12-13-2021 2:00 PM

# Learning New Words from TED Talks: Strategic use of L1 subtitles and L2 captions

Injung Wi, *The University of Western Ontario*

Supervisor: Dr. Frank Boers, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Arts degree in Education

# Abstract

In recent years, studies examining the effectiveness of audiovisual input and on-screen text for incidental vocabulary learning have proliferated. However, no studies have explored the potential of repeated viewing with an alternation of L1 subtitles and L2 captions for incidental vocabulary learning although both types of on-screen text have been proved to be beneficial for vocabulary acquisition. Given this gap in the literature, we designed the present study, the rationale for which was guided by the notion of desirable difficulty, the role of retrieval from memory, and conflicting findings regarding the benefits of trial-and-error learning. The research questions were whether using an alternation of L1 subtitles, L2 captions, and no onscreen text (henceforth "none") leads to greater vocabulary learning compared to using only L2 captions repeatedly and whether the sequence of the different kinds of onscreen text makes a difference to learning gains in the case of repeated viewing.

The participants ($N$ = 30) were upper intermediate to advanced ESL learners. They were randomly assigned to one of three conditions, which were watching a TED talk video three times with the sequence of 1) none-subtitles-captions ($n$ = 10), 2) subtitles-captions-none ($n$ = 11), and 3) captions-captions-captions ($n$ = 9). Eleven target words were selected from the video. A meaning recall test format was adopted for a pre-test, an immediate post-test, and a delayed post-test. The tests were administered at 1-week intervals. A listening comprehension test was administered after the first viewing to ensure the participants attended to the content of the TED talk video and vocabulary learning could be ascribed to incidental learning. A meaning recognition test was administered as part of delayed post-testing as well. Finally, a questionnaire elicited the participants' perceptions of the usefulness

of the different viewing sequences. The output of mixed-effects logistic regression analysis revealed that incidental vocabulary acquisition definitely happened through the repeated viewing, but no significant difference was found in the effectiveness of the three viewing conditions. That significance was not reached is unsurprising considering the small study sample. However, the descriptive statistics and the questionnaire responses suggested that using a sequence of subtitles, captions, and none may facilitate word learning at the meaning recall level compared to using captions only. The results thus call for more research on the merits of this sequence of viewing a video with decreasing support from onscreen text.

**Keywords**: repeated viewing, L1 subtitles, L2 captions, incidental vocabulary learning, TED Talks, desirable difficulty, retrieval and trial and error.

# Summary for Lay Audience

Researchers have focused on the potential benefit of audiovisual material such as TV shows, movies, and online video for vocabulary learning. They also have examined how using subtitles shown in the first language and captions shown in the second language contribute to learning. Studies have indeed found that second language learners can pick up new words from watching videos, and that using subtitles and captions can facilitate this learning process.

We found that it is common to play or watch the same video more than once both inside and outside L2 classrooms. However, no studies have examined whether using an alternation of subtitles and captions will be more effective than watching the same video more than once with the same type of onscreen textual support. Neither have any studies examined whether the precise sequence of using subtitles, captions or no onscreen text (henceforth "none") might make a difference. This exploratory study sought preliminary answers to these questions.

30 English as a Second Language (ESL) learners were assigned to one of three groups who were all requested to watch a TED Talk video three times, but under different conditions: 1) none-captions-subtitles, 2) subtitles-captions-none, and 3) captions-captions-captions. They were asked to take a test prior to watching the video, a test immediately after watching the video the third time, and a test one week later. The test was about 11 target words selected from the video. The participants were also asked to complete a listening comprehension test after the first viewing of the video. In addition, they completed a

questionnaire about their own perceptions of the benefits of repeated viewing and the usefulness of onscreen text at the end of the study.

The findings revealed that repeated viewing indeed fosters vocabulary learning. The findings also tentatively suggest that using a sequence of subtitles, captions, and none may better facilitate word learning compared to a sequence that using captions three times. The findings can inform language teachers and learners on how to use videos like TED Talks strategically for the purpose of vocabulary acquisition.

# Acknowledgments

I would like to express my deepest appreciation to my supervisor, Dr. Frank Boers. Without his expertise, the completion of this project could not have been possible. His support and guidance carried me through all the stages of conducting the project and writing the thesis. I would like to say that I was the lucky one having you as my supervisor.

I would also like to thank the professors and colleagues in the faculty of education at Western University who co-operated with me on the recruitment and data collection during this challenging time.

I would love to give special thanks to Yan (Lily) Su, a PhD student, Mengxue (Alyssa) Li, an MA student, in the faculty of education at Western University, and Mojgan Abbasi, a language instructor, at YMCA. Without their support, scoring the tests would have been impossible.

I am particularly grateful to my husband and my family in Korea and Canada as a whole for their continuous support and understanding throughout my academic journey.

Lastly, I would like to thank my dear son, Ethan, for being patient with me while working on the study.

# Table of Contents

# List of Tables

# List of Figures

# List of Appendices

# Chapter 1 Introduction

Developing vocabulary knowledge is a pivotal aspect of second and foreign language acquisition. However, it is a demanding task that requires a considerable investment of time and effort. Hence, a substantial amount of research on how to promote L2 vocabulary learning has been conducted for many years, yielding a great deal of pedagogical theories and implications for effective and efficient vocabulary acquisition. Research has stated that L2 vocabulary learning occurs when learners receive input of new words. That is, L2 learners are likely to acquire novel words when they are exposed to them and the more exposure to new words, the more probabilities of learning them (Webb & Nation, 2017). In addition, it has been claimed that it may be hard to acquire an adequate amount of vocabulary knowledge required to be able to understand spoken and written discourse in the target language by solely relying on deliberate vocabulary learning which often occurs for the sake of tests (Webb & Nation, 2017). Researchers, therefore, have endeavoured to explore potential sources of input that allow L2 learners to pick up new lexical items incidentally both inside and outside the classroom context. The sources of input found to be beneficial for incidental vocabulary learning are meaning-focused activities such as 1) reading (e.g., Horst et al., 1998; Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010; Webb & Chang, 2015), 2) listening (e.g., van Zeeland & Schmitt, 2013a), 3) reading while listening (e.g., Brown et al., 2008; Teng, 2018; Valentini et al., 2018), and 4) viewing audio-visual materials (e.g., Montero Perez, 2020; Neuman & Koskinen, 1992; Rodgers & Webb, 2011).

Among them, studies on reading as a mode of input for vocabulary acquisition have been the most common. In recent years, however, viewing has come into the spotlight in the realm of incidental vocabulary learning given the broad availability of online videos (YouTube, TED Talks, and Netflix, etc.) and an increasing amount of time spent by people viewing them. Zenith, a media agency, reported that people spent an average of 84 minutes a day viewing online videos globally in 2019. This figure is expected to continue to rise, reaching an average of 100 minutes per day by 2021 (Zenith, 2019). Additionally, audiovisual materials have been widely used in a classroom setting as a powerful language teaching tool because they can offer authentic spoken language input along with visual and contextual cues. These cues are found to assist learners' comprehension of the content and unfamiliar lexical items (Danan, 2004; Webb, 2015).

Empirical studies have revealed that audiovisual input indeed is beneficial for L2 vocabulary acquisition (e.g., Montero Perez, 2020; Peters & Webb, 2018; Puimège & Peters, 2019). Moreover, it has been reported that the benefits can be increased when audiovisual input is accompanied with on-screen text: L2 captions (e.g., Baranowska, 2020; Danan, 2004; Jelani & Boers, 2018; Majuddin et al., 2021; Markham, 1999; Montero Perez et al., 2014; Peters, 2019; Peters et al., 2016; Winke et al., 2010) and L1 subtitles (e.g., Danan, 1992, 2004; Frumuselu et al., 2015; Koostra & Beentjes, 1999; Kuppens, 2010). These findings have generated valuable pedagogical implications because L2 learners and teachers nowadays have easy access to the function of switching on and off captions and subtitles through multiple online video platforms (e.g., TED Talks, YouTube, and Netflix, etc.). It should be noted that, in the present project, captions refer to intra-lingual subtitles where L2 spoken discourse is shown in L2 written form,

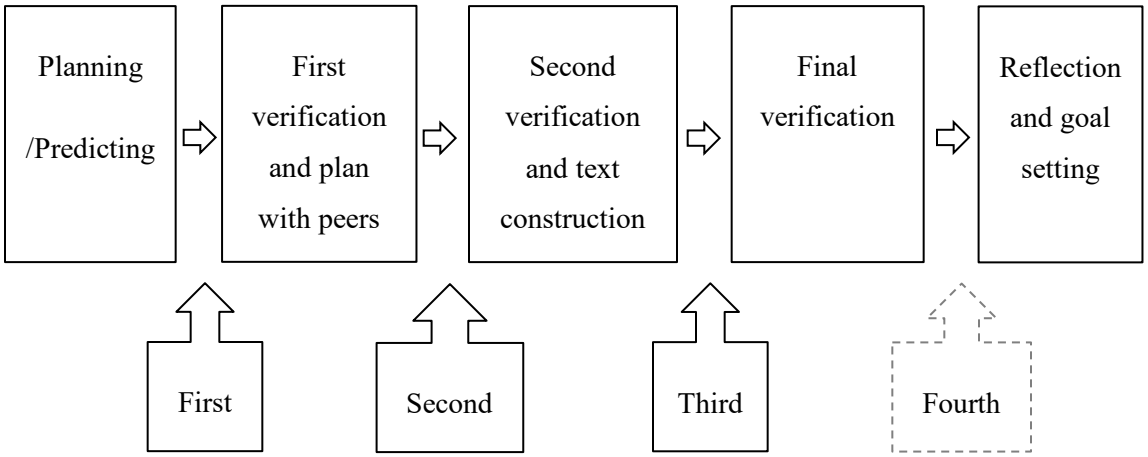
while subtitles refer to inter-lingual subtitles in which L2 spoken form is presented in L1 written form.

Most of the above studies either investigated respective effects of captions and subtitles or compared differential effects of them on L2 vocabulary development (e.g., Jelani & Boers, 2018; Peters, 2019; Peters et al., 2016; Montero Perez et al., 2014; Winke et al., 2010). However, little is known about how to use these potential language-learning aids strategically to make the best of them as an ensemble. That is, although previous research has revealed that both subtitles and captions can render the audiovisual input more comprehensible and facilitate L2 word learning, no studies have yet tackled the potential cumulative benefits of combining the use of captions and subtitles. Motivated L2 learners may watch the same video more than once with different types of on-screen text (when they are available) in the hope of aiding and verifying their understanding of the content. For instance, learners might watch a video with captions for the first viewing and with subtitles for the second viewing, and vice versa. Another possible scenario is that L2 learners watch an English video with subtitles to fully support their comprehension of the content from the beginning (especially when the input seems difficult), then watch again with L2 captions with the intention to visualize the L2 forms of what they heard, and finally watch with neither type of assistance in order to see if there is progress in their listening comprehension. Alternatively, others might choose to first watch the non-subtitled version to test and challenge their L2 listening ability without any assistance, then watch the subtitled version afterwards to confirm or rectify their understanding.

These scenarios then raise the question of whether the different sequences of adding / removing on-screen text will predict vocabulary acquisition. To the best of my knowledge, no studies have explored the potential of varied sequences of using subtitles, captions, and no on-screen text on L2 lexical development. There is one study by Winke et al. (2010) which investigated whether the different order of captions when viewing a video twice (captions and none versus none and captions) would predict vocabulary learning and found that it indeed affected leaners' performance on a test that asked them if they recognized the oral forms of the words they had been exposed to. However, this research did not include the option of L1 subtitles.

Playing audio or audiovisual materials more than once is frequently seen in L2 listening classes. Commonly, listening lessons comprise successive phases to facilitate leaners' development of L2 listening skills (Newton et al., 2018). One example is a listening lesson adopting the framework of a metacognitive pedagogical sequence suggested by Vandergrift (2004) (See Figure 1). This framework of a listening lesson constitutes phases of metacognitive processes, such as prediction, verification and evaluation with a listening phase in between them as seen in Figure 1. According to this lesson design, the same text is repeated at least three times. It is stated that repeated listening enables learners to "verify their understanding of the text as a way of gradually increasing their comprehension of the text content and control over their listening process" (Newton et al., 2018, p. 151).

**Figure 1** *A Metacognitive Pedagogical Sequence.*




(Adapted from Vandergrift & Goh, 2012, p. 109.)

Unquestionably, repetition is a crucial learning condition, and "the more repetitions there are the more likely learning is to occur" (Webb & Nation, 2017, p. 67). This repeated listening particularly involves verbatim repetition, where target words are encountered repeatedly in the same context (Durrant & Schmitt, 2010; Webb & Nation, 2017). This type of repetition reduces L2 learners' cognitive burden by familiarizing them with the context at the subsequent listening, thus freeing up attentional resources to recognize unknown words as ones they have encountered before. Consequently, the verbatim repetition promotes acquisition of at least the form of words, and the growing familiarity with the semantic context in which they occur may free up cognitive resources for learners to ponder the meaning of the words as well (Webb & Nation, 2017). Durrant and Schmitt (2010) conducted an empirical study to compare the effectiveness of single exposure, verbatim repetition and varied repetition (target words are encountered

repeatedly in different contexts) on retention of collocation knowledge. The target collocation items were 20 adjective-noun pairs. The participants in the single exposure condition saw the target items once, and the participants in the other two conditions saw them twice. In the test phase, all the participants were asked to recall the noun parts of the pairs by looking at their adjectives and the onset of the nouns. The results demonstrated that the learning gains in the verbatim repetition condition were significantly greater than the gains in the single exposure and varied repetition conditions, hence, proving the claim that verbatim repetition strengthens the memory of learning. A recent study by Majuddin et al. (2021) examined the benefits of repeated viewing (watching the same video twice). They compared the effect of single viewing and repeated viewing on learners' acquisition of multiword expressions (MWEs). The findings showed that repeated viewing led to better learning of MWEs than single viewing. Nonetheless, no empirical studies of L2 vocabulary acquisition through repeated viewing have involved more than two viewings, which is surprising after the above observation that watching the same audiovisual material three times or more is not uncommon in language courses.

Given the aforementioned areas remaining to be examined, the present research project aimed to explore whether 1) combining three possible options (subtitles, captions, and no on-screen text) generates better learning gains than conditions where only captions, subtitles, or no onscreen text are used, and whether 2) manipulating sequences of these options (e.g., sub.-cap.-none, cap.-sub.-none, none-cap.-sub.) makes a difference in acquiring new words when viewing a video three times. As will be explained further below, the actual study was for reasons of feasibility narrowed down to three repeated viewing conditions: sub-cap-none, none-cap-sub, and cap-cap-cap.

