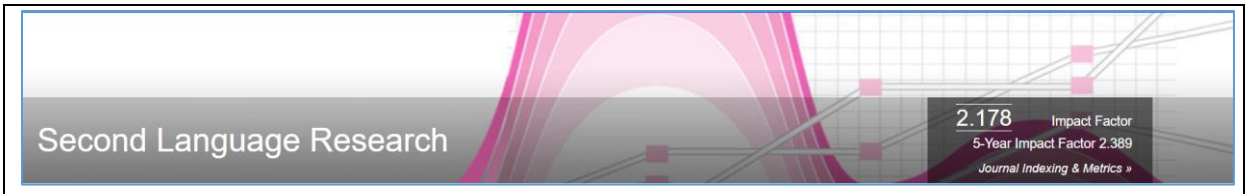

Puimège, E., Montero Perez, M., & Peters, E. (accepted). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: an eye-tracking study. *Second Language Research*.

Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: an eye-tracking study

Eva Puimège (KU Leuven)
Maribel Montero Perez (UGent)
Elke Peters (KU Leuven)

This study examines the effect of textual enhancement on learners' attention to and learning of multiword units from captioned audiovisual input. We adopted a within-participants design in which 28 EFL learners watched a captioned video containing enhanced (underlined) and
unenhanced multiword units. Using eye-tracking, we measured learners' online processing of the multiword units as they appeared in the captions. Form recall pre- and posttests measured learners' acquisition of the target items. The results of mixed effects models indicate that
enhanced items received greater visual attention, with longer reading times, less single word skipping and more rereading. Further, a positive relationship was found between amount of visual attention and learning odds: items fixated longer, particularly during the first pass, were more likely to be recalled in an immediate posttest. Our findings provide empirical support for the positive effect of visual attention on form recall of multiword units encountered in captioned television. The results also suggest that item difficulty and amount of attention were more important than textual enhancement in predicting learning gains.

Keywords: formulaic language, textual enhancement, audiovisual input, eyetracking

Introduction

In order to acquire a large second language (L2) vocabulary, learners need a great amount of exposure to L2 input. However, not all input is converted to intake (Corder, 1967), and attentional mechanisms determine which elements of the input are processed and encoded in memory (e.g., Robinson, 1995; Schmidt, 2001). The mediating role of attention in vocabulary learning from meaningful input has recently been demonstrated in a number of experiments using behavioral measures of attention such as eye-tracking and verbal reports (e.g., Godfroid et al., 2018; Pellicer-Sánchez, 2016). Their findings support the notion that attention is an important, if not essential factor in vocabulary learning from L2 input.

The central role of attention also provides an explanation for the typically small gains found in studies on incidental learning. Although studies have demonstrated that learners can develop lexical knowledge as a by-product of reading (e.g., Horst, Cobb, and Meara, 1998), listening (e.g., Van Zeeland and Schmitt, 2013), or TV viewing (e.g., Peters and Webb, 2018), incidental learning has been described as a slow, error-prone process in which knowledge develops in small increments (e.g., Webb, 2020). This may be because incidental learning activities do not require learners to notice or elaborately process new vocabulary (e.g., Laufer and Hulstijn, 2001). Attentional mechanisms may also partly explain the low acquisition rate of multiword units, many of which are discontinuous, semantically transparent, and therefore less salient than unknown words (e.g., Boers et al., 2017).

A few studies have explored ways of promoting learners' attention to and learning of multiword units through input enhancement (e.g., Choi, 2017). Input enhancement involves instructional methods that render linguistic elements more salient (Sharwood Smith, 1993). Words and phrases can be made more salient through top-down methods of enhancement (e.g., pre-teaching items appearing in the input) or by increasing bottom-up salience (e.g., underlining, boldfacing). Previous research has already demonstrated a relationship between

bottom-up enhancement, henceforth textual enhancement (TE), and learning of multiword units from written texts (e.g., Boers et al., 2017; Choi, 2017; Peters, 2012).

Only one study to date has examined the effect of TE on learning multiword units from audiovisual input with captions (= L2 on-screen text) (Majuddin, Siyanova-Chanturia and Boers, 2021). Since watching L2 television may be an effective way of acquiring knowledge of multiword units (Lin and Siyanova, 2014; Puimège and Peters; 2019, 2020), it is worth investigating whether TE might also support learning of multiword units from audiovisual input. The use of TE in captioned television has proved effective in single-word acquisition (e.g., Montero Perez et al., 2014, 2015, 2018), and may also be a useful method of promoting learning of multiword units (Majuddin et al., 2021). However, to the best of our knowledge, no study has investigated how TE affects learners' processing of multiword units while watching captioned television.

The present study explores the effect of textually enhanced captions on processing and learning of multiword units while watching L2 audiovisual input. The first aim is to examine the effect of TE on form recall of multiword units. Second, using eye-tracking, we examine how TE (underlining) affects online processing of multiword units, specifically whether TE leads to an increased amount of visual attention. Finally, we aim to investigate the relationship between attention and learning, by linking on-line processing of multiword units to learning gains in the form recall posttest.

Background

Incidental learning of multiword units

To achieve fluency in their L2, learners need to acquire a large number of formulaic sequences, that is, frequently recurring words, phrases and word combinations assumed to be familiar and conventional to native speakers (Siyanova-Chanturia and Pellicer-Sánchez,

2019). In second language acquisition (SLA) research, the term formulaic sequence is commonly used to cover a wide range of constructions such as lexical bundles, collocations, and idioms. The current study focuses on multiword units (MWUs), or formulaic sequences consisting of multiple words. MWUs fulfill many communicative functions and are very widespread in language. However, despite their ubiquity, many L2 learners have difficulty using MWUs (e.g., Laufer and Waldman, 2011). This has been explained in terms of learners' limited exposure to L2 input compared to native speakers. Each individual MWU occurs less frequently than its single-word components, and especially low-frequency MWUs are unlikely to be encountered repeatedly in a short time span (e.g., Boers and Lindstromberg, 2009). Learners may therefore lack the necessary amount of L2 exposure to learn the associative links between words (e.g., Durrant and Schmitt, 2010).

Another reason why learners might struggle with MWUs is related to the role of attention and salience in learning from meaningful input. There is some evidence that MWUs can be learned incidentally through meaning-focused activities such as reading (e.g., Pellicer-Sánchez, 2017; Szudarski, 2012; Szudarski and Carter, 2016), reading-while-listening (Webb, Newton and Chang, 2013), and TV viewing (Majuddin et al., 2021; Puimège and Peters, 2019; 2020). In most studies on incidental learning (see Szudarski, 2012, for a notable exception), relatively brief L2 exposure led to significant improvements in formulaic knowledge. However, learning gains were generally small and depended strongly on factors such as frequency of encounters (e.g., Webb et al., 2013) and semantic transparency (e.g., Puimège and Peters, 2019). Because MWUs vary widely in terms of their semantic and formal properties, not all MWUs will receive the same amount of attention when encountered in L2 input. For instance, MWUs can be highly schematic (allowing for internal variation, e.g., verb-noun collocations) or discontinuous (e.g., *provide information* vs. *provide* some *information*) (Vilkaitė and Schmitt, 2019).

