	367	138	309	426	494	217	402	586
Pronoun								
First Pass (ms)	65	37	49	81	87	48	67	107
Second Pass (ms)	26	18	18	33	27	20	18	35
Total Time (ms)	91	45	71	110	114	65	87	142

<C>Visit Counts. Table 3 provides the descriptive statistics for visit counts by group. The means capture the total number of visits to the antecedents and pronouns for the 27 treatment tasks combined. The independent samples *t*-test carried out to compare the total number of visits to the antecedents revealed a significant, medium-size difference between the captions and enhanced captions groups, indicating that participants in the enhanced captions groups visited the antecedent more frequently, $t(46) = -2.80$, $p = .01$, $d = .82$. In contrast, the independent samples *t*-test conducted to compare the number of visits to the pronouns did not yield a significant difference between the two groups, $t(46) = -1.22$, $p = .23$, $d = .35$.

<INSERT TABLE 3 ABOUT HERE>

TABLE 3
Total Number of Visits to Areas of Interest by Group

	Captions Group (n = 24)				Enhanced Captions Group (n = 24)			
			95% CI				95% CI	
	M	*SD*	Lower	Upper	*M*	*SD*	Lower	Upper
Antecedent								
Number of visits	26.83	9.44	22,85	30.82	36.99	14.85	30.61	43.14
Pronoun								
Number of visits	12.17	6.05	9.61	14.72	14.58	7.54	11.40	17.77

<B>Effects of Textual Enhancement on L2 Development (RQ2)

To examine the extent to which textual enhancement facilitated development in the receptive knowledge of pronominal anaphoric reference, the captions and enhanced caption groups' pretest and posttest performance was compared on the written and oral GJTs.

<C>Written GJT Results. The descriptive statistics for participants' performance on the written GJT are presented in Table 4. To test whether there were any differences between the captions and enhanced captions groups at the time of the pretest, independent samples *t*-tests

were conducted using the written GJT pretest accuracy scores and RTs. The results indicated that the two groups did not differ significantly either in terms of total accuracy scores $t(46) = -.06$, $p = .95$, $d = .02$ or in total RTs: $t(46) = .34$, $p = .73$, $d = .10$ at the outset of the study. The independent samples *t*-test run for the accuracy scores and reaction times for the grammatical items (accuracy: $t(46) = -.31$, $p = .76$, $d = .09$; RTs: $t(46) = .18$, $p = .85$, $d = .05$) and ungrammatical items (accuracy: $t(46) = -.27$, $p = .79$, $d = .08$; RTs: $t(46) = .58$, $p = .56$, $d = .17$) separately did not reveal significant differences between the two groups either.

<INSERT TABLE 4 ABOUT HERE>

TABLE 4

Descriptive Statistics for Written GJT Accuracy Scores and RTs by Group

	Captions Group (n = 24)				Enhanced Captions Group (n = 24)			
			95% CI				95% CI	
	M	*SD*	Lower	Upper	*M*	*SD*	Lower	Upper
Grammatical Items								
Pretest								
Accuracy	13.12	2.29	12.16	14.09	12.92	2.34	11.92	13.90
RT	6865.23	2140.78	5961.26	7769.20	6767.02	1505.10	6131.47	7402.57
Posttest								
Accuracy	13.58	2.37	12.58	14.59	15.37	1.31	14.82	15.92
RT	4930.26	1404.34	4337.26	5523.26	4721.61	1238.31	4198.72	5244.51
Ungrammatical Items								
Pretest								
Accuracy	10.75	2.69	9.61	11.88	10.95	2.59	9.86	12.05
RT	8506.83	3309.09	7109.52	9904.14	8069.20	1617.78	7386.07	8752.33
Posttest								
Accuracy	12.62	2.33	11.64	13.61	13.29	2.03	12.43	14,15
RT	5871.70	1515.64	5231.70	6511.70	6057.65	1682.74	5347.09	6768.21
Total								
Pretest								
Accuracy	23.92	4.44	22.04	25.79	24.00	4.60	22.06	25.94
RT	7563.85	2466.61	6522.29	8605.41	7361.19	1522.64	6718.23	8004.14
Posttest								
Accuracy	26.38	3.73	24.80	27.95	28.96	2.26	28.01	29.91
RT	5382.17	1450.47	4769.69	5994.65	5331.31	1389.13	4744.73	5917.88

Note. Maximum score was 32 points.

