LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Incidental Vocabulary Learning From Bilingual Subtitled Viewing: An Eye-Tracking Study

Andi Wang and Ana Pellicer-Sánchez
University College London

Abstract: This study examined the effectiveness of bilingual subtitles relative to captions, subtitles, and no subtitles for incidental vocabulary learning. Learners' processing of novel words in the subtitles and its relationship to learning gains were also explored. While their eye movements were recorded, 112 intermediate to advanced Chinese learners of English watched a documentary in one of 4 conditions: bilingual subtitles, captions, L1 subtitles, and no subtitles. Vocabulary pretests and posttests assessed the participants' knowledge of the target vocabulary for form recognition, meaning recall, and meaning recognition. Results suggested an advantage for bilingual subtitles over captions for meaning recognition and over L1 subtitles for meaning recall. Bilingual subtitles were less effective than captions for form recognition. Participants in the bilingual subtitles group spent more time reading the Chinese translations of the target items than the English target words. The amount of attention to the English target words (but not to the translations) predicted learning gains.

Keywords incidental learning; vocabulary; eye tracking; viewing; bilingual subtitles

---

This study was funded by a doctoral scholarship awarded to the first author by the China Scholarship Council, No. 201808060298. We would like to thank the three anonymous reviewers as well as the Associate Editor Professor Kara Morgan-Short for providing insightful comments and detailed feedback throughout the review process.

Correspondence concerning this article should be addressed to Andi Wang, University College London Institute of Education, 20 Bedford Way, London WC1H 0AL, UK. E-mail: andi.wang.16@ucl.ac.uk

The handling editor for this article was Kara Morgan-Short.

DOI: 10.1111/lang.12495

Introduction

Vocabulary knowledge is a key component of language proficiency, and second language (L2) learners need to build large vocabulary to be able to communicate successfully. In order for learners to reach the lexical goals necessary for successful comprehension of spoken and written input, they need to supplement deliberate learning with incidental learning (Webb, 2020). Incidental vocabulary acquisition can accrue as a result of reading, listening, reading-while-listening, and viewing, with research suggesting that watching television programs and movies is the most common source of L2 input outside of the classroom (Peters & Webb, 2018; with online activities also having a positive effect on vocabulary knowledge, Peters, Noreillie, Heylen, Bulté, & Desmet, 2019). Previous studies have shown that viewing audio-visual material (henceforth, viewing) facilitates L2 vocabulary learning (e.g., Peters & Webb, 2018; Rodgers & Webb, 2019) and that the use of captions and subtitles supports this process (e.g., Koolstra & Beentjes, 1999; Montero Perez, Van Den Noortgate, & Desmet, 2013; Peters, 2019). The majority of studies in this area have examined the effectiveness of captions (i.e., L2 subtitles) and/or first language (L1) subtitles for learning because they have been claimed to be the ones most frequently encountered by L2 learners (Muñoz, 2017). However, in certain multilingual contexts where two or more languages are spoken, bilingual subtitles (i.e., simultaneous presentation of L1 and L2 subtitles) are used to serve a wider population (Bartolomé & Cabrera, 2005). Despite the fact that mainland China is a monolingual region with Mandarin as the official language, the use of bilingual subtitles has become increasingly popular in the past two decades (Li, 2016; Liao, Kruger, & Doherty, 2020). In spite of their popularity, very little research has been conducted to examine their benefits for language learning.

Bilingual subtitles are believed to be conducive to vocabulary learning because the L1 lines provide the translation of unknown L2 words and the L2 lines provide the form of unknown words, enabling learners to connect the L2 unknown form with its correct meaning (Li, 2016). However, the use of L1 translations in viewing may also prevent learners from processing the L2 unknown words (Peters, 2019) and may also increase the cognitive load for some learners (Lwo & Lin, 2012). Empirical evidence examining the effectiveness of bilingual subtitles for vocabulary learning has remained scarce, and available research has yielded conflicting results (e.g., Li, 2016; Lwo & Lin, 2012).

Eye-tracking studies have suggested that vocabulary gains from viewing are related to the amount of time spent processing novel words during viewing (Montero Perez, Peters, & Desmet, 2015). This is particularly relevant for bilingual subtitles where learners can choose how they want to allocate their

attention (to the L2 novel word or/and translation). However, little is known about how learners process novel words in bilingual subtitles and about how that may relate to vocabulary gains. The current research aimed at addressing these gaps. The first aim of the study was to explore the relative effect of bilingual subtitles for incidental vocabulary learning (in comparison to captions, L1 subtitles, and no subtitles). Second, the study aimed to examine learners' attention to novel words in subtitles through the recording of learners' eye movements in order to investigate the relationship between the amount of attention and word learning in the different subtitling conditions.

Background Literature

Incidental Vocabulary Learning From Viewing

L2 learners are believed to gradually build up their knowledge of new words incidentally through repeated exposures to novel vocabulary in the input (Nation, 2001). Incidental learning happens while learners are engaged in various communicative activities (e.g., reading, listening, and viewing), and learners pick up vocabulary when their attention is focused on lexical meaning rather than lexical form (Hulstijn, 2003). An experimental vocabulary learning condition is considered incidental when participants receive no vocabulary posttest announcement before a treatment and when participants are only informed that they will complete a comprehension test (Hulstijn, 2003).

So far, numerous empirical studies have documented that L2 vocabulary can be acquired incidentally from reading (e.g., Pellicer-Sánchez, 2016), from listening (e.g., Pavia, Webb, & Faez, 2019), and from reading-while-listening (e.g., Teng, 2018). In the last decade, an increasing number of studies have shown that incidental vocabulary learning can also occur through viewing (e.g., Peters, 2019; Puimège & Peters, 2019; Winke, Gass, & Sydorenko, 2010). Considerable vocabulary gains have seemed to occur from viewing even a single television program (Peters & Webb, 2018).

The benefits of viewing for vocabulary learning have seemed to be further enhanced with the use of subtitles (Peters, Heynen, & Puimège, 2016). One of the most common subtitling types is captions, that is, transcriptions written in the same language as that of the spoken text in the video. Captions are helpful in drawing learners' attention to word form (Winke et al., 2010), encouraging sound-script automatization, establishing an initial form-meaning link for unfamiliar words (Montero Perez et al., 2015), reducing the effects of accent variations, and segmenting the speech stream, which can help to further improve learners' comprehension (Bird & Williams, 2002). In a meta-analysis, Montero Perez et al. (2013) analyzed 10 empirical studies exploring the use

of captions for L2 vocabulary acquisition and found a medium to large effect size, despite the different vocabulary tests used across studies and differences in participants' proficiency.

Another widely used subtitling type is L1 subtitles. Providing L1 translations facilitates learners' understanding of video content and seems to be more suitable for lower-level learners (Danan, 2004). Previous studies have shown that L1 subtitles have been effective in expanding children's L2 vocabulary knowledge (e.g., Koolstra & Beentjes, 1999). L1 subtitles also seemed to aid learners' comprehension (Pujadas & Muñoz, 2020), to help learners distinguish separate words in spoken text (Koolstra & Beentjes, 1999), to foster language learning, and to boost learners' motivation (Peters et al., 2016).

To date, empirical studies comparing captions and L1 subtitles have shown that L1 subtitles lead to better comprehension (e.g., Pujadas & Muñoz, 2020). However, their comparison in relation to their effects on incidental vocabulary learning has yielded inconclusive findings. Previous studies have reported an advantage for captions over L1 subtitles for the acquisition of the form of novel words (e.g., Peters, 2019; Peters et al., 2016). Regarding the acquisition of the meaning of novel words, some studies have reported an advantage for captions over L1 subtitles (e.g., Frumuselu, De Maeyer, Donche, & Colon Plana, 2015; Peters, 2019), whereas other studies revealed no significant difference between captions and L1 subtitles (e.g., Peters et al., 2016; Pujadas & Muñoz, 2019). These inconclusive findings were likely due to differences in participant profiles (L1 background, L2 proficiency level), number of viewing sessions (one-off viewing, longitudinal viewing), and test design.

Bilingual Subtitles

Bilingual subtitles (also called dual subtitles) involve the simultaneous presentation of L1 and L2 subtitles, with L1 presented on the first line and L2 underneath. They have been claimed to be less common than captions and L1 subtitles and used only in certain multilingual areas (Bartolomé & Cabrera, 2005). However, their use has recently gained popularity in other contexts such as in China, where bilingual subtitles have become a common subtitling type in cinemas and online L2 audio-visual material (Liao et al., 2020). Bilingual subtitles have potential benefits for vocabulary learning because they help learners make connections between the written and spoken forms of words as well as between the form of unknown words and their correct meanings (Li, 2016; Lwo & Lin, 2012). The use of L1 lexical transfer has been considered beneficial for establishing the form-meaning link of new vocabulary (Schmitt, 2010; see also Kemp & McDonald, 2021).

According to the bilingual version of the dual coding theory (Paivio, 1990; Paivio & Desrochers, 1980), there are two separate but interconnected verbal systems corresponding to a bilingual's two languages that both connect to a common nonverbal system. These three systems function independently but also interconnectedly, and the activity in either of the two language systems can be influenced by the other two systems. The associations between the three systems have an additive memory effect that could enhance individuals' memory recall, augmenting the benefits of the combination of verbal and nonverbal systems (Paivio, 1990). However, according to the depth of processing theory (Craik & Lockhart, 1972), it could also be argued that having the translations of L2 unknown words might reduce learners' cognitive analysis of their meanings and might lead to shallower memory traces that are then reflected in smaller gains. Importantly, according to the redundancy principle (Sweller, 2005), presenting the same information in multiple forms might result in cognitive overload that could be detrimental for learning. Therefore, bilingual subtitled viewing, by providing images along with the L1 and the L2 in both aural and written forms, could enhance the interconnections between the nonverbal and two verbal systems, enabling more access routes to be established for information retrieval. Information would thus more likely be activated through the three systems that facilitate learning. However, the presentation of L1 in bilingual subtitles may compete with learners' attention to the L2 input. Moreover, the L1 and L2 lines convey the same meaning in different forms, which may overload learners' cognitive capacity for information processing and impede their learning. In spite of the increasing popularity of bilingual subtitles and the controversy around their potential benefits, very few studies have investigated their effectiveness for vocabulary learning.

In the context of engineering education, García (2017) examined users' opinions about the effectiveness of bilingual subtitles for incidental L2 vocabulary learning and comprehension. Participants' responses ($N = 62$) to an online questionnaire showed that learners who had used bilingual subtitles believed that they were helpful for their development of L2 vocabulary for form, meaning, and use. However, the results were only based on participants' self-assessment that may not have reflected the actual effectiveness of bilingual subtitles.