Increasing the salience of multiword units: textual enhancement

Recent studies (Boers et al., 2017; Choi, 2017; Peters, 2012; Sonbul and Schmitt, 2013; Szudarski and Carter, 2016; Toomer and Elgort, 2019) have used TE to promote learners' noticing of unknown MWUs in L2 written input. These studies all found a positive relationship between TE and incidental learning, but usually only when knowledge of form was measured. To give one example, Szudarski and Carter (2016) found superior learning outcomes for enhanced collocations in a form recognition and a form recall test, but not in a meaning recognition test. The positive effect of TE on form learning suggests that drawing learners' attention to the target forms, and the lexical makeup of the MWUs, may result in a more durable memory trace (Boers et al., 2017). However, increased attention to form does not necessarily result in retention of semantic knowledge, which requires a greater level of analysis or more elaborate processing (e.g., Leow and Martin, 2017).

One study by Choi (2017) used eye-tracking to measure learners' online processing of unknown MWUs during reading in a textually enhanced (boldfaced) and an unenhanced condition. Advanced learners of English (L1 = Korean) read a text containing 14 semantically transparent phrases. Learners in the enhanced group performed significantly better on a posttest of form recall than learners in the baseline text group, who read the same text without enhancement one week after the experiment. Choi also found an interaction between pretest knowledge and TE in predicting learners' visual processing of the collocations. Items that were not known in the form recall pretest received more fixations and were fixated longer in both conditions, but more so in the TE group. This finding suggests that TE helped learners notice MWUs that were unknown. However, results on a separate cued recall test indicated a trade-off between attention to the enhanced and unenhanced content in the reading passage, with lower retention of unenhanced information in the TE group.

The results of previous research indicate that TE can be used to promote learners' attention to unknown MWUs. All of the aforementioned studies employed TE in written texts, but little is known about the use of TE to increase the salience of MWUs in multimodal input, e.g., television with captions.

Vocabulary learning from audiovisual input

There has been a recent surge in research investigating L2 vocabulary acquisition from audiovisual input (see, e.g., Peters and Webb, 2018). There is growing evidence that vocabulary learning can occur from watching L2 television (e.g., Feng and Webb, 2020; Peters and Webb, 2018). Further, captions, or subtitles in the language of the audio track, have been found to promote vocabulary acquisition from TV viewing (e.g., Peters, 2019; Peters, Heynen and Puimège, 2016; Pujadas and Muñoz, 2019).

A number of studies have demonstrated that enhanced captions can increase the salience of unknown vocabulary in audiovisual input (Cintrón-Valentín, García-Amaya and Ellis, 2019; Montero Perez et al., 2014, 2015, 2018). For example, Montero Perez et al. (2015) used vocabulary tests and eye-tracking to measure the relationship between learners' attention to, and learning of unknown vocabulary for two types of audiovisual input (full captions, keyword captions). The keyword captions group outperformed the full captions group in a form recognition test, but not in a form-meaning test. The authors only found a significant positive relationship between eye movement indices (second pass time, total reading time) and form recognition in the full captions group. They concluded that their eye-tracking measures may only have captured low-level attention, but not elaborate processing of meaning.

We should mention that some of the assumptions underlying eye-tracking research in reading may not be directly applicable to captioned audiovisual input, due to its multimodal,

dynamic nature (for eye-tracking research on learners' caption use, see e.g., Bisson, Van Heuven, Conklin, and Tunney, 2014; Gass, Winke, Isbell, and Ahn, 2019; Muñoz, 2017; Winke, Gass, and Sydorenko, 2013). First, the relationship between reading times and L2 learning may be very different when processing is constrained by the pace of the video. Further, due to the overlap between written and spoken input, processing may happen in various input modes simultaneously. This means that eye movement data in captioned audiovisual input have to be interpreted with caution and may not always be comparable to findings in reading research.

To the best of our knowledge, only one study has investigated the effect of TE on learning MWUs from captioned television. Majuddin, Siyanova-Chanturia and Boers (2021) examined the effect of two types of captions on learners' recall of MWUs. Participants were divided into six groups, based on the number of repetitions (viewing the same video once or twice) and captioning type (normal captions, enhanced captions, and no captions). In the enhanced condition, MWUs were bolded and underlined. A positive relationship was found between captioning and form recall in a cued gap fill test, with higher gains made from pre- to posttest in the enhanced and unenhanced captions groups compared to the no-captions group. However, enhancing the items did not further increase learning gains. The authors argued that the real-time nature of video (as opposed to written text), as well as the length of the MWUs (some of their target items contained five words), might explain why TE did not have a noticeable effect on learning.

Rationale and research questions

Previous research has shown that TE can be used to draw learners' attention to unknown vocabulary in captioned television (e.g., Montero Perez et al., 2015). However, still little is known about how TE affects learners' processing of MWUs during TV viewing. Current evidence suggests that TE may not support learning of MWUs from captioned

audiovisual input, possibly due to the brief, fleeting presentation of the items (Majuddin et al., 2021). A closer examination of learners' attention to unknown MWUs with and without TE could provide more insight into how learners interact with enhanced items in captioned video. This might in turn improve our understanding of how TE can be used in different input modalities to support learning.

The present study examines in what way TE affects learners' processing and learning of MWUs when watching captioned television. First, we use eye-tracking to examine how learners' visual processing of MWUs in captioned television is affected by TE. A second aim is to study the relationship between TE, visual attention, and learning gains. We aim to determine whether TE and attention are positively associated with learners' performance on a form recall posttest. The current study addresses the following research questions:

1. Does TE affect learners' visual attention to MWUs in the captions of audiovisual input?
2. Do TE and visual attention affect form recall of MWUs encountered in captioned audiovisual input?

Method

Participants

Thirty Flemish students (L1 = Dutch, M_{age} = 22, 7 males, 23 females) were recruited for the experiment. Participants were enrolled in an applied language studies program (with the exception of four students from other programs). Sixteen participants were majoring in English. The lexical profile of the documentary (see Materials) suggests that knowledge of the 3,000 most frequent English words was necessary to reach 95% coverage of the documentary, which is assumed to be sufficient for adequate comprehension in television (Durbahn et al., 2020; Webb and Rodgers, 2009). We excluded data from four participants who did not reach

a score of 27/30 on the 3K level of the Vocabulary Levels Test (VLT; Schmitt, Schmitt and Clapham, 2001), leaving a sample of 26 participants. Participants' results per frequency band in the VLT are presented in Table 1. The students either received course credit or monetary reimbursement for participating.

