



Visual saliency in captioned digital videos and learning of English collocations: An eye-tracking study

Sungmook Choi, Kyungpook National University

Abstract

This study explored how visual input enhancement impacts caption-reading behaviors, the acquisition of English collocations, and the recall of onscreen captions. The participants comprised 53 Korean undergraduate students at a high-intermediate level of English proficiency. They were assigned to either a baseline or an enhancement group. The baseline group viewed a digital video with unenhanced captions, whereas the enhancement group watched the same video with enhanced captions (i.e., captions including yellow-colored collocations). The eye movements of the participants were measured using an eye tracker. Thereafter, they completed a collocation test and a caption recall test. The results showed that the baseline and enhancement groups did not vary in their caption-reading behaviors. Conversely, the enhancement group significantly outperformed the baseline group on the collocation test. In the caption recall test, the enhancement group recalled significantly more target collocations than the baseline group, whereas the two groups did not differ in recalling unenhanced captions. Finally, correlational analyses revealed nonsignificant correlations between attention to target collocations and collocation test scores in both groups. This evidence suggests that enhanced video captions may be an effective means of stimulating collocational competence that is not at the expense of second language learners' ability to learn video content.

Keywords: *Caption, Collocation, Eye Tracking, Visual Input Enhancement*

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

APA Citation: Choi, S. (2023). Visual saliency in captioned digital videos and learning of English collocations: An eye-tracking study. *Language Learning & Technology*, 27(1), 1–21.
<https://hdl.handle.net/10125/73536>

Introduction

The way in which visual input enhancement in written input affects second language (L2) acquisition has become a topic of interest among many researchers. Visual input enhancement, a type of constructed salience (i.e., salience created by an extrinsic source), refers to the typographical modification of target linguistic features (e.g., grammar and collocations) using various cues, such as **boldfacing**, *italicizing*, color-coding, and CAPITALIZING (for more information regarding *constructed salience*, refer to Gass et al., 2017). It has generally been assumed that the perceptual saliency afforded by such modifications implicitly encourages L2 learners to pay additional attention to enhanced linguistic forms, which generates more favorable learning outcomes. This assumption is endorsed by the literature on the Focus-on-Form approach and the Noticing Hypothesis, which highlight the critical role of attention in language learning (Schmidt, 2001).

Over the past three decades, a plethora of empirical studies have been conducted to determine the degree to which enhanced written input impacts the acquisition of grammatical forms and lexical knowledge (e.g., Lee & Révész, 2020). Earlier studies on visual saliency have focused on the learning of various types of grammatical forms and have reported mixed results: while some studies have reported the positive effects

of visual input enhancement on grammar learning (Lee & Révész, 2020), others have affirmed null (Winke, 2013) or mixed (Cintrón-Valentín et al., 2019; Cintrón-Valentín & García-Amaya, 2021) effects of enhanced input in grammar acquisition.

More recently, several researchers have shifted their attention from grammar learning to the learning of multi-word units, such as collocations (e.g., Choi, 2017). Collocations refer to the frequently recurring combinations of two or more words (e.g., *make money* or *burst into tears*), and the knowledge of these combinations is crucial in the development of receptive and productive language skills (Benson et al., 1997). Unfortunately, collocations are difficult to acquire for L2 learners as they tend to process co-occurring combinations as separate units, rather than single units (Wray, 2002). Although the number of investigations on the role of visual input enhancement in collocation learning is limited, the findings thus far have consistently given support to using visual input enhancement to nurture collocational competence in L2 learners (e.g., Peters, 2012). These positive outcomes suggest that visual input enhancement may help L2 learners to strengthen associations between constituent words (Choi, 2017).

The present study expands upon this line of inquiry by examining whether the benefits of visually enhanced input on collocation learning extend to enhanced captions added to digital videos. Addressing this point is imperative because L2 learners access a wide selection of digital videos which can afford authentic language learning contexts. Specifically, this study primarily aims to elucidate (a) how visually enhanced captions (i.e., captions showing target collocations in yellow font) affect caption-reading behaviors and the acquisition of English collocations, and (b) how caption-reading behaviors correlate with collocation learning. As in Choi (2017), this study also explores whether a tradeoff exists between the recall of enhanced target collocations and unenhanced captions as a function of visual input enhancement. That is, this study also explores whether visual input enhancement may foster the memory of enhanced collocations at the expense of unenhanced caption contents.

Owing to the different input modalities between the previous (unimodal text) and the present (multimodal video) studies, it does not necessarily follow that enhanced captions would lead to similar findings as those demonstrated in enhanced text. Reading enhanced text appears to differ substantially from reading enhanced captions and may entail distinct strategic processes (Montero Perez, 2019). For instance, when reading an enhanced text, learners engage in one type of information processing. In other words, learners can allocate most of their attentional resources to creating a coherent mental representation of the text and at least some resources to processing enhanced targets. Conversely, when reading enhanced captions, the same learners may have difficulty paying attention to captions because the captioned video includes concurrent, multisensory input streams that have visual (images and captions) and verbal (speech and sound/audio effects) information, which compete for limited attentional resources. Thus, learners must scan or split their visual attention between images and onscreen captions, rendering them less capable of paying attention to the captions. If they attempt to process concurrent multiple modalities simultaneously, then the processing demands of visual memory may surpass their processing capacity, resulting in cognitive overload in these learners, one of the central challenges of using captioned videos, as argued in Mayer and Moreno (2003).

This study is the first eye-tracking study to empirically investigate the role of visual input enhancement in video-based collocation acquisition. Accordingly, the findings may fill an important gap in research on visual input enhancement. Furthermore, the results have essential pedagogical implications for video-based language learning and the design of video materials for L2 learners.

Literature Review

Empirical Studies on Visual Input Enhancement and L2 Acquisition

One key question in L2 acquisition research is how to effectively draw the attention of L2 learners to linguistic elements as they read a text (Choi, 2017). While reading, learners typically utilize most of their cognitive resources to make sense of the text and create a coherent mental representation of it. This, in turn,

may leave few attentional resources available for processing and learning discrete linguistic elements.

Numerous studies on L2 acquisition have delved into whether visual input enhancement can be used as an effective tool to compensate for the constraints of meaning-oriented input (e.g., Lee, 2007). The review of these studies indicates mixed findings in terms of grammatical forms (e.g., Lee, 2007; Winke, 2013). For instance, in Lee (2007), Korean high-school students ($N = 259$) read two English passages containing either enhanced or unenhanced passive constructions and responded to post-reading tasks (a form correction task and a free-recall task). In the form correction task, participants were instructed to provide correct forms to erroneous sentences, whereas in the free-recall task, they were instructed to list all the ideas they could recollect from experimental passages. According to the form correction results, the participants who read enhanced passages performed significantly better than those who read unenhanced passages, and the magnitude of the difference was large ($d = 1.02$). Conversely, the participants exposed to enhanced passages recalled significantly less of the unenhanced text content than those presented with unenhanced passages; the magnitude of the effect was small-to-medium sized ($d = .57$). The detrimental effects of visual input enhancement on text recall were also demonstrated in Overstreet (1998) and Choi (2017), whereas such negative effects have not been observed in other studies (Leow et al., 2003).

