Spring 3-15-2021

Incidental acquisition of multiword expressions through audiovisual input: The role of repetition and typographic enhancement

Elvenna Majuddin

Anna Siyanova-Chanturia

Frank Boers
fboers@uwo.ca

Citation of this paper:
Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual input: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition.* Online First.

This is a pre-final version of:
Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual input: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition.* Online First.

Incidental acquisition of multiword expressions through audiovisual input: The role of repetition and typographic enhancement

Elvenna Majuddin, Anna Siyanova-Chanturia and Frank Boers

Abstract:

There has been limited research on the efficacy of captioned second language (L2) television in facilitating the incidental acquisition of multiword expressions (MWEs). The present study aims to fill this gap. Additionally, this study examined the role of typographic enhancement and repetition. One-hundred and twenty-two L2 learners were assigned to one of six conditions which differed in terms of caption condition (no captions, normal captions, enhanced captions) and the number of times they watched the same video (once, twice). The participants took a cued MWE form recall test before watching the video, and immediately and two weeks after watching it. A content comprehension test was also administered. Repetition resulted in better content comprehension as well as better acquisition of MWEs. Both caption types positively influenced MWE recall relative to watching the video without captions, but typographic enhancement reduced the benefits of captions for content comprehension.

INTRODUCTION

The last decade has seen a proliferation of research on pedagogic interventions intended to improve second language (L2) learners' mastery of *multiword expressions* (MWEs). This line of research has gained unprecedented momentum by the recognition that MWE competence is an integral part of proficiency and is associated, for instance, with fluent language use (e.g., Boers et al., 2006; Tavakoli & Uchihara, 2019). MWEs encompass a large set of expression types, such as collocations (*strong wind*), idioms (*tie the knot*), binomials (*time and money*), lexical bundles (*one of the*), proverbs (*better late than never*), and so on (Siyanova-Chanturia & Van Lancker

Sidtis, 2019). Unlike another commonly used term, *formulaic language* (or *formulaic sequence*, Wray 2002), which may refer to both single words and multi-word items, the term MWE[i] necessarily implies a unit longer than a single word (e.g., Siyanova-Chanturia & Omidian, 2020; Siyanova-Chanturia & Pellicer-Sánchez, 2019).

It is well-established in the literature that a large proportion of language is made up of MWEs (e.g., Erman & Warren, 2000; Hill, 2000). It follows that, for L2 learners to benefit from MWE knowledge on par with native speakers, they need to master a large repertoire of MWEs. Given the limited classroom time in many L2 contexts (such as English as a Foreign Language learning contexts), only a fraction of such a large MWE repertoire can realistically be acquired through explicit MWE focused instruction. It is, therefore, unsurprising that researchers have explored ways to support incidental MWE acquisition; that is, acquisition as a by-product of activities where learners attend primarily to the content of messages rather than their linguistic packaging. This specific line of research on the development of L2 MWE knowledge, however, has focused almost exclusively on written input (see further below), not audio-visual input.

A substantial number of studies have indicated that audio-visual input can be beneficial for vocabulary acquisition (e.g., Peters & Webb, 2018). One strand of this research has furnished evidence that viewing captioned videos typically leads to superior uptake of new words compared to viewing the same videos without captions (see Montero Perez et al., 2013, for a meta-analysis). Whether the benefits of L2 viewing with captions extend to incidental acquisition of MWEs remains underexplored, however. The present study aims to fill this gap, and addresses two additional questions. One is whether the benefits of captions could be boosted if the MWEs in the captions are typographically enhanced. The other question is whether viewing the same video twice instead of once positively affects MWE uptake regardless of caption condition. Although research on incidental MWE learning has attested to the positive effects of typographic enhancement (e.g., Boers et al., 2017; Choi, 2018), especially when combined with repetition (Szudarski & Carter, 2016), these factors have been investigated almost exclusively in reading studies so far. It seems plausible that they also contribute to incidental MWE learning through L2 captioned viewing, and it is this possibility that the present study seeks to explore. At the same time, watching a captioned video is different from the self-paced reading of written texts, because the processing of captions happens under time pressure (unless one pauses the video)

and captions do not stay available for re-examination (unless one rewinds the video). Moreover, when watching captioned video, more modalities than just written text invite attention. It cannot be taken for granted, therefore, that a text modification technique which has been found beneficial for MWE uptake from self-paced reading will be equally beneficial when it is applied to captioned videos. In addition to evaluating the role of typographic enhancement and repetition in MWE learning, the present study also investigates the effect of these two factors on L2 learners' comprehension of the content of a video.

BACKGROUND

INCIDENTAL VOCABULARY LEARNING AND L2 VIEWING

The potential benefits of audio-visual input for vocabulary uptake have received increasing attention in the recent years. This shift from a focus on written input is timely, as recent surveys have revealed that people now tend to spend more time watching television than reading books. In a recent survey on media consumption by Roy Morgan Research (2015), for example, data collected from eleven countries across the Asia-Pacific region showed that people spend an average of 8.2 hours to 29.5 hours a week watching television. In her survey that investigates Flemish EFL learners' exposure to English language media, Peters (2018) found that more than 40% of these learners report regular watching of English TV with or without subtitles. In addition, these learners, aged 16 and 19, report seeking only limited exposure to written input such as books and magazines. This echoes Lindgren and Muñoz (2013), who found that young EFL learners rely more on subtitled movies than books for foreign-language exposure.

The potential of authentic audio-visual materials, such as movies and TV shows, for vocabulary acquisition has been pointed out through lexical analyses of large samples of such materials (Webb & Rodgers, 2009a). Especially if learners watch materials that are thematically related will they stand a good chance of encountering the same mid and low frequency words multiple times (Rodgers & Webb, 2011). Clearly, there are merits in watching TV for vocabulary learning and this has prompted some researchers to advocate extensive TV viewing to bolster incidental L2 vocabulary learning (Webb, 2015), analogous to earlier proposals for extensive reading programs.

If authentic audio-visual input, including TV shows, is a good resource for vocabulary acquisition, then the next question is what can be done to make optimal use of this resource. A substantial body of empirical research has gauged the benefits of captions, or on-screen text in the same language as the audio, in comparison with uncaptioned input. Two main benefits of captioned video have been reported. Firstly, most studies have shown better content comprehension for caption groups compared to no caption groups (e.g., Gass et al., 2019; Winke et al., 2013; Montero Perez et al., 2013 for meta-analysis). Secondly, they have also almost consistently furnished evidence in favour of captioned viewing for vocabulary learning (e.g., Cintrón-Valentin et al., 2019; Markham, 1999; Neuman & Koskinen, 1992; Winke et al., 2010; for a meta-analysis, see Montero Perez et al., 2013). One explanation for the positive effect of captions on vocabulary acquisition is that the orthographic representation of the words helps learners to recognize word boundaries, which is more challenging when only aural input is available (Bird & Williams, 2002; Winke et al., 2010). Surprisingly, whether the same benefit for vocabulary acquisition extends to the learning of MWEs has hardly been investigated. Thanks to a recent study by Puimège and Peters (2019), we know incidental uptake of MWEs from uncaptioned audio-visual material is possible. The question, then, is whether adding captions harnesses its potential. For example, MWEs often contain function words (e.g., articles, prepositions) and these tend to be phonologically reduced in natural speech. Seeing these words written in the captions that accompany speech may help learners notice these small constituents of MWEs, distinguish them from the content words in the MWEs, and thus aid uptake of the precise lexical composition (or form) of MWEs. The chances of learners attending to MWEs and their components in captions may be further improved through modifications to the captions that make the MWEs stand out. This possibility is what we turn to in the following section.

