

THIS IS AN EARLIER VERSION OF THE MANUSCRIPT. FOR THE FINAL VERSION, PLEASE CHECK THE JOURNAL WEBSITE: The published version of this paper should be considered authoritative, and any citations or page references should be taken from it: *Studies in Second Language Acquisition*: <https://www.cambridge.org/core/journals/studies-in-second-language-acquisition>

Examining the Effectiveness of Bilingual Subtitles for Comprehension: An Eye-Tracking Study

Andi Wang^a and Ana Pellicer-Sánchez^b

^a Beijing Foreign Studies University; ^bUniversity College London

Abstract

The present study examined the relative effectiveness of bilingual subtitles for L2 viewing comprehension, compared to other subtitling types. Learners' allocation of attention to the image and subtitles/captions in different viewing conditions, as well as the relationship between attention and comprehension, were also investigated. One-hundred-and-twelve Chinese learners of English watched an English documentary clip in one of 4 conditions (bilingual subtitles, captions, L1 subtitles, no subtitles) while their eye movements were recorded. The results revealed that bilingual subtitles were as beneficial as L1 subtitles for comprehension, which both outscored captions and no subtitles. Participants using bilingual subtitles spent significantly more time processing L1 than L2 lines. L1 lines in bilingual subtitles were processed significantly longer than in L1 subtitles, but L2 lines were processed significantly shorter than in captions. No significant relationship was found between the processing time and comprehension for either the L1 or L2 lines of bilingual subtitles.

Keywords: bilingual subtitles; captions; comprehension; eye tracking; viewing

INTRODUCTION

Watching audio-visual materials, such as television shows and movies, has become a popular entertainment activity among second language (L2) learners in the past few decades.

Authentic audio-visual materials are widely available, can be easily accessed, and provide valuable L2 input for language learning (Rodgers & Webb, 2011). Research has shown that the use of on-screen text facilitates learners' comprehension of audio-visual material, further enhancing the benefits of viewing for language learning (Webb & Rodgers, 2009). Studies exploring the effectiveness of different types of on-screen text for comprehension have shown that the use of captions (i.e., on-screen text in the same language as the soundtrack) leads to significantly better comprehension than no captions (e.g., Gass et al., 2019; Montero Perez et al., 2014; Montero Perez et al., 2013; Winke et al., 2010), and that first language (L1) subtitles (i.e., on-screen text in viewers' L1) tend to have an advantage over captions and no subtitles (e.g., Markham et al., 2001; Pujadas & Muñoz, 2020). However, previous research has mainly focused on the comparison between L1 subtitles and captions. Notably, bilingual subtitles (i.e., simultaneous presentation of L1 and L2 subtitles), one type of on-screen text being increasingly used by L2 learners, have received little research attention.

The use of bilingual subtitles has gained popularity among Chinese learners of English during the past decade (Liao et al., 2020). Bilingual subtitles have been claimed to be particularly beneficial for L2 learners, as they combine the advantages of captions and L1 subtitles by providing the L1 translation of the L2 input and making it easier to match the L1 and L2 input (Lunin & Minaeva, 2015). However, it could also be argued that the simultaneous presentation of identical information in multiple forms can be detrimental for comprehension due to increased cognitive burden (Sweller, 2005). Empirical evidence has indeed yielded conflicting findings, with some studies reporting an advantage of bilingual subtitles over captions and L1 subtitles for comprehension (e.g., Dizon & Thanyawatpokin,

2021), and others reporting no significant differences (e.g., Lwo & Lin, 2012). Importantly, the benefits of bilingual subtitles for comprehension and their advantages over other subtitling types may depend on how learners allocate their attention to the different sources of input that are presented simultaneously. However, to the authors' knowledge, Liao et al. (2020) is the only study that has examined learners' processing of input during bilingual subtitled viewing, and although informative, important methodological constraints (e.g., sample size, length of videos, research design) limit the validity of results. Previous eye-tracking studies have reported a relationship between learners' amount of attention to text in L2 multimodal materials and comprehension (e.g., Gass et al., 2019; Pellicer-Sánchez et al., 2020). However, the relationship between the processing of bilingual subtitles and viewing comprehension is yet to be examined. Investigating this relationship is crucial to understand the facilitative (or detrimental) role of bilingual subtitles to support comprehension.

The current study aimed at examining the relative effectiveness of bilingual subtitles for comprehension, compared to other common subtitling types, i.e., captions, L1 subtitles, and no subtitles, as well as learners' attention allocation to the image and on-screen text in each subtitling condition. We also investigated the potential relationship between learners' attention allocation to different subtitling areas and their comprehension scores. This study presents the most comprehensive examination of the role that bilingual subtitles have in supporting L2 comprehension.

BACKGROUND LITERATURE

VIEWING FOR LANGUAGE LEARNING

Watching audio-visual materials in an L2 has been advocated as an effective way to increase L2 learners' exposure to authentic L2 aural input (Webb & Rodgers, 2009). L2 learners also seem to be highly motivated to use audio-visual materials to facilitate their L2 learning

(Rodgers & Webb, 2011). The advantages of learning from viewing are supported by Paivio's (1986) dual-coding theory, which suggests that there are two channels which are responsible for aural and visual input that can function independently and interactively. By presenting information in aural and written modes, both visual and verbal channels can be activated, augmenting memory and leading to deeper and longer-lasting learning than when receiving information from a single mode. This theory is also the basis for the cognitive theory of multimedia learning which predicted that "people learn better from words and pictures than from words alone" (Mayer, 2009a, p. 223).

A great number of previous studies have shown that viewing facilitates L2 development, especially vocabulary (e.g., Montero Perez et al., 2014), grammar (e.g., Lee & Révész, 2018), and listening comprehension (e.g., Montero Perez et al., 2013). Importantly, the use of on-screen text, especially captions and L1 subtitles, makes viewing materials more comprehensible for L2 learners (Webb & Rodgers, 2009). The use of on-screen text could ease the burden of comprehension and increase learners' willingness to engage with authentic L2 input (Danan, 2004; Webb & Rodgers, 2009), which will in turn facilitate language learning.

Previous research on the use of on-screen text has mainly examined the effectiveness of captions and L1 subtitles to support language learning and comprehension. Regarding their effect on language learning, previous studies have shown that captions are helpful in segmenting the speech streams into meaning components and also increasing learners' attention to the unfamiliar L2 input (Gass et al., 2019; Winke et al., 2010), improving L2 learners' vocabulary learning (Montero Perez et al., 2013). Captions also seem to be more beneficial for L2 learners with larger vocabulary sizes (e.g., Montero Perez et al., 2014; Pujadas & Muñoz, 2020). However, for lower proficiency level L2 learners, or when the viewing materials are beyond learners' proficiency, L1 subtitles are more commonly used and

can also increase vocabulary learning gains (Danan, 2004). However, a potential disadvantage of L1 subtitles is that they might distract learners' attention from the written L2 input, being less effective than captions for vocabulary learning (e.g., Peters, 2019). Concerning the effectiveness of captions and L1 subtitles for comprehension, empirical evidence has shown that captions facilitate comprehension, compared to viewing without captions (e.g., Gass et al., 2019; Montero Perez et al., 2013; Winke et al., 2010), and that L1 subtitles tend to be more beneficial for comprehension than captions and no subtitles for beginner and intermediate learners (e.g., Markham et al., 2001; Pujadas & Muñoz, 2020). Importantly, most previous studies have examined the effectiveness of L1 subtitles, compared to captions or no subtitles, when they are the only on-screen text presented. However, L1 subtitles also often appear in combination with captions in bilingual subtitles.

BILINGUAL SUBTITLES

Bilingual subtitles, also called *dual subtitles*, are a subtitling type that presents both L1 and L2 lines simultaneously at the bottom of the screen usually with the L1 on the first line (Gesa Vidal, 2019; Liao et al., 2020). They are often used in certain multilingual regions where two or more languages are spoken (Gesa Vidal, 2019). In the past two decades, as a consequence of increased use of the internet and imported foreign language videos and films, they have rapidly gained popularity in China (Liao et al., 2020; Wang, 2019).