TABLE 1

Mean Vocabulary Levels Test Scores and Standard Deviations per Frequency Band (n = 26)

	2K	3K	Academic	5K	10K	Total
M (SD)	29.50 (0.76)	28.92 (0.8)	28.96 (1.15)	27.38 (1.81)	17.42 (5.89)	132.19 (8.14)

Materials

Video and captions. As audiovisual input we used the first 30 minutes of an episode of Fry's *Planet Word* (BBC), a documentary series about language. The episode *Uses and Abuses*, about swear words, slang and taboos, was chosen for its relevance to language students, to ensure that participants would pay attention to the content. Running the script of the documentary through the Compleat Lexical Tutor (Cobb, n.d.) showed that 95% of the running words in the text belonged to the 3,000 most frequent English word families. Captions were taken from the original DVD track using Aegisub, and were adapted to a word-for-word transcription of the spoken input. Of the 549 captions appearing in the video, 35 (6%) were presented across two lines. Spacing between the two lines was 1 cm. Average caption duration was 3285 ms (*SD* = 1088 ms). Each caption contained 9 words on average (*SD* = 3).

Target items. Multiword units were taken from Puimège and Peters' (2020) study, who used the same viewing material. From their sample, we selected 22 items from a wide range of MWUs, including idioms (e.g., *beyond the pale*), collocations (e.g., *supernatural powers*),

compounds (e.g., *guinea pig*), and phrasal verbs (e.g., *tap into*). All items have an MI score higher than 3 in the COCA (Davies, 2012), which is a commonly used cut-off of collocation strength (e.g., Durrant and Schmitt, 2010). MI reflects how strongly two words attract by comparing their co-occurrence rate in a corpus to their co-occurrence rate by chance (Schneider, 2018). Semantic decomposability was rated on a seven-point scale by 33 learners from the same study program as the participants. Raters were given the following question in English: "How easy is it to guess the meaning of these phrases based only on their single-word components, on a 7-point scale? 1 = impossible, 7 = very easy". They were instructed to only rate items for which they knew the meanings of both single word components. This resulted in missing values for some items (minimum number of ratings was 18 for *racial epithet*). Because of the similarity in participant profile in the current study and in Puimège and Peters' (2020) study, we discarded items that were known in the pretest by 80% of their sample of participants. To reach a sample of 28 items, we added low-frequency MWUs not appearing in Puimège and Peters' (2020) sample. Two target items (*took umbrage, abusive language*) appeared in two-line captions. All of the MWUs appeared only once in the input. The full list of items can be found in Table 2. Information on item variables (corpus frequency, mutual information, and semantic decomposability) can be found in the OSF: https://osf.io/ypmdg/?view_only=4dd24442829f4c8ab0d0c0e07eeeb209.

Experimental conditions. To represent the two experimental conditions, a counterbalanced within-participants design was adopted so that all participants were exposed to both conditions (enhanced and unenhanced) and all items appeared in both conditions in the full data set (Godfroid, 2020; Nicklin & Vitta, 2021). Participants were allocated to one of two versions of the same video. Items that were enhanced in version 1 were unenhanced in version 2, and vice versa. By adopting a counterbalanced design, we could ensure that there was no confound between our treatment and knowledge of constituent words or linguistic

complexity of the items (see also Godfroid, 2020 on the value of within-subject designs in eye-tracking research). Instead of dispersing or alternating the conditions chronologically, for instance by underlining every other item in the captions, we split each version of the video into an unenhanced and an enhanced part (see Table 2). Participants assigned to version 1 saw the first 14 items underlined, and participants assigned to version 2 saw the last 14 items underlined. This approach was chosen to avoid an attentional trade-off effect. In a study on L2 reading (Choi, 2017), TE was not just associated with increased attention to enhanced items, but it also led to decreased attention to unenhanced information appearing near enhanced items. Splitting the video in two parts could ensure that a greater amount of attention in the TE condition would reflect enhanced attention compared to attention in an unenhanced captioned video.

Form recall test

Learning of MWUs was measured by means of a form recall test. We used only one measure of knowledge to avoid a learning effect from one test to another, as was found in previous studies (e.g., Puimège and Peters, 2020). The form recall format was chosen to minimize an effect of the pretest on learners' attention to target items during the experiment. Participants had to provide the form of the English MWUs based on a Dutch translation. The first letter of each single word component was given to avoid elicitation of other plausible word combinations. Participants were asked to give the written and spoken form of each item. Spoken responses were recorded. Participants were also asked to provide single word components where they could, even if they did not know the full MWUs. To make sure that learning could be ascribed to the treatment, 30 MWUs from Puimège and Peters' (2020) study which did not appear in the part of the documentary used in the current study were included as distractor items. Learning of these items would signal potential test effects (e.g., guessing based on the first letter of the single word components) or learning outside the treatment (e.g.,

looking up MWUs at home). Two versions of the test were made in which the items appeared in a different, semi-random order. Both versions included all target items. Participants who completed version 1 in the pretest, received version 2 in the posttest, and vice versa. The test was not timed and took on average 18 minutes to complete. The full tests are available in the OSF.

Examples from the form recall test:

lichaamssappen (bloed, zweet, enz.) b _ _ _ _ _ f _ _ _ _ _ *(bodily fluids)*

een spier verrekken p _ _ _ a m _ _ _ _ _ *(pull a muscle)*

Procedure

One week before watching the documentary, participants filled out an informed consent form and completed the form recall pretest and the VLT. They were not forewarned of a vocabulary posttest, nor that any of the MWUs in the pretest would appear in the video. The eye-tracking experiment was conducted using an Eyelink Portable Duo, Version 1.0.2 (SR Research). The video was presented on a 1280 x 1024 monitor with a refresh rate of 60 Hz. Display dimensions were 19 x 33 cm. Captions appeared in Arial (proportional), with a character size of 38pt, corresponding to approximately 0.5 cm on the monitor and 0.39° of visual angle[1]. Participants were seated in front of the monitor at a viewing distance of 72 cm. A desk-mounted chin rest was used to stabilize head position. Participants' dominant eye was tracked. Sampling rate was 500-2000 Hz.[2] After setup, a nine-point calibration and validation procedure was performed. The maximum calibration error was 0.8° of visual angle. To keep track of accuracy during the experiment, the 30-minute documentary was split at scene changes into seven short (3-6 minute) video clips, each representing a separate trial in the experiment. The target items were distributed between the shorter videos as shown in Table 2.

At the start of the experiment, participants watched a practice video with captions, to let them adjust to the experimental procedure and to correct their behavior (e.g., head movements). After each trial, a drift check was performed and the calibration procedure was repeated when necessary. A 5-minute break and recalibration was inserted for each participant after the first 4 videos (approximately 15 minutes). This is also where the captions changed from enhanced to unenhanced or vice versa.

TABLE 2

Distribution of target items across the input in the two counterbalanced conditions.