Li (2016) reported advantages for bilingual subtitles for incidental vocabulary learning. English major students from a Chinese university ($N = 120$) were asked to watch three 20-minute British Broadcasting Corporation (BBC) documentary clips (three sessions) with captions, L1 subtitles, and bilingual subtitles in a counterbalanced design. A control group with no subtitles was

also included. Vocabulary meaning recognition and meaning recall tests were administered immediately after each viewing session and three weeks later. Participants' opinions toward the use of different subtitles on their vocabulary learning and comprehension were also collected in a final questionnaire. The results showed that the bilingual subtitles group outperformed the other three groups in meaning recognition and meaning recall on both immediate and delayed posttests. The questionnaire results showed that the majority of participants chose bilingual subtitles as the best subtitling type for comprehension and vocabulary learning.

Lwo and Lin (2012) reported findings different from those of Li (2016). In Lwo and Lin's study, 32 young Chinese L1 learners of English watched two animations in one of four conditions (captions, L1 subtitles, bilingual subtitles, and no subtitles). Participants' comprehension of the video was checked during their viewing by their answering questions orally after each scene. After each viewing session, participants completed a comprehension test and two vocabulary tests: a multiple-choice meaning recognition test and a fill-in-the-blank test. Semistructured interviews were held after the viewing sessions to explore participants' attention allocation during the viewing and their attitudes toward different subtitling types for comprehension. Results showed no significant advantages of bilingual subtitles over other subtitling types for vocabulary learning and comprehension. Despite the lack of group differences, lower-level learners seemed to benefit more from the use of bilingual subtitles relative to other types of subtitling, especially in recalling more complex sentences. Together with the interview data, the authors thus postulated that lower-level learners tended to use more selectively the information during viewing to meet their needs for comprehension or language learning (i.e., not process all cues in the input, but only those that were helpful), but more proficient learners were more easily distracted by the L1 lines in bilingual subtitles. However, as Lwo and Lin pointed out, the findings should be interpreted with caution because the viewing material used was specifically designed for English language teaching and contained simple sentences. The procedure crucial to this study required participants to stop and answer comprehension questions after each scene and sentence, which interrupted the viewing process.

In sum, although there has been some initial evidence suggesting that learners prefer bilingual subtitles and have positive opinions about their effectiveness (García, 2017; Li, 2016), very few studies have directly examined their effectiveness for vocabulary learning, and available studies have yielded conflicting findings (e.g., Li, 2016; Lwo & Lin, 2012). Moreover, these studies did not explore the relative effectiveness of bilingual subtitles in facilitating

form knowledge. Thus, understanding of the effectiveness of bilingual subtitles relative to the effectiveness of other subtitles has remained limited. The study by Lwo and Lin (2012) suggested a relationship between the way learners make use of the information in bilingual subtitles and learners' gains, but these initial findings were based on self-reports and focused on the general processing of the subtitles without particular reference to unknown vocabulary. Learners' allocation of attention to the different lines in bilingual subtitles when they encounter unknown words and the relation of bilingual subtitles to vocabulary learning has yet to be examined.

Eye-Tracking Studies on Learning From Viewing

Several studies have used eye-tracking to explore learners' processing of subtitled videos and variables affecting this process (e.g., Muñoz, 2017; Winke, Gass, & Sydorenko, 2013). Previous research has shown that in general adult learners process both animation and subtitles regardless of the language of the subtitles, and similar processing patterns have been reported for L1 subtitles and captions (e.g., Bisson, van Heuven, Conklin, & Tunney, 2014).

Liao et al. (2020) explored the online processing of bilingual subtitles. They asked 20 Chinese postgraduates to watch four 5-minute English documentary clips using captions, L1 subtitles, bilingual subtitles, and no subtitles while their eye movements were recorded. Participants' cognitive load was measured in a postviewing questionnaire and their comprehension was assessed using a free recall test after each viewing session. The eye-movement results showed that the time spent on bilingual subtitles (33.62%) was longer than that on L1 subtitles (21.55%), but similar to captions (32.15%). When using bilingual subtitles, participants spent less time on the L2 lines (15.29%) than when using captions (32.15%). However, similar total reading time was reported for the L1 lines of bilingual subtitles (18.33%) and L1 subtitles (21.55%). The questionnaire data also showed that, when using bilingual subtitles, participants tended to choose one language as a dominant source and used the other for supporting information. The cognitive load questionnaire indicated that the use of bilingual subtitles reduced participants' cognitive load compared to the use of monolingual subtitles. No significant differences in comprehension were revealed across groups.

Liao et al. (2020) study provided useful information about the processing of the entire subtitling area in bilingual subtitles, but it did not examine the processing of unknown vocabulary. To our knowledge, only one previous study has examined the processing of unknown vocabulary in subtitled viewing. Montero Perez et al. (2015) examined the relationship between processing

unknown words in captions and learning unknown words under different conditions (intentional vs. incidental) and with different types of captioning (keyword captions and full captions). Results showed that only fixation times on the full captions predicted word learning. A significant positive effect for total fixation duration on form recognition gains was revealed only for the intentional group. First-pass reading time positively correlated with learning gains for the incidental group. Second-pass reading time had a positive and significant correlation with learning gains for the intentional group but a negative and significant correlation for the incidental group. Montero Perez et al. interpreted this finding as an indication that second-pass reading time in the incidental group reflected participants' failure to successfully integrate knowledge of the word.

It is notable that no previous studies have examined processing unknown words in a bilingual subtitles condition or the relationship of bilingual subtitles to learning. It has been suggested that participants using bilingual subtitles might turn to their L1 as a shortcut to facilitate comprehension (Lwo & Lin, 2012), which may result in their paying less attention to the form of unknown words. However, empirical evidence to support this claim has yet to be provided.

The Present Study

The review of the literature has shown that, despite the popularity of bilingual subtitles and learners' preference for this subtitling type in certain contexts, research on the effectiveness of bilingual subtitles for vocabulary learning has remained scarce. Given the very few studies available and the methodological limitations outlined in the Background Literature section, the potential benefits of bilingual subtitles for vocabulary learning have remained unclear. Although there is some crucial, initial evidence suggesting that processing target words in subtitles might be related to learning gains (e.g., Montero Perez et al., 2015), this connection has not been studied for bilingual subtitles. Exploring the processing of unknown words in bilingual subtitles and its relation to learning gains is particularly important because L2 learners have the option to choose how they want to allocate their attention to the different sources of input presented (i.e., L2 word and/or L1 translation). The present study aimed to fill these gaps. The following research questions were addressed:

1. To what extent does the use of bilingual subtitles increase learners' vocabulary knowledge compared to the use of captions, L1 subtitles, and no subtitles?

2. How do learners allocate their attention to unknown target words and to L1 translations of unknown target words when presented with bilingual subtitles?
3. How does the processing of unknown target words and of the L1 translations of unknown target words in a bilingual subtitles condition compare to the processing of unknown target words in captions and of the L1 translations of unknown target words in L1 subtitles?
4. To what extent does the online processing of the unknown target words and of the corresponding L1 translations of unknown target words predict vocabulary gains in different subtitling conditions (i.e., captions, bilingual subtitles, and L1 subtitles)?

In order to address these questions, we asked L2 learners to view while their eye movements were recorded a 23-minute clip in one of four experimental conditions: captions, L1 subtitles, bilingual subtitles, no subtitles. We assessed the participants' knowledge of the target vocabulary included in the video before and after the viewing session. Concerning vocabulary gains (Research Question 1), based on previous findings, we hypothesized that bilingual subtitles might lead to higher vocabulary gains in meaning than would the other subtitling conditions (Li, 2016). Regarding the processing of unknown words in the bilingual subtitles condition (Research Question 2) and relative to the other subtitling conditions (Research Question 3), in the absence of relevant research, we formulated hypotheses based on previous findings on the processing of the whole subtitling area (Liao et al., 2020; Lwo & Lin, 2012). We hypothesized that the participants using bilingual subtitles would spend more time processing L1 translations than L2 unknown target words (Research Question 2) and that the amount of attention that the participants paid to target words in the bilingual subtitles condition would be shorter than in the captions condition, but that the amount of attention that the participants paid to L1 translations would be similar to the amount paid in the L1 subtitles condition (Research Question 3). Finally, based on the findings in Montero Perez et al. (2015) study, we hypothesized that longer time spent on the L2 target words would lead to higher vocabulary gains regardless of subtitling types and vocabulary test types (Research Question 4).

Method

Participants

The participants in this study were 112 Chinese learners of English (98 females and 14 males) ranging in age from 18 to 34 years ($M = 23.42$ years,

SD = 2.47, 95% CI [22.93, 23.87]). They were all students at a university in the United Kingdom with various academic backgrounds. Participation was voluntary, and the learners received a small compensation for their participation. The participants' proficiency level was determined from their self-reported International English Language Testing System (IELTS) scores and their scores for the Vocabulary Size Test (Nation & Beglar, 2007; VS: Test your word knowledge, n.d.). The participants' mean IELTS score was 6.84 (*SD* = 0.61, 95% CI [6.67, 6.90]), which approximately corresponded to B2 to C1 levels in the Common European Framework of Reference for Languages according to the IELTS official guidelines (IELTS, n.d.), and their mean receptive vocabulary size was 6,274.31 word families (*SD* = 1,704.65, 95% CI [5,950.67, 6,597.95]). Descriptive statistics for proficiency scores are provided in Appendix S1 in the online Supporting Information. We accounted for differences in proficiency in the analyses. We conducted Pearson correlations for the participants' vocabulary size scores and their overall IELTS scores. The results revealed a significant and large correlation, $r = .61$, 95% CI [.48, .72], $p = .002$, providing convergent evidence for the test validity.

We also administered the 3,000-word-level Vocabulary Levels Test (Cobb, n.d.; Schmitt, Schmitt, & Clapham, 2001) to ensure the comprehensibility of the selected viewing material for participants; 86% of the participants showed mastery of the 3,000 word level, indicating that they would be familiar with the vocabulary in the viewing materials. Sixteen participants failed to meet the mastery threshold. Thirteen participants reported no difficulty in understanding the content of the video. We conducted statistical analyses including and excluding these 13 participants and, because there were no differences in the results, these 13 participants were retained in the final analyses. We discarded data from the remaining three participants who had not shown mastery of 3,000-word-level vocabulary and who reported difficulty in understanding the content of the video. We also discarded data from three participants who did not complete the posttests. Finally, we removed data from another six participants from the analyses of the eye-movement data due to poor data quality. Poor data contained sizable track loss or problematic drift as demonstrated in the temporal graph and spatial overlay view by plotting the raw data in the EyeLink Data Viewer (2018) software. In total, we included 106 participants in the analyses of offline data (i.e., vocabulary tests) and 100 participants in the analyses of online (i.e., eye-tracking) data. Results of the multiple-choice comprehension questions (Cronbach's alpha at .83) completed after the viewing session further confirmed the participants' adequate understanding of the

video, with the average scores above 60% across all conditions ($M = 73.28\%$, $SD = 15.71\%$, 95% CI [68.98%, 75.38%]).