Next, to examine the effects of textual enhancement on participants' gains on the written GJT, separate mixed-model ANOVAs were conducted for the written GJT total accuracy scores and RTs as well as for the grammatical and ungrammatical items separately. The within-subjects variable in the analyses was time (pretest versus posttest), and the between-subjects factor was group (captions versus enhanced captions group). As shown in Table 5, a significant interaction effect emerged between time and group for the total and grammatical GJT accuracy scores, with the interaction accounting for 2% and 14% of the variation in the overall models respectively, including within and between-subject variables. That is, the enhanced captions group achieved greater gains in accuracy than the captions group on the grammatical written GJT items and in total. However, no significant interaction effect was found for reaction times in any of the analyses, that is, the two groups did not show a

significantly different pretest–posttest decrease in the speed they responded to the written GJT items, grammatical or ungrammatical.

<INSERT TABLE 5 AROUND HERE>

TABLE 5
Results for Interactions from Mixed-Model ANOVAs for the Written GJT scores and RTs

		F	*p*	η^2	η_p^2
Grammatical Items					
Time * Group	Written GJT Score	12.005	.001	.143	.207
	Written GJT RT	.067	.796	.000	.001
Ungrammatical Items					
Time * Group	Written GJT Score	.552	.461	.006	.012
	Written GJT RT	.795	.377	.009	.017
Total					
Time * Group	Written GJT Score	9.454	.004	.021	.170
	Written GJT RT	.089	.767	<.001	.002

<*C*>*Oral GJT Results.* Table 6 provides the descriptive statistics for the oral GJT scores and reaction times. The independent samples *t*-tests, which were carried out to examine whether there were initial differences between the captions and enhanced captions groups, yielded no significant difference for either the pretest oral GJT total accuracy scores, $t(46) = .46$, $p = .65$, $d = .13$, or total RTs, $t(46) = 1.24$, $p = .22$, $d = .36$. Neither were significant differences found for the accuracy and RTs of the grammatical items (accuracy: $t(46) = -.59$, $p = .56$, $d = .17$; RTs: $t(46) = 1.08$, $p = .28$, $d = .31$) or ungrammatical items (accuracy: $t(46) = -.87$, $p = .39$, $d = .25$; RTs: $t(46) = 1.35$, $p = .18$, $d = .39$) between the two groups.

<INSERT TABLE 6 AROUND HERE>

TABLE 6
Descriptive Statistics for Oral GJT Accuracy Scores and RTs by Group

	Captions Group (n = 24)				Enhanced Captions Group (n = 24)			
			95% CI				95% CI	
	M	*SD*	Lower	Upper	*M*	*SD*	Lower	Upper
Grammatical								
Pretest								
Accuracy	10.46	2.34	9.47	11.45	10.87	2.58	9.79	11.96
RT	7291.81	979.75	6878.10	7705.53	7000.86	877.06	6630.50	7371.21
Posttest								
Accuracy	11.62	1.97	10.79	12.46	13.83	1.52	13.19	14.48
RT	6954.19	983.81	6538.76	7369.61	6289.77	612.69	6031.05	6548.48
Ungrammatical								
Pretest								
Accuracy	7.21	1.50	6.57	7.84	7.71	2.39	6.70	8.71
RT	8847.88	1342.83	8280.85	9414.91	8325.39	1330.99	7763.37	8887.42
Posttest								
Accuracy	7.96	2.44	6.93	8.99	10.00	2.28	9.03	10.96

RT	7745.26	1163.20	7254.08	8236.43	7100.31	869.30	6733.24	7467.38
Total								
Pretest								
Accuracy	18.00	3.75	16.42	19.58	17.50	3.74	15.92	19.08
RT	7911.82	1004.29	7487.74	8335.86	7551.51	1005.54	7126.91	7976.12
Posttest								
Accuracy	18.92	3.91	17.27	20.57	25.21	3.15	23.88	26.54
RT	7264.47	1008.27	6838.71	7690.22	6617.73	654.46	6341.38	6894.09

Note. Maximum score was 32 points.

A separate mixed-model ANOVA was also performed for the participants' oral GJT accuracy scores and RTs in order to see whether textual enhancement had an influence on students' oral GJT gains in total and the grammatical and ungrammatical items separately (see Table 7). A significant time-by-group interaction emerged for all the accuracy scores with a small effect size. The interaction explained 13% of the variation in the overall model for the total scores, including within and between-subject variables. For both grammatical and ungrammatical items, 7% of the variation in the overall model could be attributed to the interaction. No interaction effect was found for reaction times (see Table 7). This means that the participants exposed to enhanced captions achieved significantly greater gains in accuracy on the oral GJT than participants who viewed captions without textual enhancement. The effect size for this difference was in the small range for all accuracy scores (total, grammatical, and ungrammatical). Textual enhancement, however, did not influence the extent to which participants demonstrated a decrease in the time taken to make grammaticality judgments.