Video	Target items	Condition version 1	Condition version 2
Video 1	*highest echelons, foul language, bodily fluids, common denominator, supernatural powers, sexual depravity, mutual pleasure, heck of a lot, pass over (into)*	enhanced	unenhanced
Video 2	*tell off*		
Video 3	*tap into, fair description, end up, evolutionary advantage*		
BREAK			
Video 4	*abusive language, guinea pig, spark your interest, unleash a torrent of, subliminal effect*	unenhanced	enhanced
Video 5	*take into account, pain relievers*		
Video 6	*jab line, turning point, win the right*		
Video 7	*beyond the pale, sheer coincidence, racial epithet, take umbrage*		

Timed interest areas were created for the 28 MWUs as a whole, and their single word components. Each interest area remained on screen for the duration of the caption in which it

appeared (*M* = 3523 ms, *SD* = 849 ms). Margins of approximately 0.6 cm were added at the top and bottom of each interest area. Interest areas for target items had an average size of 6 x 2.5 cm. Five eye-tracking measures were used to examine online processing of MWUs: first pass reading time, rereading (binary measure), rereading time, single-word skipping (binary measure), and total reading time. First pass reading time is a durational measure which sums all fixations during the first visit, before the eye gaze leaves the interest area. It captures early stages of processing, and may be sensitive to low-level visual, orthographic and frequency-related factors (Conklin, Pellicer-Sánchez, and Carrol, 2018; Godfroid, 2020). Textual enhancement could therefore be expected to affect first pass reading time, although previous studies investigating the effect of TE on grammar learning did not find such an effect (Lee and Révész, 2018, 2020). Rereading often results from processing difficulties related to comprehension or contextual integration (Conklin et al., 2018; Godfroid, 2020), and might also be affected by caption duration. We analyzed rereading both as a binary event and as a durational measure. Our analysis of binary rereading captures the odds of rereading an item, rather than the amount of time spent on rereading. Rereading time sums all fixation durations in an interest area after the first pass. We included a binary variable for single-word skipping, to examine whether participants fixated both single-word components of each MWU. This measure was included because TE might cause learners to distribute their attention more evenly across both words in a MWU. Final-word skipping has been interpreted as an indication of more fluent reading (see for example Carrol and Conklin, 2019), but in the context of the current study, single-word skipping could also reflect the amount of attention to the lexical composition of a MWU. Total reading time was included as a late measure encompassing both first pass reading time and rereading time. This measure was included because it has produced strong associations with learning gains in previous studies (e.g., Godfroid et al., 2018).

After watching the documentary, participants completed a short questionnaire about the content of the video and about their general viewing habits. The questionnaire can be found in the OSF. The questionnaire was immediately followed by the form recall posttest. After the posttest, participants were interviewed about their explicit recall of target items, based on their responses on the form recall test. This was done to gain more insight into participants' conscious noticing of the target items, and to check if learners had guessed in the form recall test, or had learned items outside the experiment. The interviews were not recorded, but the interviewer took notes which were used to help interpret the quantitative results of eye-tracking measures and learning gains. Finally, participants completed another short questionnaire about their awareness of the purpose of the experiment and a potential effect of the pretest on their conscious viewing behavior.

Data preprocessing and cleaning

Eye movements were parsed according to the default cognitive configuration of Eyelink. Following Godfroid and Hui (2020), the output of the algorithm was visually inspected in the DataViewer Temporal Graph Trial View. The initial pool of eye-tracking data contained 728 data points at phrase level (28 MWUs x 26 participants). Trials where track loss occurred were removed (10%). Data were cleaned using the default settings in the four-stage cleaning procedure of Eyelink Data Viewer. Items with a fixation time of 0 ms for the full phrase (not at word level) (n = 60) were excluded from the analyses because they would lead to skewed reading times, even after log transformation[3]. The analysis of rereading time also excluded zeroes (n = 400) to reduce skew.

Scoring and analysis

Eye tracking measures as outcome variable. In all analyses, the binary factor TE was the main independent variable of interest. Form recall pretest score (binary), phrase frequency

per million, mutual information (MI), length, caption duration, frequency of the least frequent single word component, and semantic decomposability were entered as control variables. Any continuous variables that were not normally distributed, were log transformed with base 2. Continuous control variables were also centered around the grand mean.

For the analysis of first pass reading time, rereading time, and total reading time, linear mixed-effects models were fit using the lmer() function in the package lme4 (Version 1.1-21) in R (Version 3.6.1). Because distributions of these measures were positively skewed, the data were log transformed. In each of the mixed effects models, the same procedure was followed. First, a null model was constructed containing only random intercepts for item and subject. Fixed effects and an interaction term for TE and pretest score variables were then added to the model. Any non-significant variables that did not improve the model fit were removed one by one. Model fit was estimated through log-likelihood ratio tests and comparison of AIC values. The final model was the most parsimonious model (i.e., with the fewest covariates) with the lowest AIC value. Restricted Maximum Likelihood was used for model fitting. Because our regression models for different eye-tracking measures could be said to test the same hypothesis (Godfroid and Hui, 2020; Von der Malsburg and Angele, 2017), we applied a Bonferroni adjustment, $\alpha = 0.01$. After adding the fixed effects, we added a random slope for TE at subject-level, and a correlation between the random slope and the random intercept. If the random slope (+ correlation) did not improve the model fit, it was removed. Finally, a sensitivity analysis was performed to check the influence of outliers (Godfroid, 2020). Each lmer() model was rerun without outliers (studentized residuals with an absolute value higher than 2.5). For the binary outcome variables (skipping, rereading), generalized linear mixed models were fit using the glmer() function in the lme4 package.

Form recall as outcome variable. In the form recall test, items that were pronounced or spelled correctly, received a score of 1, incorrect items received a score of 0. Half of the

tests were scored by a second rater. Interrater agreement was 98% for both the pre- and posttests. For the other 2%, the first rater revisited the test responses to make sure the criteria described above were applied correctly. Remaining disagreements were solved through discussion. To analyze learning gain at the item level, the glmer() function in the lme4 package in R was used. First, models were fit to analyze the effect of the treatment on form recall, by comparing learning gains for target items and distractor items. Main effects for time (pretest vs. posttest) and item type (target item vs. distractor) were entered into the first model, as well as an interaction term between these two variables. We also analyzed learning of the target items and distractors in two separate models. The main analysis included only the binary posttest scores for target items. Items that were known in the pretest, and distractor items, were excluded from the analyses, leaving 484 data points. In this analysis, TE and the eye-tracking measures were the main independent variables of interest. Because previous studies did not include any eye-tracking measures, we first fit a model with only TE as independent variable. Then, total reading time was added to the model. Control variables were phrase frequency, MI score, length, caption duration, and decomposability. Participants' score on the VLT could not be added as a fixed effect, because its inclusion led to inflated odds ratios and convergence problems, possibly due to the low number of unique values for this predictor (n = 16).

Results

Eye tracking measures as outcome variable

To find out if TE affected reading of MWUs encountered in the input (see research question 1), we first analyzed the eye-tracking measures. The descriptive results are summarized in Table 3.