Materials

Viewing Material

We extracted four authentic video excerpts (in total 23 minutes, 3,488 words) from the BBC documentary *Animal Odd Couples* (BBC, 2013)[1] and put them together using the video editing software Corel VideoStudio Pro 2018 (2018). This documentary consists of several journeys taken by a wildlife biologist to find out why animals of different species build up unusual close bonds with each other. The chosen clips included four pairs of animal couples. We analyzed the lexical frequency profile of the scripts using the Range software (Nation & Heatley, 2002), with the British National Corpus (BNC Consortium, 2007) as the reference corpus. Results showed that the first 3,000 most frequent words provided a coverage of 96% of the chosen clips. We retrieved the original English video script online, and the first author translated the script into Chinese. We then compared the translation to an online amateur translation (Bilibili, n.d.), asked three native Chinese speakers fluent in English to check it, and piloted it twice, together with 13 advanced Chinese learners of English, to ensure its accuracy. The captions and L1 subtitles were added to the videos using SrtEdit (PortableSoft, 2012) and Corel VideoStudio Pro 2018 (2018) software. The design of captions and subtitles followed the BBC *Subtitle Guidelines* (BBC, 2019). All the L1 subtitles and captions were each kept within one line (such that each language had its own line), with the maximum line length being 68% of the width of the screen for each frame. In the bilingual subtitles condition, the L1 and L2 lines were presented simultaneously with the L1 above the L2 line, which is the common presentation format of bilingual subtitles in China. English text was presented in the Calibri font, and Chinese text was presented in the Songti (宋体) font, both in 35-point font size. Example screenshots of the four subtitling conditions are presented in Appendix S2 in the online Supporting Information. The mean duration of subtitle presentation was 2,168 ms ($SD = 4{,}454$, 95% CI [1,790, 2,546]). Four versions of the video were created, one for each of the subtitling conditions (i.e., captions, L1 subtitles, bilingual subtitles, no subtitles). The subtitles (Wang & Pellicer-Sánchez, 2021a) are openly available on IRIS (https://www.iris-database.org).

Target Words

In order to maintain the ecological validity of the study, we used the original content and audio of the video without any manipulations. We initially

inspected the script and created a list of 66 potentially unknown words. We presented this list to seven experienced Chinese IELTS teachers, whom we asked to select the words that they thought would not be known by the participants. Then, we asked 13 Chinese learners of English with similar characteristics to the participants in the study to indicate their knowledge of the words in the list. This resulted in a final selection of 24 single words, including 10 nouns, 10 verbs, and four adjectives (see Appendix S3 in the online Supporting Information. The word lists are openly available at https://www.iris-database.org). The use of authentic videos made it impossible to control for item-level differences such as frequency of occurrence, word length, and part of speech. Therefore, we included these for the 24 unknown words as covariates in the analyses.

In order to reduce the potential test effects and make the target words less salient, we included 33 distractors in the vocabulary pretest and posttest. The distractors included 10 lower-frequency words (ranked above the 5,000 word level) from the same documentary series and 23 higher-frequency words (within the first 3,000 word level) to make the test less challenging and reduce guessing. The distractors shared the same semantic domain and part of speech as the target words (see Appendix S4 in the online Supporting Information. Target words and distractors are openly available at https://www.iris-database.org).

Vocabulary Tests

In line with previous studies tapping into different aspects of word knowledge (e.g., Mohamed, 2017; Montero Perez et al., 2015; and see Yanagisawa & Webb, 2021, for a meta-analysis finding that different test formats moderated incidental vocabulary learning), we used form recognition, meaning recall, and meaning recognition tests in pencil-and-paper format as both pretests and posttests, randomizing the order of items. Each test comprised 57 vocabulary items including the 24 target words and 33 distractors.

In order to reduce the number of exposures to the target words and minimize the potential test effects (see also Peters, 2019; Peters & Webb, 2018), we combined the form recognition test with the meaning recall test. Each word was presented in both spoken and written form and the audio recording of each word was played twice. After the presentation of each word, the participants were asked to indicate whether they had previously seen/heard the word by ticking Yes or No in the answer sheet. They were then asked to provide a translation/synonym/explanation for the words that they had ticked. The meaning recognition test was a written multiple-choice test, with each item

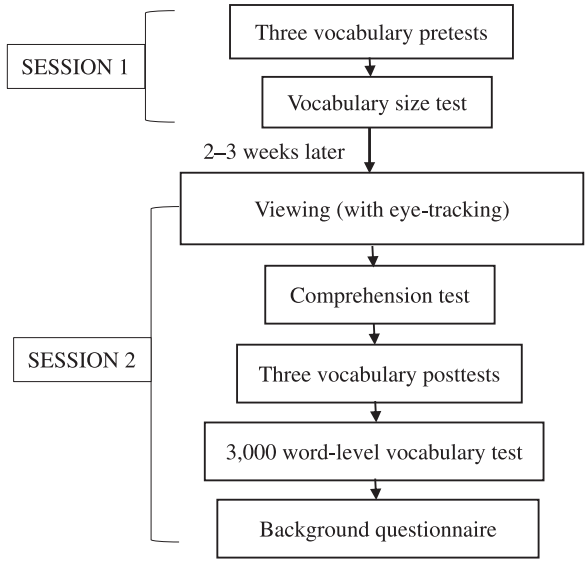

Figure 1 Visual diagram of the research procedure.

accompanied by four options in Chinese: the key, three distractors, and an "I don't know" option to minimize guessing (Peters & Webb, 2018). All the distractors shared the same part of speech with the test items and were relevant to the theme of the video. Following the procedure used by Rodgers (2013), the first distractor option was the translation of another target word. The second and third distractors were randomly chosen from a distractor pool that consisted of 34 synonyms of the 24 target words together with the 10 low-frequency nontarget distractors. We examined the internal consistency of all tests via Cronbach's alpha; results showed that the coefficients were all above .80, indicating good reliability: pretest and posttest form recognition, $\alpha = .83$ and .89, respectively; pretest and posttest meaning recall, $\alpha = .83$ and .84, respectively; pretest and posttest meaning recognition, $\alpha = .85$ for both tests. For the complete vocabulary tests (Wang & Pellicer-Sánchez, 2021b), see Appendix S5 in the online Supporting Information and they are openly available IRIS (https://www.iris-database.org).

Procedure

Data was collected individually in an eye-tracking lab in two sessions with a time gap of two to three weeks between the two sessions (see Figure 1). In the

first session, the participants were first provided with the information about the study. They were told that the purpose of the research was the examination of viewing comprehension and that they would be asked to watch the video for comprehension as they usually did in their leisure time. They were informed of the comprehension test but not of the upcoming vocabulary tests. After receiving the instructions and giving their signed consent, the participants completed the three vocabulary pretests and the vocabulary size test. They were told these tests would be used to measure their vocabulary size. The first session lasted on average 45 minutes. In the second session, the participants were randomly assigned to one of four groups: a captions ($n = 27$), L1 subtitles ($n = 24$), bilingual subtitles ($n = 30$), or no subtitles ($n = 25$) group. The eye-tracking experiment was designed and presented with Experiment Builder (2011). Eye movements were recorded with EyeLink 1000 Plus (2016), in desk-mounted mode. Recording was monocular (right eye). An adjustable head and chin rest was installed 60 cm in front of the monitor to minimize head movements. The stimulus was presented on a 19-inch monitor with a 1920 × 1080 screen resolution. A short practice session was included to acquaint the participants with the procedure. A 9-point calibration was conducted before the practice and another one before the viewing session. The participants were asked to wear headphones during the viewing session. After the short practice, they were given an opportunity to ask questions.

As Figure 1 shows, after the viewing, the participants were asked to complete 34 multiple-choice comprehension questions to check their adequate comprehension of the video. They were then asked to complete the pencil-and-paper vocabulary form recognition, meaning recall, and meaning recognition posttests. The 3,000-word-level vocabulary test and a background questionnaire were also completed at the end of the second session. The second session lasted on average 85 minutes. The procedure was the same for all the participants in the four experimental groups. The specific purpose of this research, that is, examining incidental vocabulary learning, was disclosed at the end of the experiment.

Scoring and Analyses

The 24 target words in form recognition, meaning recall, and meaning recognition tests were scored dichotomously with 0 for an incorrect response and 1 for a correct response. For meaning recall, 1 point was given for the answers with exact meanings or close synonyms. The meaning recall test was scored by two raters. Interrater reliability for both the pretest and the posttest was very

high: pretest, Cohen's kappa = .98, $p = .002$; posttest, Cohen's kappa = .99, $p = .001$.

For the eye-tracking data, separate dynamic areas of interest (AOIs) were created for each occurrence of the target words and their corresponding translations (height = 100 pixels, width varied according to item length). The average size of the AOI of an English target word was 24,338.54 pixels ($SD = 5{,}826.55$, 95% CI [21,878.21, 26,798.88]), and 20,845.83 pixels ($SD = 8{,}144.99$, 95% CI [17,406.51, 24,285.16]) for a Chinese translation of the target word. The dynamic AOIs were only activated during the time the target words and/or corresponding translations were presented. The average time duration of each target word was 3,294 ms ($SD = 1{,}911$, 95% CI [2,488, 4,101]). For each individual participant, only the unknown target words, that is, those with a score of 0 in the form recognition pretest, were included in the eye-tracking data analyses. Fixations shorter than 50 ms were merged if they were within 1° of visual angle (0.34% of the data), and those that were still below 50 ms were removed from the dataset (8.35% of the data).

We conducted the statistical analyses with R (R Core Team, 2019). Because the participants' answers for each target word item in the vocabulary tests and their eye-movement data on each item were nested in a hierarchical fashion within each participant and within each subtitling condition, we analyzed offline (i.e., vocabulary tests) and online (i.e., eye-tracking) data with mixed-effects models due to the advantages of these models for accommodating nested data and including fixed effects, covariates, and random effects (Cunnings, 2012; Gries, 2021). We fitted linear, logistic, and Poisson mixed-effects models according to the type of dependent variables with the lmer or glmer function in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). We took into account by-item and by-participant variations in all models. We checked the collinearity, normal distribution of residuals, and homoscedasticity assumptions for all linear mixed-effects models using the sjPlot package (Lüdecke, 2020), and we used the glmmTMB package (Version 1.0.1; Brooks et al., 2017) for generalized linear mixed models. We first visually inspected potential outliers and identified those with absolute standardized residuals exceeding 2.5 standard deviations ($2.5 < |z|$) using "model criticism" (Godfroid, 2020, p. 267) using the romr.fnc() function in LMERConvenienceFunctions package (Tremblay & Ransijn, 2020) in R. We removed outliers after running sensitive analyses that showed significant differences. For the cross-group comparisons, we fitted separate models for each vocabulary posttest. We ran Tukey posthoc tests using the multcomp package (Version 1.4-13; Hothorn, Bretz, & Westfall, 2008) for pairwise comparisons.