MAKING MULTIWORD EXPRESSIONS STAND OUT

Incidental learning is usually defined as learning that takes place as a by-product of activities where learners focus primarily on content (Schmitt, 2010). However, because it is hard to determine to what extent learners spontaneously turn their attention to language items or patterns as study objects during content-oriented activities, Hulstijn (2001) suggests that a critical operational feature that distinguishes intentional learning from incidental learning conditions is learners' expectation of a language test. When input is processed for content, the

precise wording of that content does not usually attract much attention. Yet, attention is vital for learning (Schmidt, 2001). To direct learners' attention to certain language features or elements, researchers have proposed to use typographic enhancement, such as the use of underlining and bolding, to make these features or elements stand out (e.g., Sharwood-Smith, 1993). A few studies have investigated whether this kind of modification of reading texts improves the chances of MWEs being noticed and remembered, and the results have been positive (Bishop, 2004; Boers et al., 2017; Choi, 2018; Sonbul & Schmitt, 2013; Toomer & Elgort, 2019). In addition, a study by Szudarski and Carter (2016), which we will return to below, showed that a combination of typographic enhancement and repeated encounters with the items led to better collocational knowledge compared to repetition alone.

Given that typographic enhancement promotes MWE learning in the context of written texts, it is conceivable that it is beneficial also in the case of captioned videos. However, there has been limited research on this, and the available research (see below) has examined uptake of words rather than MWEs. Another modification to captions which has attracted some interest from researchers is to limit the captions to key words instead of including all the words present in the audio-recording. Keyword captioning is another means to make selected words stand out, because they are visually foregrounded relative to what is not included in the captions. Montero Perez et al. (2015; 2018) included a comparison of vocabulary uptake from full captions and from keyword captions, and found an advantage for the latter, albeit mostly in posttests asking participants if they recognized the lexical items as ones they had encountered in the videos. Posttests about the meaning of the items yielded much poorer results overall, which is not surprising since captions (in L2) represent lexical forms, not the meanings of these forms. The target lexical items in Montero Perez et al. included both single words and MWEs, but no separate analyses were reported for the two categories. Teng (2019) also included a comparison of full captions and keyword captions, and his target items were all MWEs (verb-noun collocations). Rather surprisingly, the full captions were found to generate the better learning gains in this study.

To the best of our knowledge, there have only been two studies thus far that have examined the effect of typographically enhancing segments of full captions. In Montero Perez et al. (2014), L2 learners of French watched a video under one of four conditions: without captions,

with regular captions, with keyword captions, and with regular captions in which the key items were highlighted. All three caption groups outperformed the no-caption group in a test which asked them if they recognized the items. They also did slightly better on a multiple-choice meaning-recognition test, but here only the keyword caption and enhanced caption groups outperformed the no-caption group at a statistically significant level. There were no statistically significant differences between the test scores of the three caption groups. Montero Perez et al. (2014) also included tests about the content of the videos. Scores on these tests revealed no significant differences among the participant groups. It is worth mentioning that the content-related questions did not concern text portions featuring the target vocabulary. The second study that has examined the effect of typographically enhanced captions is Cintrón-Valentin et al. (2019). Animated video clips were integrated into an L2 Spanish course in three versions: uncaptioned, captioned with target vocabulary highlighted, and captioned with target grammar patterns highlighted. The captions with typographically enhanced vocabulary items generated significantly better vocabulary uptake than the other two viewing conditions, and even the caption condition where grammar patterns instead of vocabulary items were enhanced led to better vocabulary uptake than the condition without any captions, which again demonstrates that captions are to some extent beneficial for vocabulary uptake even without special steps to direct viewers' attention to lexical items.

The studies reviewed above furnish no evidence of a trade-off effect whereby typographic enhancement benefits uptake of what is enhanced but detracts attention from other elements. However, evidence of such an effect did emerge from a study by Choi (2018) of MWE uptake from reading. In this study, the students who read a text in which target collocations were enhanced recalled more of these target collocations compared to their peers who read the unenhanced version. However, the latter outperformed the former by 48% in the recall of content words which had been left unenhanced in both text versions. This suggests that typographically enhanced portions of a text can attract attention at the expense of the other portions of the text. This is reminiscent of the findings of some studies on typographic enhancement to assist grammar learning, which found that the use of enhancement promoted the uptake of grammar patterns, but impaired recollection of text content (e.g., Lee , 2007). It seems plausible that this trade-off may also occur with captioned video, perhaps especially because this type of input requires a greater

distribution of attentional resources. In short, the role of typographic enhancement in fostering MWE uptake as well as content uptake from audio-visual input warrants further research.

THE ROLE OF REPETITION

Repetition, or frequency of occurrence, has been established to be beneficial for the learning of unknown words (Uchihara et al., 2019, for a meta-analysis). Provided there are enough encounters with the same words, learning can occur (Horst et al., 1998; Rott, 1999; Webb, 2007). While there is a substantial number of reading studies on the effects of repetition on the acquisition of single words, research investigating the effects of repetition on incidental MWE acquisition is far more limited.

An example of such a study is Pellicer-Sánchez (2017), who found that encountering the same collocations in a text eight times led to better posttest scores than encountering these collocations 'only' four times, although the difference fell short of statistical significance. It is worth mentioning that the collocations in Pellicer-Sánchez (2017) were made up of adjective–pseudo-noun combinations. The positive role of repetition for the learning of real collocations was shown in a study by Durrant and Schmitt (2010). Unlike Pellicer-Sánchez's study, repetition in Durrant and Schmitt's study was not operationalised as repeated encounters of target items within a text. Instead, the participants were exposed to adjective-noun pairs under one of the following three conditions: (i) single exposure to the target collocations embedded in a sentence, (ii) verbatim repetition, i.e. repeated exposure to the same sentences containing the target collocations, or (iii) varied repetition, i.e. repeated exposure to different sentences containing the target collocations. Both repetition conditions yielded superior levels of recall compared to the single exposure. Further, verbatim repetition was found to lead to higher gains than varied repetition. In essence, Durrant and Schmitt's study provides insight into how repetition or frequency of occurrence could be operationalised when using authentic unmodified materials, and suggests that exposure to target items in the same context rather than different contexts is useful for MWE learning (at least at an early stage of learning the given MWEs). This informs our design feature, in that repetition is operationalised as exposure to target MWEs twice, through repeated viewing of the same input video. This approach was also taken by Winke et al. (2010), one of whose findings was that watching a video twice with captions led to better vocabulary learning than watching it twice without captions.

Evidence of the role of repetition has been found in the context of bimodal input as well. In Webb et al. (2013), participants read short stories while listening to an audio-recording under four treatment conditions that varied in number of encounters (in various contexts) with the target collocations. Depending on the text version, the collocations occurred up to 15 times over the course of approximately 30 minutes of reading. As expected, repetition led to more collocational knowledge.

The role of repetition in engendering MWE knowledge has thus far only been investigated in the context of written and bimodal input. To our knowledge, no study to date has investigated the role of repeated viewing in MWE learning. At least three viewing studies, however, have found a positive relationship between frequency of occurrence and the learning of single words. As part of a longitudinal study, Rodgers (2013) compared the effects of captioned and uncaptioned viewing on Japanese EFL learners' incidental vocabulary learning. The learners encountered new words repeatedly through watching multiple episodes of a TV program. A positive correlation emerged between the number of encounters and learning, but, surprisingly, there was no significant difference between the relative vocabulary gains for the caption and no-caption conditions. A study which did find an effect for captions as well as frequency was conducted by Peters et al. (2016), but in this study participants watched videos with either captions or subtitles (L1 captions). The frequency of occurrence of the words and learners' prior vocabulary knowledge were found to be positively associated with the learning gains. A study by Peters and Webb (2018) also found positive associations between incidental vocabulary learning and both the number of encounters with the target words and the learners' prior vocabulary knowledge. The likelihood that more proficient learners and learners with comparatively good vocabulary knowledge pick up new words from audio-visual materials faster than less proficient learners has been demonstrated in other recent studies (e.g., Montero Perez, 2020; Pujadas & Muñoz, 2019).