Contrarily, Winke (2013), in a partial replication of Lee (2007), reported that visual input enhancement failed to facilitate form learning, although visual input enhancement significantly drew the attention of L2 learners to the target forms (passive construction). Concerning this finding, the author concluded that “enhanced forms may attract attention but fail to be processed more highly” (Winke, 2013, p. 341). Furthermore, Izumi (2002) concluded, “visual input enhancement encourages sensory detection by means of highlighting, but this does not necessarily lead to integrative processing. The automatic link from sensory detection to further processing cannot be assumed” (p. 572).

In terms of collocation learning, empirical studies to date have yielded favorable findings (e.g., Choi, 2017; Peters, 2012). Peters (2012) was among the first researchers to explore the effects of visual input enhancement on collocation learning. In her small-scale study, 28 undergraduate students read a German text embedded with 12 German collocations and 12 German words. Half of the target collocations and words were enhanced (boldfaced and underlined), whereas the rest were not enhanced. The analyses of immediate recall scores revealed that the participants recalled significantly more enhanced collocations ($M = 2.29$) than unenhanced ones ($M = 1.11$).

Expanding upon previous collocation studies that have focused on offline measures (e.g., Peters, 2012), Choi (2017) utilized online (eye-movement recording) and offline behavioral measures to unearth the role of enhancement in collocation learning. Korean undergraduate students ($N = 38$) were instructed to read a text that contained either unenhanced or enhanced (boldfaced) English collocations. The eye-tracking results confirmed that the participants who read the enhanced text fixated significantly longer on the target collocations than those who read the unenhanced text. The collocation test results also validated that the enhancement group significantly outperformed the baseline group. In terms of the recall of the text, there was a significant tradeoff between the recall of enhanced collocations and unenhanced text. In the recall of enhanced collocations, the enhancement group outperformed the baseline group, whereas in the recall of text content, the baseline group displayed a superior performance.

Empirical Studies on Captioning Benefits

Captions, or the verbatim transcription of auditory information, play a crucial role in our lives (*subtitles*, written first-language (L1) renderings of audio input, are not the focus of this study). Primarily targeted to individuals with hearing impairments, captions are of particular use to these individuals because they ensure the accessibility of the audio content of videos. Moreover, captions boost accessibility to L2 input because they provide L2 learners with all the auditory information needed for those who may experience difficulties decoding or processing speech components in media. Many studies have demonstrated robust evidence that the synchronized presentation of spoken and written input is more likely to lead to comprehensible input,

which is a critical component in L2 development (Montero Perez et al., 2013).

Furthermore, captioned videos have been found to be superior to non-captioned videos in terms of the initial stage of incidental vocabulary acquisition (e.g., Montero Perez et al., 2014; Rodgers & Webb, 2019; Sydorenko, 2010). Neuman and Koskinen (1992) were among the first researchers to report the benefits of captions in L2 vocabulary acquisition. The results showed that the participants in the captioned television condition made significantly higher vocabulary gains than those in the other three conditions.

Employing a small study sample, Sydorenko (2010) examined the effects of three different input modalities (video + audio vs. video + captions vs. video + audio + captions) on the vocabulary acquisition of adult L2 learners. The participants were undergraduate students in the United States who had been studying Russian as a foreign language. The participants watched an experimental video that corresponded to their condition and completed post-viewing tests (e.g., word recognition and vocabulary knowledge tests). The results revealed that the captioned video was superior for the recognition of visual word forms and word meanings.

Expanding upon earlier studies, Winke et al. (2010) compared the effects of captioning and its ordering effect. From 151 undergraduate participants, 67 Spanish participants were assigned to one of the following conditions: (a) viewing twice without captions, (b) viewing twice with captions, (c) viewing first without captions and viewing again with captions, and (d) viewing first with captions and viewing again without captions. The participants watched three documentary videos (3–5 min in length) about three distinct animals (bears, dolphins, and salmon). Subsequently, they completed two types of vocabulary tests: one test with written input (target words presented on an answer sheet) and another with aural input (i.e., target words presented aurally). The results confirmed that regardless of the order in which they watched the videos, those who watched captioned videos scored significantly higher on the vocabulary test than those who watched the same videos without captions and that the magnitude of the difference was large as measured by Cohen's *d*.

Montero Perez et al. (2014) explored how three types of captions (full captions, keyword captions, and full captions with highlighted keywords) affected L2 vocabulary acquisition in Flemish undergraduate students ($N = 133$). The results elucidated that although the three caption groups performed significantly better than the control group in the measures of form recognition, no significant differences were found in meaning recall. These results suggest that the benefits of captions may vary depending on various aspects of word knowledge.

While the aforementioned studies focused on the effects of short videos, two recent studies have investigated the effects of a full-length TV series and a full-length documentary on incidental vocabulary acquisition (Peters & Webb, 2018; Rodgers & Webb, 2019). Although the findings of these studies parallel those of previous ones, the rate of retention was generally low. For instance, Rodgers and Webb (2019) reported (out of 15 target words) 2.97 words in the multiple-choice form recognition test and .31 words in the meaning recall test.

Other lines of research have indicated that the benefits of captions may be modulated by various learner factors, including age; language proficiency and concomitant lexical knowledge (Gass et al., 2019); instructional intervention, such as pre-learning vocabulary prior to viewing a captioned video (Montero Perez, 2019); caption characteristics, such as frequency of word occurrence and cognate status (Peters & Webb, 2018); prior vocabulary knowledge (Montero Perez, 2020); memory span (Gass et al., 2019; Montero Perez, 2020); and content familiarity (Winke et al., 2013). The present study carefully controlled for various factors, such as age, language proficiency, prior vocabulary knowledge, and content familiarity (refer to the Methods section).

Eye-tracking Research on Caption-reading Behaviors

Using near-infrared light and high-resolution cameras, an eye tracker records moment-to-moment eye movements and measures what a learner is looking at and how long that learner gazes at a region. Thus, the eye tracker has been considered an objective, direct, and real-time measure of visual attention (Conklin &

Pellicer-Sánchez, 2016). On the basis of these eye-tracking benefits, eye-tracking research on L2 processing and acquisition has been growing in popularity over the past decade, particularly in the area of incidental vocabulary acquisition (e.g., Pellicer-Sánchez, 2016). The eye tracker can also be fruitfully applied to understanding captioning effects on caption-reading behaviors and the subsequent acquisition of linguistic items, such as grammar and collocations.

Although eye-tracking research on caption reading remains sparse, we have witnessed an increasing number of eye-tracking studies that have explored factors that might modulate caption-reading behaviors and subsequent learning outcomes (e.g., Muñoz Lahoz, 2017; Winke et al., 2013). Some researchers have shown interest in whether differences exist in the processing of captions (L2 text) and subtitles (L1 text; Bisson et al., 2014) and whether caption-reading behaviors may vary as a function of age (Muñoz Lahoz, 2017). Winke et al. (2013) focused on the potential L1 influence (effects of different writing systems) on L2 caption-reading behaviors. Montero Perez et al. (2015) explored how test announcement (forewarning or not forewarning the participants of forthcoming vocabulary tests) affected the processing of novel words according to caption type (keywords vs. full captions). They also explored the associative links between attention and vocabulary gains depending on the types of assessment (meaning recognition and form recognition) and caption (full captions vs. partial captions).