We constructed and have reported the best models using forward selection based on likelihood ratio tests and on Akaike information criterion scores. We always added the variables participant and item as random intercepts. We adopted the maximal random-effects structure because it has been recommended for confirmatory hypothesis testing research to strengthen the generalization of findings (Linck & Cunnings, 2015). We also included random slopes (i.e., Time × Participant, Time × Item, and Group × Item) in the models when applicable if they improved the model fit (based on model Akaike information criterion scores). We removed random effect parameters resulting in the least additional variance explained one by one when models failed to converge until we achieved convergence. We also entered the participant-level variable (i.e., vocabulary size) and item-level variables (i.e., word class, frequency of occurrence, word length, word presentation duration in the video, and AOI size) into the regression models when available because previous studies have shown that these variables might influence vocabulary learning through viewing (e.g., Peters, 2019; Puimège & Peters, 2019). Details of item-level variables are presented in Appendix S3 in the online Supporting Information. We added the participant-level and item-level variables to the models as categorical or continuous covariates. We kept these parameters only in the model if they significantly improved the model fit. We log transformed all the continuous variables before conducting the analyses. We used the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) to obtain *p* values for linear models and used a two-tailed alpha level of .05 for all statistical tests. We calculated Cohen's *d* to estimate the effect size for linear regressions. We considered *d* values of 0.60, 1.00, and 1.40 to indicate small, medium, and large effect sizes respectively for within-group contrasts, and 0.40, 0.70, and 1.00 as thresholds for between-group contrasts (Plonsky & Oswald, 2014). We used the odds ratio as an alternative applicable to logistic regression to measure the effect size (Field, Miles, & Field, 2012). We considered odds ratios greater than 3 or less than 0.33 to be strong (Haddock, Rindskopf, & Shadish, 1998), equivalent to a probability of 75%.

To answer Research Question 1, we examined the effect of the independent variable time (pretest and posttest) on the dependent variable (the participants' scores for the three vocabulary tests). Then, we further investigated the group differences by examining the effect of group (captions, L1 subtitles, bilingual subtitles, and no subtitles) on the three vocabulary posttest scores. We added the participants' pretest scores as a covariate in the analyses. Due to the binary feature of the dependent variables (i.e., 0 or 1 in the vocabulary tests), we constructed a series of logistic mixed-effects models.

To respond to Research Question 2 and Research Question 3, we analyzed the following eye-movement measures: *first fixation duration* (the duration of a reader's first fixation made on an AOI), *first-pass reading time* (the sum of all fixations before exiting), *second fixation duration* (the duration of a second fixation), *second-pass reading time* (the sum of fixations on an AOI when the eye visits the AOI a second time), *total reading time* (the sum of all fixations), *fixation count* (the total number of fixations), and *skip rate* (whether the AOI was fixated upon).

First fixation duration, first-pass reading time, and skip rate are early measures, which can inform us about learners' efforts in recognizing and retrieving the meaning of a word from the mental lexicon (Conklin, Pellicer-Sánchez, & Carroll, 2018). Second-pass reading time and second fixation duration are pure late-processing measures that reflect reanalysis when learners encounter an initial processing difficulty (Godfroid, 2020) and can inform us about more controlled cognitive processes (Conklin et al., 2018). Total reading time and fixation count are aggregate late eye-movement measures. They subsume all time and visits within an AOI and can provide a more general picture of processing patterns (Conklin et al., 2018; Godfroid, 2020). The combination of multiple early and late measures is believed to capture a more complete picture of different cognitive processes (Godfroid, 2020).

We examined the correlation of the seven eye-movement measures with the Pearson correlation procedure. Results showed that, except for skip rate, the different measures had strong and positive significant correlations with each other (all ps $< .01$; the correlation matrix appears in Appendix S6 in the online Supporting Information). Therefore, we have reported only the total reading time, first-pass reading time, second-pass reading time, fixation count, and skip rate in line with previous vocabulary learning research (e.g., Mohamed, 2017; Montero Perez et al., 2015; Pellicer-Sánchez, 2016). We fitted separate models for each eye-movement metric. We always set group as the independent variable. We conducted linear, logistic, and Poisson mixed-effects models for continuous (i.e., first fixation duration, first-pass reading time, total reading time, second fixation duration, and second-pass reading time), binary (i.e., skip rate), and count (i.e., fixation count) eye-tracking data accordingly.

In order to answer Research Question 4, we explored the predictive role of participants' eye movements (as measured by three eye-movement measures: first-pass reading time, second-pass reading time, and total reading time) to the unknown target words and their L1 translations on the participants' vocabulary gains (as measured by the three vocabulary posttests) separately for

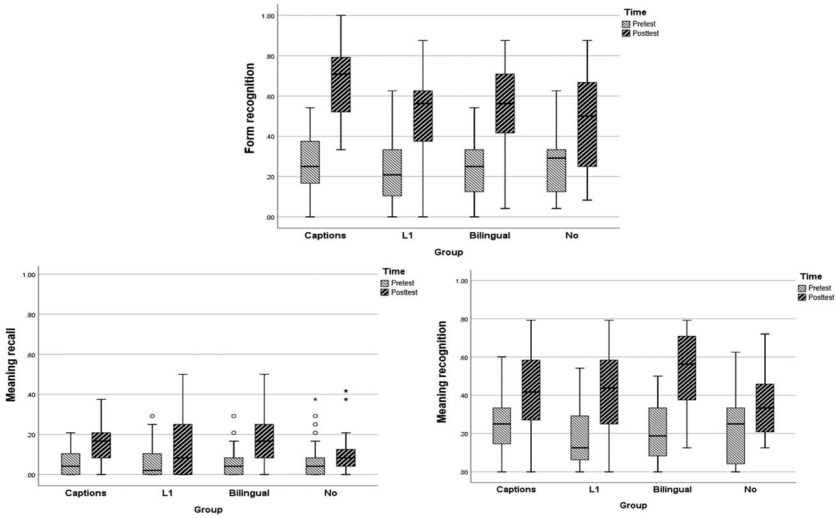

Figure 2 Mean vocabulary pretests and posttests scores by subtitle group (max. score = 1). Lines in the boxes represent median scores; boxes range from the 25th to the 75th percentile; vertical lines range from the minimum to the maximum score, with the symbol ° represents outliers. Captions = L2 subtitles group; L1 = first language subtitles group; Bilingual = bilingual subtitles group; No = no subtitles group.

the bilingual subtitles, the captions, and L1 subtitles groups through logistic mixed-effects models.

Results

Vocabulary Learning

In order to check the comparability among the groups, we first conducted one-way ANOVA analyses. No significant differences were revealed among the four experimental groups for the participants' vocabulary size, $F(3, 102) = 0.01$, $p = .99$, or their overall IELTS scores, $F(3, 102) = 0.51$, $p = .68$. We also fitted the participants' vocabulary pretest scores by three sets of logistic mixed-effects models to ensure the comparability at the onset of the study: form recognition, $\chi^2(3) = 0.72$, $p = .87$, $R^2 < .001$, meaning recall, $\chi^2(3) = 0.78$, $p = .85$, $R^2 < .001$, and meaning recognition, $\chi^2(3) = 1.79$, $p = .62$, $R^2 < .001$.

Descriptive statistics for the pretest and posttest scores by item are provided in Appendix S7 in the online Supporting Information. Figure 2 shows that the posttest scores were in general higher than the pretest scores. In

response to Research Question 1, we first checked whether all the groups had significantly improved their vocabulary knowledge after the treatment. Logistic mixed-effects models revealed that time was a strong and significant predictor of the participants' vocabulary test scores in all groups: form recognition, $\chi^2(6) = 743.08$, $p < .001$, $R^2 = .12$; meaning recall, $\chi^2(6) = 253.00$, $p < .001$, $R^2 = .09$; meaning recognition, $\chi^2(6) = 486.75$, $p < .001$, $R^2 = .09$ (see Appendix S8 in the online Supporting Information for model summaries). The odds of the participants' answering correctly in the posttests were 5.71 (95% CI [4.25, 7.82], $p < .001$), 5.28 (95% CI [3.42, 8.08], $p < .001$), and 4.26 (95% CI [3.13, 5.81], $p < .001$) times higher than those in the pretests in form recognition, meaning recall, and meaning recognition respectively. The participants' vocabulary size also significantly predicted their test scores in all tests. We explored the group differences for the three vocabulary tests in separate models, controlling for differences in pretest scores.

As Figure 2 shows, the captions group obtained the highest mean posttest scores in the form recognition test, although the bilingual subtitles group achieved the highest mean scores for meaning recall and meaning recognition. In order to examine the relative effects of bilingual subtitles compared to other subtitling types, we constructed a series of logistic mixed-effects models. Results revealed significant main effects of group for all three vocabulary posttests. The participants' vocabulary size was also a strong predictor for the three vocabulary posttest scores, but no significant interaction of group and participants' vocabulary size was revealed (see Appendix S9 in the online Supporting Information for model summaries). Results of the Tukey posthoc comparisons (see Table 1) revealed that for form recognition, the captions group significantly outperformed the no subtitles, the L1 subtitles, and the bilingual subtitles groups, with medium effect sizes. This meant that the odds of a correct answer were 2.92, 2.32, and 2.17 times higher in the captions group compared to the no subtitles, L1 subtitles, and bilingual subtitles groups respectively. No difference was revealed between the L1 subtitles, bilingual subtitles, and no subtitles groups.