<INSERT TABLE 7 ABOUT HERE>

TABLE 7
Results for Interactions from Mixed-Model ANOVA for the Oral GJT scores and RTs

		F	*p*	η^2	η_p^2
Grammatical Items					
Time * Group	Oral GJT Score	5.269	.026	.066	.103
	Oral GJT RT	1.170	.285	.021	.025
Ungrammatical Items					
Time * Group	Oral GJT Score	4.833	.033	.069	.095
	Oral GJT RT	.077	.782	.001	.002
Total					
Time * Group	Oral GJT Score	47.083	.000	.128	.506
	Oral GJT RT	.644	.426	.004	.014

<B>Relationships Between Attention and L2 Development (RQ3)

To address the relationship between attention and L2 development, a series of Pearson correlations were computed between the eye-tracking measures and gain scores on the written and oral GJT tests, for the antecedents and pronouns separately and combined. As shown in Table 8, for the antecedents and pronouns combined, a significant correlation was only identified between the GJT written scores and total reading time. The direction of this correlation was negative, and its size was in the medium range.

<INSERT TABLE 8 AROUND HERE>

TABLE 8

Pearson Correlations Between Gain Scores and Total Fixation Duration and Number of Visits for Antecedent and Pronoun Combined

	Captions (n = 24)		Enhanced Captions (n = 24)	
	Total Time (p)	Visit Count (p)	Total Time (p)	Visit Count (p)
Written GJT Gain	−.46*	−.35	−.11	.05
	(.02)	(.09)	(.60)	(.82)
Oral GJT Gain	.13	.32	−.06	−.09
	(.53)	(.13)	(.77)	(.69)

* $p < .05$

Table 9 displays the results for the antecedent and pronoun separately. For the captions group, medium to large negative correlations were identified between the written GJT gain scores and all the eye-tracking indices calculated for the antecedent (second pass reading, total fixation duration, visit counts), except for first pass reading time. The oral GJT scores of the captions group, however, were found to have medium-size positive correlations with the total duration and total visit counts on the personal pronouns.

Overall, these results mean that, in the captions group, participants achieved lower gains on the written GJT when they reread the anaphora antecedents more often and for longer periods and spent more time gazing at the antecedent and anaphora combined. On the other hand, participants who visited the pronouns more frequently and spent more time reading them displayed greater development on the oral GJT test. No significant relationships were found between the eye-tracking measures and the gain scores of the enhanced captions group.

Table 9 also demonstrates the correlations between the various eye-tracking indices. For the captions group, fewer correlations were observed, the analyses yielding no significant links between the second pass and total reading durations calculated for the antecedents and those for the pronouns. For the enhanced captions group, on the other hand, the large majority of the eye-tracking indices computed for the antecedents correlated to a medium or large degree with those computed for the pronouns, including second pass and total reading times. That is, participants in the enhanced captions group who reread and fixated longer on the anaphora antecedents also devoted more time gazing at the associated personal pronouns. In the captions group, however, those who fixated longer on the antecedents/pronouns did not pay more attention to the related pronouns/antecedents.

<INSERT TABLE 9 ABOUT HERE>

TABLE 9

Pearson Correlations Between Gain Scores and Eye-tracking Measures for Antecedents and Pronouns

	Written GJT Gain r (p)	Oral GJT Gain r (p)	First Pass Ant r (p)	Second Pass Ant r (p)	Total Time Ant r (p)	Visit Count Ant r (p)	First Pass Pronoun r (p)	Second Pass Pronoun r (p)	Total Time Pronoun r (p)
Captions Group									
First Pass	−.26	.24							