Lee and Révész (2018) conducted the first and only eye-tracking study that tested the role of visually enhanced captions added to still images. Specifically, these researchers supplied enhanced (boldfaced) captions in static illustrations and examined how this enhanced input affected the learning of pronominal anaphoric references (e.g., **My grandmother** always wanted to be a teacher when **she** was a little girl). The participants were Korean undergraduate students who were randomly assigned to either a baseline or an enhanced caption condition and had three treatment sessions while their eye movements were measured by an eye tracker. The analyses of the eye movements and two grammaticality judgment tests (written and oral) revealed that relative to unenhanced captions, enhanced captions drew more attention to target forms (anaphora antecedents) and promoted the learning of pronominal anaphoric references. According to correlational analyses, a significant relationship between attention and grammar learning was found only in the baseline (unenhanced) group. Subsequently, Lee and Révész (2020) explored how enhanced captioned videos influence the attention of L2 learners and the learning of two English constructions (past simple and present simple). The results were well aligned with those of Lee and Révész (2018).

In sum, the current review indicates that despite substantial interest in the role of visually enhanced input in L2 acquisition, little is known about the role it plays in collocation learning while L2 learners view enhanced captioned videos. Considering that audiovisual materials are typically unenhanced, addressing this problem is critical. In the present study, the eye movements of Korean undergraduate students were measured using an eye tracker while they viewed a baseline (unenhanced) captioned video or the same video with enhanced captions (i.e., captions displaying target collocations in yellow font). Thereafter, they completed a filler task, followed by an unannounced collocation test and a caption recall test. The specific research questions that guided this study were the following. Concerning the four research questions, no a priori predictions were formulated because different types of input may entail distinct processing strategies and lead to different learning outcomes, as discussed previously.

1. To what degree does visual input enhancement affect time spent reading target collocations, sentences that embed the targets, and the entirety of captions?
2. To what degree does visual input enhancement affect the learning of target collocations, as measured by a collocation test?
3. Is there a tradeoff between the recall of enhanced target collocations and unenhanced captions? In other words, do L2 learners remember more enhanced collocations at the expense of the recollection of unenhanced captions?
4. What is the relationship between the visual attention allocated to target collocations and collocation learning? For instance, if L2 learners spend more time reading target collocations, are they more

likely to learn them better?

Methods

Participants and Research Design

The sample of this study comprised 53 undergraduate students in Korea studying English as a foreign language. They have been studying English as a foreign language for approximately 10 years and pursue diverse majors (legal studies, biology, engineering, and accounting). Their reading proficiency was approximated to be that of eighth graders in the United States, as measured by the Nelson–Denny Reading Test (NDRT) ($M = 26.24$, $SD = 4.24$). Therefore, the participants were deemed high-intermediate learners of English. Each participant was compensated with approximately \$30.00 USD for their time. The purpose of this study was not disclosed to the participants until the termination of data collection to eliminate the risk of knowledge of the research purpose possibly confounding the findings. The research protocols strictly followed the ethical guidelines mandated by the Institutional Review Board of the affiliated institution.

This study utilized a between-subjects design. The independent variable, which also served as the between-subjects variable, was “GROUP,” featuring the following two groups: the baseline group ($n = 27$; 10 males and 17 females) and the enhancement group ($n = 26$; 11 males and 16 females). The participants were randomly assigned to one of the two groups. The baseline group viewed a digital video with unenhanced captions, whereas the enhancement group watched the same video with enhanced captions (i.e., target collocations made visually salient using yellow font [see [Figure 1](#)]). There were multiple dependent variables, including the eye-fixation indices and the scores obtained from the collocation and caption recall tests.

Given that the ages of learners may have mediated their caption-reading behaviors and subsequent language learning (Muñoz Lahoz, 2017), this factor was controlled for herein. The mean ages between the baseline and enhancement did not differ significantly ($t = .126$, $p = .900$). Subsequently, to ensure homogeneity between the baseline and enhancement groups in terms of English proficiency and general knowledge of English collocations, two tests were administered: (a) the NDRT (Form H) and (b) a cloze test. The NDRT includes 38 reading comprehension questions in multiple-choice format and has average internal consistency coefficients ranging from 0.85 to 0.95. The NDRT has been widely utilized by researchers to assess the English language or reading proficiency of participants (e.g., O’Reilly & McNamara, 2007). The cloze test measured the general collocational competence of the participants (see [Appendix B](#)). It included 18 test items (nine lexical and nine grammatical collocations) with their frequency ranging from 21 to 37207, according to the Corpus of Contemporary American English (Davies, 2008). None of these items appeared in the post-viewing collocation test (see [Appendix C](#)).

Table 1

Descriptive Statistics of the Baseline and Enhancement Groups (SDs in Parentheses)

Group	Age	Reading proficiency	Collocational competence	Prior knowledge of target collocations
Baseline	22.42 (1.44)	25.70 (4.05)	9.59 (3.05)	5.41 (1.22)
Enhancement	22.37 (1.59)	26.61 (4.45)	10.30 (2.89)	5.38 (1.18)

The results from the NDRT showed that the baseline and enhancement groups did not significantly differ in their language proficiency ($t_{51} = .946, p = .349$). Likewise, the analyses of the cloze test scores confirmed that the baseline and enhancement groups did not significantly differ in their general knowledge of English collocations ($t_{51} = .874, p = .386$).

The prior knowledge of target collocations was measured using a collocation knowledge survey administered during the final data collection phase. Employing the survey in lieu of a collocation pretest was critical, because a pilot study involving different participants exhibited that the collocation pretest induced the participants to allocate extra attention to the targets during video watching, thus rendering the eye-tracking data less reliable. In the survey, the participants were shown a list of all target collocations and were asked to select all the word combinations that they were already aware of prior to participating in the present study. As summarized in [Table 1](#), the survey results confirmed that the baseline and enhancement groups rarely differed in their knowledge of targets prior to participating in the present study ($t_{51} = .104, p = .917$).

Materials

The major materials comprised (a) a video with captions, (b) a post-viewing collocation test, and (c) a caption recall test. The video was obtained from Technology, Entertainment, and Design (TED) talks and served as the stimulus of an SMI RED-500 eye tracker (Teltow, Germany), a desk-mounted binocular remote eye tracker sampling at 500 Hz. Following Choi (2017), the collocation test was designed to measure the acquisition of target collocations of the participants, whereas the caption recall test measured their memory of captions (enhanced target collocations and unenhanced captions).

Selected Video and Captions

To increase the ecological validity of this study, an unmodified video from TED talks was selected. The selected video features a speaker, Rebeca Hwang, giving a talk entitled *The Power of Diversity within Yourself*. The talk is about how difficult her life was after her immigration to the United States and how she overcame major hurdles and became successful later. Therefore, an understanding of the narration did not require culture- or discipline-specific knowledge (refer to the transcription of the video in [Appendix A](#)).