For meaning recall, the bilingual subtitles group significantly outperformed the no subtitles group and the L1 subtitles group, with large and medium effect sizes, with the odds of providing a correct response being 3.42 and 2.16 times higher than in the no subtitles group and L1 subtitles group respectively. In addition, the captions group also significantly outperformed the no subtitles group in meaning recall, whereas no significant difference was observed between the captions and the bilingual subtitles groups. For meaning recognition, the bilingual subtitles group significantly outperformed the no subtitles

Table 1 Results of post hoc group comparisons for the three vocabulary tests

Group comparison	*b*	95% CI	*SE*	*OR*	95% CI	*Z*	*p*
Form recognition							
Captions > Bilingual	0.77	[0.26, 1.28]	0.26	2.17	[1.29, 3.62]	2.94	**.02**
Bilingual – L1	0.07	[−0.46, 0.60]	0.27	1.07	[0.63, 1.82]	0.25	.99
Bilingual – No	0.30	[−0.23, 0.83]	0.27	1.35	[0.80, 2.29]	1.11	.68
Captions > L1	0.84	[0.29, 1.39]	0.28	2.32	[1.35, 3.99]	3.04	**.01**
Captions > No	1.07	[0.52, 1.62]	0.28	2.92	[1.70, 5.00]	3.88	**< .001**
L1 – No	0.23	[−0.32, 0.78]	0.28	1.26	[0.73, 2.18]	0.81	.85
Meaning recall							
Bilingual – Captions	0.32	[−0.17, 0.81]	0.25	1.38	[0.84, 2.25]	1.28	.57
Bilingual > L1	0.77	[0.24, 1.30]	0.27	2.16	[1.26, 3.69]	2.80	**.03**
Bilingual > No	1.23	[0.66, 1.80]	0.29	3.42	[1.93, 6.01]	4.24	**< .001**
Captions – L1	0.45	[−0.12, 1.02]	0.29	1.57	[0.89, 2.77]	1.57	.40
Captions > No	0.91	[0.32, 1.50]	0.30	2.48	[1.38, 4.45]	3.03	**.01**
L1 – No	0.46	[−0.17, 1.09]	0.32	1.58	[0.85, 2.97]	1.45	.47
Meaning recognition							
Bilingual > Captions	0.78	[0.39, 1.17]	0.20	2.18	[1.49, 3.21]	3.99	**< .001**
Bilingual – L1	0.48	[0.09, 0.87]	0.20	1.62	[1.09, 2.39]	2.42	.07
Bilingual > No	1.08	[0.69, 1.47]	0.20	2.94	[1.99, 4.29]	5.36	**< .001**
L1 – Captions	0.30	[−0.11, 0.71]	0.21	1.35	[0.66, 2.04]	1.45	.47
Captions – No	0.30	[−0.11, 0.71]	0.21	1.35	[0.89, 2.04]	1.44	.47
L1 > No	0.60	[0.19, 1.01]	0.21	1.82	[1.20, 2.77]	2.81	**.03**

Note. Captions, $n = 27$; Bilingual, $n = 30$; L1, $n = 24$; No, $n = 25$. Values of p in boldface are significant at the .05 level of alpha. Captions = L2 subtitles; Bilingual = bilingual subtitles; L1 = first language subtitles; No = no subtitles.

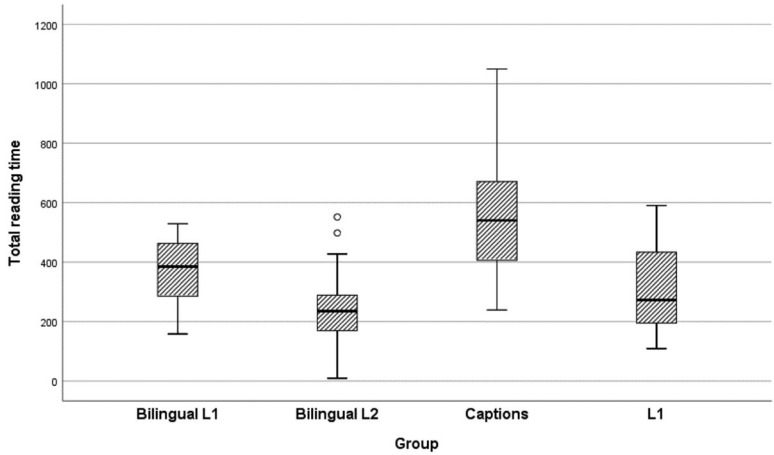

Figure 3 Mean total reading time in milliseconds by group on target items: second language (L2) unknown target words and/or corresponding first language (L1) translations. Lines in the boxes represent median scores; boxes range from the 25th to the 75th percentile; vertical lines range from the minimum to the maximum score, with the symbol ° represents outliers. Bilingual L1 = L1 translations in bilingual subtitles group; Bilingual L2 = L2 translations in bilingual subtitles group; Captions = L2 subtitles group; L1 = L1 subtitles group.

group and the captions group, with 2.94 and 2.18 times increased odds of a correct answer, respectively. Moreover, the L1 subtitles group also outperformed the no subtitles group, but no significant difference between the L1 subtitles and the bilingual subtitles groups was observed. No significant difference was revealed between the captions and L1 subtitles groups in either of the meaning tests.

Processing of Vocabulary in Different Subtitling Conditions

Table 2 summarizes the descriptive statistics for the processing of unknown target words and their corresponding L1 translations in the different subtitling conditions.

To respond to Research Question 2, we first examined the processing of the L2 target words that were unknown in the pretests and their corresponding L1 translations in the bilingual subtitles condition. Table 2 and Figure 3 show that participants in the bilingual subtitles group spent more time on L1 translations than on the L2 words. We conducted five sets of mixed-effects models (one for each eye-movement measure; see Appendix S10 in the online Supporting

Table 2 Descriptive statistics for the eye-movement measures on unknown target words and corresponding first-language translations by group

	Bilingual L1[a] ($n = 28$)		Bilingual L2[b] ($n = 28$)		Captions[c] ($n = 25$)		L1 subtitles[d] ($n = 23$)	
Variable	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI
Total reading time (ms)	355 (358)	[324, 387]	236 (347)	[206, 266]	546 (480)	[502, 591]	317 (398.23)	[278, 356]
1st-pass reading time (ms)	251 (247)	[230, 273]	179 (273)	[155, 202]	352 (368)	[318, 387]	235 (294)	[206, 263]
2nd-pass reading time (ms)	86 (164)	[72, 100]	48 (146)	[36, 61]	153 (233)	[131, 175]	59 (141)	[46, 73]
Fixation count	1.94 (1.79)	[1.79, 2.10]	1.17 (1.60)	[1.03, 1.31]	2.65 (2.15)	[2.45, 2.85]	1.56 (1.72)	[1.39, 1.73]
Skip rate	0.19 (0.40)	[0.16, 0.23]	0.48 (0.50)	[0.44, 0.53]	0.15 (0.36)	[0.12, 0.18]	0.29 (0.45)	[0.24, 0.33]

[a] First language translations in bilingual subtitles group.
[b] Second language translations in bilingual subtitles group.
[c] Second language subtitles group.
[d] First language subtitles group.

Information for model summaries). Within the bilingual subtitles group, the time spent on the L2 unknown target words was significantly shorter than that spent on their L1 translations, as revealed by total reading time, $b = -1.84$, $SE = 0.31$, 95% CI [−2.45, −1.23], $t(1007) = -5.84$, $p < .001$, $d = 0.34$, first-pass reading time, $b = -1.69$, $SE = 0.29$, 95% CI [−2.26, −1.12], $t(1007) = -5.76$, $p < .001$, $d = 0.28$, second-pass reading time, $b = -1.01$, $SE = 0.27$, 95% CI [−1.54, −0.48], $t(1007) = -3.81$, $p < .001$, $d = 0.24$, with small effect sizes, and fixation count, $OR = 0.46$, 95% CI [0.33, 0.65], $p < .001$. The odds of the fixation count on L2 target words were 54% lower than those for the fixation count for reading the L1 translations. The odds of the skip rate of the L2 unknown target words were significantly higher than the odds of the skip rate of the L1 translations, $OR = 6.17$, 95% CI [3.23, 11.79], $p < .001$.

In order to address Research Question 3, we then compared the processing of L2 unknown target words in the bilingual subtitles and captions groups as well as, separately, the processing of the L1 translations in the bilingual subtitles and the L1 subtitles groups. As Figure 3 shows, when the bilingual subtitles conditions were compared with the monolingual conditions, the participants using bilingual subtitles spent less time on the L2 target words than did the participants using captions, but more time on the L1 translations than did the participants using L1 subtitles. Results of the mixed-effects models showed that the captions group spent significantly longer time on the unknown L2 target words than did the bilingual subtitles group, as revealed by total reading time, $b = 2.18$, $SE = 0.33$, 95% CI [1.53, 2.83], $t(946) = 6.50$, $p < .001$, $d = 0.75$, with a medium effect size, first-pass reading time, $b = 1.87$, $SE = 0.32$, 95% CI [1.23, 2.50], $t(946) = 5.92$, $p < .001$, $d = 0.54$ and second-pass reading time, $b = 1.71$, $SE = 0.27$, 95% CI [1.18, 2.24], $t(946) = 6.27$, $p < .001$, $d = 0.55$, with small effect sizes, and fixation count, $OR = 2.70$, 95% CI [1.98, 3.82], $p < .001$. The odds of fixation count on L2 target words for the captions group were 2.7 times higher than the odds for the bilingual subtitles group. The odds of the skip rate in the captions group were significantly lower than the odds of the skip rate in the bilingual subtitles group, $OR = 0.08$, 95% CI [0.04, 0.17], $p < .001$.

For the participants' processing of the L1 translations of the unknown target words in the bilingual subtitles and L1 subtitles groups, results showed that the bilingual subtitles group spent significantly more time processing the L1 translations than did the L1 subtitles group, as revealed by total reading time, $b = 0.63$, $SE = 0.30$, 95% CI [0.04, 1.22], $t(910) = 2.10$, $p = .04$, $d = 0.12$, first-pass reading time, $b = 0.56$, $SE = 0.28$, 95% CI [0.01, 1.11], $t(910) = 1.99$, $p = .05$, $d = 0.07$, second-pass reading time, $b = 0.40$, $SE = 0.20$, 95%

CI [0.01, 0.79], $t(910) = 2.01$, $p = .05$, $d = 0.17$, and fixation count, $OR = 0.81$, 95% CI [0.65, 1.01], $p = .05$, with small effect sizes. Also, the odds of the skip rate in the bilingual subtitles group were significantly lower than the odds of the skip rate in the L1 subtitles group, $OR = 0.51$, 95% CI [1.03, 3.77], $p = .04$. Model summaries are presented in Appendices S11 and S12 in the online Supporting Information.

Relationship Between Processing and Vocabulary Gains

To respond to Research Question 4, we conducted logistic mixed-effects models by group (captions, L1 subtitles, and bilingual subtitles with L1 and L2 AOI separately) to explore the potential relationship between the reading of unknown target words as measured by total reading time, first-pass reading time, and second-pass reading time (following Montero Perez et al., 2015), and the vocabulary gains as measured by form recognition, meaning recall, and meaning recognition posttests scores (see Appendix S13 in the online Supporting Information for model summaries).

Results in Tables 3 and 4 show that, for the bilingual subtitles group, total reading time and first-pass reading time on the L2 unknown target words significantly predicted form recognition gains. This indicated that a 1-second increase in total time and first-pass reading time spent on a L2 unknown target word increased the odds of form recognition success by 3.01 and 5.45 times, respectively. Similarly, meaning recall scores were significantly predicted by total reading time and first-pass reading time on the L2 target words, with a 1-second increase in reading leading to 3.09 and 3.38 times higher odds of gains, respectively. However, none of the measures predicted the meaning recognition gains. Second-pass reading time was not a significant predictor of the vocabulary gains in the bilingual subtitles group.