The studies reviewed here suggest that repeated viewing will improve the chances of incidental MWE acquisition from audio-visual input. Repeated viewing has also been reported to benefit text comprehension, as listening a second time enhances content comprehension (Lund, 1991; Nguyễn, 2017; Sakai, 2009). This study seeks further evidence that repeated viewing benefits both MWE learning and comprehension.

RATIONALE AND RESEARCH QUESTIONS

The present study aims to investigate the effects of L2 viewing, with and without captions, on the incidental learning of MWEs. Previous research on incidental MWE learning has mostly focused on written input, and (to a much lesser extent) on bimodal input. Although such studies have attested to the positive effects of typographic enhancement, whether this is equally beneficial for MWE uptake from audio-visual input is not yet clear. On the one hand, it is conceivable that distributing their attention among three types of stimuli (i.e., text, audio and images) may be a challenge for learners, and the presence of typographic enhancement may add to the cognitive load. On the other hand, typographic enhancement may direct learners' attention to the target MWEs, which they might otherwise not notice due precisely to the distribution of attentional resources that multimodal input requires. If typographic enhancement achieves the latter aim, the next question, however, is whether it risks causing a trade-off between attending to the enhanced elements and fully engaging with the video content.

The present study also aims to shed light on the role of repeated viewing. It is well established in the literature on incidental MWE learning that repetition tends to lead to bigger learning gains. This study is the first to examine whether the same applies in the context of L2 viewing. Moreover, this study provides insights into the effectiveness of repeated viewing for each caption condition. This, in turn, may provide valuable information on whether repeated viewing is worth the additional investment of classroom time.

The study thus addresses the following research questions:

1. Is there an effect of caption condition (i.e., enhanced captions, normal captions, no captions) on incidental learning of MWEs?

2. Is there an effect of caption condition (i.e., enhanced captions, normal captions, no captions) on content comprehension?

3. Is there an effect of repeated viewing on incidental learning of MWEs under the various caption conditions?

4. Is there an effect of repeated viewing on content comprehension under the various caption conditions?

To answer these research questions, a pretest – posttest – delayed form-recall posttest between-participant design was adopted. Incidental learning was operationalised as learning that takes place without prior test announcement, in keeping with Hulstijn's (2001) suggestion. As such, in the current study, learners were not forewarned of vocabulary tests. Instead, to ensure they would try to engage with the content of the video, it was announced that comprehension questions would follow immediately after viewing the video.

METHOD

PARTICIPANTS

One-hundred and twenty-two Malaysian L2 English learners took part in this study. The participants were college students (i.e., pre-university) working towards their Diploma in various disciplines such as Hospitality Management and Information Technology Systems. They ranged in age between 17 and 23 (M = 18.7; SD = .09). The participants had either Malay (n = 10) or a Chinese dialect (n = 116) as their L1. Their mean Vocabulary Size Test (VST) score (Nation & Beglar, 2007) was 66 out of 140 test items (SD = 21), and the range of scores indicated they had (receptive) knowledge of between 4,000 and 9,000 most frequent word families in English. The participants came from six intact classes, and these intact classes were randomly assigned to one of the six experimental conditions, differing in number of viewings (i.e., one or two viewings) and caption condition (i.e., no captions, normal captions or enhanced captions). There was no significant difference between the six groups (see Table 1) in their scores on the VST (Kruskal-Wallis $\chi 2$ (5) = 4.57; p = .47).

TABLE 1 Number of participants and the mean VST score (out of 140) under each condition

Condition	N	Mean (*SD*)
Enhanced Captions X 1	19	68.37 (11.31)
Normal Captions X 1	23	73.04 (18.38)
Uncaptioned X 1	15	74.40 (16.63)
Enhanced Captions X 2	25	65.92 (16.75)
Normal Captions X 2	22	71.09 (11.49)
Uncaptioned X 2	18	66.39 (14.02)

Note. 1 = one viewing; 2 = two viewings.

The participants were also given the Vocabulary Levels Test (Schmitt et al., 2001). A cut-off point for showing mastery of level was a score of 24 out of 30. As shown in Table 2, all six conditions had mean scores that indicate a mastery of the 2,000 word level.

TABLE 2: mean VLT scores for each condition

	VLT 2000		VLT 3000		VLT 5000	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Enhanced Captions X 1	26.00	3.51	21.89	4.46	16.32	4.62
Normal Captions X 1	26.70	3.24	24.35	5.31	18.13	8.92
Uncaptioned X 1	27.13	2.97	24.00	5.07	18.07	6.63
Enhanced Captions X 2	25.60	3.52	22.00	5.55	15.96	7.38
Normal Captions X 2	26.91	5.02	24.27	5.00	18.05	5.91
Uncaptioned X 2	25.59	4.91	22.11	5.70	15.67	5.93

MATERIALS

Audio-Visual Input

When considering the input materials for this research, a few factors had to be considered. Firstly, the video had to be appealing enough for motivational purposes. As repeated viewing is one of the variables in question, it was important that the video could sustain the learners' interest enough for them to watch the video from start to finish and for a second viewing. Another factor that influenced the choice of input material was the number of MWEs contained in the videos that the participants were unlikely to be familiar with. At the same time, the video had to be easy to follow in terms of other language dimensions, such as lexical coverage and speed of dialogue.

With these considerations in mind, an episode of an American comedy series titled *Fresh off the Boat* was chosen. American sitcom was chosen as the participants are more accustomed to the American accent than the British accent. Further, Lin's (2014) study of internet television found that the comedy genre contains a high number of everyday spoken formulaic sequences. In addition, caption studies which used comedy series (e.g., Peters et al., 2016; Sydorenko, 2010) have reported encouraging findings for single word learning. The chosen episode (Episode 2, Season 1) is 20-minutes long. An analysis of the lexical profile of the video using RANGE (Nation & Heatley, 2002) showed that, together with proper nouns and marginal words, the most frequent 1,000, 2,000 and 3,000 word families provided 91.16%, 94.87% and 96.37% coverage (including proper nouns and marginal words) of the script's total running words, respectively. To understand L2 TV, Webb and Rodgers (2009b) proposed that learners are likely to need a vocabulary size of 2,000 to 4,000 word families (plus proper nouns and marginal words). Moreover, a recent study on the relationship between vocabulary and viewing comprehension by Durbahn et al., (2020) suggests lexical coverage of around 90% is enough for adequate viewing comprehension. Thus, the participants' VLT scores (see above) suggest that they were able to follow the video.

The video was also piloted with a group of participants (n = 12) similar to those in the present study, to ascertain that the video was easy enough for the learners to follow, and interesting enough to watch a second time.

Target Items

Exposing learners to authentic material entails that they will encounter MWEs of diverse kinds. Many previous experimental studies on MWE learning focused on certain parts of speech, such as adjective + noun collocations or verb + noun collocations, but this necessitated modifying texts to include sufficient numbers of these particular MWE types. If the aim is to explore ways of improving the chances of acquisition from authentic audio-visual materials, then it seems more ecologically valid to keep those materials intact. Moreover, there are indications that the risk of inter-item interference is increased when learners are presented with sets of MWEs that are syntactically similar (Boers et al., 2014). In any case, the phraseology of a language naturally comprises diverse kinds of MWEs and there seems to be little justification for prioritising certain syntactic categories over others. It is, nonetheless, acknowledged that not all MWEs are 'equal' in the way that they are processed (e.g., Columbus, 2010), or in the challenges they pose to L2 learners (e.g., Boers, 2020). However, in the present study this inter-item variability applied across all six conditions due to the between-participant design of the experiment where all groups were exposed to the same text.