Bilingual subtitles can integrate the advantages of the two monolingual subtitles (i.e., captions and L1 subtitles) by providing an accurate L1 translation of the L2 input and enabling easier matching of the L1 and L2 words (Lunin & Minaeva, 2015). From a theoretical perspective, the potential benefits of bilingual subtitles are supported by the bilingual version of the dual coding theory (Paivio, 1986). According to this theory, apart from the imagery system, there are two verbal systems (corresponding to a bilingual's two

languages) that can function independently and are also interconnected with each other. Thus, the use of L1 translations can augment the interplay between the L2 input and the images by engaging two separate memory stores, leading to better memory recall (Paivio, 2014). In addition, the learning preferences hypothesis (Mayer, 2009b) proposed that different people learn in different ways, so it is beneficial to present information in different formats. This could reduce the possibility of information blockage due to inefficient processing in one delivery path and help to accommodate each learner's preferred ways of learning. However, it could also be argued that the use of bilingual subtitles could impede learning. From the perspective of cognitive load theory (Sweller, 1988) and the redundancy principle (Chandler & Sweller, 1991; Sweller, 2005), identical information presented in multiple forms might result in cognitive overload, which might be detrimental for learning due to the limited capacity of working memory. Bilingual subtitles present the same verbal information in both aural and written modes, together with their written L1 translations, which can be considered redundant. During fast-paced viewing, the need to process all this input under time constraints may potentially increase learners' cognitive burden and hamper information processing.

Despite the potential benefits and the widespread use of bilingual subtitles in some regions, the number of studies exploring their effectiveness for comprehension is scarce and the available research has yielded inconclusive findings. Empirical evidence for the benefits of bilingual subtitles for beginner learners was provided by Dizon and Thanyawatpokin (2021). Results of the comprehension tests (true-false and open-ended questions) showed that bilingual subtitles had an advantage over captions and L1 subtitles. However, the advantage of bilingual subtitles over L1 subtitles was attributed to the possible higher L2 proficiency of the bilingual subtitles group compared to the other groups.

Mixed findings have been reported in studies targeting intermediate and advanced learners. In the study by Wang (2019), university students from four classes were asked to

watch four excerpts of an American sitcom series in four subtitling conditions (i.e., bilingual subtitles, captions, L1 subtitles, no subtitles). Results showed that the bilingual and L1 subtitles in general outscored captions and no subtitles, but the superiority of bilingual subtitles was not observed in all four classes. Similarly, participants in Hao et al.'s (2021) study were asked to watch four 5-minute TED talk videos in one of four conditions (i.e., bilingual subtitles, captions, L1 subtitles, no subtitles) and then completed a multiple-choice comprehension test. Results revealed no superiority of bilingual subtitles for comprehension compared to other conditions. However, the bilingual subtitles used in Hao et al.'s (2021) study presented the L2 lines on top of the L1 lines, which is not the usual presentation of bilingual subtitles and could have affected results.

The inconclusive findings reported so far could potentially relate to learners' use of the different input sources in bilingual subtitles, i.e., images, audio, and on-screen text. There is more input to process in bilingual subtitles, compared to other subtitling types, and learners need to decide how they distribute their attention to the different information being presented on the screen. Lwo and Lin (2012) asked 32 Chinese junior high school students to watch two simple animations in one of four conditions (bilingual subtitles, captions, L1 subtitles, no subtitles). Results of the multiple-choice comprehension test showed no significant advantage of bilingual subtitles for comprehension over the other conditions. Semi-structured interviews exploring participants' attention allocation during subtitled viewing showed that participants paid most attention to the images, followed by on-screen text, and audio soundtrack. The bilingual subtitles group reported that L1 lines helped them the most in understanding the content, followed by images, audio soundtrack, and lastly L2 lines. It was also argued that more proficient L2 learners were more likely to be distracted by L1 lines when using bilingual subtitles, whereas lower-level learners seemed to benefit more from bilingual subtitles by selectively using the input sources to aid their comprehension.

In sum, although some studies suggested that bilingual subtitles were as beneficial as L1 subtitles and more effective than captions for comprehension (e.g., Dizon & Thanyawatpokin, 2021; Wang, 2019), other studies failed to capture this advantage (e.g., Hao et al., 2021; Lwo & Lin, 2012). These conflicting findings can be explained by participants' different proficiency level and the different viewing materials used. Some of these studies used viewing materials that were beyond participants' proficiency (e.g., Dizon & Thanyawatpokin, 2021; Hao et al., 2021), while others used too simple stimuli (e.g., Lwo & Lin, 2012). Importantly, results of some of these studies are based on comprehension measurement instruments that were not properly validated (e.g., Dizon & Thanyawatpokin, 2021; Lwo & Lin, 2012). Crucially, the inconclusive findings might have to do with the manner in which participants use the different lines in bilingual subtitles. Although Lwo and Lin (2012) attempted to explore learners' attention distribution during viewing via semi-structured interviews, self-report data may not accurately capture learners' unconscious attention distribution during viewing.

EYE-TRACKING STUDIES ON VIEWING FOR COMPREHENSION

Many studies have used eye-tracking to examine the processing of input in subtitled viewing. d'Ydewalle et al. (1991) found that the reading of on-screen text was more or less spontaneous, and viewers could switch effortlessly between the images and the subtitling area. When on-screen text is presented, learners tend to spend slightly more time processing the on-screen text than the dynamic images to aid their comprehension, especially for beginner and low-intermediate level L2 learners (e.g., Gass et al., 2019; Winke et al., 2013). Processing time on on-screen text seems to be affected by various factors. Studies examining captions have shown that learners' L2 proficiency and the orthographic distance between learners' L1 and L2 could affect their processing of captions. Learners with higher proficiency

tend to spend less time processing captions (e.g., Gass et al. 2019; Muñoz, 2017; Winke et al., 2013), and that learners seem to rely more heavily on captions when there is a great orthographic difference between their L1 and L2 (e.g., Winke et al., 2013). Previous research has also compared L2 learners' processing of captions and L1 subtitles. When learners have no knowledge of the L2, they seem to spend similar amount of time on captions and L1 subtitles (e.g., Bisson et al., 2014), but longer processing times on captions than on L1 subtitles have been reported for learners with various L2 proficiencies (e.g., Muñoz, 2017). The longer processing time on captions might indicate increased cognitive effort required for comprehension and processing difficulties (Muñoz, 2017; Winke et al., 2013).

To the best of our knowledge, only two studies have used eye-tracking to explore the processing of bilingual subtitles during L2 viewing. The study by Author and Author (2022) focused on the acquisition of new vocabulary from documentary viewing and found that intermediate to advanced Chinese learners of English spent significantly more time processing the L1 translations of unknown words than the L2 unknown words in bilingual subtitles. However, the focus of this study was on vocabulary learning and only attention to target vocabulary was examined. The study by Liao et al. (2020) is the only study that has explored the processing of bilingual subtitles with a focus on viewing comprehension. In a within-subject design, 20 intermediate level Chinese postgraduates were asked to watch four 5-minute documentary clips in four subtitling conditions (i.e., captions, L1, bilingual, and no subtitles). Participants spent less time processing the on-screen text than the images. Similar reading times were reported on bilingual subtitles (34%) and captions (32%), both being significantly longer than on L1 subtitles (22%). No significant differences were revealed between the time processing L1 and L2 lines in bilingual subtitles. However, the time spent on L2 lines in bilingual subtitles (15%) was significantly shorter than in captions (32%), whereas the processing time of L1 lines was similar for bilingual (18%) and L1 (22%)

subtitles. Results of a written, free recall comprehension test conducted in participants' L2 revealed no significant differences across conditions. However, important methodological constraints, including the small participants sample ($N = 16$), the short duration of the videos (5 minutes), and the potential order effect caused by the within-subject design implemented (i.e., participants always used bilingual subtitles immediately after using captions), limit the validity and generalisation of the findings.

Eye-tracking studies have also shown that learners' processing of the on-screen text seems to negatively relate to their comprehension (e.g., Gass et al., 2019) and that processing time on the text in multimodal texts is negatively related to comprehension (e.g., Pellicer-Sánchez et al., 2020). The longer processing time on the text was interpreted as a sign of potential processing difficulties. However, no previous studies have examined the relationship between processing time and comprehension in bilingual subtitled viewing.