As Table 4 shows, for the captions group, form recognition scores were significantly predicted by the first-pass reading time on the L2 target words. This indicated that with a 1-second increase of first-pass reading time, the odds of the participants' correctly recognizing the form of each unknown target word increased 2.45 times. The participants' time spent on L2 target words did not significantly relate to their meaning recall. Meaning recognition results pointed to a positive effect of total reading time and first-pass reading time on vocabulary scores, suggesting 1.97- and 2.21-times higher odds of meaning recognition success with one-second increase in reading. Similar to the results for the bilingual subtitles group, second-pass reading time did not significantly predict any vocabulary scores for the captions group.

Table 3 Relationship between processing of unknown target words and vocabulary gains by the bilingual subtitles group

	Bilingual subtitles – L2 unknown target words ($n = 28$)							Bilingual subtitles – L1 translations ($n = 28$)						
Variable	*b*	95% CI	*SE*	*OR*	95% CI	*z*	*p*	*b*	95% CI	*SE*	*OR*	95% CI	*z*	*p*
Form recognition														
1st-pass reading	1.70	[0.86, 2.54]	0.43	5.45	[2.42, 13.14]	3.95	**< .001**	0.16	[−0.72, 1.04]	0.45	1.17	[0.48, 2.82]	0.36	.72
2nd-pass reading	0.39	[−0.98, 1.76]	0.70	1.48	[0.36, 5.85]	0.56	.57	−0.79	[−2.06, 0.48]	0.65	0.46	[0.12, 1.62]	−1.21	.23
Total time	1.10	[0.47, 1.72]	0.32	3.01	[1.63, 5.76]	3.45	**< .001**	−0.29	[−0.72, 0.34]	0.32	0.75	[0.39, 1.39]	−0.91	.36
Meaning recall														
1st-pass reading	1.22	[0.56, 2.18]	0.49	3.38	[1.29, 9.26]	2.47	**.01**	0.64	[−0.42, 1.70]	0.54	1.90	[0.81, 2.13]	1.19	.23
2nd-pass reading	0.99	[−0.75, 2.73]	0.89	2.68	[0.39, 13.84]	1.11	.27	−0.76	[−2.76, 1.24]	1.02	0.47	[0.05, 2.91]	−0.74	.46
Total time	1.13	[0.37, 1.89]	0.39	3.09	[1.43, 6.89]	2.86	**.004**	0.16	[−0.60, 0.92]	0.39	1.18	[0.49, 2.48]	0.41	.68
Meaning recognition														
1st-pass reading	0.65	[−0.23, 1.53]	0.45	1.91	[0.80, 4.82]	1.43	.15	0.70	[−0.36, 1.76]	0.54	2.01	[0.70, 5.84]	1.31	.19
2nd-pass reading	0.59	[−0.92, 2.10]	0.77	1.81	[0.37, 8.20]	0.77	.44	0.46	[−0.93, 1.85]	0.71	1.58	[0.39, 6.60]	0.65	.52
Total time	0.50	[−0.19, 1.19]	0.35	1.65	[0.83, 3.37]	1.42	.16	0.55	[−0.19, 1.29]	0.38	1.74	[0.82, 3.74]	1.45	.15

Note. Values of *p* in boldface are significant at the .05 level of alpha. L2 = second language; L1 = first language.

Table 4 Relationship between processing of unknown target words and vocabulary gains by the captions and L1 subtitles groups

	Captions – L2 unknown target words ($n = 25$)							L1 subtitles - L1 translations ($n = 23$)						
Variable	*b*	95% CI	*SE*	*OR*	95% CI	*z*	*p*	*b*	95% CI	*SE*	*OR*	95% CI	*z*	*p*
Form recognition														
1st-pass reading	0.90	[0.14, 1.66]	0.39	2.45	[1.14, 5.47]	2.28	**.02**	−0.002	[−0.84, 0.84]	0.43	1.00	[0.42, 2.36]	−0.004	1.00
2nd-pass reading	0.08	[−0.99, 1.16]	0.55	1.08	[0.36, 3.19]	0.15	.88	−1.82	[−3.70, 0.06]	0.96	0.16	[0.02, 0.98]	−1.89	.06
Total time	0.49	[−0.12, 1.10]	0.31	1.63	[0.88, 3.04]	1.59	.11	−0.44	[−1.11, 0.23]	0.34	0.65	[0.32, 1.25]	−1.29	.20
Meaning recall														
1st-pass reading	0.54	[−0.62, 1.70]	0.59	1.72	[0.50, 5.35]	0.92	.36	0.45	[−0.90, 1.80]	0.69	1.57	[0.36, 5.77]	0.66	.51
2nd-pass reading	0.40	[−1.60, 2.40]	1.02	1.50	[0.17, 9.93]	0.40	.69	−0.32	[−3.69, 3.05]	1.72	0.72	[0.01, 13.27]	−0.19	.85
Total time	0.42	[−0.56, 1.40]	0.50	1.51	[0.52, 3.93]	0.82	.41	0.18	[−0.92, 1.28]	0.56	1.19	[0.36, 3.34]	0.32	.75
Meaning recognition														
1st-pass reading	0.79	[−0.01, 1.59]	0.41	2.21	[0.99, 5.01]	1.96	**.05**	0.02	[−0.88, 0.92]	0.46	1.02	[0.41, 2.53]	0.05	.96
2nd-pass reading	0.42	[−0.78, 1.62]	0.61	1.53	[0.44, 5.08]	0.70	.49	0.89	[−0.95, 2.73]	0.94	2.43	[0.39, 16.52]	0.95	.34
Total time	0.68	[0.03, 1.33]	0.33	1.97	[1.03, 3.83]	2.06	**.04**	0.17	[−0.52, 0.86]	0.35	1.18	[0.59, 2.37]	0.47	.64

Note. Values of *p* in boldface are significant at the .05 level of alpha. L2 = second language; L1 = first language.

For the time spent on the L1 translations of the target words, Tables 3 and 4 indicate that none of the eye-tracking metrics on the L1 translations showed significant effects on any type of vocabulary test in both bilingual and L1 subtitles groups. This indicated that in general, the processing time that the participants in these two groups spent on the L1 translation of the unknown target words did not increase their chance of learning vocabulary irrespective of the subtitling type.

Discussion

Vocabulary Learning

Research Question 1 aimed at examining the potential advantage of bilingual subtitles over other subtitling types for incidental vocabulary learning from viewing. We examined three components of lexical mastery: form recognition, meaning recall, and meaning recognition. Overall, the results showed that the participants in all subtitling conditions learned vocabulary, further supporting the effectiveness of viewing for vocabulary learning. In line with previous research, form recognition was the easiest component for the participants to acquire, followed by word meaning (either recognition or recall; e.g., Mohamed, 2017; Pellicer-Sánchez, 2016; Peters et al., 2016). Moreover, the participants' gains were higher in meaning recognition than in meaning recall, supporting earlier research findings (e.g., Peters & Webb, 2018; Peters et al., 2016).

Our research further supports the claim that captions and bilingual subtitles are beneficial for intermediate and advanced L2 learners' incidental vocabulary learning (Danan, 2004; Li, 2016). However, L1 subtitles did not show a significant advantage over captions or bilingual subtitles, a finding that is consistent with previous findings (e.g., Li, 2016; Peters et al., 2016). This finding suggests that the benefits of L1 subtitles might be clear for younger and/or less skilled learners (Danan, 2004).

For form recognition, the results demonstrated a general advantage for captions over the L1 subtitles and no subtitles conditions, in line with previous studies (e.g., Peters, 2019; Peters et al., 2016). This study showed a disadvantage for bilingual subtitles compared to captions for participants' learning word forms. These findings indicate that having the L2 written form of unknown vocabulary seems to support learning the unknown vocabulary but that the simultaneous presentation of the L1 might compete with the L2 form, leading to detrimental effects on the learning of word form in bilingual subtitles conditions. The use of the L1 might increase learners' dependence on L1 and limit their learning of L2 forms (Peters, 2019).

For the acquisition of meaning, bilingual subtitles showed an advantage over no subtitles in both meaning tests, supporting findings of Li's (2016) study. Bilingual subtitles were also significantly more beneficial than captions in meaning recognition but not in meaning recall. This finding might be due to the fact that different test constructs reflect different dimensions of word knowledge (Laufer & Aviad-Levitzky, 2017). Meaning recognition tests examine the initial stages of vocabulary learning (Schmitt, 2010), but meaning recall tests reflect deeper vocabulary knowledge. Meaning recall tests do not take into account partial knowledge and require better memory traces than do recognition tests (Laufer & Goldstein, 2004). Bilingual subtitles could help in establishing the initial form-meaning link by providing L1 translations that could be detected via meaning *recognition* tests. However, according to depth of processing theory (Craik & Lockhart, 1972), given translations may have reduced the participants' cognitive analysis and their inferring the meanings of unknown words, leading to formation of a shallower memory trace that was not sufficient for developing an ability to *recall* the meaning of the newly learned words to an extent that was superior to the recall of the captions group.

However, the participants in the bilingual subtitles condition did outperform those in L1 subtitles condition in meaning recall. This potential benefit could be attributed to the presentation of the L2 target words that could draw the participants' attention to unknown word forms (Winke et al., 2010), reduce the chance of bypassing the spoken form of the unknown words, leading to a clearer opportunity to establish the form-meaning connection (Li, 2016). These findings also seem to support Paivio's (1990) bilingual version of the dual coding theory, indicating that the simultaneous presentation of L1 and L2 input with images could provide a stronger connection for an individual's information processing, which could enhance memory recall. This connection seemed to be particularly improved for our participants when the L2 input was presented both aurally and visually.

Online Processing of Unknown Words

Research Question 2 sought to investigate the participants' processing of the unknown target words and their corresponding L1 translations during bilingual subtitled viewing. Within the bilingual subtitles group, the participants processed both the L2 words and their translations but spent significantly more time on the translations. This might have reflected the participants' reliance on the L1 for better comprehension when the aim of the activity was understanding the content. This finding also supports Lwo and Lin's (2012) claim that

learners using bilingual subtitles might turn to the L1 as a shortcut to facilitate comprehension, resulting in less attention paid to L2 forms.

Research Question 3 aimed to compare the processing of target words and their translations in the bilingual subtitles group with their processing in the captions and L1 subtitles groups. The results corroborated the findings of Liao et al. (2020). The participants using bilingual subtitles spent significantly less time on the L2 target words than did the participants in the captions group, as all the measures examined revealed. The longer processing time on the L2 target words in the captions group might have indicated the participants' attempts to guess the meaning of unknown words. Reading research has indeed suggested that longer initial reading time as well as cumulative time may reflect readers' attempts to infer the meaning of words (e.g., Godfroid et al., 2018; Godfroid, Boers, & Housen, 2013). When using bilingual subtitles, the participant viewers could directly refer to the L1 translations to understand the unknown target words, which may have accounted for the shorter reading time on the L2 forms.