The video was first screened for MWEs likely to be as yet unfamiliar to the students in this study. To ascertain that these expressions were relatively well established in (American) English, the Corpus of Contemporary American English (COCA) (Davies, 2008) was consulted. All the potential target MWEs[ii] yielded at least 100 hits in this corpus and/or were found to have a mutual information (MI) score (i.e., a measure of collocational strength) of at least 3.0.

After compiling a list of potential target MWEs from the video, four teachers of the participants were consulted on whether they believed that the participants had knowledge of these items. Guided by this information, 20 MWEs were chosen as target items, to be included in the tests. Each MWE only occurred once in the video. Two items attracted over 30% correct responses at the pretest stage. This could be because these items were highly guessable, owing to the first letter cues. As such, these items were not included as the target items. Table 3 lists the 18 final target items and their frequencies and MI scores in COCA (at the time the study was designed).

TABLE 3. Target items with corpus frequency and MI score

MWEs	COCA	MI Score
on the same page	796	9.93
turn a profit	484	10.48
on (someone's) hands	4071	4.52
tighten up	554	5.43
kill (someone) with kindness	28	11.76
root for	1845	2.48
(someone's) hands are tied	132	5.55
slippery slope	676	13.87
put (someone) on the spot	242	5.04
talk some sense into (someone)	62	5.98
beg to differ	288	10.98
whisked away	264	6.38
chip in	1279	1.54
bear with (someone)	517	0.35
go through a rough patch	28	14.51
let up on (someone)	106	1.3
look out for (someone)	2174	3.18
work (something) out with (someone)	25	8.53

For the normal captions condition, all the target items appeared without any enhancements (Figure 1). In the enhanced captions condition, the target items appeared bolded and underlined (Figure 2).

Figure 1. An example of a still from the video (*Fresh off the Boat*) with normal captions

Figure 2. An example of a still from the video (*Fresh off the Boat*) with enhanced captions

Test instruments

Of principal interest was knowledge of the lexical composition of the MWEs, that is, their 'form'. The reason for this is that typographic enhancement can render language forms more salient, but it does not as such clarify the meanings of said forms. Recall of the MWEs was measured using a pretest, an immediate posttest and a delayed posttest. A gap-fill and C-test blend format was adopted to design the pretest. The test was constructed by first taking sentences containing the MWEs from dictionaries and COCA. Contextual clues ensured that the test taker would recognize what MWEs was missing (if known by the test taker). The same format was also used for the delayed form-recall test, with the order of the items randomised. Below is an example of the test item for *'on the same page':*

> Parents should be on the s_____ p_____ about raising their children. Parents should have a similar understanding about what to expect from their children.

The conventions of C-test were initially used to guide the cut-off point. However, presenting the learners with the first half of the words could lead to successful guessing. It was then decided that only the first letter would be given, except for those words that started with consonant clusters (in which case the cluster was given, e.g., sl_____ sl_____ to elicit *'slippery slope'*). The instrument was piloted with three American native speakers and three non-native postgraduate students in Applied Linguistics, all of whom completed the instrument 100% correctly.

Recall was also tested using an immediate posttest, but the format was different from that of the pretest and the delayed posttest. In the immediate posttest, learners were given a gap-fill transcript-based test. This format was intended to test episodic memory, that is, the participants' recall of the MWEs in direct association with the context where the MWEs were encountered. Furthermore, congruency of learning opportunity and test condition leads to better test performance (Lotto & de Groot, 1998; Schmitt, 2010). The transcript was condensed to include just the main scenes, to prevent students from being demotivated by a 20-page transcript (see Supplementary Material for samples of the tests).

Apart from the MWE tests, comprehension questions in multiple-choice and True/False formats were created to ascertain that the video had been processed for content, as intended. The

twenty-two questions tapped into skills that constitute the latent ability to perform bottom up processing such as the ability to identify gist and supporting ideas, as well as skills such as making inferences about context and relationships, which demonstrate the ability to perform top-down processing (e.g., Buck & Tatsuoka, 1998; Hildyard & Olson, 1978). None of the target MWEs were used in the comprehension test.

Three additional delayed posttests, i.e., a delayed form-recognition posttest, a delayed meaning-recall posttest and a delayed meaning-recognition posttest were also administered. However, owing to lack of space, the results of these three supplementary posttests will not be discussed here. As they concern aspects of MWE knowledge that were not pre-tested, any conclusions drawn from them would need to remain very tentative at best, in any case.

PROCEDURE

The data was collected in four sessions. In the first session, the participants were briefed on the research (without disclosing its specific purpose) before they signed the consent form. In the second session, the participants took the form-recall pretest. This was immediately followed by the VLT. Two weeks later, they watched the video under their respective conditions. Prior to watching the video, the participants were informed that they would be asked some comprehension questions. The comprehension test was administered immediately after the students had finished watching the video (once or twice). Then followed the immediate MWE recall test. Two weeks later, the delayed recall posttest was administered.

SCORING AND ANALYSIS

The responses in the MWE recall tests were scored by two raters. A lenient scoring system was adopted (e.g., see Webb and Kagimoto, 2009; 2011), in which minor mistakes were marked as correct and received a full score (e.g., a word used in singular which should have been plural, or vice versa; wrongly spelled words; wrong part of speech). While the scoring procedure was relatively straightforward for a target item in which only one content word needed to be completed, it was not the case for target items in which two or three content words of the MWEs needed to be supplied. For these items, partial credit (i.e., 0.5) was awarded for two-gap responses in which one of the words supplied by the learner contained a minor mistake, while the other contained a major mistake. A major mistake is defined as spelling mistakes that constitute a

new word, affect the pronounceability of the word or generate a form that does not resemble the target item.

For an item which required participants to fill in three gaps, a two-over-three rule was imposed. This means that partial credit was given for responses in which two out of three gaps were correct or contained a minor mistake. Put differently, partial scores were only awarded for majorly inaccurate responses with two and three gaps (see Supplementary Materials for more details on the scoring procedure along with the examples). The items with two and three gaps constituted ten out of the 18 items. Because the inter-rater reliability was high (.97, .95 and .97 for the pretest, immediate posttest and delayed posttest respectively), the average score between the two raters was used. Each rater independently awarded 0 or .5 or 1 to a response. As the average score between two raters was taken, this resulted in five possible scores: 0, 0.25, 0.5, 0.75, or 1. Table 4 shows the descriptive statistics for all the MWE tests. As the pretest data was non-normally distributed with unequal variance, the data was first transformed using Tukey Ladder of Powers before running a one-way non-parametric ANOVA (Kruskal-Wallis test). The Kruskal-Wallis test revealed no significant differences between groups in their scores in the pretest (Kruskal-Wallis $\chi 2$ (5) = 4.58, p = .47).