THE PRESENT STUDY

As shown above, studies investigating the effects of bilingual subtitles on comprehension are scarce and findings are inconclusive. The inconclusive findings could be due, among other factors, to differences in learners' use of the different input sources. The study by Liao et al. (2020) provided some initial evidence for learners' allocation of attention in bilingual subtitles, but results were affected by important methodological constraints. Moreover, no research has examined the relationship between learners' processing of bilingual subtitles and comprehension. The present study addressed these gaps and aimed to answer the following research questions (RQs):

RQ1: To what extent does the use of bilingual subtitles enhance L2 learners' viewing comprehension, compared to captions, L1 subtitles, and no subtitles?

RQ2: How do learners distribute their attention during bilingual subtitled viewing compared to captions, L1 subtitles, and no subtitles, as revealed in eye-tracking data?

RQ3: Is there a relationship between processing time on the subtitling area and comprehension scores?

To address these questions, participants were asked to watch a 23-minute clip in one of four subtitling conditions (i.e., bilingual subtitles, captions, L1 subtitles, no subtitles), while their eye movements were recorded. After the viewing, they were asked to complete a comprehension test. Based on previous research (e.g., Dizon & Thanyawatpokin, 2021; Wang, 2019), it was hypothesised that bilingual subtitles would be as good as L1 subtitles for comprehension, but better than captions and no subtitles (RQ1). Regarding participants' attention allocation (RQ2), based on the findings from Liao et al. (2020), participants using bilingual subtitles were expected to spend less time processing images than the subtitling area, and that time would be equally allocated to L1 and L2 lines. Participants using bilingual subtitles and captions would spend a similar amount of time on the overall subtitling area, and would spend longer time than the L1 and no subtitles groups. The time spent on L2 lines in bilingual subtitles would be longer than in captions, while the time on L1 lines in bilingual subtitles would be similar to L1 subtitles. Regarding the relationship between on-screen processing time and comprehension (RQ3), a negative relationship was hypothesised (Gass et al., 2019).

METHODOLOGY

This study is part of a larger study that examined the effect of bilingual subtitles on various outcome measures. Results about vocabulary learning and processing of novel words are reported in Author and Author (2022), whereas the present study focuses on comprehension

and the processing of image and subtitling areas. Thus, the participants and viewing materials are the same as those in Author and Author (2022).

PARTICIPANTS

One-hundred-and-twelve Chinese learners of English from a British university (98 females and 14 males) aged between 18 and 34 years ($M = 23.42$, $SD = 2.47$, 95% CI [22.93, 23.87]) participated in this study. They had a high-intermediate to low-advanced proficiency level in English (B2 to C1 levels according to Common European Framework of Reference for Languages), as determined by their reported International English Language Testing System (IELTS) scores ($M = 6.84$, $SD = 0.61$, 95% CI [6.67, 6.90]) in the background questionnaire and vocabulary size ($M = 6274.31$, $SD = 1704.65$, 95% CI [5950.67, 6597.95]), measured by the Vocabulary Size Test (Nation & Beglar, 2007). About 80% of the participants reported that they enjoyed watching audio-visual materials as an entertainment and that they frequently used on-screen text, with bilingual subtitles being the most frequently used ($M = 4.44$, Max = 6), followed by captions ($M = 3.14$), L1 subtitles ($M = 3.03$), and no subtitles ($M = 2.18$).

We tested participants' knowledge of words at the 3,000 (3K) word-level using the Vocabulary Levels Test (Schmitt et al., 2001) to ensure the comprehensibility of the selected viewing material. We assumed that demonstrating knowledge of the 3K frequency band would also indicate their familiarity of the first 2,000 most frequent words. While 86% of the participants showed mastery of the 3K band, 16 participants failed to meet the mastery threshold (score of 24 out of 30, Xing & Fulcher, 2007). Since 13 of them reported no difficulty in understanding the content of the video and no differences were revealed by including and excluding these 13 participants in statistical analyses, they were kept in the final analysis. Data from the remaining three participants were discarded. Data from three participants who did not complete the posttests were also discarded. Finally, data from another six participants were removed from the analysis of online data due to poor calibration and

track loss. In total, 106 participants were included in the analysis of offline data, and 100 participants in the analysis of online data.

One-way ANOVA analyses showed no significant group differences concerning participants' proficiency as reported by their overall IELTS score $F(3, 102) = 0.51, p = .68$, IELTS listening $F(3, 102) = 1.44, p = .24$, IELTS reading $F(3, 102) = 0.66, p = .58$, and vocabulary size $F(3, 102) = 0.01, p = .98$. The descriptive statistics for participants' proficiency scores are presented in Appendix S1.

MATERIALS

Viewing Material

Four authentic video excerpts (in total 23 minutes, 3488 words) from the documentary *Animal Odd Couples* (Keens-Soper et al., 2013) were extracted and put together using the video editing software Corel VideoStudio Pro 2018 (2018). This material was found suitable for several reasons: documentary is appropriate for L2 learners for its clear oral presentation and rich imagery support (Rodgers, 2018); different from the typical documentary characterised by a single narrator and slow-moving pace, this documentary also included interactive interviews between different speakers, which was considered more engaging; it was long enough to obtain sufficient aural input; and importantly, knowledge of the first 3K most frequent words provided 95.57% coverage, as analysed by the Range software (Nation & Heatley, 2002) with the British National Corpus (BNC Consortium, 2007) as the reference corpus. Since learners had demonstrated knowledge of the 3K level in the Vocabulary Levels Test, and a coverage of 95% has been considered sufficient for adequate comprehension of viewing materials (Webb & Rodgers, 2009), the selected materials should not pose comprehension difficulties for learners.

The original English video script was retrieved online and translated into Chinese by the

first author. In order to ensure accuracy, modifications were made according to comparison with the online amateur translation (Bilibili, n.d.); judgments of three L1 Chinese speakers fluent in English; and evaluations of 13 advanced Chinese learners of English. The captions and L1 subtitles were added to the videos using SrtEdit (PortableSoft, 2012) and Corel VideoStudio Pro 2018 (2018) software following the *BBC Subtitle Guidelines* (BBC, 2019). All the L1 subtitles and captions were kept within one 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 was presented in Calibri font, and Chinese was presented in Songti (宋体) font, both in 35 point font size. The average duration of subtitle presentation was 1,987 ms ($SD = 812$, 95% CI [1,918, 2,056], Range = 496 – 6,868). Four versions of the video were created with captions, L1 subtitles, bilingual subtitles, and no subtitles. The subtitles are openly available on IRIS (<https://www.iris-database.org>; Author & Author, 2021).

Comprehension Tests

Comprehension was assessed by means of 34 four-option multiple-choice questions presented in Chinese to ensure that the test scores were not influenced by other intervening variables (Buck, 2001). In line with Montero Perez et al. (2014) and Rodgers (2013), the development of the comprehension test was based on Buck's (2001) "competency-based" (p. 114) default listening construct, which includes the ability to process the general information, understand the detailed content, and the ability to make inferences. The inferencing ability was not tested in this study due to the features of the viewing material, in which the certain and factual information provided left no room to infer information (see also Montero Perez et al., 2014). Global questions about the general understanding of the content, and local questions about the

detailed content, were included in the test. All items were text-based and could not be answered only by watching the images.

The design of multiple-choice items was based on *idea units* (Pellicer-Sánchez et al., 2020; Rodgers, 2013), defined as “distinct events, actions, or dialogue spoken in the course of the program” (Rodgers, 2013, p. 33). Four plausible and reasonable distractors were chosen for each stem. In order to ensure that the tests could not be answered correctly without understanding the video (Buck, 2001), the test was first piloted with six Chinese learners of English who did not watch the video. Items that had been answered correctly by all were discarded, modifications were made based on the test results and feedback received. The remaining 34 multiple-choice items were then piloted online with 38 intermediate-advanced Chinese EFL learners (Cronbach’s alpha = .67). Modifications were finally made based on the pilot results. Cronbach’s alpha coefficients for the final comprehension test completed by participants in the main study was .83, indicating good reliability. The complete comprehension test is included in Appendix S2.