Antecedent	(.22)	(.26)							
Second Pass	−.51*	−.18	.28						
Antecedent	(.01)	(.41)	(.19)						
Total Time	−.50*	.00	.73**	.86**					
Antecedent	(.01)	(.99)	(.00)	(.00)					
Visit Count	−.54*	.15	.82**	.48*	.78**				
Antecedent	(.01)	(.48)	(.00)	(.02)	(.00)				
First Pass	−.17	.38	.54*	.11	.37	.60**			
Pronoun	(.42)	(.07)	(.01)	(.59)	(.07)	(.00)			
Second Pass	.14	.38	.29	.01	.17	.04	.26		
Pronoun	(.52)	(.06)	(.16)	(.95)	(.43)	(.86)	(.22)		
Total Time	−.09	.46*	.56**	.10	.37	.50*	.92**	.61**	
Pronoun	(.69)	(.02)	(.00)	(.64)	(.07)	(.01)	(.00)	(.00)	
Visit Count	.07	.46*	.48*	−.25	.08	.45*	.68**	.18	.63**
Pronoun	(.73)	(.02)	(.02)	(.23)	(.71)	(.03)	(.00)	(.39)	(.00)
Enhanced Captions									
First Pass	−.15	−.02							
Antecedent	(.49)	(.92)							
Second Pass	−.09	−.04	.56**						
Antecedent	(.67)	(.84)	(.00)						
Total Time	−.13	−.04	.86**	.91**					
Antecedent	(.54)	(.86)	(.00)	(.00)					
Visit Count	−.04	−.11	.69**	.60**	.72**				
Antecedent	(.85)	(.62)	(.00)	(.00)	(.00)				
First Pass	.06	−.10	.58**	.54*	.63**	.78**			
Pronoun	(.78)	(.64)	(.00)	(.01)	(.00)	(.00)			
Second Pass	−.24	−.17	.53**	.71**	.71**	.66**	.76**		
Pronoun	(.26)	(.42)	(.00)	(.00)	(.00)	(.00)	(.00)		
Total Time	−.03	−.13	.60**	.63*	.69**	.79**	.98**	.88**	
Pronoun	(.89)	(.55)	(.00)	(.01)	(.00)	(.00)	(.00)	(.00)	
Visit Count	.21	−.03	.41*	.32	.41*	.75**	.81**	.56**	.77**
Pronoun	(.31)	(.88)	(.04)	(.12)	(.04)	(.00)	(.00)	(.00)	(.00)

* $p < .05$, ** $p < .01$

<A>DISCUSSION

<B>The Effects of Textual Enhanced Captions on Attentional Allocation

Our first research question explored the extent to which textual enhancement in captions can draw learners' attention to pronominal anaphoric reference, a grammatical construction. The eye-tracking data, which were utilised to assess attentional processing, revealed that textual enhancement was successful in directing learners' attention to the referential antecedents highlighted in the captions. When the antecedents were presented in boldface, the participants devoted more time to rereading the forms, and fixated longer on them overall. As compared to the unenhanced group, learners exposed to enhanced input also visited the antecedents more frequently. The effect sizes for these differences were in the small (second pass reading time, total fixation) and medium range (visit counts). Textual enhancement,

however, did not lead to longer first pass reading times. Contrary to the results for the antecedents, the enhanced visual salience of the personal pronouns did not generate more attention; none of the eye-tracking indices yielded a significant difference between the enhanced and unenhanced groups for this target area. Overall, these results suggest that textual enhancement in captions was able to trigger more attention to the target antecedents, but textual enhancement did not yield increased attention to the associated personal pronouns.

Our results for the referential antecedents indicate that raising the visual salience of target features in captions cannot only facilitate attention to lexis, as observed in Montero Perez et al. (2015), but also promote attention to grammar. It is important to note, however, that Montero Perez et al. found an advantage for enhanced captions (key word versus full captioning) only when a forthcoming test was announced. Although no test announcement was made in the present study, it is likely that the participants anticipated the forthcoming posttests, given that they had taken a battery of pretests prior to the treatment. This, in turn, might have made them more attentive to the textually enhanced constructions, as in Montero Perez et al. (2015).

It is also worth comparing our results to textual enhancement studies utilising unimodal input. As mentioned earlier, this line of research has so far yielded mixed findings, with some studies observing a benefit for attentional allocation under the enhanced condition (Simard & Foucambert, 2013; Winke, 2013), others generating null effects (Indrarathne & Kormos, 2017; Issa et al., 2015; Loewen & Inceoglu, 2016). Indrarathne and Kormos explained these mixed patterns by the differential visual salience created by the various textual enhancement techniques across studies, suggesting that underlining (Simard & Foucambert, 2013; Winke, 2013) might be more effective in creating an isolation effect than other forms of textual enhancement such as boldfacing (Indarathne & Kormos, 2017) or the use of different colour fonts (Issa et al., 2015; Loewen & Inceoglu, 2016). Although our study employed boldfacing like Indrarathne and Kormos, this technique might have been more successful in generating an isolation effect here, as the enhanced constructions appeared in sentences. This probably made the highlighted input more salient, as compared to when target features are boldfaced in a larger text.