TABLE 4. Descriptive statistics for the form recall tests

	Form recall tests		
Condition	Pretest	Immediate	Delayed
		Mean	
		(SD)	
Enhanced Captions X 1	1.79	6.15	3.38
	(1.30)	(3.51)	(2.41)
Normal Captions X 1	3.82	7.34	5.54
	(3.61)	(4.81)	(4.28)
Uncaptioned X 1	2.78	3.55	3.72
	(3.41)	(3.85)	(3.82)
Enhanced Captions X 2	2.01	9.48	4.17
	(2.34)	(4.47)	(4.66)
Normal Captions X 2	2.68	8.40	4.46
	(2.24)	(4.25)	(3.82)
Uncaptioned X 2	2.92	3.89	3.24
	(2.47)	(3.21)	(2.91)

Note. The maximum score on all tests was 18

Because the pretest and the immediate posttest were not exactly the same format, the analysis focused on the two posttests, with performance on the pretest at the item level included as one of the fixed effects in the statistical modelling. Two *clmm* analyses for the immediate and delayed data respectively were carried out in R (R Core Team, 2018). The *clmm* function in the *ordinal* package (Christensen, 2019) was used. Apart from pretest performance, the fixed effects in all analyses included caption condition, number of viewings and VST score. VST score was included as a fixed effect because the literature has shown that learners with a larger vocabulary tend to understand reading and listening texts better than learners with a smaller vocabulary, and this gives them an advantage when it comes to acquiring new lexical items from those texts (Elgort & Warren, 2014; Noreillie et al., 2018; Schmitt et al., 2011; Stæhr, 2009). What is not certain is whether this comparative advantage might be mitigated by input conditions such as the availability of captions. VST score was centered prior to analysis. There was no issue of multicollinearity between pretest score and VST score. The five possible pretest scores at the item level (i.e., 0, 0.25 , 0.5, 0.75 and 1) were multiplied by 4 (i.e., producing 0, 1, 2, 3, and 4) before being fitted into the models. An increase in one unit of pretest score would then relate to going up one score. The model comparison procedure started with the most complex model which included all the fixed effects, as well as the two-way interactions and three-way interactions between the factors. Each term was then incrementally removed. The significance level of fixed-effect predictors for all the models was assessed using model comparison using likelihood ratio tests (i.e., comparing a full model and reduced model). The comparison returned likelihood ratio statistics with a chi-square distribution. This is reported in the form (LRT $\chi 2$ ($n1$) = $n2$, p < $n3$), where $n1$= degrees of freedom, $n2$ = likelihood ratio statistic and $n3$ = p-value. To correct for multiple testing, Bonferroni adjustment of the alpha level was applied. This means that the adjusted level of *significance* is $p < .025$ rather than $p < .05$. To locate the differences between the treatment groups (e.g., between the three caption conditions), multiple comparison was carried out using the *emmeans* function in the *emmeans* package (Lenth, 2018).

As for the comprehension test about the content of the video, the data were analysed by means of a generalised linear mixed model with the *glmer* function of the *lme4* package (Bates, Maechler, Bolker & Walker, 2015). The fixed effects included caption condition, number of viewings and VST score. Model comparisons were carried out in the same way as the *clmm* analyses.

ANALYSIS AND RESULTS

Immediate MWE posttest

The results of the *clmm* analysis revealed that number of viewings ($\chi 2$ (1) = 21.9, $p < .001$), caption condition, ($\chi 2$ (2) = 52.5, $p < .001$), VST score ($\chi 2$ (1) = 66.7, $p < .001$) and pretest score ($\chi 2$ (1) =92.9, $p < .001$) significantly predicted participants' score on the immediate posttest. No significant interactions between the factors were found. Table 5 below shows the output of the best-fit model, including the odds ratios. The odds are defined as the probability of an event occurring divided by the probability of it not occurring (Field et al., 2012). As shown in the table, the odds of obtaining a higher score in the immediate posttest for participants who watched the video twice were 2.69 times the odds of those who viewed the video once. As for the effect of caption condition, multiple comparisons using the *emmeans* function in the *emmeans* package (Lenth, 2018) with Bonferroni adjustment applied (i.e., $p < .025$) revealed that the enhanced captions conditions were more likely to get a higher score compared to the uncaptioned conditions ($p < .0001$). Additionally, the normal captions conditions were more likely to get a higher score in the immediate posttest compared to the uncaptioned conditions ($p < .0001$). The difference between the enhanced and normal captions, however, failed to reach significance ($p = .04$) As to the effect of VST score, the odds of obtaining a better score became higher as participants' VST score increased by one unit. Further, a one unit-increase in the pretest score (e.g., 0.25 to 0.5) also increased the odds of obtaining a higher score in the immediate posttest.

TABLE 5. Output of best-fit model predicting a higher score in the immediate posttest

Parameter	Estimate	*SE*	*z*	*p*	*OR*
Number of viewings (Twice)	0.99	0.21	4.83	< *.001*	2.69
Caption Condition (Normal Captions)	1.48	0.26	5.73	< *.001*	4.39
Caption Condition (Enhanced Captions)	2.05	0.27	7.72	< *.001*	7.77
Pretest score	0.47	0.05	9.39	< *.001*	1.60
VST score (centered)	0.06	0.01	8.83	< *.001*	1.06

Note: OR = odds ratio. *Intercept levels: Caption condition = Uncaptioned; Number of viewings = Once

Delayed MWE posttest

The analysis revealed that VST score significantly predicted participants' score on the delayed posttest ($\chi 2$ (1) = 43.3, $p < .001$). As shown in Table 6, the odds of obtaining a better score became higher as participants' VST score increased by one unit. The pretest score ($\chi 2$ (1) = 47.2, $p < .001$) and caption condition ($\chi 2$ (2) = 13.0, $p < .01$) were also found to be significant. A significant interaction between pretest score and caption condition was also found ($\chi 2$ (2) = 13.04, $p < .01$). To illustrate this interaction, the predicted probabilities of receiving the five possible delayed posttest scores on an item for all three caption conditions and all pretest scores were generated. The predicted probabilities of receiving the five scores in the delayed posttest for participants with the VST mean score are shown in the generated plot (Plot 1). As there were far fewer occurrences of participants receiving a pretest score of 0.25 and 0.75 than the other three scores, they were omitted from the plot. As can be seen from the plot, for participants with a pretest score of 0, the probability of receiving the same score in the delayed posttest was the lowest for the enhanced captions condition, followed by the normal captions and the uncaptioned conditions. Turning to the likelihood of receiving a full score in the delayed posttest, for participants who received 0 in the pretest, the enhanced captions had the highest likelihood of receiving a full score in the delayed posttest, followed by the normal captions and the uncaptioned condition. However, as the pretest scores became higher (e.g., 0.75 and 1), the predicted probability of getting a full score was higher for the normal captions and uncaptioned conditions, compared to the enhanced