PROCEDURE

Data were collected individually in an eye-tracking lab in two sessions. Participants’ vocabulary size was measured in the first session. In the second session, participants were randomly assigned to one of four groups (i.e., captions, L1, bilingual, and no subtitles) to watch the video for comprehension, with their eye movements recorded by EyeLink 1000 plus (SR Research, 2016) in desk-mounted mode (sampling rate = 1,000 Hz; accuracy = .25 – .5°; precision < .01°). 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 conducted before the viewing session, after which participants could ask

questions about the procedure. Participants were aware of the forthcoming comprehension test. A nine-point calibration was conducted before the practice and another one before the viewing session. Participants were asked to wear the headphones during their viewing.

Participants' comprehension was measured immediately after the viewing in pencil-and-paper format with no time pressure. Participants were then asked to complete the 3K vocabulary test and an online background questionnaire. The procedure was the same for all participants.

SCORING AND ANALYSES

For the comprehension test, one point was given for each correct response and zero for incorrect response, resulting in a maximum of 34 points. For the analysis of eye movements, following suggestions by Godfroid (2020), fixations shorter than 50 ms were first 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). The analysis of the eye-movement data was performed at two levels: the overall subtitling area and the L1/L2 subtitle line. Different area of interests (AOIs) were created for each level:

Level 1: The overall subtitling area in the four conditions

The aim of Level 1 analysis was to explore potential differences in processing the overall subtitling area across the four subtitling conditions. In order to ensure the comparability between groups, the bilingual subtitles group (with the largest subtitling area size) was set as the baseline group in deciding a size of the AOI for all groups. The overall subtitling area covered 1920×270 pixels, including the whole width of the screen and the height between the on-screen text and the bottom of the screen (Figure 1). The rest upper part of the screen (1920×810 pixels) was taken as the image AOI. The same AOI was applied to all groups.



Figure 1. Illustration of level 1 area of interest for eye-movement data analysis in the four groups. Top left: bilingual subtitles; top right: L1 subtitles; bottom left: captions; bottom right: no subtitles.

Level 2: The L1/L2 subtitle line area in the subtitled conditions

The aim of Level 2 analysis was to further investigate the processing of the different lines in bilingual subtitles and compare them with the other two monolingual subtitling conditions.

For Level 2, AOIs with the same size of 1920×100 pixels were created for L1 and L2 lines covering the subtitling area for the three subtitled conditions (Figure 2).



Figure 2. Illustration of level 2 area of interest for eye-movement data analysis in three subtitled groups. Top left: bilingual subtitles; top right: L1 subtitles; bottom: captions.

For these analyses, 535 interest periods (IPs) were generated manually according to the presentation time of the on-screen text. Only the eye-movement data that occurred within the AOIs and during the 535 IPs were included in the analysis (e.g., Winke et al., 2013). Following previous subtitle processing research (e.g., Bisson et al., 2014; Liao et al., 2020; Muñoz, 2017), four eye-tracking measures were used: *total reading time %* (i.e., the percentage of all summed fixation durations on an AOI within the defined IP), *fixation %* (i.e., the percentage of the total number of fixations on an AOI within the defined IP), *average fixation duration* (i.e., the average duration of fixations on an AOI within the defined IP), and *skip rate* (i.e., an AOI is considered skipped if no fixation occurred in the AOI within the defined IP).

All statistical analyses were performed using R (v 3.6.1; R Core Team, 2019). To answer RQ1 and explore the effect of group on participants' comprehension scores, linear regression analyses were conducted using the *lm* function from the base R stats package. Since previous studies have shown that participants' vocabulary knowledge is related to

viewing comprehension scores (e.g., Montero Perez et al., 2014; Pujadas & Muñoz, 2020), participants' vocabulary size scores were added as a covariate. The multcomp package (v 1.4-13; Hothorn et al., 2008) was used for Tukey post-hoc pairwise comparisons. No outliers were detected (taken as $|z| \geq 3$; see Field et al., 2012).

In response to RQ2, regarding participants' eye movement data, for level 1, we examined the effect of the independent variable, group (i.e., captions, L1, bilingual, and no subtitles) on the dependent variables (i.e., four eye-movement measures), through mixed-effects models. For level 2, we examined the effects of subtitling line (i.e., bilingual L1, bilingual L2, captions, L1 subtitles) on the four eye-movement measures separately. Since learners' eye-movement data during each IP were nested in a hierarchical fashion within each participant and within each subtitling condition, they were analysed with mixed-effects models to accommodate nested data and include fixed effects, covariates, and random effects, which enabled the findings to be more generalisable to different viewers and watching materials (Cunnings, 2012). Based on the types of dependent variables, linear mixed-effects models were built for continuous dependent variables (i.e., total reading time %, fixation %, and average fixation duration), and logistic mixed-effects models were constructed for binary dependent variables (i.e., skip rate) with *lmer* or *glmer* function in the lme4 package (Bates et al., 2015). The continuous outcome variables were all log-transformed and 1 was added to address the skewness problem which is common in eye-movement data (Godfroid, 2020). Participant and IP were always added as random intercepts. Group was also checked as random slope by IP. Participants' log-transformed vocabulary size was added as a covariate. Random slope and covariates were only kept in the model when they improved the model fit. The best models were constructed using forward selection method and reported based on likelihood ratio tests with the *anova* function and on Akaike information criterion scores. The normality, linearity, homoscedasticity and independence of residuals assumptions for the

models were checked for all linear mixed-effects models using sjPlot package (v 2.8.4; Lüdtke, 2020), while glmmTMB package (v 1.0.1; Brooks et al., 2017) was used for generalized linear mixed-effects models. Outliers were identified using “model criticism” (Godfroid, 2020, p. 267) after fitting the best models using the *romr.fnc* function in LMERConvenienceFunctions package (Tremblay & Ransijn, 2020) and were removed from the analyses when they changed the statistical significance of the fixed effects in the models. Tukey post-hoc tests were ran using the multcomp package (v 1.4-13; Hothorn et al., 2008) for pairwise comparisons. For linear mixed-effects models, LmerTest package (Kuznetsova et al., 2017) was used to obtain *p*-values. Cohen’s *d* was used to calculate the effect size using the *cohensD* function in the lsr package (Navarro, 2015). *D* values of .40, .70, 1.00 were considered to be small, medium and large effect sizes, respectively (Plonsky & Oswald, 2014). Odds ratio (OR) was used for generalized linear mixed-effects models to measure the effect size (Field et al., 2012). An OR larger than one indicates positive relationship and an OR less than one indicates a negative relationship. ORs greater than 3 or less than 0.33 are considered to be strong (Haddock et al., 1998).

For RQ3, the eye-movement data for level 2 were used to determine whether processing time could predict participants’ comprehension scores in three subtitled conditions. Total reading time % and average fixation duration for each IP were averaged for each participant to reveal their average processing time on different subtitling line. Participants’ comprehension scores were transformed into accuracy percentage. Since the values were averaged across IPs for each participant, linear mixed-effects models were no longer viable as a statistical method, therefore, we opted to fit simple linear regression models to the response accuracy data using the *lm* function. Emmeans package (Lenth, 2020) was used to run post-hoc analysis.

RESULTS

RQ1 – COMPREHENSION

Both the bilingual and L1 subtitles groups achieved the highest mean scores, with average scores around 80% (Table 1). Linear regression was used to test if subtitling condition significantly predicted participants' comprehension scores while controlling for participants' vocabulary size. Subtitling condition was found to significantly predict comprehension scores, $R^2 = .34$, $F(4, 101) = 14.72$, $p < .001$, suggesting significant group difference. Participants' vocabulary size was also found to positively relate to their comprehension scores, $b = 0.001$, $SE < 0.001$, $t = 2.71$, $p = .01$. However, the addition of the interaction between subtitling condition and vocabulary size did not increase the goodness of fit of the model, $\chi^2(2) = 39.44$, $p = .56$. Thus, the interaction was not included in the model. Model details are summarised in Appendix S3.