The length of the video was 6 min 47 sec; the speech rate was 135.55 words per minute (919 words; 58 sentences). The Flesh–Kincaid Grade Level was 7.3, meaning that the comprehensibility level of the talk (i.e., captions) approximated that of native English speakers who have received seven years of schooling. These indices reveal that the linguistic difficulty level of the video was appropriate for the participants.

As shown in [Figure 1](#), there were two versions of this TED talk video: the baseline version (unenhanced captions) and the enhanced version (captions including enhanced collocations). These two versions were identical, except the target collocations were made visually salient only in the enhanced version. Unlike Lee and Révész (2018) who used boldface font to enhance target grammatical units, color font was utilized in this study because using boldface font may lead to slight differences in the size of the regions of interest (ROIs).

Figure 1

Image Samples of Baseline (Left) and Enhanced (Right) Videos



For both versions, full English captions were added using encoding software. As noted above, the full captions included 18 target grammatical (a combination of function and content words) and lexical (a combination of function and content words) collocations with a frequency ranging from 5 to 17105, according to the Corpus of Contemporary American English (Davies, 2008): (a) eleven lexical collocations (*dying wish, secret weapon, a big deal, make a living, social innovator, rapidly changing, reinvent myself, switch majors, lifelong quest, average income, and get a tutor*) and (b) seven grammatical collocations (*sense of fashion, drop out, stare at, ring in my ears, in conflict with, become addicted to, and upon arrival*). Both types of collocations are included in the present study because they play a significant role in receptive and productive language skills (Benson et al., 1997). Most of these collocations were presented in the video once; the only exception was *average income* that appeared twice. Each scene of the video included a single-line caption displayed at the bottom of the screen. The font used was a proportional 40-point Arial. In the caption, there was no fixed location of the target collocations—that is, the target collocations appeared according to the flow of the narration.

Collocation Test

The collocation test assesses how enhanced captions affect the learning of target collocations. In the test, there were 18 cloze-format items. Each item was composed of a sentence in Korean (e.g., 나는 전공을 물리학에서 화학으로 바꿨다) and an English translation equivalent with one blank to fill in (e.g., I s_____ my major from physics to chemistry; *switched* is the correct answer). As the English sample sentence illustrates, as in previous studies (e.g., Choi, 2017; Peters, 2012), the initial letter was provided in some cases to avoid accurate alternatives. The given blank in each multi-word combination represents a critical component to be acquired by L2 learners.

Caption Recall Test

The caption recall test is designed to determine whether there is a tradeoff between the memory of enhanced collocations and the memory of unenhanced captions. In this test, the participants were provided with an A4 sheet that displayed the entirety of the captions (i.e., transcribed text) with 28 blanks to fill in: 18 blanks for the target collocations and 10 blanks for the content words of unenhanced captions. The following is an excerpt. The participants were expected to fill in the first blank with *quest* (target collocation) and the second with *identity* (unenhanced captions).

It made me ponder where I would want to be buried someday, where home was for me, and the answer was not obvious. And this really bothered me. So this episode launched a lifelong _____ for my _____.

Procedure

The experiment was conducted individually in a sound-attenuated booth, in two phases over a three-week

period. In the first phase (1st week), the participants submitted a consent form, filled out a survey concerning their age, language background, and language proficiency, and completed two tests (a collocation competence test and the NDRT). In the second phase (3rd week), the participants viewed a video that corresponded to their condition, while their eye movements were registered by an eye tracker (SMI RED-500, Teltow, Germany). Specifically, for the eye-tracking data collection, each participant sat comfortably in a chair approximately 70 cm from the stimulus monitor and the eye-tracking cameras. To minimize head movements, all the participants used a head rest. They were told that they would view a TED talk video for approximately seven minutes and were instructed to view the video for meaning and comprehension. The forthcoming tests (the collocation and caption recall tests) were not announced to the participants to create an incidental learning context. While conducting calibration, they were informed that their eye movements would be measured by an eye tracker. No additional recalibration was performed until each participant had completed their viewing. The participants then completed a filler task (simple math quizzes) to reduce the potential effects of short-term memory. Finally, they completed a collocation test, followed by a caption recall test and a collocation knowledge survey.

Analyses of the Eye-tracking Data

The eye-fixation data were analyzed using BeGaze (Teltow, Germany) analysis software. The primary metric used to analyze the eye-movement data was total fixation time, which refers to the sum of all eye fixations made in ROIs. Additionally, fixation counts, which refer to the total number of eye fixations (i.e., visits) made within a given ROI, were computed.

There were four ROIs: target collocations, sentences that embed target collocations, sentences that do not include target collocations, and the entire captions (all 58 sentences). Specifically, *target collocations* are the main ROI. Two additional ROIs (*sentences that embed target collocations* and *the entire caption area*) were also analyzed as longer fixation times on target collocations may also increase overall reading times for these two areas. Finally, as a control measure, *sentences that do not include target collocations* were also analyzed. That is, if the baseline and enhancement groups reveal significantly different reading times for this ROI (sentences that do not include target collocations), the difference may need to be considered while explaining the eye-movement data obtained from the remaining three ROIs.

The original dataset included 61 participants. Nonetheless, through visual inspection, the eye-fixation data from four participants were eliminated due to excessive drift. Additionally, the data from four participants were excluded because the tracking ratio of these participants was less than 80%. Therefore, the final dataset involved data from 53 participants. The tracking ratio is defined as “the proportion of time that the eye tracker recorded point of gaze coordinates over the entire task” and is used to further assess data quality (Amso et al., 2014, p. 2; Kang et al., 2022).

Scoring

The collocation and recall test results were scored by two independent raters. One point was awarded for each correct response (i.e., exact forms) and for minor errors (e.g., 1–2 spelling errors, such as “inovator”). Inter-rater reliability coefficients in the collocation and recall tests were $r = .99$ and $.98$, respectively. Any discrepancies between the two raters were settled through discussion. Additionally, the two raters double-checked the scoring independently.

Results

Eye-tracking Results

The eye-tracking results are presented according to the four aforementioned ROIs. As mentioned before, the primary area of interest was the target collocations.

Target Collocations

As summarized in Table 2, the enhancement group ($M = 7092.11$ ms) spent a longer amount of time reading the target collocations than the baseline group ($M = 6252.96$ ms). Nevertheless, an independent-samples t -test revealed that the difference between the two groups ($M = 839.14$ ms) was not statistically significant ($t_{51} = .613$, $p = .543$) and showed a small effect size of $d = .17$ (Plonsky & Oswald, 2014). Similarly, the baseline ($M = 81.23$ ms) and enhancement ($M = 81.00$ ms) groups did not differ in terms of the number of visits to the target collocations ($t_{51} = .386$, $p = .701$, $d = .01$).