# Chapter 2 Literature Review

## 2.1 L2 Vocabulary Acquisition Through Viewing

Expanding one's repertoire of vocabulary is crucial for successful L2 mastery. To do so, it may be insufficient to purely resort to deliberate vocabulary learning. Although deliberate efforts to teach and learn unknown words leads to considerable and more immediate learning gains (Webb & Nation, 2017), a more expansive range of vocabulary knowledge can be acquired incidentally in the long run through consistent engagement in meaning-focused activities both inside and outside the classroom (Ellis, 1999; Nation, 2001; Schmitt, 2000; Webb & Nation, 2017).

The bulk of studies so far have examined reading as a useful source of written input for incidental L2 word learning (e.g., Horst, Cobb, & Meara, 1998; Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010; Webb & Chang, 2015). In recent years, however, researchers have focused on the potential of audiovisual input for L2 lexical development considering its prevalence in people's daily lives (e.g., OECD, 2009; Zenith, 2019) and the multiple advantages that it offers. One advantage of audiovisual material is that it is motivational and engaging (Baranowska, 2020; Baltova, 1994), which is a key factor in fostering L2 learning (Laufer & Hulstijn, 2001). Secondly, it provides a large amount of aural language input that is likely to be encountered in a real-life situation (Danan, 2004; Webb, 2015) and opportunities for repeated encounters with the same words, even ones of low frequency (Rodgers & Webb, 2011; Webb & Rodgers, 2009). Third, it presents visual and contextual clues that help learners infer meanings of unfamiliar words (Danan, 2004).

Empirical studies measuring the impact of audiovisual input on L2 vocabulary learning have almost consistently demonstrated that learners can incidentally acquire new words through viewing TV or video (e.g., Montero Perez, 2020; Neuman & Koshinen, 1992; Peters & Webb, 2018; Rodgers & Webb, 2020). The learning gains may be comparable with those from reading (Rodgers, 2013). A recent study by Montero Perez (2020) examined high-intermediate learners' incidental vocabulary learning from audiovisual input. Learners of French (*N*=63) were assigned into an experimental group and a control group. The participants in the experimental group watched a 25-minute French documentary twice without on-screen text. The material was manipulated to include 15 pseudowords. The control group only took the tests (pre-test and post-test) without viewing the documentary. The vocabulary gains were measured by means of a form recognition test, a meaning recognition test and a meaning recall test. The findings of the data analysis revealed that the experimental group markedly surpassed the control group on the form and meaning recognition tests. However, word learning did not happen at the level of meaning recall. Another recent piece of research in this vein, conducted by Puimège & Peters (2019), explored the effects of TV viewing on the incidental learning of single words and multi-word expressions (MWEs). The participants were 20 Dutch speaking learners of English whose English proficiency levels were intermediate. The material of the study was a 30-minute clip of a TV reality show. Unlike the other studies, a form recall test was adopted along with a meaning recognition test and a meaning recall test. The results surprisingly suggested that the learning of words and MWEs could happen at the level of form recall, which is regarded as a more challenging task than meaning recognition and recall (Nation, 2001), although it needs to be mentioned that the

first letter of the missing items was given on the test sheets as an extra cue. As for the results of the meaning recall test, no learning gains were found, which aligns with the findings of Montero Perez (2020). However, it should be noted that the findings of the study need to be cautiously evaluated given the small sample size.

While these two studies employed clips of a documentary and TV show, Peters and Webb (2018) designed a study to examine the effects of viewing a full-length TV program on L2 word learning. Thus, they selected a 1-hour BBC documentary on a topic that the participants, who were business students, would be interested in. The learning gains were measured by a form recognition test and a meaning recall test. It was stated that the results of the form recognition test were not valid as it turned out that learners recognized the forms of the target words through the pretest. Learning gains (albeit very modest ones) were detected at the meaning recall level. The experimental group gained an average of four out of 64 target words, whilst the control group gained only 1.5 words from the pre-test to post-test. Furthermore, Rodgers and Webb (2020) examined the effectiveness of extensive viewing on vocabulary learning. 187 Japanese university students watched 10 episodes (45 minutes each) of an English TV series. The participants' knowledge of the form-meaning connection of 60 words was measured in a pre- and post-test. The findings revealed that there was an increase in the mean scores by approximately six words from pre-test to post-test. Considering the extensive amount of viewing time, however, these learning gains seem rather marginal. It is speculated that addition of on-screen text might lead to more substantial learning gains.

## 2.2 The Use of On-Screen Text

More of the body of research into incidental vocabulary learning through viewing, in fact, has focused on whether addition of on-screen text (captions and subtitles) has a facilitative effect on lexical development. Although both types of on-screen text are known to yield more vocabulary learning gains compared to no on-screen text (Peters et al., 2016), the effects of captions have been investigated more extensively than subtitles. As already mentioned, captions or intra-lingual subtitles are presented in the same language as the soundtrack (Danan, 2004; Peters et al., 2016). They were originally invented for the purpose of aiding deaf and hearing-impaired people to access video content. In the 1980s, however, the use of captions was first adopted in foreign language classrooms with the intention to draw learners' attention, lower their affective filters, help learners immediately confirm what they heard, and promote motivation (Winke et al., 2010; Vanderplank, 1988). In fact, studies have suggested that captions help learners to visualize phonological forms of the sound rendering the input more explicit (Birds & Williams, 2002). This, consequently, creates a more precise memory trace of the words, subsequently helping learners to develop the ability to eventually identify the same sounds without textual assistance (Birds & Williams, 2002). Additionally, Vanderplank (1988) reported that captioned video assists learners in chunking the stream of speech, which in turn promotes vocabulary learning. The participants in his study commented that viewing with captions helped them consciously recognize new words and phrases and unfamiliar orthographies of proper nouns which would have otherwise been lost in the speech stream.

Montero Perez et al. (2013) conducted a meta-analysis to probe the overall efficacy of captioned video for L2 listening and vocabulary learning. They analyzed 15 studies for listening comprehension and 10 studies for vocabulary acquisition. The findings revealed the overall positive effectiveness of captions for both comprehension and vocabulary learning. A recent study conducted by Majuddin et al. (2021) compared the effects of three different caption conditions (no captions, normal caption, captions with parts underlined) on learners' uptake of multiword expressions (MWEs). The participants took a MWE form recall test. They found that both caption conditions led to better scores on this test than the viewing condition without captions. Another recent study, by Jelani and Boers (2018), compared the effects of viewing captioned and uncaptioned video on intermediate EFL learners' vocabulary uptake and the mediating effect of test modality (written or aural test prompts). The participants watched a TED Talk video twice; one group watched the captioned version, and the other group watched without captions. The results demonstrated that captions fostered the learning of new words in the aspect of meaning recall. Additionally, the test modality was proven to mediate the outcomes of the word meaning test, showing that the group in the captioned viewing condition obtained better scores especially when the test prompts were in the written modality. When test prompts were presented aurally, no advantage for captions emerged. In addition to investigating whether viewing with captions is more effective for vocabulary learning than viewing without captions, Winke et al. (2010) examined whether captioning with the first viewing is more effective than captioning with the second viewing and vice versa. The participants were native English speakers who were learning different foreign languages: Spanish, Russian, Arabic and Chinese. They were

required to translate the target words from the video into English, which is a meaning recall test. Half of the target words were presented aurally, and the other half were presented in written form. The findings of the study demonstrated that the participants who were exposed to the captioned video during the first viewing significantly outperformed those who were exposed to the captioned video during the second viewing on the aural vocabulary test. As for the written vocabulary test, the participants who saw the captioned video during the first viewing also did only slightly better than those who saw the captioned video during the second viewing, and the difference was not significant. The authors explained the findings in relation to the importance of attention in word learning. It was stated that the captions at the first viewing helped the learners notice words that they thought would be more important, leading them to pay more attention to those words at the second viewing. This is perhaps how the captions at the first viewing strengthened the participants' recognition memory of the word forms. It was also mentioned that this claim is aligned with Schmidt's *noticing hypothesis* (1990, 2010), which contends that conscious noticing is a fundamental stage required for L2 learning, particularly at the level of form.

Like captions, viewing with subtitles can promote both content comprehension and L2 vocabulary acquisition (Danan, 2004). As mentioned, subtitles provide L1 text together with L2 sound (Danan, 2004; Peters et al., 2016). Thus, viewing with inter-lingual subtitles may involve a higher level of processing by simultaneously interlinking associations of image, L2 sound, and L1 text (Danan, 1992, 2004). Some might find it more difficult to attend to the L2 soundtrack when watching subtitled video because the subtitles automatically attract a lot of attention and they can usually be relied on for

comprehension without a real need to listen to the L2 discourse. However, there is some evidence that automatic reading of subtitles does not necessarily impede the encoding of the soundtrack (Danan, 2004; d'Ydewalle & Pavakanun, 1997; Kookstra & Beentjes, 1999). Krugal et al. (2014) conducted an experimental study investigating visual attention distribution when watching a subtitled English university lecture and the impact of on-screen text on content comprehension and cognitive load. The participants were university students whose mother tongue was Sesotho. They were asked to watch a 14-minute video clip of a first-year psychology lecture in English. The first group watched the un-subtitled version, the second group watched the lecture with subtitles (Sesotho), and the last group watched it with captions (English). To measure their comprehension of the lecture, a comprehension test and self-report questionnaire were adopted. The comprehension test was administered twice: immediately after the intervention and two weeks later. The results of the comprehension test revealed that there was no significant difference in the immediate test, but the L1 subtitle group outperformed the other two groups in the delayed comprehension test. Therefore, it was tentatively suggested that the subtitles might be more beneficial for retaining acquired information. More interestingly, findings from the questionnaire showed that the subtitle group indicated the lowest comprehension effort levels and highest engagement levels. In contrast, the captions group reported the highest mental load.

Compared to studies on captions, fewer experimental studies have sought to address the potential of L1 subtitles in incidental vocabulary learning from viewing (Peters et al., 2016). One of the earliest studies, by Koostra and Beentjes (1999), investigated the effects of watching a TV program with subtitles (in Dutch) on young

learners' uptake of English words. The participants watched a 15-minute episode of an American documentary. The results showed that the children in the subtitled condition did significantly better at recognizing the English words than those in the condition without subtitles. This attested that reading subtitles did not hinder young learners from hearing the English words. Other studies pertaining to the use of subtitles mostly focused on comparing the differential effects of subtitles and captions. One of the most recent studies, by Pujadas and Muñoz (2019), investigated the effects of these three variables on adolescents' L2 vocabulary acquisition through extensive TV viewing: 1) type of on-screen text (captions versus subtitles), 2) type of instruction (pre-teaching versus no pre-teaching), and 3) the learners' proficiency level. The participants ($N$ = 74) were Catalan-Spanish bilingual secondary school learners of English. The learning gains were measured for two aspects of vocabulary knowledge: form and meaning recall. The results relating to the two types of on-screen text, captions versus subtitles, demonstrated that the group in the subtitled condition had a better performance at the meaning recall test than those in the captioned group, while no significant difference was found at the level of form recall between the two groups. Another recent study in this area, by Peters (2019), explored whether the viewing condition (captions, subtitles, and no on-screen text) would affect vocabulary learning from audiovisual input. Intermediate EFL learners of English ($N$ = 142) were randomly assigned to three experimental conditions: 1) captions 2) subtitles 3) no subtitles. The participants watched an 11-minute excerpt from a documentary twice. The findings revealed that the participants in the captioned condition outperformed those in the other groups (subtitles and no on-screen text) on both form recognition and meaning recall tests. Most recently, Baranowska (2020) carried out an

empirical study to compare the effects of subtitles and captions on cognitive load, comprehension, and incidental vocabulary learning when watching an English video. The participants were 63 Polish high school students whose English levels were intermediate. All the participants were asked to view a 12-minute clip of a TV series and they were assigned to one of three conditions: no on-screen text, English captions, and Polish subtitles. They took pre- and post- tests. The tests were designed to measure the participants' form recognition and meaning recall including 4 levels of prompts: 'I haven't seen this word before', 'I have seen this word before', 'I think it's related to the category ____', and 'I know this word. It means (answer in Polish or English)'. A text comprehension test and self-report questionnaire were used, too. For the level of cognitive load, the results demonstrated that the highest mental effort was made by the group which watched the non-subtitled video, whereas the lowest cognitive load was indicated by the group that watched the video with Polish subtitles. In terms of comprehension, the group that was exposed to Polish subtitles showed the best performance, whilst the non-subtitled group garnered the lowest scores. As for the results of the vocabulary tests, although it was reported that the caption group gained the most word items, vocabulary learning occurred in all three groups. It should be noted, however, that the pre-test scores of the subtitle group were the lowest, which indicates that the participants in this group were likely to have poorer prior vocabulary knowledge. This perhaps affected their low rate of vocabulary learning.

To sum up, it has been seen that incidental vocabulary learning can happen through viewing and its effectiveness can be enhanced by adding on-screen text (i.e., subtitles or captions). However, the existing research has either investigated the

respective effects of captions and subtitles relative to watching audiovisual material without onscreen text or compared the differential effects of captions and subtitles. That is, no studies to date have attempted to explore whether using a combination of both types of on-screen text would be more beneficial for vocabulary learning than using the same kind of on-screen text when the same video is viewed more than once. As both types of on-screen text are believed to have a facilitative effect on the learning of new words, it seems worth investigating their *cumulative* benefits. Furthermore, in case of combining different types of on-screen text in repeated viewings, the question of whether the order of adding/removing them would affect vocabulary acquisition has not been addressed. The only study which examined the ordering effects of on-screen text when viewing a video more than once was Winke et al. (2010). However, this study again looked at one kind of on-screen text, captions. Finally, as shown above, the findings pertaining to the differential effects of captions and subtitles are mixed and so remain inconclusive. This is perhaps due to the fact that there are various factors that differ from one study to another, such as learning contexts, learners' proficiency levels in English, type of audiovisual input (short or long and different genres), times of viewing (once or twice), etc. Given these gaps and inconclusive findings identified from the literature review, more empirical studies regarding the use of on-screen text are needed.

# Chapter 3 Research Questions

1. Does repeated viewing with a *combination* of L1 subtitles, L2 captions, and no onscreen text (none-sub-cap or sub-cap-none) better facilitate upper intermediate to advanced ESL leaners' incidental vocabulary learning compared to repeated viewing the same type of on-screen text (cap-cap-cap)?

2. Does the precise *sequence* of providing different on-screen text (none-sub-cap vs. sub-cap-none) during repeated viewing make a difference to upper intermediate to advanced ESL learners' incidental vocabulary learning?

3. Do the above conditions of providing on-screen text during repeated viewing differentially affect upper intermediate to advanced ESL learners' *long-term* retention of learned vocabulary knowledge (if any learning occurs)?

# Chapter 4 Theoretical Framework

The present research was guided by three theories from cognitive psychology and memory research. In this chapter, we describe the theories and their connection with the inquiries of the research.