Table 1. Descriptive statistics for comprehension scores by group

Group	Comprehension scores	
	<i>M (SD)</i>	95% CI
Captions ($n = 27$)	23.48 (4.15)	[21.84, 25.12]
L1 subtitles ($n = 24$)	27.38 (4.71)	[25.39, 29.36]
Bilingual subtitles ($n = 30$)	27.93 (4.23)	[26.35, 29.51]
No subtitles ($n = 25$)	20.48 (4.81)	[18.50, 22.46]

Note. max = 34 in all cases

Post-hoc group comparisons were conducted to further investigate group difference on comprehension. Table 2 shows that the bilingual subtitles group significantly outperformed the captions and no subtitles groups, with medium and large effect sizes respectively. The L1

subtitles group also significantly outscored the captions and no subtitles groups, with small and large effects sizes respectively. However, no significant difference was revealed between the bilingual and L1 subtitles groups. The difference between the captions and no subtitles groups was only approaching significance.

Table 2. Results of post-hoc comparisons for comprehension scores

Group	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>d</i>	95% CI
Bilingual > Captions	4.46	1.15	3.88	<.001	0.89	[2.21, 6.71]
Bilingual – L1	0.52	1.19	0.44	.97	0.26	[-1.81, 2.85]
Bilingual > No	7.48	1.17	6.38	<.001	1.47	[5.19, 9.77]
L1 > Captions	3.94	1.22	3.24	.01	0.56	[1.55, 6.33]
Captions – No	3.02	1.20	2.51	.06	0.69	[0.67, 5.37]
L1 > No	6.96	1.24	5.62	<.001	1.13	[4.53, 9.39]

RQ2 – ATTENTION ALLOCATION DURING VIEWING

Level 1 Overall subtitling area

Analyses at this level examined participants’ attention distribution to the images and subtitling areas within each group and compared participants’ attention allocation to the subtitling area across groups.

As shown in Table 3, the bilingual subtitles group recorded about 60% total reading time and fixations on the subtitling area and about 40% of the time was allocated to the image area during viewing. Participants had about 8% probability of skipping the bilingual subtitles. The overall time distribution between the image and subtitling area in bilingual subtitles was similar to the captions group. The L1 subtitles group tended to distribute their attention equally to the image and subtitling area. As expected, the no subtitles group, who did not have

any on-screen text, spent most of the time on the image area.

Table 3. Descriptive statistics for eye-movement data at level 1 for overall subtitling and image area by group

Group	Total reading time %				Fixation %				Average fixation duration (ms)				Skip rate			
	Subtitle		Image		Subtitle		Image		Subtitle		Image		Subtitle		Image	
	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI
Captions (<i>n</i> = 25)	0.64 (0.31)	[0.63, 0.64]	0.35 (0.31)	[0.35, 0.36]	0.68 (0.28)	[0.68, 0.69]	0.31 (0.28)	[0.31, 0.32]	179 (87)	[178, 181]	213 (183)	[210, 216]	0.07 (0.26)	[0.07, 0.07]	0.24 (0.42)	[0.23, 0.24]
L1 subtitles (<i>n</i> = 23)	0.44 (0.29)	[0.43, 0.44]	0.55 (0.29)	[0.54, 0.55]	0.50 (0.27)	[0.49, 0.50]	0.48 (0.27)	[0.48, 0.49]	177 (150)	[175, 180]	262 (190)	[259, 265]	0.10 (0.30)	[0.09, 0.10]	0.08 (0.28)	[0.08, 0.09]
Bilingual (<i>n</i> = 28)	0.60 (0.32)	[0.59, 0.60]	0.39 (0.32)	[0.38, 0.39]	0.63 (0.30)	[0.62, 0.63]	0.36 (0.29)	[0.35, 0.36]	177 (128)	[175, 179]	196 (168)	[193, 199]	0.08 (0.27)	[0.07, 0.08]	0.21 (0.41)	[0.20, 0.21]
No subtitles (<i>n</i> = 24)	0.04 (0.13)	[0.03, 0.04]	0.94 (0.20)	[0.93, 0.94]	0.04 (0.12)	[0.03, 0.04]	0.94 (0.20)	[0.93, 0.94]	33 (116)	[31, 35]	322 (250)	[317, 326]	0.88 (0.32)	[0.88, 0.89]	0.03 (0.17)	[0.03, 0.03]

To compare participants' processing of the subtitling area across four groups, four sets of mixed-effects models were constructed for four eye-tracking measures separately (see Appendix S4 for model summaries). Results revealed that group was a significant predictor of participants' processing of the subtitling area in all measures, while controlling for vocabulary size. Table 4 and Table 5 summarise post-hoc pairwise comparisons with Tukey correction for the four eye-tracking measures. Table 4 reveals that, concerning total reading time %, fixation %, and average fixation duration, as expected, the no subtitles group had the shortest average fixation duration and spent significantly less time on the subtitling area than the other subtitled groups, with large effect sizes. Among the subtitled groups, no significant differences were revealed concerning participants' average fixation duration. The bilingual subtitles and captions groups spent similar total reading time % and fixation % on the subtitling area, which were both significantly longer than those in the L1 subtitles group, with small effect sizes. Table 5 also shows that the three subtitled groups had significant lower odds of skipping the subtitling area than the no subtitles group, with no statistical differences among the three subtitled conditions.

Table 4. Results for post-hoc contrasts for total reading time % and fixation % at level 1 Overall subtitling area

Group	<i>b</i>	95% CI	<i>SE</i>	<i>z</i>	<i>p</i>	<i>d</i>
Total reading time %						
Captions – Bilingual	0.03	[-0.01, 0.07]	0.02	1.11	.69	0.13
Bilingual > L1	0.11	[0.07, 0.15]	0.02	4.45	<.001	0.53
Bilingual > No	0.42	[0.38, 0.46]	0.02	17.61	<.001	2.25
Captions > L1	0.13	[0.09, 0.17]	0.02	5.39	<.001	0.68
Captions > No	0.44	[0.40, 0.48]	0.02	18.20	<.001	2.53
L1 > No	0.31	[0.27, 0.35]	0.02	12.49	<.001	1.81

Fixation %						
Captions – Bilingual	0.03	[-0.01, 0.07]	0.02	1.52	.42	0.18
Bilingual > L1	0.08	[0.05, 0.12]	0.02	3.74	.001	0.46
Bilingual > No	0.44	[0.40, 0.48]	0.02	19.92	<.001	2.53
Captions > L1	0.12	[0.08, 0.16]	0.02	5.09	<.001	0.66
Captions > No	0.47	[0.43, 0.51]	0.02	20.85	<.001	2.93
L1 > No	0.36	[0.32, 0.40]	0.02	15.38	<.001	2.21
Average fixation duration						
Captions – Bilingual	0.02	[-0.25, 0.29]	0.14	0.13	.99	<.001
Bilingual – L1	0.13	[-0.14, 0.40]	0.14	0.90	.81	0.003
Bilingual > No	4.65	[4.38, 4.92]	0.14	32.76	<.001	1.77
Captions – L1	0.15	[-0.14, 0.44]	0.15	1.00	.75	0.004
Captions > No	4.67	[4.38, 4.96]	0.15	32.01	<.001	2.56
L1 > No	4.52	[4.23, 4.81]	0.15	30.37	<.001	1.58

Table 5. Results for post-hoc contrasts for skip rate at level 1 overall subtitling area

Group	<i>b</i>	<i>SE</i>	<i>z</i>	OR	OR 95% CI	<i>p</i>
Captions – Bilingual	-0.57	0.35	-1.62	0.57	[0.28, 1.13]	0.37
L1 – Bilingual	0.04	0.36	0.95	1.40	[0.70, 2.81]	0.78
Bilingual > No	-6.26	0.35	-17.82	0.002	[0.001, 0.004]	<.001
L1 – Captions	0.90	0.37	2.45	2.47	[1.20, 5.10]	.07
Captions > No	-6.83	0.37	-18.65	0.001	[0.001, 0.002]	<.001
L1 > No	-5.92	0.37	-16.04	0.003	[0.001, 0.006]	<.001

Level 2 L1/L2 line area

The aim of this analysis was to explore the reading of the L1 and L2 lines within the bilingual subtitles group and to compare it to the captions and L1 subtitles groups.