Table 2

Descriptive Statistics of the Four Regions of Interest (SDs in Parentheses)

Group	Target collocations		Sentences with target collocations		Sentences without target collocations		Entire caption area	
	TRT	FC	TRT	FC	TRT	FC	TRT	FC
Baseline	6252.96 (4587.77)	37.66 (26.34)	13308.73 (9057.12)	81.00 (53.87)	78265.60 (53942.23)	486.40 (320.50)	91574.34 (62897.18)	567.40 (373.29)
Enhancement	7092.11 (5360.32)	40.53 (27.81)	13796.81 (10004.66)	81.23 (54.59)	78435.94 (56610.09)	469.57 (313.29)	92232.75 (66497.38)	550.80 (367.30)

Note. TRT = total reading time in milliseconds, FC = fixation count

Sentences with Target Collocations

All the sentences that included target collocations were analyzed to examine whether visual input enhancement led to increased reading time for the sentences that included target collocations. The results confirmed that the enhancement group ($M = 13796.81$ ms) spent a little more time than the baseline group ($M = 13308.73$ ms) reading sentences that included target collocations. However, the difference of the means was not significant ($t_{51} = .186$, $p = .853$). Likewise, the baseline ($M = 81.00$) and the enhancement ($M = 81.23$) groups rarely differed in terms of fixation counts, and the difference between the two groups was not significant ($t_{51} = .015$, $p = .988$).

Sentences without Target Collocations

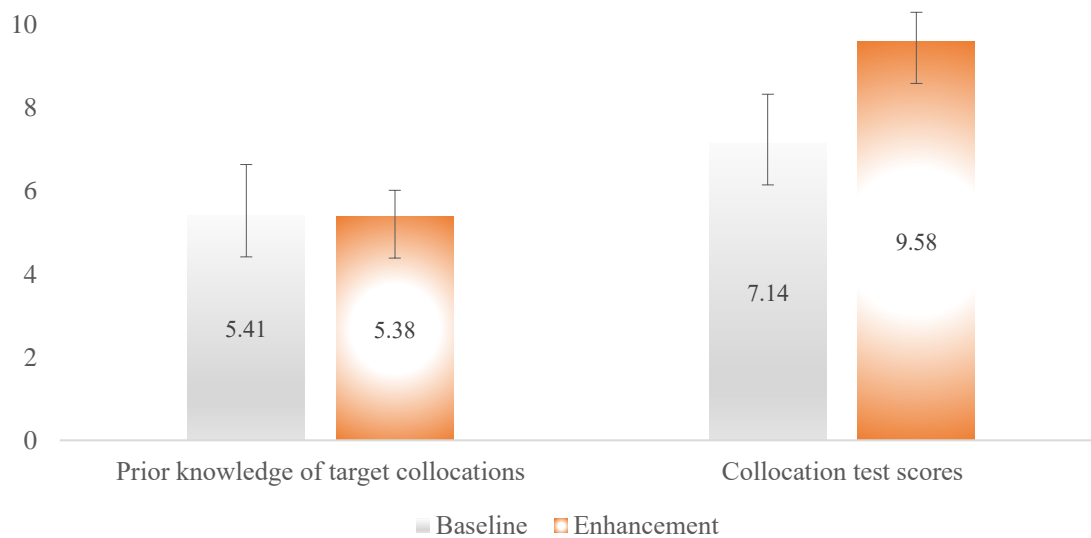
In addition to captions that included target collocations, sentences that did not include target collocations were analyzed. The results validated those participants in the baseline ($M = 78265.60$ ms) and the enhancement ($M = 78435.94$ ms) conditions spent similar amounts of time reading sentences without target collocations. An independent-samples t -test showed that the difference of means between the two groups did not reach statistical significance ($t_{51} = .011$, $p = .991$). Likewise, the analyses of the eye-fixation counts revealed that fixations on the captions embedding target collocations between the two groups were not significantly different ($t_{51} = .193$, $p = .848$).

Entire Caption Area

The last ROI comprised the entire onscreen caption area. The results elucidated that the participants in the baseline and enhanced conditions spent a similar amount of time reading the captions ($M = 92232.75$ ms, 91574.34 ms, respectively). An independent-samples t -test displayed that the two groups did not significantly differ in the amount of time spent reading the entire caption area ($t_{51} = .037$, $p = .971$). Similarly, the two groups did not differ significantly in terms of their fixation counts ($t_{51} = .163$, $p = .871$).

Figure 2

Prior Knowledge of Target Collocations and Collocation Test Scores for the Baseline and Enhancement Groups



Note. The vertical bars (error bars) indicate the standard error of means.

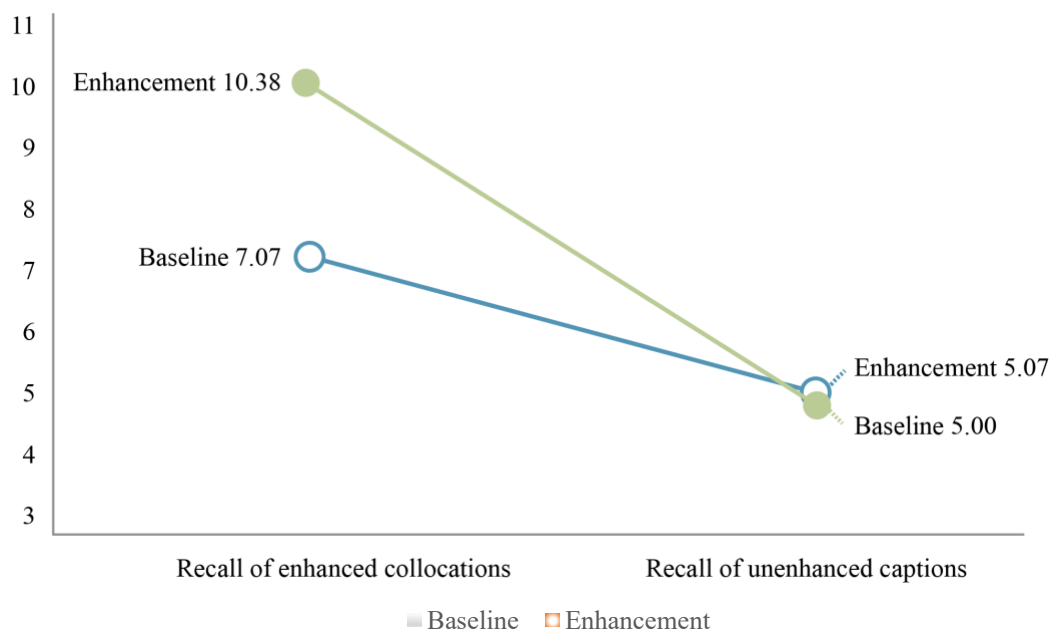
Collocation Test Results

Figure 2 illustrates that the enhancement group ($M = 9.58$, $SD = 3.21$) outperformed the baseline group ($M = 7.14$, $SD = 3.73$) on the collocation test. An independent-samples t -test was run, and the results showed that the difference of means ($M = 2.44$) between the two groups reached statistical significance ($t_{51} = 2.532$, $p = .014$). Cohen's d was computed to examine the magnitude of difference between the two groups. The results revealed a medium effect size of $d = .70$.