## 4.1 Theory of Desirable Difficulty

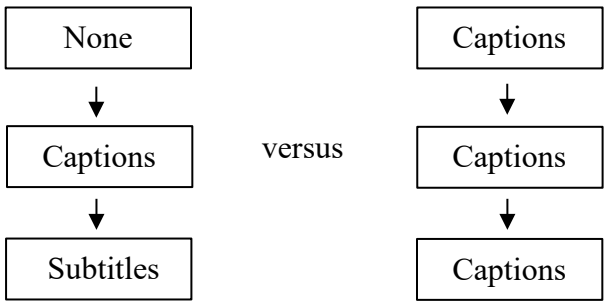
The claim of the theory is that when learners are facing difficulties that are manageable (desirable) while performing a task, the new knowledge is more likely to be stored in long-term memory and be transferable in real-life contexts (Bjork, 1994; Bjork, 2018; Schmidt & Bjork, 1992). Bjork (2018) emphasized that if the level of difficulty is beyond one's Zone of Proximal Development (ZPD) where learners are not able to "successfully respond to a given difficulty, it becomes an undesirable difficulty" (p. 147).

It is suggested that such desirable difficulties can be created by 1) varying the conditions of practice rather than keeping them consistent, 2) spacing study or practice sessions rather than massing, and 3) interweaving rather than blocking instruction of separate topics (Bjork, 1994; Bjork, 2018; Bjork & Bjork, 2011). These methods tend to pose challenges to learners. Thus, they are considered to yield "difficulties". These challenges are concurrently deemed "desirable" because they consequently foster long-term retention and transferability of learned knowledge by involving encoding and retrieval opportunities (Bjork, 1994; Bjork, 2018; Bjork & Bjork, 2011).

In light of these methods of creating desirable difficulties, the present study compares the effects of varied viewing conditions and consistent viewing conditions (See

Figure 1). Based on the theory, it is hypothesized that the varied viewing condition where the participants view a video with an alternation of captions, subtitles, and no on-screen text will promote learning because it entails the variation associated with desirable difficulties, whereas viewing a video several times with the same kind of on-screen text does not.

**Figure 2** *Example of Varied Viewing Condition and Consistent Viewing Condition*

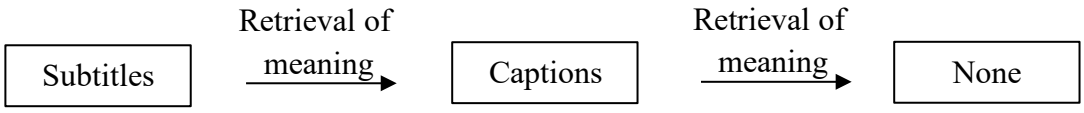



## 4.2 Retrieval Versus Trial and Error

The benefits of retrieval have been demonstrated in the domain of second language acquisition and memory research (e.g., Karpicke et al., 2014; Roediger & Butler, 2011; Strong & Boers, 2019). Retrieval is also referred to as the testing effect (Roediger & Karpicke, 2006) because retrieval happens when learners try to test their memory for previously learned knowledge. Researchers have argued that a successful retrieval episode can be more powerful than an extra study opportunity, specifically in promoting long-term memory retention (e.g., Hogan & Kintsch,1971). Barcroft (2007) conducted an empirical study on the effect of retrieval on L2 vocabulary learning. The participants in the retrieval condition first studied L2 words through word-picture pairs. After the study session, they received the pictures without their paired words, and had to

recall a word that corresponded to each picture. In the comparison group, the participants re-studied the same word-picture pairs after the first study session without the retrieval process. The outcomes of the post-test illustrated that the retrieval group outperformed the re-study group. Similarly, the sequence of on-screen text: subtitles-captions-none in this study entails opportunities to retrieve the meaning of L2 words, as subtitles (meaning) are shown at the first viewing (See Figure 2). Superior learning gains from this varied sequence could thus be explained with reference to the retrieval effect.

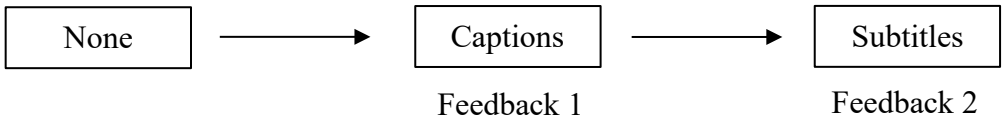
**Figure 3** *Demonstration of the Sequence Involving Retrieval of Meaning*




In contrast, a trial-and-error approach reflects the belief that we learn from our mistakes. Thus, under the learning condition of trial and error, learners attempt to find an answer based on their prior knowledge and intuition, and this is followed by feedback. Learning then happens through receiving confirmation of correct guesses and through the corrective feedback on wrong (or only partially correct) guesses (e.g., Strong & Boers, 2019). With this procedure in mind, it can be said that the condition of viewing first with no subtitles, then with captions, and last with subtitles reflects the trial-and-error procedure (See Figure 3). That is, participants need to infer the meaning of unfamiliar words in the first viewing as no subtitles are provided, possibly refine their hunch in the second viewing (perhaps assisted by the captions), and the subtitles at the third viewing

then serve as feedback (confirming, finetuning or rectifying the earlier hunches), similar to in trial-and-error learning.

**Figure 4** *Demonstration of the Sequence Involving Feedback*



# Chapter 5 Method

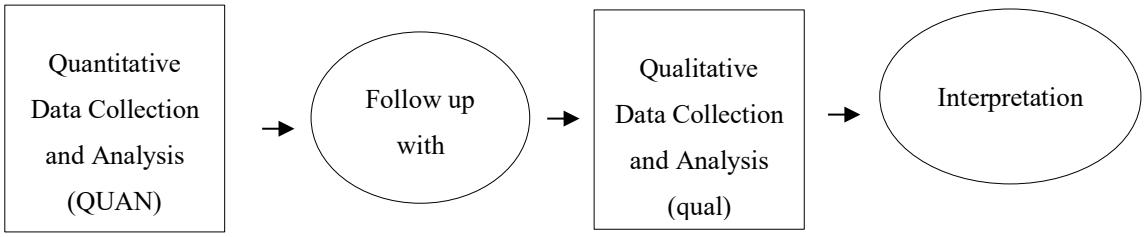
## 5.1 Research Design

A mixed methods design was adopted involving both quantitative (e.g., test scores) and qualitative (e.g., responses from the questionnaire and participants' notes during watching the video) research elements to gain expanded understanding of the research questions as described in the following definition of mixed methods research:

> Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e. g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration. (Johnson et al., 2007, p.123)

In the present study, the quantitative data were the primary components in answering the research problems, and the qualitative data provided supportive information which helped us further explore the participants' thoughts and prior experiences regarding *repeated* viewing and the use of on-screen text. For this reason, an explanatory sequential mixed methods design was employed (see Figure 5).

**Figure 5** *Explanatory Sequential Mixed Methods*




(Creswell, 2014, p.220)

It is a common strategy for mixed methods design to put emphasis on quantitative data (Creswell, 2014). This is highlighted by marking "QUAN" in upper case whereas "qual" is in lower case. Following the framework above, the vocabulary learning gains from the intervention were measured in a quantitative manner through a pretest, an immediate posttest, and a delayed posttest and analyzed. It was followed by the collection and analysis of short interviews with some of the participants, the participants' notes taken during the viewing, and the questionnaire. Subsequently, the two forms of data were merged in an interpretation phase. The pieces of information obtained from the qualitative data were embedded where it could support the quantitative data. In addition, the interpretation of the results as well as the formation of the research questions was guided by the theories mentioned in the Theoretical Framework section. Whether research is grounded on a broader theoretical viewpoint is another factor to be considered when designing a mixed methods research (Creswell, 2014).

## 5.2 Participants

It should be mentioned that our original plan was to recruit around 100 participants from the MPEd TESOL program at Western University in Canada. Due to the Covid-19 pandemic, however, enrolment in this program was greatly reduced by the time the recruitment took place (throughout the summer and fall term of 2021). There were 3 sections with about 20 to 25 students in each section whereas there used be 6 or 7 sections with about 25 to 30 students during the pre-pandemic period. Moreover, all the classes were delivered online during the summer term of 2021, which made the recruitment more challenging. As of the fall term, students started to take on-site or hybrid classes. Consequently, 34 learners of English were initially recruited within and outside the MPEd TESOL program through Zoom or in person. They were either living in China or Canada. There were participants ($n$ = 4) who did not complete both immediate and delayed posttests, and their data were excluded from the analysis. It should be noted that data from participants who had missed only the pretest or one of the posttests were retained in the analysis. As a result, data from a total of 30 participants were analyzed. The majority of them ($n$ = 23) were students from the MPEd TESOL program at Western University. The rest were former MPEd TESOL graduates ($n$ = 2), a first-year MA student ($n$ = 1) and PhD students ($n$ = 3) in the field of Applied Linguistics, and a graduate with a B.A. ($n$ = 1) at Western University. The MPEd TESOL students randomly received one of the three experimental conditions based on their section numbers, but the other seven participants were randomized into one of the three conditions (e.g., none-cap-sub, $n$ = 10, sub-cap-non, $n$ = 11, and cap-cap-cap, $n$ = 9). Except for two participants whose L1 was Korean and Farsi, Chinese was L1 for the rest

of the participants. As for English language proficiency, the participants are likely to be upper intermediate to advanced because they must obtain a minimum overall score of 6.5 on the IELTS test to enter Western University. This score falls into the borderline of B2 (upper intermediate) and C1 (advanced) in accordance with the Common European Framework of Reference (CEFR). To ascertain that the participants of the three conditions had similar prior vocabulary knowledge, the Updated Vocabulary Levels Test (Webb et al., 2017) was administered (see Table 1).

**Table 1** *Descriptive statistic of the Updated Vocabulary Levels Test (UVLT)*

| | 3000 | 4000 | 5000 | Total |
|---|---|---|---|---|
| Condition | *M (SD)* | *M (SD)* | *M (SD)* | *M (SD)* |
| None-cap-sub (*n* = 7) | 29.6 (0.53) | 27.7 (2.50) | 26.9 (4.10) | 84.1 (6.49) |
| Sub-cap-none (*n* = 9) | 28.9 (0.92) | 26.6 (1.74) | 26.7 (2.45) | 82.1 (4.40) |
| Cap-cap-cap (*n* = 5) | 28.4 (1.81) | 27.2 (1.48) | 26.6 (4.87) | 82.2 (7.25) |

*Note*: Max for each word level = 30.

Prior to looking at the descriptive data in the table, it should be mentioned why there are discrepancies in the number of the participants (21 out of 30). This is because it was administered in a voluntary manner. They were not asked to take the test during the study session. Instead, they were asked to complete the electronic version of the test on their own time by the end of the last session of the research. This decision was made with ethical consideration to help reduce the participants' fatigue during the study session as the session started immediately after their three-hour online or onsite classes. Moreover,

most of them who took the online courses were in China, so it was very late when the classes were finished (between 11:30 pm and 12 am) due to the time difference. Despite the missing data, we can estimate the participants' prior vocabulary knowledge from looking at the table. It demonstrates that the mean scores of each group are around 29/30 (or a bit lower) at the 3,000 level and 26/30 (or a bit higher) at the 4,000 and 5,000 levels. This indicates that most of the participants are likely to have acquired receptive knowledge of vocabulary at each level with reference to the cut-off point suggested by Webb et al. (2017) which is 29/30 (96.67%) for the 3,000 word level, and 24/30 (80%) for the 4,000 and 5,000 word levels. It is worth noting that the cut-off point for the 3,000 level is rather strict compared to 26/30 (86.66%) of the Vocabulary Levels Test (VLT) designed by Schmitt et al. (2001). A one-way ANOVA did not demonstrate a significant difference in the means scores of the three groups, $F(2, 8.8) = 0.24$, $p = 0.79$.

## 5.3 Input Material

The audiovisual material has been selected with two main criteria in mind: length and the number of target words. First, considering the fact that the study includes three viewings, it was important to choose a relatively short video to maintain the participants' attention throughout the repeated viewings. Thus, videos longer than 10 minutes were excluded. Secondly, the video had to contain a plausible number of words that were as yet unknown to the participants. In this regard, the participants' English proficiency levels were taken into consideration. As previously mentioned, the participants were academic ESL learners whose English proficiency levels were assumed to be upper

intermediate to advanced. Thus, videos that contain sufficient vocabulary beyond the 4,000 most frequent word families were considered as candidates for the material.

Given these considerations, a 5-minute and 16-second long video from Ted Talks entitled *A brie(f) History of Cheese* was selected for the study: https://www.ted.com/talks/paul_s_kindstedt_a_brie_f_history_of_cheese?utm_campaign=tedspread&utm_medium=referral&utm_source=tedcomshare. Analysis of the lexical profile of the video demonstrates that the video is comprised of 688 tokens and 89.7% and 92.5% lexical coverage was reached with the 4,000 and 5,000 most frequent word families respectively (see Table 2) as indicated by the VocabProfile tool in Lextutor (Cobb, n.d.). van Zeeland and Schmitt (2013b) reported that it is possible for L2 learners to understand spoken discourse with 90% lexical coverage although 95% coverage ensures more accurate listening comprehension. Additionally, Nurmukhamedov (2017), who analyzed 400 transcripts of TED Talks, suggested that knowledge of the 4,000 most frequent word families suffices to reach 95% lexical coverage. In this sense, it can be said that the difficulty of the given material was appropriate because it was found that most of the participants had vocabulary knowledge at the 5,000 level. This point was further supported by the participants' responses to the question on the difficulty level of the video in the questionnaire (see further below). The presented scales of the difficulty level were 1) very difficult, 2) difficult, 3) somewhat difficult, 4) appropriate, and 5) easy. Twelve out of 18 respondents (66.67%) reported that it was somewhat difficult, and 5 of them (27.78%) thought that it was appropriate. There was one participant (5.56%) who responded that it was difficult.