The fact that the input was presented bimodally could have further promoted the salience of the textually enhanced features in the captions. It is possible that the captioned texts, at least for some of the participants, were delivered faster than their normal reading speed. Thus, due to lack of time to read the captions in full, the learners might have devoted increased attention to the highlighted words, assuming that they contained key information. Fast presentation speed could also provide an explanation why there was no difference in the amount of attention allocated to the enhanced referential pronouns between the two groups. Even though the pronouns were made more salient to the enhanced group, participants probably had less time to revisit them after rereading the antecedents, due to the short-lived nature of the captions.

<B>The Effects of Textually Enhanced Captions on Development

Our second research question was concerned with the extent to which textually enhanced captions facilitated development in the knowledge of pronominal anaphoric reference, as measured by a written and an oral grammaticality judgment test. On both GJTs, participants exposed to enhanced captions demonstrated greater gains in accuracy than participants who viewed captions without enhanced input, with the effect size values falling into the small range. The only exception to this trend was the absence of a significant difference in ungrammatical written GJT gain scores between the two conditions. Unlike for the accuracy scores, textual enhancement did not affect the degree of decrease in reaction times from the

pretest to the posttest. These findings suggest that the increased attention that participants paid to pronominal anaphoric reference under the enhanced condition led to further processing of the target construction, resulting in greater gains in accuracy as measured by the immediate posttests.

The results of the present study are well aligned with the findings of Montero Perez et al. (2015) for lexis, indicating that enhanced visual salience in captions cannot only lead to better recognition of lexical forms but also improved receptive knowledge of grammar. Our results, however, differ from the conclusion of Lee and Huang's (2008) meta-analytic review that textual input enhancement has only a marginal impact on grammar learning. The somewhat larger effect sizes found in the present study could be attributed to the fact that participants had some prior knowledge of the target construction, as reflected in the considerably higher than chance accuracy scores on the written GJT pretest. Several researchers (e.g., Han, Park, & Combs, 2008; Lee & Huang, 2008; Park, 2004; Winke, 2013) have noted that prior knowledge is likely to be a key determinant of whether textual enhancement succeeds in promoting grammatical knowledge. An additional, or alternative, explanation for the positive findings obtained here could be that textual enhancement was used in combination with captioning. As pointed out by Leow and Martin (2017), the joint use of textual enhancement with other attention-getting tools (e.g., captioning) is likely to trigger greater depth of processing of enhanced input and thereby result in more developmental benefits than using textual enhancement alone.

<B>Relationship Between Attention and L2 Development

Our third research question addressed the relationship between attention, operationalised in terms of eye-tracking measures, and development in the knowledge of pronominal anaphoric reference. While no significant links were observed between the eye-movement indices and GJT gain scores for the enhanced captions group, the correlational analyses yielded a number of significant medium to large relationships between the gain scores of the captions group and the eye-movement measures. In the captions group, participants who visited the anaphora antecedents more often and reread and fixated longer on them exhibited less development on the written GJT, whereas participants who looked at the pronouns more frequently and devoted more time to reading them overall achieved greater gains on the oral GJT.

In light of previous research (e.g., Godfroid & Uggen, 2013; Indrarathne & Kormos, 2017; see, however, Montero Perez et al., 2015), it was expected that we would find a positive relationship between the oral GJT gains and the overall length and number of eye fixations on the target pronouns. This result, however, was anticipated for both groups, not just for the unenhanced captions condition. It is more puzzling that, in the unenhanced group, we also observed negative links between the written GJT gain scores and a number of eye-tracking indices, including total fixation durations, second pass reading times and visit counts for the antecedents as well as total fixation times combined for the antecedents and pronouns. Possibly, those participants who more frequently reread the antecedents (but not the pronouns) failed to pay attention to the anaphora construction as a whole, and, as a result, displayed less improvement on the written GJT. For these participants, lack of increased attention to the pronouns might have resulted from the fixed rate of input delivery. Due to time constraints, participants might only have had time to revisit the antecedents but not the personal pronouns in the captions. This account is supported by the fact that, for the unenhanced captions group, no significant correlations emerged between how much time participants spent rereading the antecedents and the associated pronouns.

The question also rises why we did not find any relationships between the eye-tracking indices and their GJT gain scores for the enhanced group. Probably, participants in this group naturally orientated their attention to both components of the highlighted anaphora constructions, as indicated by the strong correlations between the amount of attention participants paid to the antecedents and pronouns. Thus, despite the set speed of delivery, they might have had enough time to select both components of the anaphora construction for further processing. For this group, the degree of gains in the knowledge of pronominal anaphoric reference might have been more related to the extent to which participants engaged in a higher level of processing after the initial selection of information, rather than differences among participants in the amount of lower level of processing they performed, which was captured in the eye-tracking measures (Leow, Grey, Marijuan, & Moorman, 2014). Montero Perez et. al., when explaining a similar lack of a significant link between attention and learning in their enhanced condition, also speculated that longer fixations might not necessarily be a reflection of more elaborate processing. Future studies using verbal protocols such as the stimulated recall procedure could help shed light on the correctness of this interpretation.