Given the significant interaction, two separate independent-samples t -tests were run for the recall of enhanced targets and the recall of unenhanced caption contents, respectively. In terms of the recall of target collocations, the participants in the enhancement group ($M = 10.38$, $SD = 3.78$) recalled more target collocations than their counterparts in the baseline group ($M = 7.07$, $SD = 4.51$), and the difference of means between the two groups was statistically significant ($t_{51} = 2.891$, $p = .006$) with a medium-to-large effect size ($d = .80$). By contrast, in terms of the recall of unenhanced captions, the baseline ($M = 5.00$, $SD = 1.65$) and enhancement ($M = 5.07$, $SD = 1.73$) groups rarely differed ($t_{51} = .159$, $p = .874$, $d = .04$).

Figure 3

Recall of Enhanced Collocations and Unenhanced Caption Contents



Correlation Analyses

The correlational analyses primarily aimed to determine the degree of association between the attention of the participants to the target collocations and collocation test scores. Thus, a series of Pearson correlation coefficients were calculated separately for the baseline and enhanced captions groups (see Tables 3 and 4). In the baseline group (Table 3), the correlation coefficients between the eye-tracking indices and the test scores (collocation and recall tests) mostly dropped into the nonsignificant, negative territory. Conversely, in the enhancement group (Table 4), the correlations between the eye-tracking results and the test scores (collocation and recall tests) were all in the nonsignificant, positive territory.

Table 3

Relationship between Eye-tracking Measures and Collocation Test Scores in the Baseline Group

Variable	1	2	3
1. TRT on target collocations	–	.804**	–.334
2. FC on target collocations		–	–.315
3. Collocation test scores			–

Note. TRT = total reading time; FC = fixation counts

* $p < 0.05$, ** $p < 0.01$

Table 4*Relationship between Eye-tracking Measures and Collocation Test Scores in the Enhancement Group*

Variable	1	2	3
1. TRT on target collocations	–	.967**	.034
2. FC on target collocations		–	.097
3. Collocation test scores			–

Note. TRT = total reading time; FC = fixation counts

* $p < 0.05$, ** $p < 0.01$

Discussion

This study investigated how visual input enhancement impacts caption-reading behaviors, the learning of English collocations, and the recall of enhanced targets and unenhanced captions. To this end, online (i.e., eye tracking) and offline performance measures (i.e., paper-based tests) were utilized. The important findings of this study are discussed by relating them to the research questions posed herein.

The first research question concerned how visual input enhancement affects caption-reading behaviors. To answer this question, the eye-movement data of the baseline and enhancement groups were compared in the following four ROIs: target collocations, sentences that include target collocations, sentences that do not include target collocations, and entire caption area (all sentences). The results confirmed that the baseline and enhancement groups exhibited similar caption-reading behaviors in all ROIs. For instance, although the enhancement group spent a little more time processing the enhanced targets than the baseline group, the difference between the two groups did not reach statistical significance, indicating a small effect size ($d = .17$). Additionally, the two groups rarely varied in the amount of time spent processing the entirety of the captions ($d = .01$).

The absence of increased attention to enhanced collocations in the video captions contrasts sharply with Choi (2017), who reported that the participants in the enhanced condition fixated significantly longer on target collocations than their counterparts in the baseline condition. This contradictory finding may be attributed to using different input modalities, as previously noted. Specifically, the participants in Choi (2017) were exposed to a unimodal text, thereby having relatively more attentional resources available for processing enhanced collocations. Conversely, this study used a captioned video and included not only verbal information (speech sounds) but also two types of concurrent visual information (images and captions), which may have exceeded the processing capacity of the visual channel or the visuospatial sketch pad (Mayer & Moreno, 2003). Thus, the participants in this study had to split their limited visual attention between captions and visual images (Sweller, 1988), therefore, they may have had relatively fewer attentional resources available for the enhanced targets. Additionally, the participants in Choi (2017) were allowed to read the experimental text at their own pace, meaning that they were free to revisit the enhanced collocations as many times as they needed. By contrast, the participants in this study had to follow a speedy display of moving captions (135.55 words per minute) and were not allowed to pause or rewind the video to revisit parts of the captions.

The second research question concerned how visual input enhancement affects the acquisition of English collocations. The analyses of the test scores showed that the enhancement group scored 34.90% higher than the baseline group, and the magnitude of the difference between the groups showed a medium effect size ($d = .70$). The positive effects of visual input enhancement on collocation learning replicated those of

previous studies that have examined the effects of enhanced unimodal text on the learning of English collocations (e.g., Choi, 2017; Peters, 2012) and the role of enhanced captions added to still images in English grammar learning (Lee & Révész, 2018). Therefore, based on the present work and previous studies, it can be suggested that enhanced input may positively affect the incidental learning of English collocations in both single and multiple modalities.

The superior performance of the enhancement group over the baseline group is noteworthy because the baseline and enhancement groups did not differ significantly ($p = .543$, $d = .17$) in the amount of attention allocated to processing the target collocations. Accordingly, it can be suggested that visual attention quantified using eye-tracking measures may not adequately reflect the conscious registration of target linguistic items for further elaborate processing and eventual learning, calling for the use of a combination of the eye-tracking paradigm with complementary research methods that could refine the current findings. This interpretation is in line with Montero Perez et al. (2015) and Lee and Révész (2018), who posited that longer fixations may not reflect more elaborate processing. If this interpretation is correct, then the current findings may only partially support research on the Focus-on-Form approach and the Noticing Hypothesis, which underscore drawing the attention of learners and raising consciousness of linguistic features during meaning-focused activities (Schmidt, 2001). Alternatively, although the collocation and recall tests were not announced to the participants, perceptually salient input may have insinuated that there would be some sort of test after they finished viewing the video. Consequently, the participants in the enhanced condition may have been more conscious of and sensitive to the presence of multi-word combinations and may have stored these constituent words as a combined entity in their mental lexicon (Wray, 2002). The speculations, however, are tentative and should be further elucidated by other research methods, such as through verbal protocols or interviews (e.g., Montero Perez, 2019).

The third research question examined whether L2 learners would remember more enhanced collocations at the expense of the recollection of unenhanced captions. The results indicated that the enhancement group recalled significantly more target collocations than the baseline group, whereas the two groups did not differ in their recollection of unenhanced captions, indicating no debilitating effects of visual input enhancement on the recall of unenhanced captions. These results are supportive of the use of visual input enhancement as an implicit, unobtrusive, focus-on-form technique (Schmidt, 2001) and are in line with those of several earlier studies (Leow et al., 2003). Nevertheless, the nonsignificant tradeoff between the recall of enhanced target collocations and unenhanced captions as a function of visual input enhancement runs counter to the findings of other previous studies (Choi, 2017; Lee, 2007; Overstreet, 1998). The nonsignificant tradeoff may be explained by the current finding that the baseline and enhancement groups spent similar amount of time reading caption contents. Nonetheless, as stated above, future studies that integrate eye-tracking and additional measures (e.g., verbal reports) are warranted to elucidate this speculation.

