



2015-07-01

The Effect of Various Word Treatments on Chinese Vocabulary Comprehension During Multimedia Language Learning

Jeongwoon Kim
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Educational Psychology Commons](#)

BYU ScholarsArchive Citation

Kim, Jeongwoon, "The Effect of Various Word Treatments on Chinese Vocabulary Comprehension During Multimedia Language Learning" (2015). *All Theses and Dissertations*. 5978.
<https://scholarsarchive.byu.edu/etd/5978>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

The Effect of Various Word Treatments on Chinese Vocabulary Comprehension
During Multimedia Language Learning

Jeongwoon Kim

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Michael D. Bush, Chair
Richard E. West
Peter J. Rich
Matthew B. Christensen
Dennis L. Eggett

Department of Instructional Psychology and Technology

Brigham Young University

July 2015

Copyright © 2015 Jeongwoon Kim

All Rights Reserved

ABSTRACT

The Effect of Various Word Treatments on Chinese Vocabulary Comprehension During Multimedia Language Learning

Jeongwoon Kim

Department of Instructional Psychology and Technology, BYU

Doctor of Philosophy

The current study investigated the effect of viewing various types of word definitions—dictionary, glossed, and annotated definitions—on Chinese learners' vocabulary comprehension during video-based language learning. A total of 53 advanced Chinese learner participants were given access to various randomized word definitions while viewing the subtitled videos used in the experiment. Results showed that learners recalled words for which they viewed definitions significantly better after the experiment than before. The viewing of dictionary, glossed, or annotated definitions yielded comparable levels of word comprehension. In situations where learners did not view word definitions; however, word comprehension was not significantly higher on the posttest measures of recall. Attrition occurred whether participants viewed word definitions or not, but it occurred more severely for those words that learners chose not to look up definitions. The more a target word occurred, the better the word was comprehended. Survey results showed that participants preferred dictionary and annotated definitions more than glossed definitions. The results from this study suggest that word treatment is an effective way to comprehend vocabulary. If language learners and teachers lack resources to afford cost-intensive glosses and annotations, they can consider taking advantage of free, automatic online dictionaries, which are as effective as glosses and annotations.

Keywords: word treatment, word definition, dictionary, gloss, annotation, Chinese, advanced learner, learner perspective

ACKNOWLEDGEMENTS

I thank my Heavenly Father who guided me throughout my doctorate education with His infinite wisdom and love. I am very grateful for my family for their unconditional love and example of hard work. I thank Dr. Michael Bush for the excellent mentorship he provided with knowledge and love. I am grateful for my dissertation committee members for their wonderful expertise and feedback. I thank the ARCLITE Lab employees for aiding with my study design, Dr. Eggett for his help with data analysis, those who participated in this study, and friends who helped with editing.

TABLE OF CONTENTS

TITLE PAGE	i
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES.....	viii
LIST OF FIGURES.....	ix
Chapter 1: Introduction	1
Statement of the Problem	4
Past research on word treatments.	4
The dilemma of advanced learners.	5
Significance of the Problem	6
Research Questions.....	7
Definitions	8
Limitations of the Current Study.....	8
Chapter 2: Review of Literature	10
Language Comprehension.....	10
Non-Incidental Language Learning.....	12
Subtitles and Language Learning.....	13
Subtitles and vocabulary learning	13
Interlingual and intralingual subtitles.....	14

Reducing Cognitive Overload.....	15
Word Treatment and Vocabulary Learning	16
Dictionary use.	16
Glosses.	18
Annotations.	19
Word Frequency.....	20
Preferences	20
Advanced Language Learners.....	21
Reading Chinese	22
Summary	23
Chapter 3: Research Design & Procedures.....	24
Overview of the Study Design.....	24
Participants	24
Word as Unit of Observation	25
Materials and Instruments.....	26
Pretest and surveys.	26
Treatment.....	27
Posttests and use preference survey.....	28
Context: Ayamel	29
Video playback.	29
Activity stream.....	32