<A>IMPLICATIONS

The findings of this research are of theoretical and pedagogical significance. At the theoretical level, similar to Godfroid and Uggen (2013) and Indrarathne and Kormos (2017), our results, for the unenhanced group, provide some evidence for the claim that the amount of attention L2 learners pay to grammatical constructions is linked to the extent of grammatical development they display. However, the lack of significant links between the eye-tracking indices and development in the enhanced group implies that more attention, unless it triggers greater depth of processing, does not necessarily lead to more learning (Godfroid et al., 2013; Leow, 2015). It is also of theoretical importance that, in this study, increasing the visual salience of a grammatical construction in captions resulted in greater attention allocated to the target feature. Our results, therefore, suggest that textual enhancement in captions does not only have the capacity to direct attention to lexical (Montero Perez et al., 2015) but also grammatical features. Moving on to pedagogy, an implication of this research is that highlighting target grammatical constructions in captions can draw learner attention to target grammatical features and thereby promote L2 learning. The study also proves that textual enhancement integrated into input-based multimodal activities has the potential to facilitate L2 grammatical development.

<A>LIMITATIONS AND FUTURE DIRECTIONS

In interpreting the findings of this research, it is also important to take into account the limitations of the study. One methodological weakness lies in the fact that we did not triangulate our eye-tracking measures with verbal protocol comments. As implied previously, the combination of eye-gaze and verbal protocol data would have enabled us to tap not only the amount of attention participants paid to the enhanced features but also the level of processing in which they engaged when encountering the highlighted anaphoric constructions. As a result, our interpretations about the findings obtained here would have been less tentative. In future research, the incorporation of stimulated recall protocols would appear suitable to gain insights into the conscious operations that may be induced by textually enhanced captions. Unlike think-alouds which require verbalisation, this procedure does not interfere with the online processing of aural input, although it is potentially affected by memory decay (Gass & Mackey, 2016).

Absence of a delayed posttest is an additional weakness of this study. Administering a delayed posttest would have allowed us to determine whether increasing the visual salience of the target linguistic construction had a long-term effect on development in L2 grammatical knowledge. Ideally, the immediate posttest would also have been administered on a day different from the third treatment session to facilitate consolidation of new knowledge. To address these limitations, we recommend that further studies of textual enhancement in captions include both an immediate and delayed posttest, and that participants do not take the immediate posttest on the same day as the last treatment session.

Another limitation that deserves attention concerns the nature of the typographical modification (i.e., bolded font) that was used to enhance the target linguistic construction. When letters are presented in boldface, there is a possibility that the font size will slightly change (Winke, 2013), which could affect the size of the area of interest. Although we took great caution to ensure that the areas of interest for each target linguistic construction was the same across the two groups, future studies should consider using other types of typographical cues to avoid this potential threat to validity.

A further shortcoming of this research concerns the lack of control for individual differences among the participants. Working memory, in particular, is likely to have moderated the impact of textually enhanced captions on attentional allocation, given the multiple sources of input to which participants were exposed and the relatively limited L2 knowledge and processing skills they possessed. Working memory has also been demonstrated to be significantly linked to gains in receptive knowledge in other studies investigating the effects of textual enhancement on attentional allocation and development (e.g., Indrarathne & Kormos, 2018).

Other limitations of the study include its focus on a single grammatical feature, activity, and context. Replication of this research with constructions that are of lower physical salience and communicative value would be especially warranted, since such linguistic targets are less prone to attracting attention and thus being acquired by L2 learners from exposure to input alone (e.g., Long & Robinson, 1998). Future studies are also needed to examine whether the results found here would transfer to different activities, for example, activities involving videos rather than static images. A follow-up study could also look into whether the same developmental trends would emerge if the study was carried out in a classroom setting as opposed to individualised sessions in the laboratory.

<A>CONCLUSION

The primary goal of this study was to investigate the extent to which textual enhancement in captions can promote learner attention to and subsequent development in the knowledge of pronominal anaphoric reference, an L2 grammatical construction. In doing so, we intended to initiate research into the capacity of visual enhancement in captions to influence attention to and development in grammatical, as opposed to lexical, knowledge. Finally, using eye-tracking methodology, our aim was to expand on previous research that has examined the link between attention and L2 learning. In line with our expectations, we found that boldfacing the pronominal anaphoric reference construction in captions succeeded in directing learner attention to the referential antecedents, and led to increased gains in the receptive knowledge of the construction as a whole. Contrary to expectations, however, we only found significant links between attention and development when the target construction was left unenhanced in the captions. We interpreted this finding as suggesting that, in the enhanced group, differential gains might have been associated with differences in higher-level rather than lower-level processing, with the former not being captured by the eye-gaze indices.