Finally, the last research question probed the relationship between visual attention and collocation learning under two distinct conditions (baseline and enhanced captions). The results confirmed that the associations between attention and collocation learning were nonsignificant in the baseline and enhancement groups. However, a closer examination indicated that the direction of association appeared to differ as a function of visual input enhancement. In the baseline condition, nonsignificant, *negative* associations were found. By contrast, in the enhancement condition, nonsignificant, *positive* links were found between attention and the collocation test scores. Although it is a delicate matter to compare these findings with those of previous studies, these results are along the lines of Montero Perez et al. (2015), who reported that the links between attention and vocabulary gains differed as a function of caption type (full captions vs. partial captions). However, concerning the various correlational patterns, no logical explanations are put forth, which merits further research.

Despite the novel findings, this study seems to raise more questions than it answers. First, it was conducted in a laboratory setting. Therefore, future studies should consider non-laboratory experiments wherein learners can navigate audiovisual materials backward and forward at their own pace because such behaviors may more accurately reflect the natural video viewing of many L2 learners. Given the potentially more

frequent use of captioned videos in out-of-classroom settings rather than in-classroom settings, addressing this issue is critical. Second, if the speech rate had been faster (or slower) and if the onscreen text had been presented at a faster (or slower) rate, the findings might have differed. Third, the font size used for captions was 40, which may not reflect actual or optimal font sizes used in diverse video materials. Finally, due to the different format used in the pretest (collocation knowledge survey) and posttest (collocation cloze test), the results elicited from a direct comparison between these two types of the tests remain questionable. These issues should be examined in future studies.

Conclusion

To reiterate, the results revealed that visually enhanced captions led to sizable benefits in collocational knowledge, despite participants not spending extra time reading the targets. That is, although no additional time was spent on processing targets, the color acted as a cue to facilitate retrieval of targets, implying that the benefit of the color enhancement may not have been at encoding but at retrieval. Furthermore, visual input enhancement did not undermine the memorization of caption contents, suggesting that it may not interfere with overall meaning-making processes. Collectively, the results suggest that enhanced captions in digital videos can be employed as a useful and unobtrusive means of improving collocational competence without hampering the ability of L2 learners to learn from video contents. Given that this study represents an initial attempt to investigate the role of visual input enhancement in video-based collocation acquisition, the findings make notable contributions to the current L2 literature. From a pedagogical perspective, it is recommended that materials designers for L2 learners incorporate enhanced captions in video materials. However, the findings and implications of this study will be better understood only if future studies consider the potential limitations and mediating variables.

Acknowledgments

The author would like to thank the three anonymous reviewers for providing insightful and constructive comments. The work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (2018S1A5A2A01033017).

References

- Amso, D., Haas, S., & Markant, J. (2014). An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PloS one*, 9(1), 1–7. <https://doi.org/10.1371/journal.pone.0085701>
- Benson, M., Benson, E., & Ilson, R. (1997). *The BBI dictionary of English word combinations*. John Benjamins Publishing Company.
- Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, 35(2), 399–418. <https://doi.org/10.1017/S0142716412000434>
- Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, 21(3), 403–426. <https://doi.org/10.1177/1362168816653271>
- Cintrón-Valentín, M., & García-Amaya L. (2021). Investigating textual enhancement and captions in L2 grammar and vocabulary: An experimental study. *Studies in Second Language Acquisition*, 43(5), 1068–1093. <https://doi.org/10.1017/S0272263120000492>
- Cintrón-Valentín, M., García-Amaya L., & Ellis, N. C. (2019). Captioning and grammar learning in the L2 Spanish classroom. *The Language Learning Journal*, 47(4), 439–459. <https://doi.org/10.1080/09571736.2019.1615978>

- Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research*, 32(3), 453–467.
<https://doi.org/10.1177/0267658316637401>
- Davies, M. (2008). The Corpus of Contemporary American English. <https://www.english-corpora.org/coca/>.
- Gass, S., Spinner, P., & Behney, J. (2017). *Saliency in second language acquisition*. Routledge.
- Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology*, 23(2), 84–104.
<https://doi.org/10.125/44684>
- Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, 24(4), 541–577.
<https://doi.org/10.1017/S0272263102004023>
- Kang, H., Kweon, S.-O., & Choi, S. (2022). Using eye-tracking to examine the role of first and second language glosses. *Language Teaching Research*, 26(6), 1252–1273.
<https://doi.org/10.1177/1362168820928567>
- Lee, M., & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: An eye-tracking study. *The Modern Language Journal*, 102(3), 557–577.
<https://doi.org/10.1111/modl.12503>
- Lee, M., & Révész, A. (2020). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks [Special issue]. *Studies in Second Language Acquisition*, 42(3), 625–651. <https://doi.org/10.1017/S0272263120000108>
- Lee, S. K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning*, 57(1), 87–118.
<https://doi.org/10.1111/j.1467-9922.2007.00400.x>
- Leow, R., Egi, T., Nuevo, A., & Tsai, Y. (2003). The roles of textual enhancement and type of linguistic item in adult L2 learners' comprehension and intake. *Applied Language Learning*, 13(2), 1–16.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6
- Montero Perez, M. (2019). Pre-learning vocabulary before viewing captioned video: An eye-tracking study. *The Language Learning Journal*, 47(4), 460–478.
<https://doi.org/10.1080/09571736.2019.1638623>
- Montero Perez, M. (2020). Incidental vocabulary learning through viewing video. *Studies in Second Language Acquisition*, 42(4), 749–773. <https://doi.org/10.1017/S0272263119000706>
- Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, 18(1), 118–141. <https://hdl.handle.net/10125/44357>
- Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned Video: An eye-tracking study. *The Modern Language Journal*, 99(2), 308–328.
<https://doi.org/10.1111/modl.12215>
- Montero Perez, M., van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720–739.
<https://doi.org/10.1016/j.system.2013.07.013>

- Muñoz Lahoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System*, 67, 77–86. <https://doi.org/10.1016/j.system.2017.04.015>
- Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, 27(1), 95–106. <https://doi.org/10.2307/747835>
- O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, 43(2), 121–152. <https://doi.org/10.1080/01638530709336895>
- Overstreet, M. H. (1998). Text enhancement and content familiarity: The focus of learner attention. *Spanish Applied Linguistics*, 2(2), 229–258.
- Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading. An eye-tracking study. *Studies in Second Language Acquisition*, 38(1), 97–130. <https://doi.org/10.1017/S0272263115000224>
- Peters, E. (2012). Learning German formulaic sequences: The effect of two attention-drawing techniques. *The Language Learning Journal*, 40(1), 65–79. <https://doi.org/10.1080/09571736.2012.658224>
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551–577. <https://doi.org/10.1017/S0272263117000407>
- Plonsky, L. & Oswald, F. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Rodgers, M. P. H., & Webb, S. (2019). Incidental vocabulary learning through viewing television. *ITL-International Journal of Applied Linguistics*, 171(2), 191–220. <https://doi.org/10.1075/itl.18034.rod>
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 1–32). Cambridge University Press.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)
- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14(2), 50–73. <https://hdl.handle.net/10125/44214>
- Winke, P. M. (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(2), 323–352. <https://doi.org/10.1017/S0272263112000903>
- Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language learning activities. *Language Learning & Technology*, 14(1), 65–86. <https://hdl.handle.net/10125/44203>
- Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*, 97(1), 254–275. <https://doi.org/10.1111/j.1540-4781.2013.01432.x>
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Appendix A. Transcription of the TED Video

We're holding hands, staring at the door. My siblings and I were waiting for my mother to come back from the hospital. She was there because my grandmother had cancer surgery that day. Finally, the doors opened, and she said, "She's gone. She's gone." She started sobbing and immediately said, "We must make arrangements. Your grandmother's dying wish was to be buried back home in Korea."