Data Collection and Analysis.....	33
Chapter 4: Analysis of Data	35
Comparative Effect of Dictionary, Glosses, and Annotations.....	35
Comparative Effect of Treatment and No Treatment	39
Gain over Time.....	41
Effect of Word Difficulty.....	42
Effect of Word Occurrence.....	46
Effect of Video Presentation Order.....	47
Frequency of Usage	51
Preference and Helpfulness	53
Helpfulness of Subtitles.....	54
Chapter 5: Discussion and Conclusions.....	56
Comparative Effect of Three Word Treatment Methods	56
Comparative Effect of Treatment vs. No Treatment.....	59
Effect of Independent Variables on Word Comprehension	60
Word difficulty.	60
Word occurrence.	61
Video presentation order.....	62
Learner Perspective for Word Treatment Methods	63
Implications.....	65

Suggestions for Future Research.....	66
Conclusions	67
References	69
APPENDIX A: LANGUAGE BACKGROUND SURVEY.....	77
APPENDIX B: POST STUDY SURVEY	78
APPENDIX C: RANDOMIZATION DESIGN	79
APPENDIX D: TRANSCRIPT FOR AYAMEL VIDEO TUTORIAL.....	81
APPENDIX E: VOCABULARY TEST ITEMS SAMPLE.....	83

LIST OF TABLES

Table 1. <i>Descriptive Data on the Time-by-Treatment Effect (D, G, A, N, C)</i>	39
Table 2. <i>Descriptive Data on the Time-by-Treatment Effect (T, C)</i>	40
Table 3. <i>Descriptive Data on the Gains over Time (T, C)</i>	42
Table 4. <i>Summary Data of the Word Difficulty Effect</i>	43
Table 5. <i>Descriptive Data on the Treatment-by-Difficulty Effect</i>	44
Table 6. <i>Descriptive Data on the Time-by-Treatment-by-Difficulty Effect</i>	46
Table 7. <i>Descriptive Data on the Word Occurrence Effect</i>	47
Table 8. <i>Summary Data of Video Order Effect</i>	47
Table 9. <i>Descriptive Data on the Treatment-by-Video Order Effect</i>	48
Table 10. <i>Descriptive Data on the Time-by-Treatment-by-Video Order Effect</i>	51
Table 11. <i>Frequency of Clicks on Three Word Definition Types</i>	52
Table 12. <i>The Effect of Frequency and Length of Word Definition View</i>	52
Table 13. <i>Preference and Helpfulness Ratings</i>	53
Table 14. <i>Correlation Data on Preference and Helpfulness Ratings</i>	54

LIST OF FIGURES

<i>Figure 1.</i> Image annotations for the word hit.	4
<i>Figure 2.</i> Second language acquisition process (Chapelle, 1997).	11
<i>Figure 3.</i> Multimedia-based language-learning process.....	12
<i>Figure 4.</i> Example of dictionary definition.	30
<i>Figure 5.</i> Example of glossed definition.....	31
<i>Figure 6.</i> Example of annotation definition.....	32
<i>Figure 7.</i> Line graph showing vocabulary test means (D, G, A, N, C words).	37
<i>Figure 8.</i> Line graph showing vocabulary test means (T, C words).....	41
<i>Figure 9.</i> Line graph showing vocabulary test means (Time-by-treatment-by-word difficulty).	45
<i>Figure 10.</i> Line graph showing vocabulary test means (treatment-by-video order).	49
<i>Figure 11.</i> Line graph showing vocabulary test means (Time-by-treatment-by-video order).	50

Chapter 1: Introduction

Increasingly, more websites and commercial language teaching programs produce language-learning tools with TV programs, videos, podcasts, and other media. Media-based language learning has been attracting second language (L2) researchers' attention since the 1990s (Vanderplank, 2010). More specifically, learning a foreign language by watching subtitled films, thereby increasing the chances for incidental learning to take place, has captured the attention of language learners and second language researchers alike. A number of research studies have found that watching movies with subtitles has a positive effect on language acquisition (Danan, 2004; d'Ydewalle & Van de Poel, 1999; d'Ydewalle & Van Rensbergen, 1987; Garza, 1991; Hayati & Mohmedi, 2011; Koolstra & Beentjes, 1999; Vanderplank, 1999; Vanderplank, 2010; Zarei & Rashvand, 2011).