**Table 2** *Lexical Profile of the Input*

| Frequency level | Tokens | Cumulative tokens |
|---|---|---|
| 1k level | 66.1% | 66.1% |
| 2k level | 13.8% | 79.9% |
| 3k level | 6.7% | 86.6% |
| 4k level | 3.1% | 89.7% |
| 5k level | 2.8% | 92.5% |
| 6k level | 1.2% | 93.7% |
| 7k level | 0.6% | 94.3% |
| 8k level | 0.4% | 94.7% |
| 9k level | | |
| 10k level | 0.4% | 95.1% |
| >10k level + off-list words | 4.9% | 100% |

The rationale for choosing a video from TED Talks rests on several considerations. First, popularization of TED Talks renders this input material proper and authentic for the purpose of the study. As of September 2019, it had over 10 million subscribers and over 1.5 billion views. Furthermore, the videos have globally and

frequently been incorporated in ESL/EFL/EAP classroom settings as listening materials (e.g., Jelani & Boers, 2018; Nguyen & Boers, 2019; Takaesu, 2013). It is often considered that videos from TED Talks are suitable for L2 for academic purposes courses because they often involve scientific terms and presentation skills. In fact, the videos on TED Talks were originally intended for the general public and not for specific audiences, as implied in the slogan of TED Talks, *Ideas worth spreading*. Scotto di Carlo (2014) stated that the purpose of TED Talks is to convince non-expert audiences to 'make a change' as well as to pass on knowledge to them by dealing with issues that are closely linked with our daily concerns, aims and interests. Thus, they provide videos on a wide range of topics as indicated by the title, Technology, Entertainment, and Design (TED). This means that English learners can enjoy the videos selectively depending on individual interests. Therefore, all types of English learners, provided they have knowledge of at least the 4,000 most frequent word families according to the study of Nurmukhamedov (2017), will benefit from viewing TED Talk videos in terms of expanding their intellectual capacity as well as improving their English proficiency.

Secondly, the presenters of TED Talks often employ a variety of effective communicative strategies, such as explanations of terms, paraphrases, comparisons, exemplifications, analogies, and visuals as part of their endeavours to breach a barrier between the experts and lay audiences (Scotto di Carlo, 2014). Thus, watching the videos has the potential to positively affect L2 learners' development of effective communication skills. It is well-known that communication ability is essential in a wide range of contexts, including academic, occupational and everyday settings. Hence, it can be said that the audiovisual content provided by TED Talks is of great value.

Thirdly, TED Talks are freely available and short enough to be incorporated in educational settings and watched without extensive time commitment anywhere and anytime through any available devices.

Lastly, the function of switching on and off captions and subtitles is provided in over 40 languages. As mentioned in the Literature Review section, research has shown that captions and subtitles can scaffold L2 learners' listening comprehension and vocabulary uptake. Particularly, these scaffolding tools have the potential to facilitate self-directed learning outside the classroom context, too.

## 5.4 Target Vocabulary

Eleven target words were selected from the audiovisual input in light of the analysis of the lexical profile. The words belong to the 4,000 frequency level or beyond (see Table 3). Most of the words appear once except for *monastery* and *clump* which appear twice and *to coagulate* which occurs three times throughout the video. The mean score on the pre-test of all participants was 2.07 (*SD* = 1.69), indicating that many participants already knew 2 of the 11 target words.

**Table 3** *List of Target Words*

| Word | Frequency level | Frequency of occurrence | Part of speech |
|---|---|---|---|
| commodity | 4k level | 1 | noun |
| legacy | 4k level | 1 | noun |
| monastery | 5k level | 2 | noun |
| ration | 5k level | 1 | noun |
| famine | 6k level | 1 | noun |
| clump | 6k level | 2 | noun |
| staple | 6k level | 1 | noun |
| delicacy | 7k level | 1 | noun |
| to stockpile | 8k level | 1 | verb |
| domesticated | 8k level | 1 | adjective |
| to coagulate | 12k level | 3 | verb |

## 5.5 Procedures

Table 4 illustrates the procedures of the data collection of the research. There were three study sessions over a 3-week period. The data were collected either on-line (Zoom) or in on-site classrooms due to the nature of how classes were being delivered during the pandemic as mentioned above. The option to participate in the study sessions through Zoom was given to the students who were taking on-site classes as well to accommodate their preferences. Thus, some of them participated through individual Zoom sessions, and others participated in a hybrid manner (e.g., the first session in person and the subsequent sessions through Zoom). Providing the participants with this flexibility in selecting the venue was a strategy for increasing the rate of participation and reducing the rate of withdrawal during the challenging time of recruitment and data collection. During on-line data collection, it was ensured that all the participants turned on their cameras in all sessions to make it resemble the on-site setting and to have more control over the experiment (e.g., making sure that the participants watched the video and did not use a dictionary during the tests). The participants completed an electronic version of the tests and the questionnaire created on Western Qualtrics (see further below for descriptions of these instruments). They accessed these instruments through the given links or QR code on their own devices such as smart phones, tablet pcs, and laptops. The Ethics Approval Notice can be found in Appendices along with all the other related documents indicated in Table 4.

**Table 4** *Data Collection Procedures*

| Session | Stage | Time (minutes) | Instruments/ Activity | Procedure |
|---|---|---|---|---|
| I | 0 | 25 | (1) Letter of Information<br><br>(2) Consent form | The Letter of Information and the Consent Form was given to the potential participants. At this stage, they were told that the study was about development of listening skills. However, they were told the real purpose of the study at stage 7 through the debriefing letter. Only the participants who agreed to participate in the study proceeded to the following stages. Based on their course section, they randomly received one of the three experimental conditions. The participants recruited from outside of the MPEd |

| | | | |
|---|---|---|---|
| | | | TESOL program were randomized to one of the three conditions. |
| | 1 | 20 | (3) The Updated Vocabulary Levels Test | The participants in all three groups were asked to complete an electronic version of the Updated Vocabulary Levels Test on their own time before the last study session. The test measures the participants' receptive vocabulary knowledge up to the 5,000 level. The participants completed 3 levels from 3,000 to 5,000. The range of the time to complete the test was between 15 and 20 minutes. |
| | 2 | 20 | (4) Pre-test | Participants in each group were asked to complete an electronic version of the pretest that consists of a combination of the target words and filler words. This activity took about 15 to 20 minutes. |

| | | | | |
|---|---|---|---|---|
| II | 3 | 30 | (5) Treatment: Viewing<br><br>(6) Listening Comprehension Test | Participants in all groups watched the selected video three times under different viewing conditions. After the first viewing, they were asked to take a listening comprehension test followed by the second and third viewings in order. This process took approximately 30 minutes. |
| | 4 | 20 | (7) Immediate Post-test | At this stage, the participants were asked to complete the immediate post-test (a meaning recall test), which took approximately 15 to 20 minutes. |
| III | 5 | 30 | (8) Delayed Post-test | The delayed post-test included two tests: meaning recall and meaning recognition. The participants took the meaning recall test before the meaning |

| | | | |
|---|---|---|---|
| | | | recognition test. This stage took around 25 to 30 minutes. |
| | 6 | 15 | (9) Questionnaire | The participants were asked to complete an online questionnaire at their convenience. The questionnaire was about their viewing experiences and opinions on using on-screen text for both listening comprehension and vocabulary learning. It took about 15 minutes. |
| | 7 | 5 | (10) Debriefing Letter | Lastly, all the participants received a debriefing letter that explains the real purpose of the study and contains contact details and related references. |

# 5.6 Research Instruments

## 5.6.1 Meaning Recall Test

As a means of measuring vocabulary learning gains for the study, we adopted a meaning recall test informed by existing research (e.g., Jelani & Boers, 2018; Peters, 2019; Peters & Webb, 2018). In a study design including pre- and two post-tests, the outcomes of a form recognition test might not be valid due to a test taking effect, where participants recognize response options that they were shown as part of a pre-test battery (e.g., Peters & Webb, 2018). Thus, this type of test was not employed here.

The test consisted of 11 target words and 9 random words (see Appendix E). These random words did not appear in the video. They were from the most frequent 1,000 to 3,000 word families, so they were likely to be known to the participants. This was to help the participants maintain their motivation during the pre-test. Otherwise, they might find it frustrating as most target words were unknown to them. The test items were provided in a combination of written and aural prompts. This was to ensure equal input-modality—test-modality congruency for all three groups. Research has suggested that tests including only written prompts give an advantage to participants who have been exposed to captions owing to test-modality congruency, while no such advantage exists when aural test prompts are used (Jelani & Boers, 2018; Sydorenko, 2010). For the written prompts, the participants were asked to read a given word and write its meaning in their L1, an English synonym, or a definition. For the aural prompts, the participants were asked to listen to each word twice and type its spelling and meaning.

### 5.6.2 Meaning Recognition Test

The meaning recognition test was a multiple-choice format (see Appendix E). Only the target items were included in the test. The multiple-choice questions consisted of one correct L2 definition, three distractors, and an *I don't know* option.

The meaning recognition test was only employed at the stage of the delayed posttest. It was incorporated in case of a floor effect in the scores on the meaning recall test. A fair amount of attrition was expected to happen in the interval between the intervention and the delayed post-test. Moreover, meaning recall tasks are very demanding. Adding a recognition task (which is less demanding), therefore, was a way of capturing potential between-group differences where scores on the more challenging recall test might be too low for a difference to be noticeable. The meaning recognition test was administered after the meaning recall test to avoid a test-taking effect (where seeing the correct meaning option in the recognition test—a multiple-choice test—helps recall of that meaning later on).

### 5.6.3 Listening Comprehension Test

We incorporated a listening comprehension test to make certain that the participants would pay attention to the content of the video rather than focus just on the language itself (e.g., unknown words). Recall that this study is situated in the realm of incidental vocabulary acquisition. Prior to viewing, the participants were informed that there would be a listening comprehension test after the first viewing.

The test comprised 6 questions (see Appendix G). There were two short-answer questions, three true/false questions, and one multiple-choice question. All the questions

were about the content of the video. The questions were designed to assess the construct of listening for details. None of the target words were included within the questions and prompts. The test was administered once after the first viewing.

### 5.6.4 The Updated Vocabulary Levels Test

Learners' prior vocabulary knowledge has been reported as one of the individual differences that is positively correlated with vocabulary learning (e.g., Horst et al., 1998; Peters & Webb, 2018; Webb & Chang, 2015). Hence, the Updated Vocabulary Levels Test (Webb et al., 2017) was administered to ascertain that the participants in the three conditions had similar vocabulary knowledge (see Appendix D).

The Updated Vocabulary Levels Test is designed to measure test takers' knowledge of the form-meaning connections of words at the 1,000, 2,000, 3,000, 4,000 and 5,000 most frequent vocabulary levels. The word items for the test were adopted with reference to Nation's (2012) BNC/COCA word lists. Each level consists of 10 boxes with 3 items, which tallies to 30 items per level. One appealing feature of the test is that it can be administered as a whole, completing all levels, or it can (to save time) be done selectively in accordance with the test takers' current levels. For the current research only knowledge of the 3,000, 4,000, and 5,000 levels were measured given the participants' English proficiency level (with IELTS scores of at least 6.5 already before they entered the English-medium university).

### 5.6.5 Questionnaire

We developed a questionnaire to have a better understanding of the participants' perspectives on the use of on-screen text in repeated viewing (see Appendix H).

This was integrated as a follow up to the intervention (repeated viewing with on-screen text) and the tests in the final data-collection session. The questionnaire consisted of 17 questions, 12 of which were closed questions, and 4 were open-ended questions. It was completed in a voluntary and anonymous manner.

# Chapter 6 Data Processing

A meaning recall test in which participants need to write meanings either in L2 or L1 by looking at or hearing L2 words is regarded as a more difficult task than a meaning recognition test where participants need to choose a correct meaning among options. (i.e.,a multiple-choice test). For the meaning recall test, we therefore adopted a lenient scoringmethod test following Sydorenko (2010) and Jelani and Boers (2018). That is, 1 point for a correct response, 0.5 point for a response that signified a meaning closely related to a target item, and 0 points were awarded. For instance, the answer 'monk' in response to 'monastery' (a building occupied by a community of monks living under religious vows) received 0.5 as it indicates that the participant grasped a strongly associated meaning of monastery. This decision can be supported by Nation (2001) stating that association (what other words come into mind when seeing/hearing the word) is one aspect of knowing the meaning of a word. Because many of the L1 Chinese participants used their L1 to write the meaning of the target words on the test, two Chinese speakers (an MA anda PhD student at Western University) were recruited as a translator and second rater. Each of them participated in half of the scoring process. A native Farsi speaker who is a current ESL instructor in Canada was recruited as well to score the one Iranian participant's tests. The Korean participant's tests were scored by the student researcher (anative Korean speaker) and the Chinese PhD student. The process was guided by an answer sheet created prior to scoring in collaboration with the Chinese PhD student (see Appendix I). Any controversial responses were discussed by the two raters until agreement was reached. Responses in the

meaning recognition and comprehension tests were scored dichotomously with1 point for a correct response and 0 points for an incorrect response.The “I don’t know” option in the meaning recognition test was scored as an incorrect answer. As for the responses of the closed questions on the questionnaire, they were analyzed and reported by the default setting on Western Qualtrics. This automatically generated report provides information like frequencies, descriptive data, and visualizations of the results. For the open-ended questions, the responses were analyzed in a qualitative manner.

# Chapter 7 Analysis

We adopted a mixed model or mixed-effects model to examine whether, according to the meaning recall tests, there was a difference in vocabulary learning gains and retention rate among the three conditions of repeated viewing with 1) none-sub-cap, 2) sub-cap-none, and 3) cap-cap-cap. We selected a mixed model over a repeated measures ANOVA because recent research has reported that the former model is more robust and offers more advantages than the latter (Cunnings, 2012; Linck & Cunnings, 2015). One advantage is that mixed models enable us to account for variance associated with not only predictors of interest (fixed effects) but also random effects such as participants and items (Speelman et al., 2018). Another advantage is that mixed models are better able to process an unbalanced data set whereas ANOVA needs a balanced data set without missing cells (Linck & Cunnings, 2015; Speelman et al, 2018). Mixed models suppose that data are randomly missing, and so do not require complex imputation techniques to substitute the missing cells (Qeunce & van der Bergh, 2004, 2008). Given that the data set of the present study is incomplete with some missing responses for the none-cap-sub condition (see Table 5 further below), a mixed model indeed seems a better fit for this research.

Initially, a linear mixed model was applied with individual participants' sum of the meaning recall test scores (continuous) as outcome variables. A mixed model with continuous dependent variables assumes normality of residuals (Linck & Cunnings, 2015). However, Shapiro-Wilk's test showed that the residuals were not normally distributed ($p < .001$). In this case, research suggests a logit mixed model (or a mixed-effects logistic regression) as an alternative solution which can be used with a binary

distribution (Cunnings, 2012). Thus, we needed to recode the meaning recall responses dichotomously with 0 for incorrect responses and 1 for correct responses. To do so, we decided to apply a stricter scoring method which adjusted the partially correct responses initially rated 0.5 ($n$ = 28) to 0 points across all conditions. It should be noted that there were 4 cases of 0.5 points in the none-cap-sub condition and 12 cases in both the sub-cap-none and the cap-cap-cap conditions. It could therefore we argued that dichotomizing the scores this way skews the results somewhat in favour of the none-cap-sub condition. Finally, the mixed-effects logistic regression analysis was conducted to estimate the degree to which the differential ways of providing on-screen text affect incidental vocabulary learning and retention from repeated viewing. We entered condition (none-sub-cap, sub-cap-none, cap-cap-cap), time (pre, immediate, delayed), and test prompt type (written, spoken) as fixed effects and participants and items as random effects. We ran the analyses using GAMLj module in jamovi (The jamovi project, 2021).