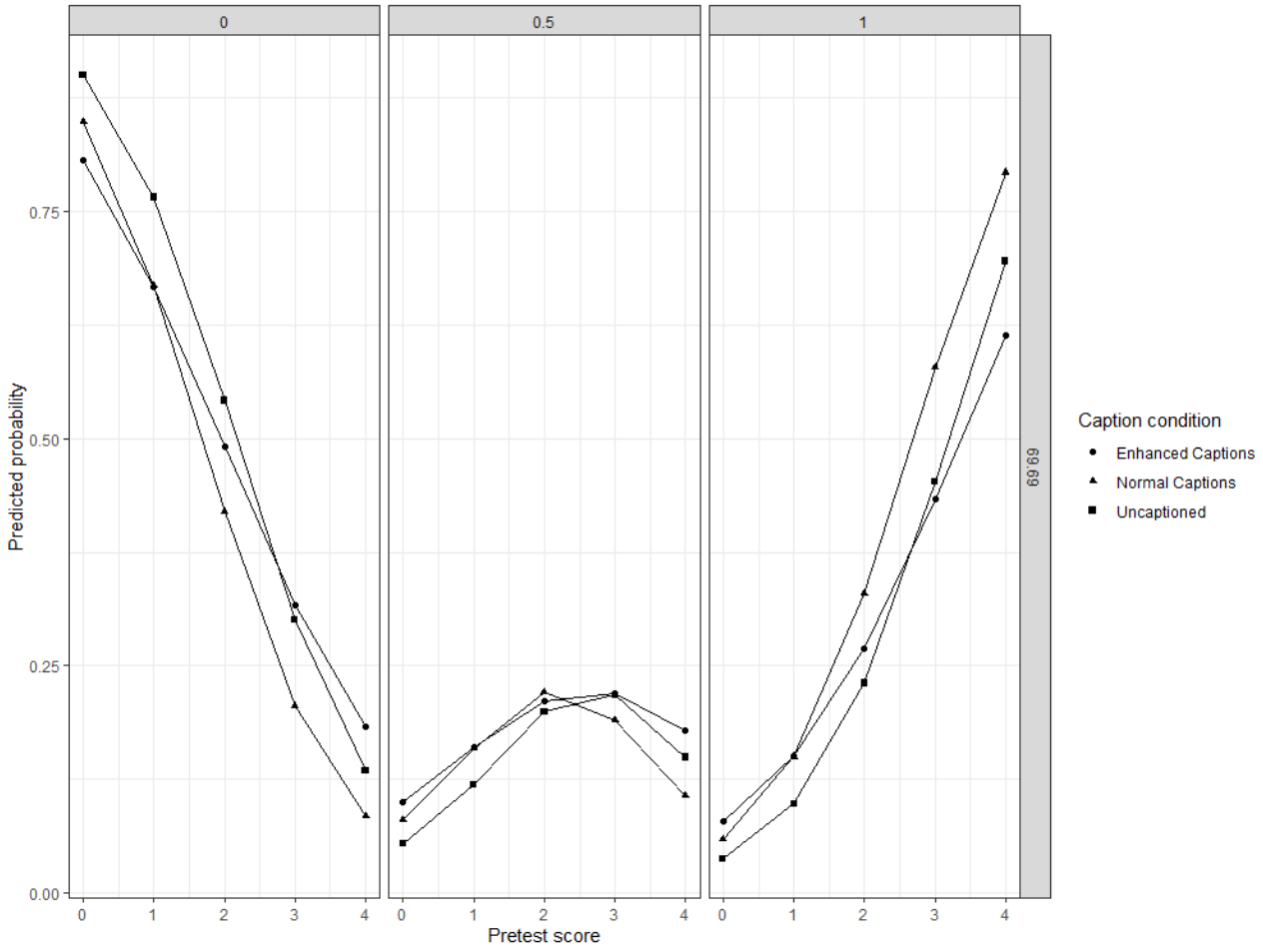
captions condition. This suggests that the effects of typographic enhancement could be stronger for participants with little or no knowledge of the target items. Put differently, participants with relatively good knowledge of the target items may not have needed typographic enhancement to pick up unknown MWES.

TABLE 6. Output of best-fit model predicting a higher score in the delayed posttest

Parameter	Estimate	*SE*	*z*	*p*	*OR*
VST score (centered)	0.05	0.01	6.94	< *.001*	1.05
Pretest score	1.05	0.10	11.57	< *.001*	2.86
Caption condition (Normal captions)	0.52	0.27	1.97	*.04*	1.68
Caption condition (Enhanced captions)	0.84	0.27	3.14	< *.01*	2.32
Pretest score X Normal captions	-0.01	0.12	-0.11	*.92*	1.00
Pretest score X Enhanced captions	-0.38	0.12	-3.10	< *.001*	0.68

Note: OR = odds ratio. *Intercept levels: Caption condition = Uncaptioned; Number of viewings = Once

Plot 1. Predicted probability of receiving a 0, 0.5 and 1 in the delayed posttest



Content comprehension test

Table 7 shows the descriptive statistics for the comprehension test. The results of the *glmer* analysis showed that the number of viewings had a significant influence ($\chi 2$ (1) = 8.87, p = < .01). That is, the odds of obtaining a correct answer on the text comprehension test were higher for those who viewed the video twice compared to those who viewed it once. Further, VST score was also a significant predictor ($\chi 2$ (1) = 25.9, p < .001), suggesting that a higher VST score led to a higher likelihood of getting an item correct. There was also a significant main effect of caption condition ($\chi 2$ (2) = 8.64, p = .01). Participants in the normal captions condition were significantly more likely to get an item correct compared to their peers in the uncaptioned condition, as the odds of a correct response in the former were 1.77 the odds in the latter (Table 8). The difference between the enhanced and normal captions conditions did not reach significance (z = 1.25, p < .20). No interactions were found significant.

TABLE 7. Descriptive statistics for the comprehension test

Condition	Mean (*SD*)
Enhanced Captions X 1	17.44 (2.28)
Normal Captions X 1	17.87 (2.36)
Uncaptioned X 1	16.67 (1.11)
Enhanced Captions X 2	18.55 (2.61)
Normal Captions X 2	18.00 (2.51)
Uncaptioned X 2	17.57 (1.91)

Note. The maximum possible score was 22

TABLE 8. Output of best-fit model predicting a correct response in the comprehension test

Parameter	Estimate	*SE*	*z*	*p*	*OR*
Intercept*	1.64	0.40	4.13		
VST score (centered)	0.03	0.01	5.39	< *.001*	1.03
Caption Condition (Enhanced Captions)	0.34	0.19	1.82	*.07*	1.40
Caption Condition (Normal Captions)	0.57	0.19	3.03	< *.01*	1.77
Number of viewings (Twice)	0.48	0.16	3.10	< *.01*	1.62

Note: OR = odds ratio. *Intercept levels: Caption condition = Uncaptioned; Number of viewings = Once

DISCUSSION

The first research question sought to find out whether there was an effect of caption condition on incidental MWE learning. The immediate and delayed posttest results suggest that caption condition did affect MWE recall. In the immediate posttest, both captioned conditions led to a higher likelihood of receiving a higher score compared to the uncaptioned condition. This corroborates previous viewing studies that found evidence in favour of captioned viewing over uncaptioned viewing for vocabulary learning (see Montero Perez et al., 2013, for a meta-analysis). Unexpectedly, the presence of typographic enhancement did not lead to significantly better recall compared to viewing with normal captions. This finding is in contrast with previous reading studies that have demonstrated that, compared to unenhanced texts, typographically enhanced items attract more attention, and in turn promote recall (e.g., Sonbul & Schmitt, 2013; Szudarski & Carter, 2016; Choi, 2018; Toomer & Elgort, 2019). There are a few plausible explanations for this finding. The first relates to the obvious difference in input modality. Compared to written input, audio-visual input offers more to look at, such as moving images, besides the words alone. Further, compared to written input, the real-time nature of viewing entails that learners have less time to fixate anything, including typographically enhanced items. Thirdly, while previous reading studies on typographic enhancement only included collocations as the target items, the present study included MWEs that consist of up to five words. In other

words, the presence of typographic enhancement may not be enough to create a significant difference in form recall when learners are presented with relatively long MWEs that appear fleetingly. Additionally, the MWEs in this study were only met either once or twice. In comparison, in most reading studies on typographic enhancement each target was met between three (e.g., Sonbul & Schmitt, 2013) and twelve times (e.g., Szudarski & Carter, 2016). In sum, the findings of the present study suggest that where incidental learning is concerned, the use of typographic enhancement leads to better MWE uptake compared to uncaptioned viewing, but it makes little difference when compared to viewing with normal captions. While caption condition also influences participants' scores on the delayed posttest, the effects of caption condition depended on the pretest scores. To illustrate, the presence of typographic enhancement had a positive impact for participants with low pretest scores, but for participants with higher pretest scores, typographic enhancement did not lead to a greater likelihood of receiving a better score compared to the uncaptioned condition. This suggests that the effects of typographic enhancement may not be as strong for delayed MWE recall compared to immediate recall.