TABLE 3

Means and Standard Deviations (in Parentheses) for the Eye-tracking Measures, per Condition

	Unknown in pretest		Known in pretest		All	
	unenhanced	enhanced	unenhanced	enhanced	unenhanced	enhanced
FPR	471.6 (332.3)	662.5 (466.5)	371.0 (217.9)	494.1 (384.4)	439.5 (303.8)	604.1 (446.4)
RRT	631.9 (401.1)	587.4 (401.5)	385.4 (225.4)	521.8 (390.7)	562.4 (376.3)	568.4 (398.1)
TRT	725.9 (468)	968.8 (500.3)	501.2 (278.7)	702.8 (476.2)	654.2 (429.5)	876.4 (507.3)
Skipping	.17 (.1)	.08 (.05)	.10 (.05)	.09 (.06)	.27 (.10)	.17 (.07)
RRR	.14 (.06)	.17 (.08)	.05 (.05)	.07 (.05)	.19 (.07)	.24 (.09)

Note. FPR = First pass reading time, RRT = rereading time, TRT = total reading time, Skipping = rate of single-word skipping, RRR = rereading rate. Rates for skipping and rereading were calculated by dividing the number of items that were reread, or in which one word was skipped, by the total number of fixated items (Conklin et al., 2018).

The Bonferroni-adjusted results of the mixed effects models (see Tables 5 and 6) indicate that TE was associated with significantly longer first pass and total reading times, as well as less single-word skipping. Enhancement also led to higher odds of rereading, but did not significantly predict rereading time. Pretest knowledge was a significant predictor of rereading time, total reading time, and single-word skipping. Items that were not known in the pretest tended to receive longer reading times. Item length and decomposability predicted first

pass reading time and total reading time. Longer and less decomposable items received longer reading times, particularly during the first pass. Caption duration predicted rereading time, binary rereading, and total reading time. Items that were unknown in the pretest and stayed on screen longer were more likely to be reread and had longer reading times, particularly after the first pass. Mutual information, frequency of the full MWU and of the least frequent component predicted single word skipping, with greater odds of skipping for higher-frequency and less strongly associated MWUs. The effect of VLT score was not significant for any of the eye-tracking measures.

TABLE 4

Best Fitting Models for the Continuous Eye-tracking Measures

	*First pass reading time (*n = *587)*				*Rereading time (*n = *254)*				*Total reading time (*n = *590)*			
Fixed effects	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Level 1												
Intercept	8.5**	0.07	116.72	< 0.001	8.93**	0.09	98.74	< 0.001	9.08**	0.09	102.89	< 0.001
TE	0.35*	0.1	3.59	0.002					0.54**	0.09	5.82	< 0.001
Pretest score					-0.35*	0.13	-2.64	0.01	-0.32**	0.08	-3.74	< 0.001
Mutual information	0.16	0.07	2.39	0.025					0.13	0.07	2.05	0.053
Length	0.66**	0.11	5.97	< 0.001					0.52**	0.11	4.81	< 0.001
Decomposability	-0.44**	0.11	-3.98	< 0.001					-0.39*	0.12	-3.63	0.002
Duration					0.59*	0.19	3.09	0.006	0.46*	0.14	3.24	0.004
Level 2												
VLT score	-1.79	0.7	-2.54	0.018	-0.51	0.89	-0.57	0.571				
Random effects	*Variance*	*SD*			*Variance*	*SD*			*Variance*	*SD*		

(1|item)	0.04	0.04	0.03	0.18	0.03	0.18
(1|subject)	0.05	0.05	0.06	0.24	0.1	0.31
TE	0.14	0.37			0.11	0.33
Residual	0.56	0.75	0.72	0.85	0.56	0.75
Marginal R^2 / conditional			0.08 /			
R^2	0.18 / 0.36		0.18		0.22 / 0.39	
AIC	1437.69		678.18		1449.05	

Note. ** $p < .001$, * $p < .01$. Level 1 = item level, level 2 = participant level. The reference level for textual enhancement was “no enhancement”, the reference level for pretest score was “no pretest knowledge”.

TABLE 5

Best Fitting Models for the Binary Eye-tracking Measures

	*Single word skipping (*n = *595)*				*Rereading (*n = *595)*			
Fixed effects	*B*	*SE*	*z*	*p*	*B*	*SE*	*z*	*p*
Intercept	-0.04	0.21	-0.21	0.835				
TE = 1	-0.96**	0.2	-4.84	< 0.001	0.63**	0.18	3.49	< 0.001
Pretest score = 1	0.66*	0.27	2.48	0.01				
Corpus frequency	0.16*	0.06	2.59	0.009				
SW frequency	0.15*	0.05	2.82	0.005				
Mutual information	-0.84**	0.24	-3.54	< 0.001				
Duration					1.20*	0.39	3.13	0.002
Random effects	*Variance*	*SD*			*Variance*	*SD*		
(1|item)	0.56	0.75			0.25	0.5		
(1|subject)					0.15	0.39		
Adjusted ICC / conditional ICC	0.15 / 0.11				0.11 / 0.10			
AIC	689.85				778.85			

Notes. ** $p < .001$, * $p < .01$. SW frequency = frequency of the least frequent single word component.

Form recall as outcome variable

In the second part of the analyses, we examined participants' scores on the form recall test, to see if (a) learning had occurred from pre- to posttest, (b) learning could be ascribed to the treatment, and (c) a relationship could be found between learning gains, TE, and amount of attention (research question 2). Scores on the form recall tests are summarized in Table 6.

TABLE 6

Mean Scores, Standard Deviations (in Parentheses), and Gains on the Form Recall Pre- and Posttests

	target items			
	unenhanced (max. = 14)	*enhanced* (max. = 14)	*all* (max. = 28)	distractors (max. = 30)
pretest score	4.58 (1.90)	4.81 (2.26)	9.39 (3.01)	10.77 (4.17)
posttest score	6.42 (2.28)	6.81 (2.73)	13.23 (3.35)	12.69 (4.43)
absolute gain	1.85 (1.57)	2 (1.33)	3.85 (2.05)	1.92 (1.52)
normalized gain	.20 (.16)	.24 (.20)	.21 (.11)	.10 (.09)

Note. Normalized gains were calculated using the following formula: (post - pre)/(total number of test items - pre) (Horst et al., 1998).

The results of the first mixed effects model show that there was a significant main effect of time on the binary outcome variable form recall score ($B = 1.19$, $SE = 0.16$, $z = 7.5$, $p < .001$), indicating that items were more likely to be known in the posttest than in the pretest. The interaction between time (reference level = pretest) and item type (reference level = target item) was also significant ($B = -0.68$, $SE = 0.21$, $z = -3.22$, $p = 0.001$). Predicted probability of form recall knowledge was similar for target and distractor items in the pretest, but

significantly higher for target items in the posttest. The main effect of item type was not significant (p = .82). In two additional models, the effect of time was analyzed separately for target items and distractors. The results confirm that the effect was stronger for target items (B = 1.2, SE = 0.16, z = 7.48, p < 0.001) than for distractors (B = 0.55, SE = 0.15, z = 3.75, p < 0.001). The results of the three models are reported in full in the OSF.