# Chapter 8 Results

Table 5 shows the mean scores of each condition group for the meaning recall test. The output of the mixed-effects logistic regression analysis revealed that the total amount of variance explained by the model was 53% (conditional $R^2 = 0.53$). 11% of the variance (marginal $R^2 = 0.11$) in the scores of the meaning recall test was associated with the fixed effects, which means that the rest of 42% variance was due to the random effects (participants and items). The omnibus (Wald) Chi-Squared test showed that time ($\chi^2$ (2) = 50.6, $p < .001$) was a significant predictor of incidental vocabulary learning at the level of meaning recall (see Table 6), indicating that learning happened through repeated viewing. However, neither statistical difference between the experimental conditions and item types nor interaction between the factors was found. Table 7 illustrates the same findings in more detail (interaction between the factors not included). There was a statistically significant difference between the pre-test and the immediate post-test ($p < .001$), and between the pre-test and the delayed post-test ($p < .001$). The odds ratio indicates that the intervention (repeated viewing) increased the probabilities of providing a correct response by 4.16 times compared to pre-test performance. Moreover, the effect of repeated viewing appeared to be lasting as the odds of providing a correct response in the delayed post-test were 4.6 times greater than in the pre-test. We performed a Bonferroni's post hoc test to determine whether the effect of the test time (repeated viewing) was significant in all conditions. The test revealed that there was a significant difference between the pre-test and the immediate post-test and between the pre-test and the delayed post-test in the sub-cap-none condition ($p < .001$ for both) and the none-cap-sub ($p = 0.004$ and $p = 0.002$ respectively). However, no significant

difference was found between the different test times in the cap-cap-cap condition ($p$ = 0.921 for both).

Although the analysis of the mixed-effects logistic reported that condition was not a significant predictor, the findings from the post hoc test imply that repeated viewing with a combination of subtitles, captions, and none might promote incidental learning of words at the meaning recall level better than repeated viewing with a single type of on-screen text. This interpretation is also supported by the descriptive statistics, notably the mean scores on the immediate post-test in each condition group (see Table 5). Figure 6 also displays that both groups that watched the video with varied types of on-screen text performed better than the group that was exposed to the same kind of on-screen text. Thus, these findings help us to tentatively answer our first research question.

**Table 5** *Descriptive Statistics of the Meaning Recall Test with Strict Scoring*

| Condition | Time | *n* | Missing | *M* | *SD* | Min. | Max. |
|---|---|---|---|---|---|---|---|
| None-Cap-Sub | pre | 99 | 11 | 1.56 | 1.35 | 0.00 | 4.00 |
| | immediate | 99 | 11 | 4.33 | 2.55 | 0.00 | 7.00 |
| | delayed | 88 | 22 | 4.25 | 2.55 | 0.00 | 7.00 |
| Sub-Cap-None | pre | 121 | 0 | 2.09 | 1.61 | 0.00 | 5.00 |
| | immediate | 121 | 0 | 4.91 | 1.78 | 3.00 | 8.00 |
| | delayed | 121 | 0 | 5.18 | 1.59 | 2.00 | 8.00 |
| Cap-Cap-Cap | pre | 99 | 0 | 2.56 | 1.78 | 1.00 | 6.00 |
| | immediate | 99 | 0 | 3.89 | 2.44 | 0.00 | 8.00 |
| | delayed | 99 | 0 | 4.00 | 2.37 | 0.00 | 7.00 |

*Note.* The maximum score is 11. *n* = number of observations in each condition.

**Table 6** *Descriptive Statistics of Meaning Recall Test with Lenient Scoring*

| Condition | Time | *n* | Missing | *M* | *SD* | Min. | Max. |
|---|---|---|---|---|---|---|---|
| None-Cap-Sub | pre | 99 | 11 | 1.61 | 1.42 | 0.00 | 4.00 |
| | immediate | 99 | 11 | 4.33 | 2.61 | 0.00 | 7.50 |
| | delayed | 88 | 22 | 4.31 | 2.62 | 0.00 | 7.50 |
| Sub-Cap-None | pre | 121 | 0 | 2.00 | 1.54 | 0.00 | 5.00 |
| | immediate | 121 | 0 | 5.27 | 1.80 | 3.00 | 8.00 |
| | delayed | 121 | 0 | 5.41 | 1.57 | 2.50 | 8.00 |
| Cap-Cap-Cap | pre | 99 | 0 | 2.61 | 1.89 | 1.00 | 6.50 |
| | immediate | 99 | 0 | 4.11 | 2.41 | 0.50 | 8.00 |
| | delayed | 99 | 0 | 4.39 | 2.58 | 0.00 | 7.00 |

*Note*. The maximum score is 11. *n* = number of observations in each condition.

**Table 7** *The Main Effects of the Independent Variables and Their Interaction*

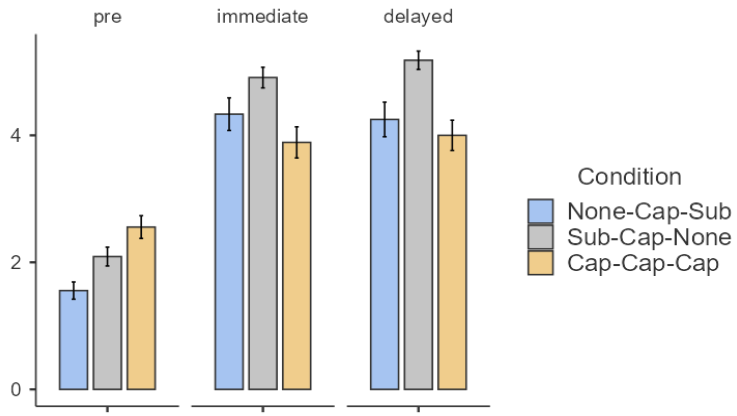
| | $X^2$ | *df* | *p* |
|---|---|---|---|
| Condition | 1.371 | 2.00 | 0.504 |
| Time | 50.588 | 2.00 | < .001 |
| Prompt type (written) | 0.575 | 1.00 | 0.448 |
| Condition ✻ Time | 5.030 | 4.00 | 0.285 |
| Condition ✻ Prompt type | 0.826 | 2.00 | 0.662 |
| Time ✻ Prompt type | 0.190 | 2.00 | 0.909 |
| Condition ✻ Time ✻ Type | 1.502 | 4.00 | 0.826 |

**Table 8** *Parameter Estimates of the Fixed Effects on the Meaning Recall Test*

| | | | | 95% CI for *OR* | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Estimate | *SE* | *OR* | Lower | Upper | *z* | *p* |
| condition (Sub-Cap-None) | 0.6388 | 0.559 | 1.894 | 0.634 | 5.663 | 1.143 | 0.253 |
| condition (Cap-Cap-Cap) | 0.2268 | 0.586 | 1.255 | 0.398 | 3.958 | 0.387 | 0.699 |
| Time (immediate) | 1.4255 | 0.229 | 4.160 | 2.656 | 6.515 | 6.228 | < .001 |
| Time (delayed) | 1.5269 | 0.234 | 4.604 | 2.908 | 7.287 | 6.516 | < .001 |
| Prompt type (Written) | -0.6009 | 0.792 | 0.548 | 0.116 | 2.590 | 0.758 | 0.448 |

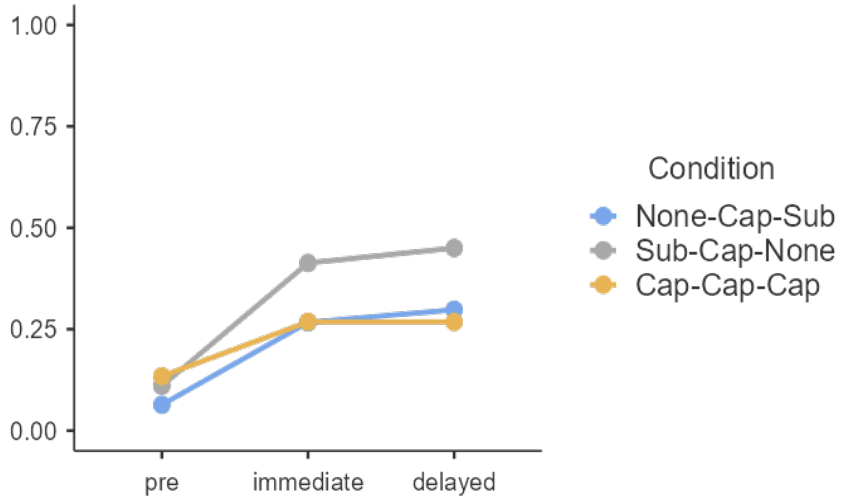
*Note*. Intercept levels: condition = None-Cap-Sub; time = pre(test); type = spoken. OR = odds ratio.

**Figure 6** *Bar Plot Comparing the Mean Scores of Each Condition for the Meaning Recall Test*




As for the second research question regarding whether there would be an ordering effect (e.g., the none-cap-sub vs. the sub-cap-none), the mixed-effects logistic regression model revealed that there was no significant difference between the two conditions ($p$ = .253). However, the descriptive data show that the sub-cap-none group ($M$ = 4.91) performed slightly better than the none-cap-sub group ($M$ = 4.33) in the immediate post-test. It should be mentioned that the gap in the mean scores between the sub-cap-none group ($M$ = 5.27) and the none-cap-sub ($M$ = 4.33) was bigger when the test was scored in a lenient manner (e.g., 0.5 points awarded for partially correct responses) (see Table 6). Because there were more cases ($n$ = 8) of 0.5 in the post-test of the sub-cap-none condition, the mean scores decreased when the strict scoring method was applied. Interestingly, the group nonetheless gained better scores than the other group. This effect is visualized in the effects plot generated by the mixed-effects logistic regression model (see Figure 7). Therefore, we can cautiously suggest that this sequence sub-cap-none has the greater potential to promote incidental vocabulary learning in repeated viewing.

**Figure 7** *The Success Rates of Each Condition Over Time*




The last research question is whether there would be a differential effect in retaining the learned vocabulary knowledge among the three conditions. Figure 6 demonstrates that there was not a noticeable decrease in the mean scores from the immediate post-test to the delayed post-test across all groups. The results of Bonferroni's post hoc test revealed that there was no significant difference between the immediate post-test and the delayed post-test ($p = 1$) for all conditions, which implies that all conditions contributed to the retention of the learned knowledge at the meaning recall level. This might be attributed to the repeated viewing.

We administered the meaning recognition test as well in preparation for a possible floor effect on the scores on the delayed post-test. The participants completed the meaning recognition test after the delayed meaning recall test. Table 8 presents the mean scores of all groups. The performance of all groups on the meaning recognition test was better compared to their performance on the delayed post-test (meaning recall), which we anticipated as a meaning recognition test is less demanding. Interestingly, however, the

cap-cap-cap group performed as well as the sub-cap-none group and outperformed the none-cap-sub group. This is different from the trend attested in the meaning recall tests and will be returned to further below in the Discussion section.

We also ran a mixed-effects logistic regression to detect any significant difference among the experimental conditions and to examine whether the responses on the delayed meaning recall post-test (taken just before the meaning recognition multiple-choice test) was a significant predictor of the performance on the meaning recognition test. The latter possibility is quite plausible—if you recall the meaning of a word in one test, then you are likely to recognize that meaning in a subsequent test. The dependent variables were the responses (correct and incorrect) on the meaning recognition test. Condition and the response on the delayed post-test were entered as fixed effects, and participant and item were input as random effects. The analysis showed that the participants' responses on the delayed meaning recall post-test ($\chi^2$ (1) = 40.418, $p < .001$) were a significant predictor of performance on the meaning recognition test. No significance difference was found among the three conditions, and no interaction between the factors was identified. The participants had a 16.75 times greater chance of giving correct responses on the items whose meanings they recalled correctly in the delayed post-test (see Table 9). Figure 8 illustrates the positive relationship between the scores on the delayed meaning-recall post-test and the meaning recognition test. The explanation about the unexpected results of the cap-cap-cap group will be discussed in the next chapter.

**Table 9** *Descriptive Statistics of the Meaning Recognition Test*

| Condition | *n* | Missing | *M* | *SD* | Min. | Max. |
|---|---|---|---|---|---|---|
| None-Cap-Sub | 88 | 22 | 5.50 | 2.41 | 1 | 8 |
| Sub-Cap-None | 110 | 11 | 7.20 | 2.45 | 4 | 11 |
| Cap-Cap-Cap | 99 | 0 | 7.44 | 1.90 | 4 | 10 |

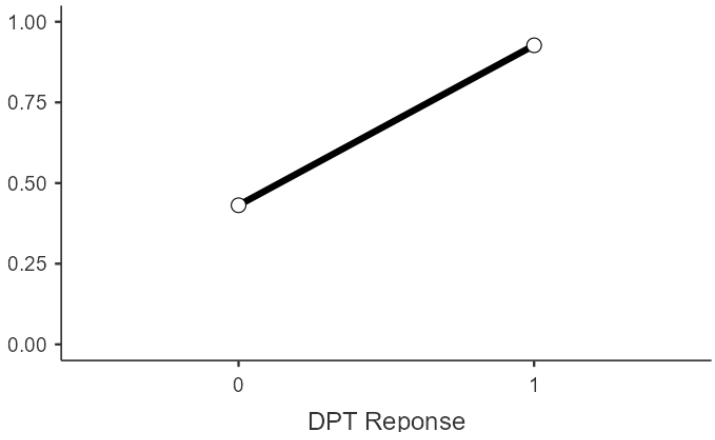
*Note*. The maximum score is 11. *n* = number of observations in each condition.

**Table 10** *Parameter Estimates of the Fixed Effects on the Scores of the Meaning Recognition Test*

| | | | | 95% CI for *OR* | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Estimate | *SE* | *OR* | Lower | Upper | *z* | *p* |
| condition (Sub-Cap-None) | 0.854 | 0.587 | 2.35 | 0.743 | 7.43 | 1.454 | 0.146 |
| condition (Cap-Cap-Cap) | 1.221 | 0.647 | 3.39 | 0.954 | 12.05 | 1.888 | 0.059 |
| Response on DPT (correct) | 2.818 | 0.443 | 16.75 | 7.026 | 39.94 | 6.358 | < .001 |

*Note*. Intercept levels: condition = None-Cap-Sub; response on DPT = incorrect. DPT = delayed post-test; OR = odds ratio.

**Figure 8** *Positive Correlation Between the Response on the Delayed meaning recall Post-test and the Response on the Meaning Recognition Test*



# Chapter 9 Discussion

In this chapter, the findings from the data analysis are discussed in connection with the guiding theories and the participants' responses on the questionnaire.