Table 6 shows that when using bilingual subtitles, participants spent less time reading the L2 lines than L1 lines. Results of linear mixed-effects models (see Appendix S5 for model summaries) confirmed this difference for total reading time %, $b = -0.16$, $t(534) = -33.45$, $p < .001$, $d = 0.82$, fixation %, $b = -0.17$, $t(534) = -34.47$, $p < .001$, $d = 0.87$, and average fixation duration, $b = -1.89$, $t(534) = -55.03$, $p < .001$, $d = 0.63$, with small effect sizes. In addition, results of the generalised mixed-effects models for the skip rate (see Appendix S5 for model summary) also showed that the L2 lines were skipped more than the L1 lines in bilingual subtitles ($OR = 9.69$, 95% CI [8.73, 10.75], $p < .001$). Participants' vocabulary size did not significantly contribute to the difference in reading L1 and L2 lines in bilingual subtitles.

Table 6. Descriptive statistics for eye-movement data at level 2 for L1 and L2 lines in three subtitled groups

Subtitle lines	Total time %		Fixation %		Average fixation duration (ms)		Skip rate	
	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI
Bilingual L1 lines ($n = 28$)	0.42 (0.28)	[0.41, 0.42]	0.44 (0.26)	[0.43, 0.44]	174 (129)	[171, 176]	0.11 (0.31)	[0.10, 0.11]
Bilingual L2 lines ($n = 28$)	0.20 (0.24)	[0.20, 0.21]	0.22 (0.24)	[0.21, 0.22]	97 (117)	[95, 99]	0.43 (0.50)	[0.42, 0.43]
L1 subtitles ($n = 23$)	0.34 (0.27)	[0.34, 0.35]	0.39 (0.27)	[0.39, 0.40]	148 (148)	[146, 151]	0.18 (0.38)	[0.17, 0.18]
Captions ($n = 25$)	0.56 (0.32)	[0.55, 0.56]	0.59 (0.31)	[0.58, 0.59]	175 (93)	[174, 177]	0.11 (0.31)	[0.10, 0.11]

However, the above results should be treated with caution because they compared the reading behaviour of two different languages. Comparisons were then made for the different lines in bilingual subtitles with their corresponding lines in the monolingual subtitle groups. The reading of L2 (English) lines in the bilingual subtitles and captions groups was firstly compared. The descriptive statistics in Table 6 show that the captions group spent more time on L2 lines than the bilingual subtitles group, and this was statistically significant for total reading time %, $b = 0.39$, $t(49) = 6.98$, $p < .001$, $d = 1.24$, fixation %, $b = 0.26$, $t(49) = 7.15$, $p < .001$, $d = 1.36$, with large effects sizes, and for average fixation duration, $b = 1.97$, $t(52) = 7.96$, $p < .001$, $d = 0.57$, with a small effect size. Moreover, participants' vocabulary size negatively predicted their processing of the L2 lines in total reading time % and fixation %, but it was only significant for the captions group. The odds of skipping the L2 lines in the captions group were significantly lower than those in the bilingual subtitles group ($OR = 0.01$, 95% CI [0.003, 0.021], $p < .001$). Participants' vocabulary size also revealed a significant positive effect on skipping rate in the captions group (see Appendix S6 for model summaries).

Comparing the processing of L1 (Chinese) lines, results showed significantly longer total reading time %, $b = 0.06$, $t(49) = 2.65$, $p < .001$, $d = 0.28$, and average fixation duration, $b = 0.69$, $t(49) = 3.75$, $p < .001$, $d = 0.18$, on the L1 lines in bilingual subtitles than in the L1 subtitles group, with small effect sizes. However, no group difference was revealed in terms of fixation %, $\chi^2(1) = 2.48$, $p = .12$, $R^2 < .001$. In terms of skip rate, the odds of skipping the L1 lines in bilingual subtitles were significantly lower compared to the L1 subtitles group ($OR = 0.49$, 95% CI [0.26, 0.84], $p = .01$). Participants' vocabulary size did not show significant effects on participants' use of the L1 lines (see Appendix S7 for model summaries).

RQ3 – RELATIONSHIP BETWEEN ON-SCREEN TEXT PROCESSING AND COMPREHENSION

Linear regression models were constructed to determine whether comprehension scores were predicted by the degree of attention allocated to the L1 and L2 line areas in bilingual subtitles, captions, and L1 subtitles groups, as measured by the mean total reading time % and average fixation duration:

Model 1 Comprehension ~ Average Total Reading Time% * Subtitle Line in Groups

Model 2 Comprehension ~ Average Fixation Duration * Subtitle Line in Groups

Model 1 revealed significant main effects of total reading time % on comprehension scores, $b = -0.32$, $t(96) = -2.68$, $F(7, 96) = 5.05$, $p = .01$, $R^2 = .22$. The interaction between total reading time % and subtitling lines in different groups was also significant. Post-hoc analysis of this interaction only produced a significant effect in the captions condition, showing that longer total reading time % on captions led to lower comprehension scores, $b = -0.32$, $t(96) = -2.68$, 95% CI $[-0.55, -0.08]$, $p = .01$. As can be observed in Figure 3, negative effects were also revealed in the processing of L1 lines in the L1 subtitles condition, $b = -0.23$, $t(96) = -0.95$, 95% CI $[-0.72, 0.25]$, $p = .34$, and in the bilingual subtitles condition, $b = -0.37$, $t(96) = -1.84$, 95% CI $[-0.77, 0.03]$, $p = .07$, but neither of them reached statistical significance. Notably, although non-significant, only the total reading time % on the L2 lines in bilingual subtitles was positively related to comprehension scores, $b = 0.30$, $t(96) = 1.18$, 95% CI $[-0.20, 0.80]$, $p = .24$.

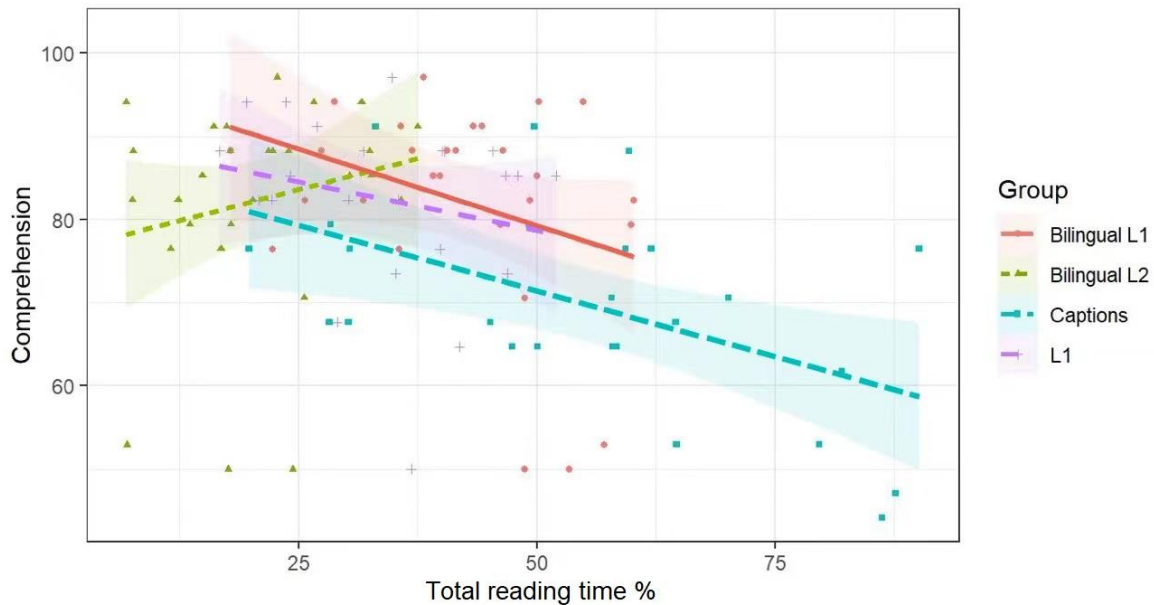


Figure 3. Relationship between mean total reading time percentage on the subtitling lines and participant's comprehension test accuracy. Shaded areas represent 95% confidence intervals.