It is interesting that the participants in the bilingual subtitles group also spent significantly more time reading the L1 translations than did the participants in the L1 subtitles group. This reading of the L1 translations could have signaled the participants' attempt to build form-meaning connections between the auditory and/or written L2 form and the L1 meaning. This finding confirms the benefits of using L1 in establishing the initial form-meaning link (Schmitt, 2010). This finding is also in line with Li's (2016) finding that bilingual subtitles have a building connection function, as mentioned by one of Li's participants: "It is easier to combine the two languages and built (sic) a connection between them by bilingual subtitles" (p. 198). Importantly, this pattern of eye movements helps to explain the advantage of bilingual subtitles for learning the meaning of unknown words, as the results of the vocabulary tests showed.

Relationship Between Vocabulary Learning and Online Processing

As for the relationship between the participants' attention allocation to the unknown target words (or their translations) and their vocabulary gains, the predictive role of eye-fixations differed by vocabulary knowledge component and by subtitling group. Longer reading time on the L2 target words significantly predicted gains in form recognition for both the bilingual subtitles and captions groups. This is in line with previous findings showing that longer time spent on unknown words is related to successful recognition on posttests, both in reading (Godfroid et al., 2013; Mohamed, 2017) and in viewing (Montero Perez et al., 2015). In line with results of the study by Montero Perez et al.

(2015), the first-pass reading time on the L2 words significantly predicted form recognition gains for the captions group, but total reading time failed to reach significance.

Processing time on the L2 words was also a significant predictor of meaning recall gains in the bilingual subtitles condition. This relationship has also been reported in reading studies (e.g., Godfroid et al., 2018; Mohamed, 2017; Pellicer-Sánchez, 2016). For the captions group, longer time spent on the L2 target words led to higher meaning recognition scores, which also supported findings of reading studies (e.g., Godfroid et al., 2018; Mohamed, 2017; Pellicer-Sánchez, Conklin, & Vilkaitė-Lozdienė, 2020). Time spent on target words only predicted meaning recall for the bilingual subtitles group, whereas it predicted meaning recognition for the captions group. The discrepancy may possibly relate to the involvement of different cognitive processes. The presence of L1 translations in the bilingual subtitles might have facilitated the participants' mapping of L1 translations onto the corresponding L2 forms. Longer processing time on the L2 forms might have reflected this mapping process and might have supported the development of meaning recall ability. However, when the participants used captions, where no L1 translations were available, the longer processing time on L2 forms might have reflected the participants' effort to figure out the meaning of unknown words. Their meaning inferences might have been only strong enough to manifest in meaning recognition but not in the more demanding meaning recall. Second-pass reading time was not a useful predictor for any vocabulary gains. As Montero Perez et al. (2015) argued, longer second-pass reading time might indicate processing difficulty rather than successful learning.

Overall, the results of our study provide further evidence to support the role of eye-movement measures on L2 unknown vocabulary in predicting vocabulary gains, in line with previous studies (e.g., Godfroid et al., 2013, 2018; Mohamed, 2017; Montero Perez et al., 2015; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2020). It is interesting that processing the L1 translations failed to predict any of the vocabulary scores in either the L1 subtitles or bilingual subtitles groups. This points to an interesting contradiction. The participants in the bilingual subtitles group spent more time on the L1 subtitles, and we have interpreted that extra time as a reflection of their attempts to build the form-meaning connections, which was then reflected in an advantage of bilingual subtitles over captions in the participants' meaning recognition scores. However, time processing the L1 was not a significant predictor of the participants' vocabulary gains. This might suggest that it was not only the amount of attention allocated to the translations, as measured by the fixation durations, but

what participants did when processing those translations and the underlying cognitive processes involved. As Montero Perez et al. (2015) argued, eye-tracking data cannot provide a full picture of learners' engagement with the unknown words, and it is not clear whether the reading time reflects learning process, learning difficulty, or just superficial viewing behavior. This further attests to the complexity of the relationship between processing times and outcome measures and points to the need to combine eye-movement data with other types of data such as stimulated recall to further uncover the different subprocesses involved (Pellicer-Sánchez, 2020).

Limitations and Future Research

There are several limitations to our studies that we must acknowledge. First, the results for the vocabulary tests, in particular recognition scores, could have been influenced by test effects. We attempted to control for this by examining the improvements in knowledge of the low-frequency distractors that appeared in the tests but not in the video. The results of the analyses indicated that there were significant gains from the treatments beyond the possible test effects (see Appendix S14 in the online Supporting Information for analyses). A control group who only completes the tests should be included in future research to control for potential test effects. However, it is important to note that any test effects would equally apply to all conditions and therefore, the results of the group comparisons reported in this study would still hold.

Second, the participants in this study were intermediate to advanced Chinese learners of English who had experience using bilingual subtitles. Thus, the findings might not be well-generalized to L2 learners with different proficiency levels and L1 backgrounds. Bilingual subtitles might be more suitable for higher level learners or for learners with experience using bilingual subtitles due to the potential processing burden of the timed-reading of bilingual subtitles.

Third, previous research has shown that audio-visual material with different genres may also affect learning gains (Webb, 2011). Thus, more research targeting different L2 learners with different types of audio-visual material should be conducted. Vocabulary scores could have also been affected by image support (Peters, 2019). Future studies should consider including this variable as a covariate in analyses.

Fourth, in our study, we administered immediate posttests because they were important for allowing us to explore the relationship between processing and learning gains. However, the absence of delayed posttests meant that no claims could be made about the retention of the vocabulary that was learned.

Future studies should examine the retention of vocabulary learned incidentally from bilingual subtitled viewing.

Finally, our study examined the benefits of bilingual subtitles in incidental learning conditions. The effectiveness of bilingual subtitles for vocabulary learning might be maximized in intentional conditions where different form-focused techniques (e.g., textual enhancement or prereading instruction, as in Pellicer-Sánchez et al., 2020) could be employed to direct learners' attention to both L2 forms and L1 translations in a more planned way. Future research should examine the effectiveness of bilingual subtitles in various intentional conditions.

Conclusion

This study provides further evidence for the effectiveness of viewing for vocabulary learning as well as for the benefits of bilingual subtitles in facilitating meaning knowledge. Bilingual subtitles were superior to other forms of subtitling in the acquisition of meaning, whereas they were less effective than captions for form recognition. The participants paid more attention to the L1 translations of unknown vocabulary than to L2 forms in the bilingual subtitles condition, and they spent more time reading the L1 translations of the unknown words than did the L1 subtitles group, which might imply that the participants were attempting to establish form-meaning links. However, additional data (e.g., debriefing interview, stimulated recalls) would be needed to confirm this. Time spent processing the L2 target words predicted form and meaning gains. However, processing time on the L1 was not a significant predictor of learning gains, which further confirmed the need to empirically distinguish amount of attention, as measured by eye-movement measures, from underlying cognitive processes.

Final revised version accepted 23 November 2021

Open Research Badges

This article has earned an Open Materials badge for making publicly available the components of the research methods needed to reproduce the reported procedure. All materials that the authors have used and have the right to share are available at http://www.iris-database.org. All proprietary materials have been precisely identified in the manuscript.

Note

1 Information about the timings of the videos and about the transcripts is available at https://www.iris-database.org. A preview of the clips is available (BBC, 2013), but we have not made the complete videos available because they are proprietary and protected by BBC copyright.

References

Bartolomé, A. I. H., & Cabrera, G. M. (2005). New trends in audiovisual translation: The latest challenging modes. *Miscelánea: A Journal of English and American Studies*, *31*, 89–104. Retrieved from https://www.miscelaneajournal.net/index.php/misc/article/view/438

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

BBC (2013). Animal odd couples. Retrieved from https://www.bbc.co.uk/programmes/b03jp46s

BBC (2019). Subtitle guidelines (Version1.1.7). Retrieved from https://bbc.github.io/subtitle-guidelines/

Bilibili. (n.d.). 奇特的动物伙伴 1 [Odd Animal Couples 1]. Retrieved from https://www.bilibili.com/video/av21620515/

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*, 509–533. https://doi.org/10.1017/s0142716402004022

Bisson, M. J., van Heuven, W. J., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, *35*, 399–418. https://doi.org/10.1017/s0142716412000434

BNC Consortium. (2007). *The British National Corpus* (XML edition; Oxford Text Archive). Retrieved from http://hdl.handle.net/20.500.12024/2554

Brooks, M. E., Kristensen, K., van Benthem, K. J., Berg, A. M. C. W., Nielsen, A., Skaug, H. J., … Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R* Journal, 9, 378–400. https://doi.org/10.3929/ethz-b-000240890

Cobb. T. (n.d.). *Vocabulary Levels Test* (Version 2) [Measurement instrument). Retrieved from https://www.lextutor.ca/tests/vlt2/?mode=test

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

Corel VideoStudio Pro 2018 [Computer software]. (2018). Retrieved from https://www.videostudiopro.com/en/pages/videostudio-2018/

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*, 369–382. https://doi.org/10.1177/0267658312443651

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta*, 49, 67–77. https://doi.org/10.7202/009021ar

Experiment Builder (Version 1.10.165) [Apparatus and software]. (2011). Mississauga, Ontario, Canada: SR Research. Retrieved from https://www.sr-research.com/experiment-builder/

Eyelink 1000 plus [Apparatus and software]. (2016). Mississauga, Ontario, Canada: SR Research. Retrieved from https://www.sr-research.com/eyelink1000plus.html

EyeLink Data Viewer (Version 3.1.246) [Apparatus and software]. (2018). Mississauga, Ontario, Canada: SR Research. Retrieved from https://www.sr-research.com/data-viewer/

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

Frumuselu, A. D., De Maeyer, S., Donche, V., & Colon Plana, M. D. M. G. (2015). Television series inside the EFL classroom: Bridging the gap between teaching and learning informal language through subtitles. *Linguistics and Education*, *32*, 107–117. https://doi.org/10.1016/j.linged.2015.10.001

García, B. (2017). Bilingual subtitles for second-language acquisition and application to engineering education as learning pills. *Computer Applications in Engineering Education*, *25*, 468–479. https://doi.org/10.1002/cae.21814

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

Godfroid, A., Ahn, J., Choi, I. N. A., Ballard, L., Cui, Y., Johnston, S., … Yoon, H.-J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism*, *21*, 563–584. https://doi.org/10.1017/S1366728917000219

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*, 483–517. https://doi.org/10.1017/s0272263113000119

Gries, S. T. (2021). (Generalized linear) Mixed-effects modeling: A learner corpus example. *Language Learning*, *71*, 757–798. https://doi.org/10.1111/lang.12448

Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological method*s, *3*, 339–353. https://doi.org/10.1037/1082-989X.3.3.339

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*, 346–363. https://doi.org/10.1002/bimj.200810425

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Oxford, UK: Blackwell.