## 9.1 Does repeated viewing with a *combination* of L1 subtitles, L2 captions, and no onscreen text better facilitate incidental vocabulary learning compared to repeated viewing with the same type of on-screen text?

In this study, we found that test time was a significant predictor of incidental vocabulary learning at the meaning recall level, which indicates that incidental vocabulary learning occurred through repeated viewing. However, the learning gains were not equal across the three viewing conditions. The results of post hoc tests showed that there was significant difference between the pre-test and the immediate post-test performance for the two groups who watched the video with an alternation of L1 subtitles, L2 captions, and none of them, but no significant difference was found between the pre-test and the immediate post-test for the group who watched the video with captions only. It was found that on average 2.3 words (20.9%) out of 11 words were learned from the pre-test to the post-test across all groups. As for the average number of words gained from pre-test to immediate post-test per condition, these were 2.82 words for the sub-cap-none group, 2.77 words for the none-cap-sub group, and 1.33 words for the cap-cap-cap group. From these findings, it can tentatively be concluded that repeated viewing with a mixture of L1 subtitles, L2 captions, and no onscreen text leads to more incidental vocabulary learning than repeated viewing with captions only.

If variation in viewing conditions is more effective for learning than keeping the viewing condition constant, then this may in part be in accordance with the notion of desirable difficulty theory. The word desirable refers to an appropriate level of difficulty, a level that is not too easy and too difficult. In other words, when the level of a learning activity is desirably difficult, learning is more likely to happen and the knowledge will be transferable and stored in the long-term memory (Bjork, 1994; Bjork, 2018; Schmidt & Bjork, 1992). One way of creating the desirable difficulty is varying the learning conditions rather than keeping them consistent. Given this point of view, we can understand how the two groups who watched the video with an alternation of onscreen textual support and absence of this support outperformed the group who watched the video repeatedly with the same onscreen support throughout. Including a viewing without onscreen textual support makes the task more challenging and this may thus be a way to create the desirable difficulty. The level of difficulty will of course also depend on where in the repeated-viewing sequence the onscreen textual support is present or absent. It stands to reason that watching the video without subtitles or captions at the very beginning of the sequence is the hardest. In comparison, watching it after first having processed the contents with support from subtitles and captions will be much less challenging. The desirable-difficulty account therefore appears more compelling with regards to the none-cap-sub sequence than with regards to the sub-cap-none sequence.

Another explanation for the better performance from the varied repeated viewing conditions can be that subtitles and captions supplemented each other. This is supported by the participants' comments from the questionnaire which the participants completed at the end of the study. Their comments showed that they indeed benefited from both types

of on-screen text. Most comments indicated that subtitles helped learn meanings of unknown words, and captions helped recognize forms (spoken and written) of unknown words. One participant's comment demonstrated that both subtitle and captions were helpful for learning meanings of new words. Below are the original excerpts from the questionnaire:

> 1. Chinese subtitles help me understand the main idea of this video and after this, I know what kind of words should be focus. Then the English subtitles help me learn the special words' form such as "clump" and "whey".
>
> 2. It helps me learn the meaning first and then get to know the spelling for the words.
>
> 3. Captions help me learn the written form of unknown words and subtitles help me learn the meaning of the unknown words.
>
> 4. It helps me to notice some vocabulary that I do not know before, and the subtitles with L1 and L2, help me to know the meaning of the new word. The last time without the subtitles can help me to ensure I know the meaning of the new words.

We asked the participants to rank the following possible conditions of watching the same video three times for their likely benefits regarding vocabulary acquisition: 1) subtitles – captions – none, 2) none – captions – subtitles, 3) captions – captions-captions, 4) subtitles – subtitles – subtitles, and 5) none-none-none. 16 Participants answered this question, and their intuitions also indicated that the varied repeated

viewing would be more beneficial than the consistent repeated viewing condition: 50% of the respondents picked the sequence of none-cap-sub and 37.5% of them chose the sequence of sub-cap-none as the most beneficial order for L2 word learning when watching the same L2 video three times. No participant ranked the cap-cap-cap sequence the first. Instead, this order was ranked the third with 56.25%. The option of watching the video without any on-screen text was ranked the last with 68.75%.

Another relevant finding from the questionnaire is that a majority of the participants (17 out of 18 accounting for 94.44%) indeed reported that repeated viewing (watching the same video more than once) was a common activity that they do outside the classroom and that they have in fact tried the above sequences in doing that. The repeated viewing conditions examined in the study were the top 3 ways that they regularly made use of with none-sub-cap at 30.43%, and sub-cap-none and cap-cap-cap at 26.09% respectively. This indeed shows that it is important to explore to what extent the effectiveness of each condition differs.

Nonetheless, this is the first study which explored the potential benefit of using a combination of subtitles and captions. Existing studies mostly focused either on the effect of captions or subtitles separately or on their differential effects. Our findings suggest that it is worthwhile to pay attention to the benefit of using varied on-screen text for incidental vocabulary learning.

## 9.2 Does the precise *sequence* of providing different on-screen text during repeated viewing make a difference to incidental vocabulary learning?

We found that there was no significant difference in vocabulary learning gains on the immediate post-test between the sub-cap-none and the none-sub-cap groups. The average number of learned words immediately after the intervention was 2.82 ($M$ = 4.91) and 2.77 ($M$ = 4.33) respectively. However, it should be noted that the partial knowledge gained by the sub-cap-none group was hidden in the above mean scores because we recoded the responses of 0.5 into 0 to run the logit mixed model. Prior to applying the stricter scoring method, the mean score of the sub-cap-group was 5.27 whereas that of the other group was 4.33. In this case, the former group learned approximately 1 more word on average than the other group. Interestingly, there was visible difference in the scores on the meaning recognition test between the two groups. The sub-cap-none group ($M$ = 7.20) outperformed the none-cap-sub ($M$ = 5.20). Perhaps, this may be attributed to partial knowledge that was gained by the participants in the sub-cap-none group but that was hidden in the meaning recall test due to the stricter rating. This may indicate that the sequence of sub-cap-none is more beneficial than that of none-cap-sub.

The advantage of the sup-cap-none sequence can be explained in relation to the retrieval effect. Retrieval happens when learners recall previously learned knowledge. Each retrieval helps strengthen the knowledge. Given that the sub-cap-group used the L1 subtitles at the first viewing, they likely learned word meanings early on, and the subsequent viewings, without subtitles, provided opportunities to retrieve these meanings

from memory. Consequently, their memory of the meanings of some target words might have become stronger than was the case for the participants in the other group, who first needed to establish word meanings through inferencing before these meanings could be consolidated in memory through retrieval episodes. Studies comparing the benefits of vocabulary learning procedures where word meanings are given to the learners versus ones where the learners are first prompted to try and guess the meanings have yielded mixed findings, but it is clear that the latter procedures (i.e., inferencing) take longer and are error prone (e.g., Elgort, 2017; Mondria, 2003; Nassaji, 2003). It is possible that participants' initial attempts at inferring word meanings in the absence of subtitles were unsuccessful and that wrong guesses needed to be amended or rectified in a subsequent viewing. This may explain why the none-cap-sub condition yielded slightly poorer learning outcomes than the sub-cap-none condition. The following is the original excerpt of one participant's feedback on the sub-cap-none viewing sequence that they provided on the questionnaire:

> I understood most of the content with the help of Chinese subtitles, and then I tried to match the English captions with the Chinese meaning. I also learned the form (written and spoken) and meaning at the same time. (I need to take notes.)

This participant's response describes that they tried to recall the meaning at the second viewing (more likely at the third viewing as well). Therefore, it is confirmed that this specific sequence involves retrieval of meaning, which can further facilitate incidental vocabulary learning through repeated viewing.

As mentioned above, however, incidental vocabulary learning also happened in the none-cap-sub group, and it was greater than that of the cap-cap-cap group. Perhaps the subtitles at the last viewing greatly contributed the learning gains as they confirmed or rectified the participants' initial right or wrong guesses of unknown words. In other words, in this condition, the subtitles played a role as feedback like in trial-and-error learning, which is assumed to stimulate learning as well. The following is another excerpt from the questionnaire, from a participant in the none-cap-sub group:

> When I heard the new words from the TED talk, I have no clue what it is. But seeing subtitles makes these words become "visible" to me. I think it also gives me a sense of "confirmation" since I only guess those words when the first time I saw this video.

This feedback shows that the subtitles in the condition indeed helped notice unknow words and confirm the guesses made at the first viewing. Thus, it is more likely that feedback (subtitles) were the factor that promoted vocabulary learning for this condition.

Our findings are compatible with the findings of a recent study conducted by Strong and Boers (2018). They discovered that exercises on phrasal verbs were implemented in ESL/EFL textbooks either as retrieval practice or for learning through trial and error. Then they compared the effectiveness of these two approaches to phrasal verbs. The participants in the retrieval condition studied the form and meaning of target phrasal verbs prior to the exercises, whereas the procedure was reversed for the participants in the trial-and-error condition (i.e., they had to guess and then received

feedback). The results demonstrated that the participants in the retrieval condition provided more correct answers on the exercises, the post-test, and the delayed post-test than the ones in the trial-and-error condition, indicating that the retrieval effect further promotes learning compared to the effect of feedback in the trial-and-error procedure.

## 9.3 Does the degree to which the above conditions of providing on-screen text during repeated viewing make a difference to *long-term* retention of learned vocabulary knowledge?

The findings showed that there was not a noticeable decrease in the mean scores from the immediate post-test to the delayed post-test across all groups, indicating that all three conditions helped the participants remember the learned vocabulary from watching the TED Talk video until the following week. Considering that time was a significant predictor of the test score, however, the high retention rate might be due to the effect of repeated viewing. A recent study by Majuddin et al. (2021) compared the effect of single viewing and repeated viewing (viewing the same video twice) on acquisition of multiword expressions (MWEs). The findings showed that repeated viewing led to better learning of MWEs than single viewing. In light with this finding, it can be speculated that viewing the same video three times boosted the memory of the acquired vocabulary. It should be mentioned that, in fact, there was a slight increase in the mean scores on the delayed post-test for the sub-cap-none group and the cap-cap-cap group. This may be because some of the students looked up some of the target words after the immediate post-test. This behavior is often found in longitudinal studies including a pre-test, an immediate post-test, and a delayed post-test. However, the increase is very marginal.

Looking at the results of the meaning recognition test, which was administered after the delayed post-test, the sub-cap-none group (*M* = 7.20) garnered better scores than the none-cap-sub group (*M* = 5.50). However, the cap-cap-cap (*M* = 7.44) performed as well as the sub-cap-none group and better than the none-cap-sub group. This was an unexpected result given that the former group had obtained the poorest score in the immediate meaning recall post-test. First, the sub-cap-none group obtained higher scores on the meaning recognition test likely because the partial knowledge that was not taken into account in the mean scores on the meaning recall tests helped the participants recognize meanings of the target words when they saw these in the multiple choice test. Probably, the retrieval effect enabled the participants to successfully recall the partially learned knowledge and match this to the exact meanings in the meaning recognition test considering that they recognized meanings of 7.2 words correctly on average out 11 words. That number is significant accounting for 65.45%.

As for the cap-cap-cap group, they performed the best (M = 7.44) on the meaning recognition test, in stark contrast with their performance on the meaning recall test. The following scenario might explain this result. The input that they received three times was L2 sound and L2 text, and the option given in the multiple-choice meaning recognition test were presented in L2. This may have given them an advantage when processing the L2 definitions of target words in the test, which often contained words that appeared concurrently with the target words in the video. Recognizing these related words in the definition may have helped them to guess the correct matches. A good example is the test item for the target word *ration*. Its matching definition on the test was “b. a fixed amount of food for each **soldier** in an army”. This target word appears once in this part of the

video: Under Roman rule, “dry cheese” or “caseus aridus,” became an essential **ration** for the nearly 500,000 **soldiers** guarding the vast borders of the Roman Empire. As can be seen in the example, it is possible that participants who had seen *ration* together with *soldiers* three times in the captions, determined that the definition containing *soldier* was the most plausible option. If so, these participants may have established the meaning of *ration* in the meaning recognition test rather than through watching the captioned video. Indeed, we found that there was only one correct response for *ration* on the delayed meaning recall post-test in the cap-cap-cap group, but there were 5 correct responses on the meaning recognition test in this group. This compares to respectively 2 and 4 participants who chose the correct definition of ration in the none-cap-sub and sub-cap-none groups. This suggests that input–test congruency should be taken into account when developing test instruments. Research has indeed found that input modality–test modality congruency can affect test performance (Jelani & Boers, 2018; Sydorenko, 2010).

# Chapter 10 Conclusions and Pedagogical Implications

To the best of our knowledge, this is the first study which explored the potential benefit of using the combination of L1 subtitles, L2 captions, and none of them compared to using L2 captions only during repeated viewing. In addition, this is the first study to examine the effect of manipulating the sequence of alternating the different viewing conditions when a video is used three times. The research questions were developed to fill the gap in the literature, informed by theories in the field of cognitive psychology and memory research, and motivated by the student researcher's personal L2 learning habits.

Our findings from the quantitative data were in line with the ones from the qualitative data, hence, we were able to gain broader understanding of the results. The findings from both forms of data suggest that varying the repeated viewing condition (adding a mixture of subtitles, captions, and no onscreen text) has the potential to promote incidental vocabulary acquisition more effectively at the level of meaning recall than keeping the repeated viewing condition consistent (providing captions only). Interpreting the findings with the theoretical lens of desirable difficulty, the alternation of different viewing conditions, including a condition without onscreen textual support, creates desirable difficulty that can stimulate the learners' curiosity and requires varied levels of effort.

As for the long-term retention, we tentatively suggest that the sequence of watching with subtitles first, then captions, and lastly without onscreen textual support has the greatest potential because this was the sequence that generated the best performance on both the delayed meaning recall post-test and the meaning recognition

test. Apart from the likely advantage of variation, this viewing sequence creates the best conditions for retrieval because the word meanings are established early on thanks to the subtitles. The subsequent viewings are increasingly challenging and stimulate recall of the meanings introduced in the first viewing. However, no significant attrition occurred in the other two groups either after the 1-week period of time, which implies that repetition is an important factor in strengthening memories.

In line with the motivation of the research, the questionnaire revealed that repeated viewing is a common activity that L2 learners engage in outside the classroom. The reasons of repeated viewing varied: to learn new words and phrases (35.48%), to understand the content of the video better (29.03), to improve pronunciation (16.13%), and to simply enjoy the video again (16.13%). Given the fact that the respondents were successful L2 learners, these responses are not surprising.