I was barely 12 years old, and when the shock wore off, my mother's words were ringing in my ears. My grandmother wanted to be buried back home. We had moved from Korea to Argentina six years prior, without knowing any Spanish, or how we were going to make a living. And upon arrival, we were immigrants who had lost everything, so we had to work really hard to rebuild our lives. So it hadn't occurred to me that after all these years, back home was still in Korea. It made me ponder where I would want to be buried someday, where home was for me, and the answer was not obvious. And this really bothered me. So this episode launched a lifelong quest for my identity.

I was born in Korea -- the land of kimchi; raised in Argentina, where I ate so much steak that I'm probably 80 percent cow by now; and I was educated in the US, where I became addicted to peanut butter.

During my childhood, I felt very much Argentinian, but my looks betrayed me at times.

I remember on the first day of middle school, my Spanish literature teacher came into the room. She scanned all of my classmates, and she said, "You -- you have to get a tutor, otherwise, you won't pass this class." But by then I was fluent in Spanish already, so it felt as though I could be either Korean or Argentinian, but not both. It felt like a zero-sum game, where I had to give up my old identity to be able to gain or earn a new one.

So when I was 18, I decided to go to Korea, hoping that finally I could find a place to call home. But there people asked me, "Why do you speak Korean with a Spanish accent?"

And, "You must be Japanese because of your big eyes and your foreign body language." And so it turns out that I was too Korean to be Argentinian, but too Argentinian to be Korean.

And this was a pivotal realization to me. I had failed to find that place in the world to call home. But how many Japanese-looking Koreans who speak with a Spanish accent -- or even more specific, Argentinian accent -- do you think are out there? Perhaps this could be an advantage. It was easy for me to stand out, which couldn't hurt in a world that was rapidly changing, where skills could become obsolete overnight. So I stopped looking for that 100 percent commonality with the people that I met. Instead, I realized that oftentimes, I was the only overlap between groups of people that were usually in conflict with each other.

So with this realization in mind, I decided to embrace all of the different versions of myself -- even allow myself to reinvent myself at times. So for example, in high school, I have to confess I was a mega-nerd. I had no sense of fashion -- thick glasses, simple hairstyle -- you can get the idea. I think, actually, I only had friends because I shared my homework. That's the truth. But once at university, I was able to find a new identity for myself, and the nerd became a popular girl. But it was MIT, so I don't know if I can take too much credit for that. As they say over there, "The odds are good, but the goods are odd."

I switched majors so many times that my advisors joked that I should get a degree in "random studies."

I told this to my kids.

And then over the years, I have gained a lot of different identities. I started as an inventor, entrepreneur, social innovator. Then I became an investor, a woman in tech, a teacher. And most recently, I became a mom, or as my toddler says repeatedly, "Mom!" day and night. Even my accent was so confused -- its origin was so obscure, that my friends called it, "Rebecanese."

But reinventing yourself can be very hard. You can face a lot of resistance at times. When I was nearly done with my PhD, I got bitten by that entrepreneurial bug. I was in Silicon Valley, and so writing a thesis

in the basement didn't seem as interesting as starting my own company. So I went to my very traditional Korean parents, who are here today, with the task of letting them know that I was going to drop out from my PhD program. You see, my siblings and I are the first generation to go to university, so for a family of immigrants, this was kind of a big deal. You can imagine how this conversation was going to go. But fortunately, I had a secret weapon with me, which was a chart that had the average income of all of the graduates from Stanford PhD programs, and then the average income of all the dropouts from Stanford graduate programs.

I must tell you -- this chart was definitely skewed by the founders of Google.

But my mom looked at the chart, and she said, "Oh, for you -- follow your passion."

Appendix B. General Collocational Competence Cloze Test

1. 그 남자는 그녀에게 꽃 한 다발을 주었다.	= The man brought her a b_____ of flowers.
2. 그는 내 안경을 일부러 깨뜨렸다.	= He broke my glasses _____ purpose.
3. 우리는 기차를 타고 유럽을 여행했다.	= We traveled across Europe _____ train.
4. 그는 기타를 치는 데에 특별한 재능이 있다.	= He has special ability _____ playing the guitar.
5. 그녀는 개를 매우 좋아한다.	= She is very fond _____ dogs.
6. 무슨 일이 있어도, 난 너를 믿어.	= Whatever happens, I believe _____ you.
7. 나는 그녀를 우연히 또 만났다.	= I met her again _____ chance.
8. 커피 한 잔을 미리 주문하시겠어요?	= Would you order a cup of coffee _____ advance?
9. 어떻게 대처했니?	= How did you cope _____ it?
10. 나는 공부를 열심히 하기로 결정했다.	= I _____ a decision to study hard.
11. 나는 위험을 감수하고 싶지 않다.	= I don't want to t_____ a risk.
12. 그는 깊이 몰입한 것처럼 보였다.	= He looked d_____ absorbed.
13. 그 다리는 상당히 안전한가요?	= Is the bridge q_____ safe?
14. 그레이스는 연한 차 한잔을 따랐다.	= Grace poured a cup of _____ tea.
15. 아기는 위층에서 깊이 잠들어 있었다.	= The baby was s_____ asleep upstairs.
16. 자신감을 얻기 위해서는, 노력하는 것 밖에 없다.	= To gain confidence, you just have to _____ an effort.
17. 내 부탁 하나만 들어줄래?	= Will you _____ me a favor?
18. 나는 곧 삼촌 집을 방문할 예정이다.	= I'm going to p_____ a visit to my uncle's home soon.

Appendix C. Non-target Collocations Used in the Collocational Competence Cloze Test; Target Collocations Used in the Collocation Knowledge Survey and Collocation Test

Non-target collocations		Target collocations	
collocations	frequency	collocations	frequency
deeply absorbed	21	social innovator	5
weak tea	69	switch majors	9
quite safe	154	ring in my ears	9
pay a visit	369	get a tutor	24
sound asleep	490	lifelong quest	33
take a risk	520	sense of fashion	64
ability in	764	reinvent myself	74
by train	874	dying wish	140
make an effort	1057	become addicted to	221
by chance	2452	average income	464
do me a favor	3329	upon arrival	689
make a decision	3572	in conflict with	792
on purpose	4227	secret weapon	806
fond of	5348	rapidly changing	958
cope with	7003	make a living	3149
in advance	13356	drop out	3421
a bunch of	30487	stare at	4978
believe in	37207	a big deal	17105
Mean frequency	6183		1830

* Note. Frequency data obtained from the Corpus of Contemporary American English (Davies, 2008).