IELTS. (n.d.). *IELTS in CEFR scale*. Retrieved from https://www.ielts.org/about-ielts/ielts-in-cefr-scale

Kemp, L. S., & McDonald, J. L. (2021). Second language vocabulary acquisition: The effects of semantic relatedness, form similarity, and translation direction. *Language Learning*, *71*, 716–756. https://doi.org/10.1111/lang.12449

Koolstra, C. M., & Beentjes, J. W. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development*, *471*, 51–60. https://doi.org/10.1007/BF02299476

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest: Tests in linear mixed effects models (Version 3.1-0) [Computer software]. Retrieved from https://cran.r-project.org/src/contrib/Archive/lmerTest/lmerTest_2.0-29.tar.gz

Laufer, B., & Aviad-Levitzky, T. (2017). What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning ecognition? *The Modern Language Journal*, *101*, 729–741. https://doi.org/10.1111/modl.12431

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*, 399–436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Li, Mingyue. (2016). *In vestigation into the differential effects of subtitles (first language, second language, and bilingual) on second language vocabulary acquisition* (Unpublished doctoral dissertation). Edinburgh, UK: The University of Edinburgh.

Liao, S., Kruger, J.-L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation*, *33*, Retrieved from https://www.jostrans.org/issue33/art_liao.php

Linck, J., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*, 185–207. https://doi.org/10.1111/lang.12117

Lüdecke, D. (2020). sjPlot: Data visualization for statistics in social science. R package (Version 2.8.4) [Computer software]. Retrieved from https://CRAN.R-project.org/package=sjPlot

Lwo, L., & Lin, M. C.-T. (2012). The effects of captions in teenagers' multimedia L2 learning. *ReCALL*, *24*, 188–208. https://doi.org/10.1017/s0958344012000067

Mohamed, A. A. (2017). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, *40*, 269–293. https://doi.org/10.1017/S0272263117000092

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

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*, 720–739. https://doi.org/10.1016/j.system.2013.07.013

Muñoz, C. (2017). The role of age and proficiency in subtitle reading: An eye-tracking study. *System*, *67*, 77–86. https://doi.org/10.1016/j.system.2017.04.015

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *317*, 9–13. Retrieved from https://jalt-publications.org/files/pdf/the_language_teacher/07_2007tlt.pdf

Nation, I. S. P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts (Version 3) [Computer software]. Retrieved from https://www.lextutor.ca/cgi-bin/range/texts/index.pl.

Paivio, A. (1990). Bilingual cognitive representation. In A. Paivio (Ed.), *Mental representations: A dual coding approach* (pp. 239–257). Oxford, UK: Oxford University Press.

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology*, *34*, 388–399. Retrieved from https://psycnet.apa.org/doi/10.1037/h0081101

Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning through listening to songs. *Studies in Second Language Acquisition*, *41*, 745–768. https://doi.org/10.1017/S0272263119000020

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

Pellicer-Sánchez, A. (2020). Expanding English vocabulary knowledge through reading: Insights from eye-tracking studies. *RELC journal*, *51*, 134–146. https://doi.org/10.1177/0033688220906904

Pellicer-Sánchez, A., Conklin, K., & Vilkaitė-Lozdienė, L. (2020). The effect of pre-reading instruction on vocabulary learning: An investigation of L1 and L2 readers' eye movements. *Language Learning*, *71*, 162–203. https://doi.org/10.1111/lang.12430

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

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

Peters, E., Noreillie, A.-S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning*, *69*, 747–782. https://doi.org/10.1111/lang.12351

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*, 551–577. https://doi.org/10.1017/s0272263117000407

Plonsky, L., & Oswald, F. L. (2014). How big is "Big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

PortableSoft (2012). SrtEdit (Version 6.3) [Computer software]. Retrieved from http://www.portablesoft.org/srtedit-portable/

Puimège, E., & 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. https://doi.org/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*, 479–496. https://doi.org/10.1080/09571736.2019.1616806

Pujadas, G., & Muñoz, C. (2020). Examining adolescent EFL learners' TV viewing comprehension through captions and subtitles. *Studies in Second Language Acquisition*, *42*, 551–575. https://doi.org/10.1017/S0272263120000042

R Core Team. (2019). R: A language and environment for statistical computing (Version 3.6.1) [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). Wellington, New Zealand: Victoria University of Wellington.

Rodgers, M. P. H., & Webb, S. (2019). Incidental vocabulary learning through viewing television. *ITL - International Journal of Applied Linguistics*, *171*, 191–220. https://doi.org/10.1075/itl.18034.rod

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. London, UK: Palgrave Macmillan.

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

Sweller, J. (2005). The redundancy principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159–168). Cambridge, UK: Cambridge University Press.

Teng, F. (2018). Incidental vocabulary acquisition from reading-only and reading-while-listening: A multi-dimensional approach. *Innovation in Language*

Learning and Teaching, *12*, 274–288. https://doi.org/10.1080/17501229.2016.1203328

Tremblay, A., & Ransijn, J. (2020). LMERConvenienceFunctions: Model selection and post-hoc analysis for (G)LMER models. R package (Version 3.0). Retrieved from https://cran.r-project.org/web/packages/LMERConvenienceFunctions/LMERConvenienceFunctions.pdf

VS: Test your word knowledge [Measurement instrument]. (n.d.). Victoria University of Wellington. Retrieved from https://my.vocabularysize.com

Wang, A., & Pellicer-Sánchez, A. (2021a). *Subtitles. Materials from "Incidental vocabulary learning from bilingual subtitled viewing: An eye-tracking study"* [Collection: Stimuli and experiment files]. IRIS Database, University of York, UK. https://doi.org/10.48316/yk9k-sq16

Wang, A., & Pellicer-Sánchez, A. (2021b). *Vocabulary tests. Materials from "Incidental vocabulary learning from bilingual subtitled viewing: An eye-tracking study*" [Language test]. IRIS Database, University of York, UK. https://doi.org/10.48316/0b9c-3h43

Webb, S. (2011). Selecting television programs for language learning: Investigating television programs from the same genre. *International Journal of English Studies*, *11*, 117–135. https://doi.org/10.6018/ijes/2011/1/137131

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). London, UK: Routledge.

Winke, P., Gass, S. M., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*, 65–86. Retrieved from https://doi.org/10.10125/44203

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

Yanagisawa, A., & Webb, S. (2021). To what extent does the Involvement Load Hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning*, *71*, 487–536. https://doi.org/10.1111/lang.12444

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Descriptive Statistics for Participants' Proficiency.
Appendix S2. Screenshots of Four Subtitling Conditions.
Appendix S3. Characteristics of the 24 Target Words (in Alphabetical Order).
Appendix S4. Distractors Used in the Vocabulary Tests (in Alphabetical Order).
Appendix S5. The Offline Vocabulary Tests.

Appendix S6. A Correlation Matrix of Seven Eye-tracking Measures.
Appendix S7. Descriptive Statistics for the Vocabulary Pretests and Posttests by Group (Research Question 1).
Appendix S8. Mixed-Effects Models of Learning Gains in Three Offline Vocabulary Tests (Research Question 1).
Appendix S9. Mixed-Effects Models for Three Offline Vocabulary Test by Subtitling Group (Research Question 1).
Appendix S10. Mixed-Effects Models for Eye-Movement Measures within the Bilingual Subtitles Group (Research Question 2).
Appendix S11. Mixed-Effects Models for Eye-Movement Measures Between the Bilingual Subtitles and the Captions Group (Research Question 3).
Appendix S12. Mixed-Effects Models for Eye-Movement Measures Between the Bilingual Subtitles and the First-Language Subtitles Group (Research Question 3).
Appendix S13. Mixed-Effects Models of the Relationship Between Offline and Online Measures (Research Question 4).
Appendix S14. Post-Hoc Analyses of the Test Effects.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Using Bilingual Subtitles While Watching Videos to Facilitate Vocabulary Learning

What This Research Was About and Why It Is Important

Language learners can expand their vocabulary knowledge through watching subtitled foreign language videos as entertainment. Bilingual subtitles, which simultaneously present the first language (L1) subtitles and captions (in the second language, L2), have become particularly popular in certain contexts. Despite their popularity, very little is known about their effectiveness for vocabulary learning or how learners make use of the different information presented. This research investigated the relative effects of bilingual subtitles for vocabulary learning compared to other kinds of subtitles. Moreover, learners' eye movements on the unknown words and L1 translations were recorded to explore how learners distributed their attention during viewing. The relationship between eye movements and word learning gains was also investigated. Bilingual subtitles were more effective than other subtitles for facilitating the learning of word meanings, and the learning gains were higher when longer time was spent processing the L2 unknown words during viewing.

What the Researchers Did

- 112 adult Chinese L2 speakers of English completed three vocabulary tests before and after watching an English video clip in one of four subtitling conditions. Their eye movements were recorded during viewing.
- Three vocabulary tests were used to assess learners' knowledge of 24 potentially unknown single words that were selected from the video. The tests assessed learners' ability to recognize the form and meaning of the word and to recall its meaning.
- Four subtitling conditions were compared: bilingual subtitles, captions, L1 subtitles, and no subtitles.
- Five eye-movement measures were used to capture learners' processing of the L2 words and L1 translations.

What the Researchers Found

- Bilingual subtitles seemed to be superior to other subtitling types in learning word *meaning*, but they were less effective than captions for learning word *form*.
- When using bilingual subtitles, learners paid more attention to the L1 translations of the unknown L2 words than to the unknown L2 words themselves.
- The learners with bilingual subtitles spent more time reading the L1 translations than did the learners in the L1 subtitles group, and they spent less time on the L2 unknown words than did the learners who had captions.
- In general, longer time spent on the L2 unknown words was related to higher vocabulary learning gains, whereas longer time spent on the L1 translations of the unknown words did not lead to higher gains.

Things to Consider

- Bilingual subtitles could facilitate the development of meaning knowledge.
- However, learners tended to be attracted by the L1 translations in bilingual subtitles, resulting in less attention on the unknown L2 word forms. Therefore, the benefits of bilingual subtitles could be maximized using techniques to direct learners' attention to the L2 word forms too.
- The current findings are limited to proficient learners who already had experience using bilingual subtitles. Care should be taken when introducing bilingual subtitles to learners with different viewing habits or lower proficiency levels.

Materials and data: Materials are publicly available at https://www.iris-database.org.

How to cite this summary: Wang, A., & Pellicer–Sánchez, A. (2022). Using bilingual subtitles while watching videos to facilitate vocabulary learning. *OASIS Summary* of Wang & Pellicer–Sánchez (2022) in *Language Learning*. https://oasis-database.org