<A>ACKNOWLEDGMENTS

We would like to thank Ron Leow, Carmen Muñoz and Ana Pellicer-Sánchez for their insightful comments on the research on which this article is based. Any errors, of course, are our own. This research was supported in part by The International Research Foundation for English Language Education (TIRF) Doctoral Dissertation Grant and the Language Learning Dissertation Grant.

REFERENCES

An, S. (2008). The function of third-person '*Ku, Kunye*'. *Korean Linguistics, 38*, 145–164.
Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Linguistics, 23*, 509–533.
Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics, 36*, 1–22.
Conklin, K, & Pellicer–Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research, 32*, 453–467.
Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching, 5*, 161–170.
Gass, S. M., & Mackey, A. (2016). *Stimulated recall methodology in applied linguistics and L2 research*. New York: Routledge.
Godfroid, A., Boers, F., & Housen, A. (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.
Godfroid, A., & Uggen, M. S. (2013). Attention to irregular verbs by beginning learners of German. *Studies in Second Language Acquisition, 35*, 291–322.
Grabe, W. (2012). *Reading in a second language: Moving from theory to practice*. Cambridge: Cambridge University Press.
Guillory, H. C. (1998). The effects of key word captions to authentic French video in foreign language instruction. *CALICO Journal, 15*, 89–108.
Han, Z., Park, E. S., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics, 29*, 597–618.
Huang, Y. (2005). Anaphora, cataphora, exophora, logophoricity. In K. Brown (Ed.) *Encyclopaedia of language and linguistics* (pp. 231–238). New York: Elsevier.
Indrarathne, B., & Kormos, J. (2017). Attentional processing of input in explicit and implicit conditions: An eye-tracking study. *Studies in Second Language Acquisition, 39*, 401–430.
Indrarathne, B., & Kormos, J. (2018). The role of working memory in processing L2 input: Insights from eye-tracking. *Bilingualism: Language and Cognition, 21*, 355–374.
Issa, B., Morgan–Short, K., Villegas, B., & Raney, G. (2015). An eye-tracking study on the role of attention and its relationship with motivation. *EUROSLA Yearbook, 15*, 114–142.
Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York: Routledge.
Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology, 8*, 441–480.
Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research. *Studies in Second Language Acquisition, 37*, 1–32.
Lee, S. K., & Huang, H. T. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition, 30*, 307–331.
Leow, R. P. (2015). *Explicit learning in the L2 classroom*. New York: Routledge.
Leow, R. P., Grey, S., Marijuan, S., & Moorman, C. (2014). Concurrent data elicitation procedures, processes, and the early stages of L2 learning: A critical overview. *Second Language Research, 30*, 111–127.
Leow, R. P., & Martin, A. (2017). Enhancing the input to promote salience of the L2: A critical overview. In S. Gass, P. Spinner & J. Behney (Eds.), *Salience in SLA* (pp. 167–186). New York: Routledge.

Loewen, L., & Inceoglu, S. (2016). The effectiveness of visual input enhancement on the noticing and L2 development of the Spanish past tense. *Studies in Second Language Learning and Teaching*, *6*, 89–110.

Long, M. H. (2000). Focus on form in task-based language teaching. In R. D. Lambert & E. Shohamy (Eds.), *Language policy and pedagogy: Essays in honor of A. Ronald Walton* (pp. 179–192). Philadelphia: Benjamins.

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

Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals, 32*, 321–328.

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language, Learning & Technology, 1*8, 118–141.

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned Video: An eye- tracking study. *The Modern Language Journal, 99*, 308–328.

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*, 720–739.

Norouzian, R., & Plonsky, L. (2017). Eta-and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. *Second Language Research*. doi:10.1177/0267658316684904

Park, E. S. (2004). Constraints of implicit focus on form: Insights from a study of input enhancement. *Teachers College, Columbia University Working Papers in TESOL and Applied Linguistics, 4*, 1–30.

Pellicer–Sánchez, A. (2015). Learning L2 collocations incidentally from reading. *Language Teaching Research, 19*, 1–22.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*, 878–912.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition, 30*, 333–357.

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678). Oxford: Blackwell.