The second research question, which concerned the effects of caption condition on video comprehension, can be answered positively as caption condition predicted video comprehension. Specifically, learners who watched the video with normal captions performed significantly better in the comprehension test compared to those who viewed the uncaptioned video. This is consistent with the findings of previous viewing studies such as Gass et al., (2019), Winke et al., (2013) and Montero Perez et al., (2013). It is then rather surprising that captioned viewing with typographically enhanced items did not lead to significantly better scores on the comprehension test compared to the uncaptioned condition. Put differently, while captions without enhancements resulted in the learners taking in the video content better than those who watched the uncaptioned video, the presence of typographic announcement appears to have reduced this advantage of captioned over uncaptioned viewing. This suggests a possible trade-off between a positive effect of typographically enhanced captions on the learning of MWEs and a negative effect on comprehension, which is reminiscent of Choi's (2018) findings.

The third and fourth research question concern the effects of repeated viewing on MWE learning and content comprehension, respectively. Repetition was found to positively affect MWE recall, at least according to the immediate recall tests. Additionally, as no interaction was

found between number of viewings and caption condition, this suggests that repeated viewing is beneficial for all three caption conditions. These findings corroborate previous reading studies (Durrant & Schmitt, 2010) and reading-while-listening studies (Webb et al., 2013) which furnished positive evidence for the role of repetition for incidental learning of MWEs. These studies, however, only employed immediate posttests, and so they reveal little about the durability of the reported gains. In the present study, repeated viewing did not emerge as a strong predictor of delayed form recall. It is worth reiterating that the immediate recall test presented the participants with excerpts from the video transcript while the delayed test used new, decontextualized, sentence prompts. Recognizing a given segment from a video (and recalling a missing MWE associated with it from episodic memory) is probably easier after repeated viewing of the video. Transferring this knowledge to a different-format test after a considerable delay is less straightforward, however, and this may then compromise the benefits of repeated viewing. It might be interesting to explore whether additional opportunities for re-viewing the same video could reveal stronger long-term effects for repetition. As to comprehension, repetition also emerged as a significant predictor. That is, viewing twice led to significantly higher odds of getting an item correct in the comprehension test. This is in line with previous studies that have demonstrated that listening twice significantly enhanced content comprehension compared to listening once (Lund, 1991; Sakai, 2009). Additionally, compared to other techniques that have been proposed to boost listening comprehension, such as previewing of comprehension questions and activation of background knowledge, repeated viewing has been shown to enhance understanding more (Chang & Read, 2006). Further, the present study lends support to Nguyễn (2017), in which repeated viewing of TED talks led to better comprehension as compared to viewing once.

Though not part of the research question, vocabulary knowledge emerged as a significant predictor in both the MWE recall and content comprehension analyses. For instance, the higher the participants' VST score, the higher the likelihood of receiving a better score in the immediate and delayed posttests. This corroborates the results of recent studies which have demonstrated that learners with good vocabulary knowledge stand a better chance of learning new lexical items from audio-visual input compared to their less proficient counterparts (e.g., Montero Perez, 2020; Pujadas & Muñoz, 2019). Similarly, the odds of learners getting an item correct in the comprehension test increased as their VST score became higher. Our finding thus supports the

"Matthew effect", a phenomenon where weaker students learn less while stronger students learn more (e.g., Stanovich, 1986). The Matthew effect has been observed in the context of vocabulary learning through reading (e.g., Elgort et al., 2015; Horst et al., 1998) as well as L2 viewing (e.g., Feng & Webb, 2020; Montero Perez et al., 2014; Peters et al., 2016; Peters & Webb, 2018; Puimège & Peters, 2019). Additionally, prior knowledge of target items also led to a higher likelihood of receiving a higher score in the immediate posttest. This is in line with previous viewing studies which also found a positive relationship between prior knowledge and incidental vocabulary (single and/or MWEs) learning through L2 viewing (Montero Perez et al., 2014; Peters et al., 2016; Peters & Webb, 2018; Puimège & Peters, 2019).

LIMITATIONS

Several limitations have to be acknowledged. The first concerns the number of participants within each condition. Originally, 133 participants were recruited, but some had to be excluded from the analyses due to the unavailability of their VST score. As a result, one of the conditions had 15 participants, which is on the low side. It should be noted, however, that the number of participants in each condition in the present investigation is similar to other studies that investigated vocabulary learning through L2 viewing (e.g., Feng & Webb, 2020). The second limitation concerns the number of target MWEs (n =18). The authentic nature of the video did not allow us to test as many MWEs as we would have liked. Repetition was then operationalised as repeated viewing. Using more extensive video input (as used in Rodgers, 2013) can increase the number of target items and the likelihood of their reoccurrence in a single viewing. However, that would mean that participants in the repetition conditions would need to sit through a very long treatment (possibly several hours). To reduce the risk of fatigue, we had to settle for a shorter video, which naturally entails fewer potential targets. Still, the number of items used in the present investigation is similar to that used in earlier studies (e.g., Webb et al., 2013). We could have increased the number of encounters with the target MWEs by playing the same video multiple times, but this would probably have compromised the ecological validity of the learning experience. For one thing, teachers are unlikely to spend such a substantial amount of time on L2 viewing. For another, learners are likely to lose interest if they are required to watch a video three or more times.

Another aspect of the research design that could be considered a limitation pertains to the unidentical format of the MWE recall tests. As mentioned earlier, the format of the immediate posttest probably tested the learners' episodic recall of the MWEs, whereas in the pretest and delayed posttest, the learners had to insert the MWEs in a context not entirely the same as the video. Future studies could make use of the methodology in Peters and Webb (2018), in which different tests were carried out with different groups of learners. This may give us a finer-grained picture of the extent to which MWE learning is promoted. Finally, including a control group or control items would further clarify the effects of the treatment, as including these would help to further determine how much learning resulted from the treatment as such.

CONCLUSION AND FUTURE DIRECTIONS

Our study aimed to investigate whether caption condition and repetition had an effect on incidental MWE uptake through L2 viewing. Of further interest was whether comprehension would also be enhanced with the use of captions and repetition. Given previous research findings suggesting that typographic enhancement of language items or features may distract from text content, we considered it possible that learners watching a video with enhanced captions might not take in the content of the video as well as learners watching the video with normal captions. MWE recall was tested at three different time points: two weeks before, immediately after, and two weeks after watching the video. A comprehension test was administered immediately after viewing. Our findings provide affirmative evidence that L2 viewing promotes the incidental uptake of MWEs (at least at the level of form recall). Specifically, both normal and enhanced captions led to better short-term form recall knowledge compared to uncaptioned viewing. However, enhanced captions did not hold an advantage over normal captions. The presence of captions also benefitted comprehension. Interestingly, applying typographic enhancement to captions did not lead to better comprehension. In fact, our findings suggest that the addition of typographic enhancement may cause captioned viewing to lose an advantage over uncaptioned viewing. Our findings also revealed that repetition had a positive influence, with viewing twice leading to better scores in the immediate form recall test, as well as better content comprehension, compared to viewing once. Finally, another factor that influenced both MWE recall and content comprehension was learners' vocabulary knowledge. The Matthew effect was

observed, and learners with a bigger vocabulary size obtained bigger gains in MWE knowledge and better odds of getting comprehension questions correct.

This study is a step towards gaining better understanding of the potential use of audio-visual resources as a way of enhancing MWE knowledge. More research is needed to establish the effects of caption conditions and number of viewings in other ESL and EFL contexts. Finally, the effects of these variables might be different when learners are aware of the impending tests. As posited by Webb and Nation (2017), "in principle, fewer repetitions are needed in deliberate learning than in incidental learning" (p.65). Therefore, when learners are aware of forthcoming tests, it is plausible that a second viewing may not be needed. Similarly, whether the use of typographically enhanced captions makes a difference when L2 learners expect a test warrants further investigation. In sum, future studies could explore the role of repetition and captioning in the learning of MWEs under intentional learning conditions.