We took the significant interaction between time and item type as evidence for learning from watching the captioned video, and went on to analyze the relationship between TE and learning by comparing learning gains between the enhanced and unenhanced target items. Results of the second mixed effects model show that TE significantly predicted learning from pre- to posttest, with greater odds of learning in the enhanced condition than in the unenhanced condition (B = 0.66, SE = 0.26, z = 2.51, p = .01). When total reading time was added to the model, TE was no longer significant. In the final model, summarized in Table 7, total reading time significantly predicted learning. A doubling in total reading time (the variable was log transformed with base 2) was associated with 63% (OR = 1.63) higher odds of learning an item. Other variables that significantly predicted learning were item length, caption duration, and decomposability.

TABLE 7

Best-fitting Model for Form Recall (*n* = 397)

Fixed effects	*B*	*OR*	*SE*	*z*	*p*
Intercept	-1.40***	0.25	0.24	-5.73	< 0.001
Total reading time	0.49**	1.63	0.17	2.89	0.004
Duration	-1.24*	0.29	0.61	-2.02	0.04
Length	-1.61**	1.2	0.51	-3.14	0.002
Decomposability	2.25***	9.49	0.54	4.15	< 0.001

Random effects		*Variance*	*SD*
(1|item)		0.64	0.8
(1|subject)		0.22	0.47
Adjusted ICC / conditional ICC	0.21 / 0.17		
AIC	392.26		

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. The reference level for textual enhancement was "no enhancement", the reference level for pretest score was "no pretest knowledge".

Questionnaires and interview data

A questionnaire concerning learners' TV viewing habits and input comprehension was used to check for obvious comprehension issues with regard to the content of the episode. This helped us identify participants who had trouble understanding the input (see Participants section). The questionnaire also showed that all participants were used to watching subtitled or captioned television.

A second questionnaire was used to check if the pretest had affected learners' conscious attention to target items during the experiment. Participants were also asked if they had looked up any of the target items after the pretest. The results showed that the majority of participants (n = 16) had been aware while watching the documentary that some items had appeared in the pretest. This was true for both enhanced and unenhanced items. Further, almost half of the participants (n = 12) indicated that this had made them pay more conscious attention to MWUs.

Finally, notes taken by the researcher during the interviews were examined to gain more insight into learners' noticing of items that were learned from pretest to posttest. A few trends emerged from these data. First, all participants reported that they remembered seeing items underlined in the input. However, most of the time, participants could not remember whether a specific item had been underlined. Often participants could not explain why they had been able to recall an item in the posttest. In some cases, participants reported that they had guessed the correct response in the form recall test based on the first letters of the single word components. For example, one participant had guessed the distractor *white lie* in both the pretest and posttest. Some participants reported that they already knew items before the experiment, but had not been able to recall them in the pretest.

Discussion

Research question 1: does TE affect learners' attention to MWUs in captioned audiovisual input?

To address the first research question, we used five eye-tracking indices measuring learners' processing of 28 MWUs encountered with or without TE in the captions of a 30-minute documentary. Eye-tracking measures in the two conditions were compared using mixed effects models. The results show that participants spent significantly more time fixating

on target items that were underlined in the captions. Enhanced items were also more likely to be reread, although rereading time was not significantly correlated with TE. The odds of fixating both single word components were also higher for enhanced items.

The positive relationship between TE and first pass reading time suggests that underlined items caught learners' attention, leading to increased initial processing. This interpretation is supported by the results of the questionnaires and interviews, in which participants reported noticing underlined items in the input. It is important to note that studies investigating learners' attention to grammatical structures in captioned input did not find a significant effect of TE on first pass reading time (Lee and Révész, 2018, 2020). A possible explanation is that analyzing a grammatical structure in parallel with sentence comprehension may be more cognitively demanding than processing a MWU during the first pass. The effect of TE may therefore occur later for grammatical structures (Alsadoon and Heift, 2015; Lee and Révész, 2018, 2020; Winke, 2013). It would be interesting to further investigate how learning burden interacts with TE in different stages of processing.

TE also significantly predicted rereading. Learners were more likely to revisit enhanced items, suggesting that the salience-raising effect of TE led to increased reanalysis. Despite its effect on rereading, no significant relationship was found between TE and rereading time. This may be due to the limited amount of time learners had to read the captions. Visual inspection of the eye movement data confirmed that, in many cases, learners could only briefly revisit the target items, limiting the opportunity for late processing. Caption duration was a significant predictor of both rereading measures, which indicates that rereading time depended first and foremost on how long an item remained on screen.

Total reading time was also positively associated with TE. This is perhaps not surprising, given that the measure incorporates first pass reading time and rereading time, and correlated with both of these measures (total reading time and first pass reading time: r = .61,

$p < .001$; total reading time and rereading time: $r = .73$, $p < .001$). Total reading time was also associated with all variables that predicted first pass reading time and rereading time.

Finally, single-word skipping was affected by TE as well as by frequency-related variables (item frequency, mutual information, and single word frequency). In line with findings in unimodal reading (e.g., Rayner, 1998), highly frequent component words (e.g., *into, off,* …) tended to be skipped more often. The effect of mutual information suggests that a weak association between the words in a MWU was also likely to result in skipping. TE was associated with increased odds of reading both words in the MWU, which suggests that TE led to increased visual attention to the written form of the entire phrase.

Taken together, the results for the duration measures indicate that learners spent more time fixating underlined target items, but due to time limits, they mainly processed them during the first pass. It is important to mention that our analysis only focused on visual attention, although learners could also process the target items in spoken form. We need to be careful in drawing strong conclusions about early and late processes in caption reading, as the assumptions about different stages of processing may not hold when auditory input is presented simultaneously (see for example Conklin et al., 2020). For example, we did not analyze to what extent learners integrated information from both input modalities, and how this may have affected word identification and meaning integration. It is possible that the auditory support changed the degree or quality of processing of the target items. Caption duration was an important predictor of reading times, but exactly which elements of word processing could occur in the limited presentation time of the captions remains open for investigation. Finally, in line with previous findings (Choi, 2017), our results show that pretest knowledge of the MWUs affected reading times. Items unknown in the pretest tended to be fixated longer, which suggests that the novelty of certain phrases may have rendered them more salient. Novel words have been shown to attract attention in reading (e.g.,

Godfroid et al., 2013), and our results suggest that novelty may affect multiword processing as well. However, many of the MWUs contained low-frequency words (e.g., *epithet, depravity*), which may have contributed to this novelty effect. Another explanation for the effect of pretest knowledge could be that learners paid special attention to previously (partially) unknown items because they remembered them from the form recall pretest. In the form recall test, we found that knowledge improved for some of the distractor items, indicating learning from the pretest itself. In addition, participants reported that the pretest caused them to pay closer attention to certain MWUs, sometimes in anticipation of an (unannounced) posttest. These findings indicate that pretesting can affect learners' engagement with linguistic items during the learning treatment, and may, to some extent, even undermine the construct validity of incidental learning, which is often defined in terms of a primary focus on meaning (e.g., Swanborn & de Glopper, 1999). In the current study, the pretest (in addition to TE) may have enhanced learners' attention to MWUs in the input, even with a one-week interval between the pretest and the treatment.