Model 2 revealed no significant relationship between participants' average fixation duration and their comprehension scores, $b = -0.11$, $t(96) = -1.45$, $F(7, 96) = 3.51$, $p = .15$, $R^2 = .15$ (see Appendix S8 for model summaries).

DISCUSSION

RQ1 aimed at investigating the effects of bilingual subtitles on comprehension compared to other subtitling conditions. The comprehension test revealed that, in line with previous studies (e.g., Wang, 2019), bilingual subtitles were as effective as L1 subtitles, and both of them were significantly more beneficial than captions and no subtitles for facilitating comprehension. These findings also support previous studies showing the advantage of L1 subtitles over captions to facilitate comprehension (e.g., Markham et al., 2001; Pujadas & Muñoz, 2020). The presence of the L1, in either L1 subtitles or bilingual subtitles, seems to support comprehension. The effectiveness of bilingual subtitles also suggests that the redundancy principle (Chandler & Sweller, 1991) does not seem to apply in this L2 learning

context. The presence of the L1 translations seems to facilitate comprehension and the L2 written input simultaneously presented did not limit the benefits of the L1 lines. This could be attributed to participants' use of the different input sources in bilingual subtitles, which will be discussed in relation to RQ2.

These findings contradict results of studies that have reported no superiority of bilingual subtitles over captions on comprehension (e.g., Hao et al., 2021; Liao et al., 2020; Lwo & Lin, 2012). This discrepancy could be attributed to the different materials and research designs adopted by these studies. The bilingual subtitles used in Hao et al.'s (2021) research presented the L2 lines over the L1 lines, which could lead to participants' different use of the bilingual subtitles, affecting their comprehension. In the study by Lwo and Lin (2012) participants were interrupted during viewing and asked to answer interview questions about their attention allocation and understanding. Thus, their findings might not well represent learners' natural viewing processes. The non-significant findings reported by Liao et al. (2020) could be due to a potentially challenging comprehension test. Comprehension was assessed by a free recall test conducted in participants' L2, and performance in the test might have been affected by participants' L2 writing competence, as reflected in the relatively low comprehension scores.

RQ2 explored L2 learners' attention distribution during bilingual subtitled viewing and compared that to other subtitling conditions. Participants' visual attention to the subtitling and image areas was investigated via eye-tracking. The eye-tracking findings revealed that, when using bilingual subtitles, participants spent about 60% of their time on the overall subtitling area and 40% on images when the on-screen text was presented. In addition, more time was spent on the L1 lines (42%) than L2 lines (20%), and the L2 lines were more likely to be skipped. These findings showed that participants in bilingual subtitles relied more on the L1 lines, which could explain the higher comprehension over captions. This finding is

understandable in the context of viewing for comprehension. Participants were asked to watch the video for comprehension rather than language learning, therefore, participants relied more on the L1 for comprehension and enjoyment. However, these findings are different from those by Liao et al. (2020). First, participants in Liao et al.'s (2020) research spent less time reading the overall bilingual subtitling area (34%) than the images (64%). Second, Liao et al. (2020) found no significant difference between the processing of L1 (18%) and L2 (15%) lines in bilingual subtitles. It should also be noted that the participants in Liao et al.'s (2020) study demonstrated a within group variation in terms of their processing of the L1 and L2 lines in bilingual subtitles, with half of the participants spending more time on L1 lines and the other half spending more time on L2 lines. This could potentially be attributed to the within-subject design adopted by Liao et al. (2020), in which three out of four groups used bilingual subtitles immediately after using captions. This potential order effect might have influenced participants' use of the bilingual subtitles, which might not have accurately represented participants' natural viewing behaviour. This group variation was not observed in the present study. As noted in the literature review section, care should be taken when interpreting Liao et al.'s (2020) results due to the limited sample size ($N = 16$) and limitations in their research design.

Comparing the processing of bilingual subtitles to monolingual subtitles, it is interesting to note that, despite more on-screen text being presented in bilingual subtitles, participants did not spend more time on the overall subtitling area in bilingual subtitles than in captions. A similar finding was also reported by Liao et al. (2020). This indicates that participants did not process all the information available, but selectively attended to the information. Similar to Liao et al.'s (2020) findings, participants using bilingual subtitles spent shorter time on the L2 lines than the captions condition. In contrast to Liao et al.'s findings, participants also spent longer time processing the L1 lines in bilingual subtitles than

in L1 subtitles. Overall, there was a stronger reliance on the L1 lines in the bilingual condition. This explains why the redundant information provided in bilingual subtitles did not impede comprehension. Sweller (2005) argued that the best strategy to deal with redundant information is to ignore it. The L2 lines in bilingual subtitles, which were the written forms of the soundtrack and were less effective than L1 lines for comprehension, received less attention during viewing. The processing of the L2 lines seems to have been more selective, possibly reflecting automatic subtitle reading behaviour which happens just due to the presence of on-screen text (e.g., Bisson et al., 2014; d'Ydewalle et al., 1991), or reflecting participants' attempts to match the L1 translations to the L2 input, as has been documented in previous research (e.g., Author & Author, 2022; Lwo & Lin, 2012). Reliance on the L1 in bilingual subtitles facilitated comprehension, and selective attention to the L2 lines might also support the learning potential of subtitled viewing. As shown in Author and Author (2022), the presence of the L2 input in bilingual subtitles could better support the establishment of the form-meaning connection for vocabulary learning. Presenting the information in different formats allows learners to choose how they want to selectively use it to support comprehension and learning, supporting the learning preferences hypothesis (Mayer, 2009b).

Finally, RQ3 explored the potential relationship between participants' processing of the different subtitling lines and their comprehension. Higher percentage of total reading time on the L2 lines was significantly associated with lower comprehension scores in the captions group. This negative relationship was also documented in previous eye-tracking research, where longer processing time on written L2 input in multimodal materials was interpreted as a reflection of participants' processing difficulty, which was reflected in lower comprehension scores (e.g., Gass et al., 2019; Pellicer-Sánchez et al., 2020). However, the relationship between processing time and comprehension accuracy was not significant in either bilingual

subtitles or L1 subtitles conditions.

It is interesting to note that, although non-significant, when using bilingual subtitles, longer total reading time on the L2 lines showed a tendency to be related to better comprehension. One possible explanation is that the time spent on the L2 lines in bilingual subtitles could signal learners' extra cognitive capacity to process the L2 after obtaining sufficient understanding of the content. It could be the case that only when participants have understood the input, they could devote the remaining time and cognitive resources to process the redundant L2 lines in bilingual subtitles, which was then reflected in higher comprehension scores. Another possibility is that the longer processing time on the L2 lines might imply participants' language learning motivation, where some of them might have referred to the L2 lines in an attempt to learn language. Participants' processing of the L2 lines in bilingual subtitles could imply their attempts to match the L2 lines with L1 translations or with the L2 auditory input, which further facilitated their comprehension. However, this finding should be interpreted with caution due to the lack of statistical significance. Moreover, the above explanations should also be taken with caution due to the absence of verbal reports which could further explore participants' level of processing.

The percentage of total reading time on the L1 lines demonstrated negative but non-significant relationships with comprehension scores in both the bilingual and L1 subtitles groups. Previous eye-tracking studies in L1 have shown that longer processing time could signal more processing effort and seems to be more common when reading difficult texts (Rayner et al., 2009), and it also seems to be negatively related to L1 comprehension (e.g., Pellicer-Sánchez et al., 2021). However, it should be noted that this relationship did not reach significance in either group in the present research, suggesting a difference in reading and viewing studies.

In terms of the average fixation duration on the subtitling lines, no significant

relationships were reported in any conditions. Average fixation duration should be interpreted with caution in viewing research. The time-limited nature of on-screen text affects participants' reading behaviour and limits the maximum fixation duration, resulting in shorter average fixation durations compared to those in L1 text reading research (225–250 ms; Rayner, 1998), which might have accounted for the lack of significant results.