---

[i] The term *MWE* was also chosen because Wray's (2002) definition of *formulaic language/sequence* suggests holistic processing and was proposed with L1 speakers in mind. The issue of holistic storage is a complex one, with no clear evidence pointing to formulaic language being stored holistically in either the L1 or L2 mental lexicon (Siyanova-Chanturia, 2015).

[ii] One of the planned posttests (not reported in this study) was intended to gauge the participants' comprehension of the MWEs. Therefore, totally transparent phrases (i.e., semantically compositional expressions, whose meaning can be computed based on the literal meaning of their constituent words) were excluded.

REFERENCES:

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. Retrieved from http://CRAN.R-project.org/package5lme4

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4), 509-533. doi:10.1017/S0142716402004022

Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 227-248). John Benjamins.

Boers, F. (2020). Factors affecting the learning of multiword items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 143-157). Routledge.

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb–noun collocations. *Language Teaching Research, 18*(1), 54-74. doi:10.1177/1362168813505389

Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics, 27*(2), 448-469. doi:10.1111/ijal.12141

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10*(3), 245-261. doi:10.1191/1362168806lr195oa

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing, 15*(2), 119-157. doi:10.1177/026553229801500201

Chang, A. C. S., & Read, J. (2006). The effects of listening support on the listening performance of EFL learners. *TESOL Quarterly, 40*(2), 375-397. doi:10.2307/40264527

Choi, S. (2018). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research, 21*(3), 403-426. doi:10.1177/1362168816653271

Christensen, R. H. B. (2019). ordinal - Regression Models for Ordinal Data. R package version 2019.12-10. Retrieved from http://www.cran.r-project.org/package=ordinal/

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. doi:10.1080/09571736.2019.1615978

Columbus, G. (2010). Processing MWU: Are MWU subtypes psycholinguistically real? In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 194-210). Continuum.

Davies, M. (2008). *Corpus of Contemporary American English (COCA)*. Retrieved from: https://corpus.byu.edu/coca/

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

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*(2), 163-188. doi:10.1177/0267658309349431

Elgort, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language, Cognition and Neuroscience, 30*(5), 506-528. doi:10.1080/23273798.2014.942673

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning, 64*(2), 365-414. doi:10.1111/lang.12052

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29-62. doi:10.1515/text.1.2000.20.1.29

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, 1-25.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London, UK: Sage.

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.

Hildyard, A., & Olson, D. R. (1978). Memory and inference in the comprehension of oral and written discourse. *Discourse Processes, 1*(2), 91-117.

Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 47-69). Language Teaching Publications.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207-223.

Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge University Press.

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.

Lin, P. M. S. (2014). Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System, 42*(1), 164-176. doi:10.1016/j.system.2013.11.010

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism, 10*(1), 105-129. doi:10.1080/14790718.2012.679275

Lotto, L., & de Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning, 48*(1), 31-69. doi:10.1111/1467-9922.00032

Lund, R. J. (1991). A comparison of second language listening and reading comprehension. *Modern Language Journal, 75*(2), 196-204. doi:10.1111/j.1540-4781.1991.tb05350.x

Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals, 32*(3), 321-328.

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, 1-25. doi: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 and Technology, 18*(1), 118-141.

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *Modern Language Journal, 99*(2), 308-328. doi:10.1111/modl.12215

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning, 31*(1-2), 1-26. doi:10.1080/09588221.2017.1375960

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. doi:10.1016/j.system.2013.07.013

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Nation, I. S. P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts. Retrieved from http://www.vuw.ac.nz/lals/staff/paulnation/nation.aspx

Neuman, S., & 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. doi:10.2307/747835

Nguyễn, C. Đ. (2017). *Fostering incidental vocabulary uptake from audio-visual materials: The role of text comprehension.* (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington, New Zealand.

Noreillie, A., Kestemont, B., Heylen, K., Desmet, P., & Peters, E. (2018). Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages: An approximate replication study of Stæhr (2009). *ITL - International Journal of Applied Linguistics, 169*(1), 214-233.

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research, 21*(3), 381-402. doi:10.1177/1362168815618428

Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *International Journal of Applied Linguistics*, *169*(1), 142-168.

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

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. doi:10.1017/S0272263117000407

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audio-visual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal, 47*(4), 424-438. doi:10.1080/09571736.2019.1638630

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, *47*(4), 479-496. doi:10.1080/09571736.2019.1616806

R Core Team. (2018). R: A Language and environment for statistical computing (Version 3.4.4) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington, New Zealand.

Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, *45*(4), 689-717. doi:10.5054/tq.2011.268062

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, *21*(4), 589-619.

Roy Morgan Research. (2015). *Delhi to Dunedin, Beijing to Ballarat: The time spent with TV, internet, newspapers and radio across Asia*. Retrieved from: http://www.roymorgan.com/findings/6277-time-spent-with-television-radio-internet-newspapers-across-asia-december-2014-201506090624

Sakai, H. (2009). Effect of repetition of exposure and proficiency level in L2 listening tests. *TESOL Quarterly*, *43*(2), 360-372. doi:10.1002/j.1545-7249.2009.tb00179.x

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

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York, NY: Palgrave Macmillan.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in text and reading comprehension. *Modern Language Journal*, *95*(1), 26-43. doi:10.1111/j.1540-4781.2011.01146.x

Schmitt, N., Schmitt, D., & 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*(2), 165-179. doi:10.1017/S0272263100011943

Siyanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language. *Corpus Linguistics and Linguistic Theory*, *11*(2), 285-301. doi:10.1515/cllt-2014-0016

Siyanova-Chanturia, A., & Omidian, T. (2020). Key issues in researching multiword items. In S. Webb (Ed.), (pp. 529-544). Routledge.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2018). Formulaic language: Setting the scene. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 1-15). Routledge.

Siyanova-Chanturia, A., & Van Lancker Sidtis, D. (2018). What online processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 31-61). Routledge.

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*(1), 121-159. doi:10.1111/j.1467-9922.2012.00730.x

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, *31*(4), 577-607. doi:10.1017/S0272263109990039

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360-407. doi:10.1598/RRQ.21.4.1

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50-73.

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, *26*(2), 245-265. doi:10.1111/ijal.12092

Tavakoli, P., & Uchihara, T. (2019). To what extent are multiword sequences associated with oral fluency? *Language Learning*, 1-42. doi:10.1111/lang.12384

Teng, F. (2019). The effects of video caption types and advance organizers on incidental L2 collocation learning. *Computers & Education, 142*. doi:10.1016/j.compedu.2019.103655

Toomer, M., & 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. doi:10.1111/lang.12335

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning, 69*(3), 559-599. doi:10.1111/lang.12343

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(2), 45-65. doi:10.1093/applin/aml048

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 175-184). Routledge.

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly, 43*(1), 55-77. doi:10.5054/tj.2010.215611

Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics, 32*(3), 259-276. doi:10.1093/applin/amq051

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning, 63*(1), 91-120. doi:10.1111/j.1467-9922.2012.00729.x

Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407-427. doi:10.1093/applin/amp010

Webb, S., & Rodgers, M. P. H. (2009b). Vocabulary demands of television programs. *Language Learning, 59*(2), 335-366. doi:10.1111/j.1467-9922.2009.00509.x

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology, 14*(1), 65-86.

Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *Modern Language Journal, 97*(1), 254-275. doi:10.1111/j.1540-4781.2013.01432.x

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.