Research question 2: do TE and visual attention affect form recall of MWUs encountered in captioned audiovisual input?

To examine the effects of TE and visual attention on learners' recall of unknown MWUs encountered in the input, we used mixed effects logistic regression. The results show that TE only contributed to learning odds when reading times were not taken into account. Once total reading time was entered into the model, the effect of TE was no longer significant. This suggests that the amount of (visual) attention was more important than the experimental manipulation, or that the effect of the learning intervention depended on how learners engaged with the items in the input. As discussed above, TE can clearly affect visual processing, and therefore has the potential to promote vocabulary learning. However, remembering the form and meaning of MWUs from a single exposure is likely to rely on the

degree and quality of processing, which cannot be controlled directly by means of TE (see also Leow and Martin, 2017). Other variables (such as the inclusion of a pretest in the current study) may contribute to depth of processing as well.

Nevertheless, the significant relationship between TE and learning gains before total reading time was entered into the regression models confirms that TE can support incidental learning of MWUs under certain conditions (e.g., Choi, 2017). In Majuddin et al.'s (2021) study, higher average scores were found for enhanced captions compared to unenhanced captions, but the difference did not reach significance when other variables (number of viewings, VST score) were taken into account. The researchers offered a number of explanations why the effect of TE might not be as outspoken in audiovisual input, such as the limited amount of time for caption reading, the length of the MWUs in their experiment (which included items of 5 words), and the distribution of attentional resources between the captions and imagery. These explanations are consistent with the findings of the current study, and they reveal the complex and multifaceted nature of audiovisual input. In line with studies that emphasize the importance of cognitive load in multimodal processing (e.g., Gass et al., 2019), the effectiveness of enhanced captions for learning MWUs is likely to depend on variables related to the input and the learners. It appears that TE might benefit learning provided that learners can fluently read the L2 captions, and distribute their attention efficiently.

Overall, our findings suggest that TE can promote learners' attention to unknown MWUs encountered in captioned video, but that engagement with the input more strongly affects learning than TE by itself. Further, because semantic decomposability played an important role in predicting learning gains, we cannot draw strong conclusions about form-meaning mapping. Our results show that participants could remember the form of MWUs that were semantically transparent, such as *evolutionary advantage*. However, MWUs with low

decomposability, which in our sample were generally idiomatic items or items containing low-frequency single words (e.g., *beyond the pale*), may require more contextual support and possibly also longer exposure time to allow for more elaborate semantic processing. The small gains found for non-decomposable items suggest that TE alone was insufficient to overcome the high learning burden of semantically less accessible MWUs from a single exposure in captioned video. One unexpected finding was a negative correlation between caption duration and learning odds. This finding may seem counterintuitive, especially as longer reading times led to higher gains. We cannot offer a clear explanation based on our data, but it is possible that auditory and visual information were integrated differently in shorter versus longer captions. In multimodal input, L2 learning may not just rely on the amount of (visual) attention to relevant information, but also on the way in which information from different input modalities is combined in memory. Currently, there is hardly any research that has investigated the effects of auditory processing on reading patterns and learning gains in multimodal input with moving imagery (see Wisniewska & Mora, 2018 for an exception). This could be an interesting avenue for future research.

Limitations

Our study has several limitations. Although the goal of the experiment was to measure meaning-focused or incidental learning, and how TE affects this process, the results of our questionnaire show that some participants noticed items from the pretest while they were watching the documentary. We suggest that future studies consider other methods to control for prior item knowledge (see e.g., Sonbul and Schmitt, 2013).

Another limitation is our reliance on a small sample of participants and a small, varied sample of MWUs. Further, although we focused on learners' visual processing of MWUs, the target items were also presented in spoken form, which may have affected processing. In addition, although the input chosen in the current study did not contain any explicit visual

cues to the meaning of the target items, transitions between the imagery region and the caption region may have affected processing of the target items (e.g., Bisson et al., 2014).

Conclusion

Despite these limitations, our study provides further empirical support for the beneficial effect of TE in captioned audiovisual input (Cintrón-Valentín et al., 2019; Lee and Révész, 2018, 2020; Montero Perez et al., 2014, 2015, 2018), and extends the findings of previous studies by examining how TE can affect learners' visual processing of MWUs in captioned audiovisual input. It seems that the attention-raising effect of TE has the potential to increase the likelihood that MWUs are picked up. However, the effectiveness of TE in captioned television may depend on factors related to item difficulty and processing load. Further research is needed to examine how different input modalities are integrated during L2 processing of captioned audiovisual input, and how this might affect the acquisition of MWUs.

Notes

1. Because captions were presented in a proportional font instead of a monospaced font, spatial dimensions of characters had to be estimated (see Godfroid, 2020: 175–176).
2. Lowering the sampling rate made it easier to track the eyes of some participants. This means that the sampling error was not the same for all participants, increasing the amount of individual variability in the data. However, simulations show that differences in fixation durations at different sampling rates tend to be negligible (Andersson et al., 2010). Further, because conditions were compared within participants, we do not expect that different sampling rates led to any systematic differences that might confound the effect of the treatment.

3. Of the 28 MWUs, 22 were not fixated by at least one participant. We could not discern any patterns in full phrase skipping (at the item nor the participant level), and the number of cases (n = 60) was too small for a statistical analysis.

Acknowledgements

We thank the anonymous reviewers for their constructive feedback on earlier versions of the paper. We also thank the participants for their time and efforts.

References

Alsadoon R and Heift T (2015) Textual input enhancement for vowel blindness: A study with Arabic ESL learners. *The Modern Language Journal* 99: 57–79.

Bisson MJ, Van Heuven WJB, Conklin K and Tunney RJ (2014) Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics* 35(2): 399–418.

Boers F and Lindstromberg S (2009) *Optimizing a lexical approach to instructed second language acquisition.* Basingstoke: Palgrave Macmillan.

Boers F, Demecheleer M, He L, Deconinck J, Stengers H and Eyckmans J (2017) Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics* 26: 448–469. https://doi.org/10.1111/ijal.1214

Choi S (2017) Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research* 21(3): 403–426. doi: 10.1177%2F1362168816653271

Cintrón-Valentín M, García-Amaya L and Ellis NC (2019) Captioning and grammar learning in the L2 Spanish classroom. *The Language Learning Journal* 47(4): 439–459. doi: 10.1080/09571736.2019.1615978

Conklin K, Alotaibi, S, Pellicer-Sánchez, A and Vilkaitė-Lozdienė, L (2020) What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research.* https://doi.org/10.1177/0267658320921496.