These findings have pedagogical value. In general, the findings inform L2 teachers and learners on how to use a myriad of audiovisual materials with subtitles and captions. For instance, videos from the TED Talks are freely available and they provide subtitles and captions in more than 20 languages. In fact, the videos from the TED Talk or TED Ed are frequently being used in L2 classroom as a source of authentic spoken input. However, using captions and especially subtitles in L2 class is not a common teaching strategy yet, although they have been shown to facilitate vocabulary learning and listening comprehension. If using the L1 subtitles does not meet the purpose of the activity (e.g., testing listening skills as if in the real situation), it can be incorporated at the last viewing as feedback. Alternatively, L2 teachers can encourage their students to

use the sequence of subtitles, captions, and no onscreen text when they watch the same video outside the classroom.

There was one participant who approached me after completing all study sessions and told me that she loved to watch TED Talks but did not know how to use them effectively. Then she asked me to share the results of the study with her. I share this story because I believe that individual participants' voices are as important data as test scores for researchers. The feedback from the participant confirmed that there is need for such information regarding viewing or repeated viewing. As a result, the strategy of using subtitles and captions should be encouraged both inside and outside classrooms given the findings from quantitative data and the fact that this is indeed part of successful learners' language learning habits.

# Chapter 11 Limitations and Future Research

The present study has some major limitations that need to be acknowledged. The first limitation of the study is the small number of participants ($N = 30$). Although 946 observations were included in the mixed-effects logistic regression, it was hard to detect significant difference in the effects of the independent variables of interest (repeated viewing conditions). However, it should be reiterated that recruitment of participants was severely compromised owing to the pandemic. With the findings of the present study as a groundwork, a larger-scale replication study could be conducted in the future. Such a larger- scale project could also include additional treatment conditions. For example, the current study did not include a group who watched the video three times without any type of on-screen text. Thus, it was hard to measure the pure effect of the differential ways of providing on-screen text on incidental vocabulary acquisition through repeated viewing. This is also a limitation caused by the small number of participants—initially, more treatment conditions were planned. Neither did we manage to include a condition in which the video is watched three times with subtitles only. In our original plan, this condition was included, but we had to exclude it, again due to the small number of participants.

Another limitation concerns the research instruments. We used only one test format (meaning recall) as pre-test, immediate post-test, and delayed post-test. Other test formats could perhaps have revealed more differences in learning outcomes. We did also administer a multiple-choice meaning recognition test as a delayed measure, but this test format may have given an advantage to the cap-cap-cap group if the repeated exposure to the captions helped them to determine the correct test responses by recognizing the L2

words in the meaning definitions that co-occurred with the target words in the captions. It is possible that the input-test congruency affected their scores on the test. Researchers should think of ways of eliminating such compound variables for future research. For instance, we can present the multiple choices in both L1 and L2 and add some of words that appeared with target words (e.g., soldier for ration) in other options as well not to give a better clue to the cap-cap-cap group.

# References

Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian modern language Review*, *50*(3), 507-531.

Baranowska, K. (2020). Learning most with least effort: subtitles and cognitive load. *ELT Journal, 74*(2), 105–115. https://doi.org/10.1093/elt/ccz060

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning, 57*, 35–56. https://doi.org/10.1111/j.14679922.2007.00398.x

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

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.185-205). MIT Press.

Bjork, R. A. (2018). Being suspicious of the sense of ease and undeterred by the sense of difficulty: Looking back at Schmidt and Bjork (1992). *Perspectives on*

*Psychological Science, 13*(2), 146–148. https://doi.org/10.1177/1745691617690642

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*, 136–163.

Cobb, T. (n.d.). Vocabprofile [Computer program]. Retrieved August 4, 2020, from https://www.lextutor.ca/vp/comp/

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.

Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning*, *42*(4), 497–527. https://doi.org/10.1111/j.1467-1770.1992.tb01042.x

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: Journal Des Traducteurs, 49*(1), 67–77. https://doi.org/10.7202/009021ar

d'Ydewalle, G., Pavakanun, U. (1997). Could enjoying a movie lead to language acquisition?. In Winterhoff-Spurk P., van der Voort T.H.A. (Eds.), *New Horizons in Media Psychology*. VS Verlag für Sozialwissenschaften, Wiesbaden. (pp. 145-155). https://doi.org/10.1007/978-3-663-10899-3_10

d'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research, 28*(3), 227–244. https://doi.org/10.1023/A:1023202130625

Cunnings, I. (2012). An overview of mixed-effects Statistical models for Second language researchers. *Second Language Research*, *28*(3), 369–382. https://doi.org/10.1177/0267658312443651

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*(2), 123.

Elgort, I. (2017). Incorrect inferences and contextual word learning in English as a second language. *Journal of the European Second Language Association, 1*, 1–11. https://doi.org/10.22599/jesla.3

Ellis, R. (1999). *Learning a second language through interaction*. John Benjamins.

Frumuselu, A., De Maeyer, S., Donche, V., & Colon Plana, M. (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

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*(5), 562-567.

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

Jelani, N. A. M., & Boers, F. (2018). Examining incidental vocabulary acquisition from captioned video. *ITL-International Journal of Applied Linguistics, 169*(1), 169-190. https://doi.org/10.1075/itl.00011.jel

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation*, *62*, 237-284. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Koolstra, C., & Beentjes, J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development, 47*(1), 51–60. https://doi.org/10.1007/BF02299476

Kruger, J. L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research, 7*, 1-15.

Kuppens, A. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology, 35*(1), 65–85. https://doi.org/10.1080/17439880903561876

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1–26. https://doi.org/10.1093/applin/22.1.1

Linck, J. A., & Cunnings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research. *Language Learning*, *65*(S1), 185–207. https://doi.org/10.1111/lang.12117

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition*, 1–24. https://doi.org/10.1017/S0272263121000036

Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, *32*(3), 321–328. https://doi.org/10.1111/j.1944-9720.1999.tb01344.x

Mondria, J. (2003). The effects of inferring, verifying, and memorizing on the retentionof L2 word meanings: An experimental comparison of the "meaning-inferred method" and the "meaning-given method." *Studies in Second Language Acquisition, 25*, 473–499. https://doi.org/10.1017/S0272263103000202

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*(3), 720–739. https://doi.org/10.1016/j.system.2013.07.013

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology, 18*(1), 118–141. http://dx.doi.org/10125/44357

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition,* 1-25. https://doi.org/10.1017/S0272263119000706

Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly, 37*, 645–670. https://doi.org/10.2307/3588216

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

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59–82.

Nation, I. S. P. (2012). The BNC/COCA word family lists. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>

Newton, J. M., Ferris, D. R., Goh, C. C. C., Grabe, W., Stoller, F. L., & Vandergrift, L. (2018). Teaching English to second language learners in academic contexts. Routledge.

Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly, 27*, 94–106.

Nguyen, C., & Boers, F. (2019). The effect of content retelling on vocabulary uptake from a TED Talk. *TESOL Quarterly, 53*(1), 5–29. https://doi.org/10.1002/tesq.441

Nurmukhamedov, U. (2017). Lexical coverage of TED talks: Implications for vocabulary instruction. *TESOL Journal*, *8*(4), 768–790. https://doi.org/10.1002/tesj.323

OECD (2009). Special Focus: Measuring Leisure in OECD Countries. *Society at a Glance 2009: OECD Social Indicators* (pp. 19-49). OECD Publishing. https://doi.org/10.1787/soc_glance-2008-3-en.

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., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*(3), 551-577. https://doi.org/10.1017/S0272263117000407

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

Pellicer-S´anchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition, 38*, 97–130. https://doiorg.proxy1.lib.uwo.ca/10.1017/S0272263115000224

Pellicer-S´anchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language, 22*, 31–55.

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: Video and Language Learning, 47*(4), 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*(4), 479-496. https://doi.org/10.1080/09571736.2019.1616806

Quené, H & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Communication*, *43*(1), 103–121. https://doi.org/10.1016/j.specom.2004.02.004

Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*, 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger III, H. L., & Karpicke, J. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Rodgers, M. P. (2013). English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions. Unpublished PhD thesis, Victoria University of Wellington.

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

Rodgers, M. P. H., & Webb, S. (2020). Incidental vocabulary learning through viewing television. *ITL-International Journal of Applied Linguistics*, *171*(2), 191-220. https://doi.org/10.1075/itl.18034.rod

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological science*, *3*(4), 207-218.

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

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.

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

Scotto di Carlo, G. (2014). The role of proximity in online popularizations: The case of TED talks. *Discourse Studies*, *16*(5), 591–606. https://doi.org/10.1177/1461445614538565

Speelman, D., Heylen, K., & Geeraerts, D. (Eds.). (2018). *Mixed-effects regression models in linguistics*. Springer Publishing Company.

Strong, B., & Boers, F. (2019). The Error in Trial and Error: Exercises on Phrasal Verbs. *TESOL Quarterly*, *53*(2), 289–319. https://doi.org/10.1002/tesq.478

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology, 2*(14), 50–73. http://dx.doi.org/10125/44214

Takaesu, A. (2013). TED Talks as an extensive listening resource for EAP students. Language Education in Asia, 4, 150–162.

Teng, F. (2018). Incidental vocabulary acquisition from reading-only and reading-while-listening: a multi-dimensional approach. *Innovation in Language Learning and Teaching*, *12*(3), 274–288. https://doi.org/10.1080/17501229.2016.1203328

The jamovi project (2021). *jamovi*. (Version 1.8) [Computer Software]. Retrieved from https://www.jamovi.org.

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics, 24*, 3–25. DOI: 10.1017/S0267190504000017.

Vandergrift, L., & Goh, C. (2012). *Teaching and learning second language listening: Metacognition in action.* Routledge.

Vanderplank, R. (1988). The value of teletext subtitles in language learning. *English Language Teaching Journal, 42*(4), 272–281. http://search.proquest.com/docview/58186229/

Valentini, A., Ricketts, J., Pye, R., & Houston-Price, C. (2018). Listening while reading promotes word learning from stories. *Journal of Experimental Child Psychology*, *167*, 10–31. https://doi.org/10.1016/j.jecp.2017.09.022

van Zeeland, H., & Schmitt, N. (2013a). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609–624. https://doi.org/10.1016/j.system.2013.07.012

van Zeeland, H., & Schmitt, N. (2013b). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457–479. https://doi.org/10.1093/applin/ams074

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159-168). Routledge.

Webb, S., & Chang, A. C. S. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition, 37*, 651-675. https://doi.org/10.1017/S0272263114000606

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

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated vocabulary levels test. *ITL-International Journal of Applied Linguistics, 168*(1), 33-69.

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

Winke, P., Gass, S., & Syodorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology, (14)*1, 65–86. http://dx.doi.org/10125/44203

Zenith. (2019, October). Global online video viewing and AD spending forecast. Marketing Chart. https://www.marketingcharts.com/digital/video-110520

# Appendices

## Appendix A: Ethics Approval Notice

Western Research

**Date:** 18 May 2021

**To:** Professor Frank Boers

**Project ID:** 116673

**Study Title:** Learning new words from TED Talks: Strategic use of L1 subtitles and L2 captions

**Short Title:** Strategic use of L1 sub. and L2 cap.

**Application Type:** NMREB Initial Application

**Review Type:** Delegated

**Full Board Reporting Date:** June 4 2021

**Date Approval Issued:** 18/May/2021 16:48

**REB Approval Expiry Date:** 18/May/2022

---

Dear Professor Frank Boers

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. **All other required institutional approvals and mandated training must also be obtained prior to the conduct of the study.**

**Documents Approved:**

| Document Name | Document Type | Document Date | Document Version |
|---|---|---|---|
| Selected TED Talk video for the project | Other Data Collection Instruments | | |
| Debriefing Letter to Participants | Debriefing document | | |
| Link to Online Questionnaire | Online Survey | 12/Apr/2021 | online version |
| The Updated Vocabulary Levels Test PDF | Other Data Collection Instruments | 29/Apr/2021 | |
| Meaning Recall Test | Other Data Collection Instruments | 29/Apr/2021 | |
| Meaning Recognition Test | Other Data Collection Instruments | 29/Apr/2021 | |
| Listening Comprehension Test | Other Data Collection Instruments | 29/Apr/2021 | |
| Questionnarie | Online Survey | 29/Apr/2021 | |
| Letter of Information and Consent | Written Consent/Assent | 11/May/2021 | 11/05/2021 |
| Zoom Oral Recruitment Script | Recruitment Materials | 11/May/2021 | Revised |

**Documents Acknowledged:**

| Document Name | Document Type | Document Date | Document Version |
|---|---|---|---|
| Procedures of the Study | Supplementary Tables/Figures | | |

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

## Appendix B: Letter of Information

### Letter of Information - Student

**Project Title:** Improving listening skills from TED Talks: Strategic use of on-screen text

**Principal Investigator**

Dr. Frank Boers

**Student Investigator**

Injung Wi

Faculty of Education

The University of Western Ontario, London, Canada

Thank you for being interested in this research project. Please read the following Letter of Information and decide whether you would like to participate in this project or not. If you decide to participate in this study, we will sincerely appreciate your help. If you decide not to take part in the study, we will also be thankful for your interest.

**Invitation to the study project**

You are invited to participate in the current study about how to use videos from TED Talks for development of English listening skills as you are a potential English language teacher as well as a learner. It is expected that you will be in this study for three weeks. It will take approximately one hour per week. The aim of this study is to investigate how to effectively use TED Talks to enhance listening skills, particularly through the use of on-screen text (L1 subtitles and L2 captions).

**The rationale of the study**

Audiovisual materials (e.g., movies, TV shows, and online videos) have been widely used in L2 classroom settings as authentic sources of listening. Even outside the classroom, more and more L2 learners are watching online videos. When watching them, nowadays the function of switching on and off on-screen text is often available. Recent studies have found that addition of on-screen text has a facilitative effect on listening comprehension while viewing a video in the target language. However, little is known about how to use the function strategically to maximize its effectiveness on listening comprehension. Therefore, this research project aims to examine various ways to make use of on-screen text to best support L2 learners' listening comprehension.

**The assignment of groups**

If you decide to participate then your will receive one of three conditions based on your course number. All three conditions include viewing the same TED Talk video with different types and orders of on-screen text. All three conditions include viewing the same TED Talk video with different types and orders of on-screen text.

**The procedures of the study**

The study involves three sessions over a three-week period. Each session will take approximately thirty minutes. Those of you who give consent to participate in the study will be asked to perform a series of tests related to English and watch a TED Talk video. After that, you will be asked to complete a short questionnaire. At the end of the study session, you will receive a debriefing form that explains the purpose of the study in more detail.

**The risks and harms of participating in the study**

We do not anticipate any risks or discomfort related to participating in this study project, but you may feel tired while performing the required activities. However, the researcher will create a comfortable environment, give support, and answer potential questions. The study sessions are well assigned in order to decrease your fatigue.