It is important to acknowledge the limitations of the present study. First, this research only focused on high-intermediate to low-advanced L2 learners who were also experienced users of bilingual subtitles. Thus, the findings might not be generalised to L2 learners of lower proficiency or who lack experience in using bilingual subtitles. More research is also needed to better understand the effectiveness of bilingual subtitles on L2 learning in other languages. Importantly, the bilingual subtitles used in the present study adopted the most common bilingual subtitles format in mainland China, presenting L1 lines on top of L2 lines. Future research should also explore whether presenting the L2 line on top of the L1 line would affect participants' viewing behaviour and comprehension performance. Thirdly, this study has only focused on one relative short documentary clip in the context of viewing for entertainment. Replication studies using audio-visual materials of different genres or lengths, and designed with different learning purposes are therefore needed. Lastly, the present study mainly focused on participants' processing of the on-screen text without taking into account the effects of images on comprehension. Although bilingual subtitles did not seem to hinder participants' processing of the images, follow-up interviews would be useful to further explore participants' experience of using bilingual subtitles and their underlying cognitive processes.

The results of the present study have important pedagogical implications. In the context of viewing for entertainment and comprehension, results have shown that bilingual subtitles do not lead to increased cognitive overload and that they seem to support comprehension.

Thus, bilingual subtitles might allow learners to engage with authentic material, increasing their exposure to L2 aural input (Webb & Rodgers, 2009), which is crucial for L2 development. It is true, however, that the stronger reliance on the L1 lines in bilingual subtitles does not seem to promote a focus on the L2 written form. However, the L2 auditory input is still available and the L1 lines can support learners in processing the L2 auditory input and making a connection between the L1 meaning and the L2 auditory forms. If bilingual subtitled videos are used for the purpose of language learning, it might also be useful to implement other techniques to direct learners' attention to the written L2 form.

CONCLUSION

This study provides a comprehensive investigation on the effects of bilingual subtitles for comprehension. The results of the present study showed that bilingual subtitles had an advantage over captions and no subtitles for facilitating comprehension and were as beneficial as L1 subtitles. The presence of the L1, either in L1 subtitles or in bilingual subtitles, supported comprehension. The eye movement data showed that the benefits of bilingual subtitles for comprehension were explained by the clear reliance on the L1 lines over L2 lines. When using bilingual subtitles, despite the presentation of more on-screen text, participants did not seem to spend more time processing the subtitling area than the captions group, but selectively used the bilingual subtitles to aid their comprehension, relying more on the L1 lines than L2 lines. This study provided further evidence to show that longer processing time on captions is related to lower comprehension scores. However, when the L1 input is also available (i.e., when using bilingual and L1 subtitles), there seems to be no relationship between the processing time on on-screen text and comprehension.

REFERENCES

Author & Author (2021).

Author & Author (2022).

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

Keens-Soper, A., Revill, B., Collins, P., Ord, L., & Laurie, K. (Executive Producers). (2013). *Animal odd couples* [TV series]. British Broadcasting Corporation.

BBC. (2019). Subtitle guidelines (Version 1.1.7). <https://bbc.github.io/subtitle-guidelines/>

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

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.

BNC Consortium. (2007). The British National Corpus (XML edition; Oxford Text Archive). <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.

Buck, G. (2001). *Assessing listening*. Cambridge University Press.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8, 293–332.

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

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28, 369–382.

- d'Ydewalle, G., Praet, C., Verfaillie, K., & Rensbergen, J. V. (1991). Watching subtitled television: Automatic reading behavior. *Communication Research, 18*, 650–666.
- Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta, 49*, 67–77.
- Dizon, G., & Thanyawatpokin, B. (2021). Language learning with Netflix: Exploring the effects of dual subtitles on vocabulary learning and listening comprehension. *Computer Assisted Language Learning Electronic Journal, 22*, 52–65.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE.
- Gass, S. M., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology, 23*, 84–104.
- Gesa Vidal, F. (2019). *L1/L2 subtitled TV series and EFL learning: A study on vocabulary acquisition and content comprehension at different proficiency levels*. (Unpublished doctoral dissertation). University of Barcelona, Barcelona, Spain.
- Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge.
- 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 methods, 3*, 339–353.
- Hao, T., Sheng, H., Ardasheva, Y., & Wang, Z. (2021). Effects of dual subtitles on Chinese students' English listening comprehension and vocabulary learning. *The Asia-Pacific Education Researcher, 31*, 529–540.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*, 346–363.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in

- linear mixed effects models. *Journal of Statistical Software*, 82, 1–26.
- Lee, M., & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: An eye-tracking study. *Modern Language Journal*, 102, 557–577.
- Lenth, R. (2020). emmeans: Estimated marginal means, aka least-squares means. (Version 1.4.5) [Computer software]. <https://CRAN.R-project.org/package=emmeans>
- 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, 70–98.
- Lüdtke, D. (2020). sjPlot: Data visualization for statistics in social science. (Version 2.8.4) [Computer software]. <https://CRAN.R-project.org/package=sjPlot>
- Lunin, M., & Minaeva, L. (2015). Translated subtitles language learning method: A new practical approach to teaching English. *Procedia - Social and Behavioral Sciences*, 199, 268–275.
- Lwo, L., & Lin, M. C.-T. (2012). The effects of captions in teenagers' multimedia L2 learning. *ReCALL*, 24, 188–208.
- Markham, P. L., Peter, L. A., & McCarthy, T. J. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals*, 34, 439–445.
- Mayer, R. E. (2009a). Multimedia principle. In R. E. Mayer (Ed.), *Multimedia learning* (2nd ed., pp. 223–241). Cambridge University Press.
- Mayer, R. E. (2009b). Redundancy principle. In R. E. Mayer (Ed.), *Multimedia learning* (2nd ed., pp. 118–134). Cambridge University Press.
- Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, 18, 118–141.

- 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.
- Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System*, *67*, 77–86.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*, 9–13.
- Nation, I. S. P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts (Version 3) [Computer software]. *Lextutor*. <https://www.lexutor.ca/cgi-bin/range/texts/index.pl>
- Navarro, D. J. (2015). Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.5) [Lecture notes]. School of Psychology, University of Adelaide, Adelaide, Australia.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.
- Paivio, A. (2014). Bilingual dual coding theory and memory. In R. Heredia & J. Altarriba (Eds.), *Foundations of bilingual memory* (pp. 41–62). Springer.
- Pellicer-Sánchez, A., Conklin, K., Rodgers, M., & Parente, F. (2021). The effect of auditory input on multimodal reading comprehension: An examination of adult readers' eye movements. *The Modern Language Journal*, *105*, 936–956.
- Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M. P. H., Serrano, R., & Llanes, Á. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. *Studies in Second Language Acquisition*, *42*, 577–598.
- Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, *53*, 1008–1032.
- Plonsky, L., & Oswald, F. L. (2014). How big is “Big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912.

- PortableSoft. (2012). SrtEdit (Version 6.3) [Computer software].
<http://www.portablesoft.org/srtedit-portable/>
- Pujadas, G., & Muñoz, C. (2020). Examining adolescent EFL learners' TV viewing comprehension through captions and subtitles. *Studies in Second Language Acquisition*, 42, 551–575.
- 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.
<https://www.R-project.org/>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2009). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10, 241–255.
- Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington, New Zealand.
- Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and vocabulary. *ITL - International Journal of Applied Linguistics*, 169, 191–211.
- Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, 45, 689–717.
- 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.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285.

- Sweller, J. (2005). The redundancy principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159–168). Cambridge University Press.
- Tremblay, A., & Ransijn, J. (2020). LMERConvenienceFunctions: Model selection and post-hoc analysis for (G)LMER models. R package (Version 3.0). <https://cran.r-project.org/web/packages/LMERConvenienceFunctions/LMERConvenienceFunctions.pdf>
- Wang, Y. (2019). Effects of L1/L2 captioned TV programs on students' vocabulary learning and comprehension. *CALICO Journal*, 36, 204–224.
- Webb, S., & Rodgers, M. P. H. (2009). The lexical coverage of movies. *Applied Linguistics*, 30, 407–427.
- 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.
- 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.
- Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System*, 35, 182–191.