Rost, M. (2011). *Teaching and researching listening*. London: Longman.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 1–32). Cambridge: Cambridge University Press.

Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty & M. H. Long, (Eds.), *The handbook of second language acquisition* (pp. 382–408). Malden, MA: Blackwell.

Simard, D., & Foucambert, D. (2013). Observing noticing while reading in L2. In J. M. Bergsleithner, S.N. Frota & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 207–226). Honolulu, HI: National Foreign Language Resource Center, University of Hawai`i at Mānoa.

Tobii Studio. (2014). *User Manual – Tobii Studio (Version 3.3.0)*. Retrieved from http://www.tobii.com/Global/Analysis/Downloads/User_Manuals_and_Guides/Tobii_X2-30_EyeTrackerUserManual_WEB.pdf

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*, 191–210.

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT journal*, *42*, 272–281.

Williams, R. S., & Morris, R. K. (2004). Eye movements, word familiarity and vocabulary acquisition. *European Journal of Cognitive Psychology, 16*, 321–339.

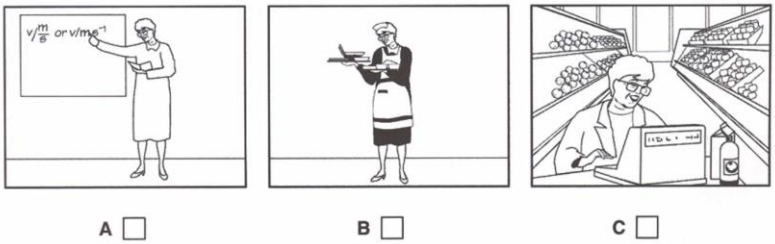
Winke, P. (2013). The effects of input enhancement on grammar learning and comprehension: A modified replication of Lee (2007) with eye-movement data. *Studies in Second Language Acquisition, 35*, 323–352.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology, 14*, 65–86.

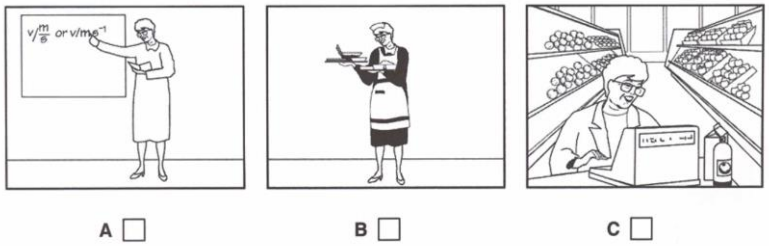
Yang, J. C., & Chang, P. (2014). Captions and reduced forms instruction: The impact on EFL students' listening comprehension. *ReCALL, 26*, 44–61.

APPENDIX A
A Sample of Multimodal Input-based Activity

What is the grandmother's job now?



My grandmother always wanted to be a teacher when **she** was a little girl,



but she had to leave school when she was fourteen and help her mother clean offices and shops.

A ☐ B ☐ C ☐

When she was in her thirties she went to college, but she had to work as a waitress in the evenings to pay for her studies.

A ☐ B ☐ C ☐

A few years later she finally got the job she'd always wanted and she's done it ever since.

APPENDIX B
Results of Kolmogorov-Smirnov Tests of Normality

	Captions Group (n = 24)			Enhanced Captions Group (n = 24)		
	Statistic	*df*	*Sig*	*Statistic*	*df*	*Sig*
Oxford Placement Test						
Listening	.109	24	.200	.127	24	.200
Grammar	.086	24	.200	.097	24	.200
Eye-movement data						
First pass reading - ANT	.157	24	.132	.127	24	.200
Second pass reading - ANT	.114	24	.200	.114	24	.200
Total fixation duration - ANT	.106	24	.200	.164	24	.095
Number of visits - ANT	.140	24	.200	.173	24	.061
First pass reading - PRO	.169	24	.073	.128	24	.200
Second pass reading - PRO	.109	24	.200	.125	24	.200
Total fixation duration - PRO	.144	24	.200	.121	24	.200
Number of visits - PRO	.132	24	.200	.158	24	.126
PRE GJT						
Written – Score	.131	24	.200	.128	24	.200
Written – Reaction Time	.164	24	.096	.085	24	.200
Oral – Score	.147	24	.196	.155	24	.140
Oral – Reaction Time	.168	24	.079	.132	24	.200
POST GJT						
Written – Score	.148	24	.189	.174	24	.058
Written – Reaction Time	.151	24	.169	.109	24	.200
Oral –Score	.147	24	.194	.174	24	.059
Oral – Reaction Time	.136	24	.200	.109	24	.200