**The benefits of participating in the study project**

You are invited to participate in this study because you are currently an English language learner and potential English teacher. This study will be beneficial for you as it will allow you to (1) gauge your current vocabulary level, (2) learn new word items, and (3) develop learning and teaching strategies for listening skills and vocabulary. At the same time, you will be helping with research that is useful for both learners and teachers. More specifically, the results of the study will inform learners' and teachers' decision making regarding their use of audiovisual materials inside and outside the classroom.

**The option of leaving the study**

As your participation in this study is voluntary, you can leave the study at any time. We can also remove your information from the study if you would like us to. If so, you can simply send us an email to let us know of your decision. However, a month after the end of data collection (after the last session), your data cannot be removed any longer, because we will have started processing the data by then.

**Data privacy**

All the data collected from you will be kept confidential. We will keep the data for nine years. Only the student investigator and her supervisor (the principal investigator) will have access to the data collected from you, and the data will only be used for the research purposes outlined above. The results of the research project will be

reported in the dissertation and possibly in journal articles and conference presentations. No names of any individual students will be mentioned in these reports.

**The rights of participants**

Your participation in this study is voluntary. You may decide not to be in this study. Even if you consent to participate you have the right to not answer individual questions or to withdraw from the study at any time. If you choose not to participate or to leave the study at any time, it will have no effect on your school grade. You do not waive any legal right by consenting to this study. We will give you any new information that may affect your decision to stay in this study.

**Contact for questions**

If you have questions about this research study, please contact Dr. Frank Boers or Injung Wi, MA Student.

If you have any questions about your rights as a research participant or the conduct of this study, you may contact the Office of Human Ethics. This office oversees the ethical conduct of research studies and is not part of the study team. Everything that you discuss will be kept confidential.

**This letter is yours to keep for your future reference.**

## Appendix C: Consent Form

### Consent Form - Student

**Project Title:** Improving listening skills from TED Talks: Strategic use of on-screen text

**Principal Investigator:** Dr. Frank Boers

**Student Investigator:** Injung Wi

**For participants**

I have read the Letter of Information, and I have understood the nature of the study. All the questions regarding the research project were explained to my satisfaction, therefore, I agree to participate in this research project. I have been provided a copy of the Information Letter and the Consent Form. I voluntarily and freely consent to participate in this study.

| _____ | _____ | _____ |
|---|---|---|
| *Print Name of Participant* | *Signature* | *Date (DD-MM-YYYY)* |

**For person obtaining consent**

My signature means that I have explained the study to the participant named above. I have answered all questions.

| _____ | _____ | _____ |
|---|---|---|
| *Print Name of Person* | *Signature* | *Date (DD-MM-YYYY)* |

*Obtaining Consent*

# Appendix D: A Sample of the Updated Vocabulary Levels Test

This is test that looks at how well you know useful English words. Put a check under the word that goes with each meaning. Here is an example.

| | game | island | mouth | movie | song | yard |
|---|---|---|---|---|---|---|
| land with water all around it | | | | | | |
| part of your body used for eating and talking | | | | | | |
| piece of music | | | | | | |

It should be answered in the following way.

| | game | island | mouth | movie | song | yard |
|---|---|---|---|---|---|---|
| land with water all around it | | ✓ | | | | |
| part of your body used for eating and talking | | | ✓ | | | |
| piece of music | | | | | ✓ | |

**1,000 Word Level**

| | choice | computer | garden | photograph | price | week |
|---|---|---|---|---|---|---|
| cost | | | | | | |
| picture | | | | | | |
| place where things grow outside | | | | | | |

| | eye | father | night | van | voice | year |
|---|---|---|---|---|---|---|
| body part that sees | | | | | | |
| parent who is a man | | | | | | |
| part of the day with no sun | | | | | | |

| | center | note | state | tomorrow | uncle | winter |
|---|---|---|---|---|---|---|
| brother of your mother or father | | | | | | |
| middle | | | | | | |
| short piece of writing | | | | | | |

| | box | brother | horse | hour | house | plan |
|---|---|---|---|---|---|---|
| family member | | | | | | |
| sixty minutes | | | | | | |
| way of doing things | | | | | | |

## Appendix E: Meaning Recall Test

*Instructions: Try to give the meaning of the words. You can choose to give a translation in your first language, give an English synonym, or give a definition in English (See the example below). The test is divided into two parts. In the first part of the test (from number 1 to 10), the words are presented in written forms. In the second part of the test (from 11 to 20), the words are presented in spoken forms.*

*Example: to explain: 解释, to describe, or to make things clear to someone*

| No. | Word | Give the meaning of this word in L1 translation, English synonym, or definition |
|---|---|---|
| 1 | to stockpile | |
| 2 | to accumulate | |
| 3 | staple | |
| 4 | cooperation | |
| 5 | monastery | |
| 6 | to regulate | |
| 7 | domesticated | |
| 8 | famine | |
| 9 | to manufacture | |
| 10 | delicacy | |
| *This is the second part of the test. You will hear each word twice. After that, please write down the spelling of the word that you heard in the first column and give its meaning in the second column.* | | |

| | | |
|---|---|---|
| 11 | (creation) | |
| 12 | (legacy) | |
| 13 | (development) | |
| 14 | (ration) | |
| 15 | (to absorb) | |
| 16 | (clump) | |
| 17 | (legal) | |
| 18 | (to coagulate) | |
| 19 | (nation) | |
| 20 | (commodity) | |

## Appendix F: Meaning Recognition Test

*Instructions: Circle the correct definition of the given word. If you do not know the meaning, please make sure to circle the 'I don't know' option.*

1. staple:

    a. A famous or talented performer in the world of entertainment or sports

    b. A food or product that is basic and important in people's everyday lives

    c. moving or capable of moving at high speed

    d. lacking in importance or significance

    e. I don't know

2. to coagulate:

    a. to work together toward the same goal

    b. to change a fluid to a solid state

    c. to think carefully about something before making a decision

    d. to make food by preparing and heating the ingredients

    e. I don't know

3. famine:

    a. a very large quantity of something

    b. the quality of being hot

    c. the state of being free from illness

    d. extreme shortage of food

    e. I don't know

4. domesticated:

    a. living or growing in the natural environment

    b. feeling or showing unhappiness

    c. kept as a pet or on a farm

d. far away from other places or people

e. I don't know

5. clump:

a. a compacted lump or glob of something

b. a substance that flows freely

c. a substance that does not have shape and volume

d. a small piece of bread, cake, or cracker

e. I don't know

6. monastery:

a. a building where products are made by machine

b. a building used by religious men called monks

c. a place where meals are prepared and served to customers

d. a place where products are sold

e. I don't know

7. to stockpile:

a. to exchange with another for money

b. to become solid from cold temperatures

c. to accumulate a large amount of goods or materials

d. to take from one place to another

e. I don't know

8. legacy:

a. a book used for teaching a particular subject

b. something a person gives without wanting anything in return

c. anything that contains nutrients

d. anything that is passed down from ancestors

e. I don’t know

9. ration:

a. a food made by baking flour dough

b. a fixed amount of food for each soldier in an army

c. any liquid for drinking except water or medicine

d. a drug or other substance used to treat a disease or injury

e. I don’t know

10. delicacy:

a. a sweet and sour food

b. a frozen food

c. a cheap food

d. a choice or expensive food

e. I don’t know

11. commodity:

a. goods that can be bought and sold

b. animals used for food

c. a type of food made from milk

d. food or other thing that is thrown away

e. I don’t know

## Appendix G: Listening Comprehension Test

Comprehension Test for The Brie(f) History of Cheese

*Instructions: The following questions are about the video (The Brie(f) History of Cheese) that you have just watched. Write your answers or circle the correct answer.*

1. After milk is left in warm conditions, it becomes sour and forms soft yellowish globs curds. You can get these edible curds after draining the remaining liquid. What is the name of the remaining liquid?

Your answer: _____

2. What are the three beneficial nutrients that milk contains according to the video?

Your answer: _____ , _____ , _____

3. Lactose is a type of sugar that exists in milk and it is easy to digest. Is this true or false?

Your answer: _____

4. Cheese contains less lactose than milk but provides the same nutrients as milk does. Is this true or false?

Your answer: _____

5. Which of the following statements is False according to the video?

a. People in Egypt ate cottage cheese made from goats' milk.

b. Yak's milk was used to make hard cheese in Mongolia.

c. Feta cheese was produced in Greece.

d. Feta cheese became the essential food for Roman soldiers guarding the borders.

6. Since the Industrial Revolution, cheese making has been mechanized. As a result, old ways of hand crafting cheese used by Neolithic people have disappeared. It true or false?

Your answer: _____

## Appendix H: Questionnaire

*Instructions: Please read each question carefully to answer them. If you have any questions, please raise your hand for assistance.*

1. Had you watched the video, *The Brie(f) History of Cheese*, before this class?

    a. yes

    b. no

2. How did you like the video, *The Brie(f) History of Cheese*?

    a. very interesting

    b. interesting

    c. not bad (okay)

    d. boring

    e. very boring

3. What was the difficulty level of the video?

    a. Very difficult

    b. difficult

    c. somewhat difficult

    d. appropriate

    e. easy

4. Do you think that repeated viewing (viewing the video more than once) helped your comprehension of the video?

    a. yes

    b. no

5. Do you think that repeated viewing helped you learn new vocabulary in the video?

    a. yes

b. no

6. Do you think that the addition of captions (English) or subtitles (Chinese) helped you understand the video?

a. yes

b. no

7. If yes, please describe how you think they helped with your listening comprehension. If no, please describe why not.

8. Do you think that the addition of captions or subtitles helped you learn new words in the video?

a. yes

b. no

9. If yes, please describe how you think they helped you learn new words. If no, please describe why not.

10. Which of the following do you think is more helpful for learning new words when viewing an English video?

a. L2 captions (English)

b. L1 subtitles (Chinese)

c. both L2 captions and L1 subtitles

e. none

11. Have you tried to watch the same English video (movies, TV shows, and online videos like TED Talks, etc.) more than once?

a. yes

b. no

12. If yes, what is the reason for doing so? (You can select more than one option that applies to you.)

a. to understand the content of the video better

b. to learn new words and phrases

c. to improve pronunciation

d. to simply enjoy the video one more time

e. others (please describe your own reason):

13. When viewing the same video in English three times, which of the following sequence of adding on-screen text (captions, subtitles, or none) do you think will be the most beneficial for L2 vocabulary learning? Please rank them by writing a number from 1 to 5 in the brackets. 1 means the most beneficial and 5 means the least beneficial.

a. subtitles – captions – none ( )

b. none – captions – subtitles ( )

c. captions – captions – captions ( )

d. subtitles – subtitles – subtitles ( )

e. viewing all three times without captions and subtitles ( )

14. If you answered that you have watched the same video more than once, have you ever tried any of the above sequences of adding on-screen text or other sequences?

a. yes

b. no

15. If yes, what was(were) the sequence(s) that you have tried? (You can select more than one option that applied to you.)

a. subtitles – captions – none

b. none – captions – subtitles

c. captions – captions – captions

d. subtitles – subtitles – subtitles

e. viewing all three times without captions and subtitles

16. If you have tried a different sequence, please describe the sequence.

17. If you have any additional opinions that you would like to share about the use of L1 subtitles and L2 captions during viewing in relation to listening comprehension and vocabulary learning, please feel free to provide them here.

## Appendix I: Answer Sheet for the Meaning Recall Test

1. **to stockpile**: accumulate a large stock of (goods or materials) -> 储存

2. **staple**: n. a main or important element of something, especially in terms of consumption; a main item of trade or production ->主要产品，主题

3. **monastery**: n. a building or buildings occupied by a community of monks living under religious vows. ->修道院

4. **domesticated**: a. tame and kept as a pet or on a farm -> 被驯化的

5. **famine**: n. extreme scarcity of food. -> 饥荒

6. **delicacy**: n. a delicacy is usually a rare or expensive food item that is considered highly desirable, sophisticated or peculiarly distinctive, within a given culture. -> 美味精致的食物

7. **legacy**: n. an amount of money or property left to someone in a will; a thing handed down by a predecessor. -> 遗产

8. **ration**: n.an amount of food supplied on a regular basis, especially to members of the armed forces during a war; food, provision -> 定量配给, 食粮

9. **clump**: n. a compacted mass or lump of something. -> 块

10. **to coagulate**: v. (of a fluid, especially blood) change to a solid or semisolid state. -> 凝固

11. **commodity**: n. a raw material or primary agricultural product that can be bought and sold, such as copper or coffee; a useful or valuable thing, such as water or time. -> 商品

## Appendix J: Debriefing Letter

### Debriefing Letter

**Project Title:** Learning new words from TED Talks: Strategic use of L1 Subtitles and L2 Captions

**Principal Investigator:** Dr. Frank Boers

**Student Investigator:** Injung Wi

Thank you for participating in this research project. The purpose of this project was to examine (1) whether different orders of adding on-screen text in repeated viewings have an effect on L2 vocabulary learning gains and (2) whether viewing with a combination of captions, subtitles and no on-screen text results in more vocabulary learning than viewing with only the same kind of on-screen text (e.g., captions only). Given the broad availability of online videos and the increasing availability of captions and subtitles, we believe that it is worth investigating how we can make the best use of these functions to help foster L2 vocabulary acquisition. The outcome of this type of research can be informative to L2 learners as well as L2 teachers (including future ESL/EFL teachers).

Your results will be kept confidential to the researchers, and all your data will be kept anonymous in any publications. If you have any questions or concerns, please feel free to contact the investigators, Dr. Frank Boers, fboers@uwo.ca, or Injung Wi, iwi@uwo.ca.

Here are some references related to this topic if you want to read more:

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: Journal Des Traducteurs, 49*(1), 67–77.

Jelani, N. A. M., & Boers, F. (2018). Examining incidental vocabulary acquisition from captioned video. *ITL-International Journal of Applied Linguistics, 169*(1), 169-190.

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

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

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

Thank you for your participation.

# Curriculum Vitae

| | |
|---|---|
| **Name:** | Injung Wi |
| **Post-secondary Education and Degrees:** | Kyungpook National University<br>Daegu, Kyungpook, South Korea<br>2004-2009 B.A.<br><br>The University of Western Ontario<br>London, Ontario, Canada<br>2018-2019 MPEd<br><br>The University of Western Ontario<br>London, Ontario, Canada<br>2019-2021 M.A. |
| **Honours and Awards:** | Western Entrance Scholarship<br>2019-2021 |
| **Related Work Experience** | Research Assistant<br>University of Western Ontario<br>2019-2021<br><br>Freelance Interpreter<br>MCIS Language Solutions<br>Toronto, Ontario, Canada<br>2018-present<br><br>On-line English Teacher<br>Small Bridge Inc.<br>Tokyo, Japan<br>2015-2018<br><br>English Instructor<br>Samsung Electronics<br>Gumi, Kyungpook, South Korea<br>2013-2014<br><br>English Teacher<br>Otae Elementary School<br>Gumi, Kyungpook, South Korea<br>2010-2014 |