Conklin K, Pellicer-Sánchez A and Carrol G (2018) *Eye-tracking: A guide for applied linguistics research.* Cambridge University Press.

Corder SP (1967) The significance of learners' errors. *International Review of Applied Linguistics* 5: 161–169. doi: 10.1515/iral.1967.5.1–4.161

Durbahn M, Rodgers M and Peters E (2020) The relationship between vocabulary and viewing comprehension. *System* 88. https://doi.org/10.1016/j.system.2019.102166.

Durrant P and Schmitt N (2010) Adult learners' retention of collocations from exposure. *Second Language Research* 26: 163–188.

Feng Y and Webb S (2020) Learning vocabulary through reading, listening and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition* 1–22. doi: 10.1017/S0272263120000297

Gass S, Winke P, Isbell DR and Ahn J (2019) How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology* 23(2): 84–104.

Godfroid A and Hui B (2020) Five common pitfalls in eye-tracking research. *Second Language Research* 36(3): 277–305. doi: 10.1177/0267658320921218

Godfroid A, Ahn I, Choi I, Ballard L, Cui Y, Johnston S, Lee S, Sakar A and Yoon H (2018) Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition* 21(3): 563–584.

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

Horst M, Cobb T and Meara P (1998) Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language* 11: 207–223.

Laufer B and Hulstijn J (2001) Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics* 22(1): 1–26.

Laufer B and Waldman T (2011) Verb–noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning* 61: 647–672. doi: 10.1111/j.1467–9922.2010.00621.x

Lee M and Révész A (2018) Promoting grammatical development through textually enhanced captions: An eye-tracking study. *The Modern Language Journal* 102(3): 557–577. doi: 10.1111/modl.12503

Lee M and Révész A (2020) Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition* 42(3): 625–651. doi: 10.1017/S0272263120000108

Leow RP and Martin A (2017) Enhancing the input to promote salience of the L2: A critical overview. In: Gass S, Spinner P, and Behney J (eds) *Salience in SLA*. New York: Routledge, pp.167–186.

Lin PMS and Siyanova A (2014) Internet television for L2 vocabulary learning. In: Nunan D and Richards JC (eds) *Language learning beyond the classroom*. London: Routledge: pp.149–158

Majuddin E, Boers F and Siyanova-Chanturia A (2021) Incidental acquisition of multiword expression through audiovisual materials: the role of repetition and typographic

enhancement. *Studies in Second Language Acquisition*. Advance online publication. https://doi.org/10.1017/S0272263121000036

Montero Perez M, Peters E, Clarebout G and Desmet P (2014) Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology* 18: 118–41.

Montero Perez M, Peters E and Desmet P (2015) Enhancing vocabulary learning through captioned video: an eye-tracking study. *The Modern Language Journal* 99: 308–28. doi:10.1111/modl.12215.

Montero Perez M, Peters E and Desmet P (2018) Vocabulary learning through viewing video: the effect of two enhancement techniques. *Computer Assisted Language Learning* 31: 1–26. doi:10.1080/09588221.2017.1375960.

Muñoz C (2017) The role of age and proficiency in subtitle reading: An eye-tracking study. *System* 67: 77–86.

Nicklin C and Vitta J P (2021) Effect-Driven Sample Sizes in Second Language Instructed Vocabulary Acquisition Research. *The Modern Language Journal* 105(1): 218-236. doi: 10.1111/modl.12692

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

Pellicer-Sánchez A (2017) Learning L2 collocations incidentally from reading. *Language Teaching Research* 21: 381–402. doi:10.1177/1362168815618428

Peters E (2012) Learning German formulaic sequences: The effect of two attention-drawing techniques. *The Language Learning Journal* 40: 65−79.

Peters E (2019) The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly* 53(4): 1008–1032. doi: 10.1002/tesq.531

Peters E and Webb S (2018) Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition* 40: 551–77. doi:10.1017/S0272263117000407

Peters E, Heynen E and Puimège E (2016) Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System* 63: 134–148. doi: 10.1016/j.system.2016.10.002

Puimège E and Peters E (2019) Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal* 47: 424–438. doi: 10.1080/09571736.2019.1638630

Puimège E and Peters E (2020) Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition* 42(3): 525-549. doi: 10.1017/S027226311900055X

Pujadas G and Muñoz C (2019) Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal* 47: 479–496. https://doi.org/10.1080/09571736.2019.1616806

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

Robinson P (1995) Attention, memory and the “noticing” hypothesis. *Language Learning* 45: 283–331.

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

Schmitt N, Schmitt D and Clapham C (2001) Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* 18(1): 55–88. doi: 10.1177/026553220101800103

Sharwood Smith M (1993) Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition* 15: 165–179.

Siyanova-Chanturia A and Pellicer-Sánchez A (eds) (2019) *Understanding formulaic language: A second language acquisition perspective.* London, UK: Routledge.

Sonbul S and Schmitt N (2013) Explicit and implicit lexical knowledge acquisition of collocations under different input conditions. *Language Learning* 63: 121–159. https://doi.org/10.1111/j.1467–9922.2012.00730.x

Swanborn M S L and de Glopper K (1999) Incidental word learning while reading: A meta-analysis. *Review of Educational Research* 69: 161–285.

Szudarski P and Carter R (2016) The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics* 26: 245–265. https://doi.org/10.1111/ijal.12092

Szudarski P (2012) Effects of meaning- and formed-focused instruction on the acquisition of verb–noun collocations in L2 English. *Journal of Second Language Teaching and Research* 1*:* 3–37.

Toomer M and Elgort I (2019) The Development of Implicit and Explicit Knowledge of Collocations: A Conceptual Replication and Extension of Sonbul and Schmitt (2013). *Language Learning* 69(2): 405–439.

Van Zeeland H and Schmitt N (2013) Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System* 41: 609–24.

Vilkaitė L and Schmitt N (2019) Reading Collocations in an L2: Do Collocation Processing Benefits Extend to Non-Adjacent Collocations? *Applied Linguistics* 40(2): 329–354.

Von der Malsburg T and Angele B (2017) False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language* 94: 119–133.

Webb S and Rodgers MPH (2009) Vocabulary demands of television programs. *Language Learning* 59: 335–366.

Webb S, Newton J and Chang A (2013) Incidental learning of collocation. *Language Learning* 63: 91–120. doi:10.1111/j.1467–9922.2012.00729.x

Webb S (2020) Incidental vocabulary learning. In: Webb S (ed) *The Routledge handbook of vocabulary studies*. New York, NY: Routledge, pp. 225–239.

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 and Sydorenko T (2013) Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal* 97: 254–75.

Wisniewska N and Mora JC (2018) Pronunciation learning through captioned videos. In: Levis J (ed) *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference.* Ames, IA: Iowa State University, pp. 204–215.