Watching programs with subtitles/captions has been shown to be beneficial for language learning in studies conducted in countries where English TV programs are prevalent (Koolstra & Beentjes, 1999; Kuppens, 2010; Lommel, Laenen, & d'Ydewalle, 2006; Vanderplank, 1999; Vanderplank, 2010). Among various aspects of second language (L2) learning, vocabulary acquisition has been shown to be highly correlated to the use of subtitled foreign language videos (Koolstra & Beentjes, 1999; Neuman & Koskinen, 1992; Zarei & Rashvand, 2011).

Several researchers, however, also assert that language acquisition does not occur automatically (incidental acquisition) through watching foreign films. Rather, in order to learn L2 through subtitles, viewers must pay more attention to the subtitles and be aware of how they process subtitles (Vanderplank, 1999). Danan (2004) suggests that viewers employ relevant strategies in order to better process subtitles and thus enhance L2 learning.

Given that access to subtitles provides the opportunity for an important activity that precedes acquisition, L2 reading research has important implications for research on language acquisition through subtitled videos. Scholars involved in reading research have been studying the use of various word treatments as a strategy to enhance vocabulary learning, providing solid reasons for the use of word treatment. For example, Coady (1993) argues that there is only a 5–15% probability of learning words in context at first exposure. Other researchers, on the other hand, found that the rate of word acquisition increases when actual objects and imagery techniques are used (Kellogg & Howe, 1971; Terrell, 1986). Likewise, while regular reading only allows a linear approach to text, hypermedia-annotated text provides readers with a more global or multidimensional approach to the text (Martínez-Lage, 1997).

Based on these findings, the growing interest in using word treatment to facilitate L2 vocabulary learning in reading research has been spreading to L2 vocabulary learning via subtitled videos. Considering the low probability of word acquisition at first exposure, researchers hypothesize that appropriate word treatment will enhance word acquisition for students who use video subtitles. As explained in the review of the literature that follows, three methods of word treatment have been studied: online dictionaries, glosses, and annotations. The following definitions for each word treatment type will be useful for understanding the goals and methods of this study:

- Dictionaries: online dictionaries such as WordReference.com or Dictionary.com that enable instantaneous lookup of definition of words.
- Glosses: short definitions or notes (Nation, 1983), or semantic explanations (Lomicka, 1998; Nation, 1983) for unfamiliar words in the specific context where those words appear.

- Annotations: more extensive notes on linguistic or cultural issues.

For example, the sentence, “I am going to hit the party” could confuse English as a Second Language (ESL) learners. In order to help them understand the meaning of this sentence, the word *hit* could be treated in the above-mentioned three ways:

- [Online Dictionary] **hit** /hit/

vb. (**hits, hitting, hit**)(mainly tr)

1. (also intr) to deal (a blow or stroke) to (a person or thing); strike
2. to come into violent contact with: *the car hit the tree*
3. to reach or strike with a missile, thrown object, etc: *to hit a target*

n.

1. an impact or collision
2. a shot, blow, etc, that reaches its object
3. an apt, witty, or telling remark

Note: For the purpose of this document, only the first three definitions from each part of speech were included. To view the complete definition, please refer to <http://www.wordreference.com/definition/hit>

- [Gloss] Hit: attend
- [Annotations] To “hit” informally and figuratively means to “go to” or “attend.”

For example, to “hit the town” means to arrive at a town (or to spend time in town), and “to hit the store” suggests a quick visit to a store (see Figure 1). “To hit” also means “to do something extensively.” For example, “hit the books” means to study hard and to “hit the bottle” means to drink alcohol excessively and habitually.

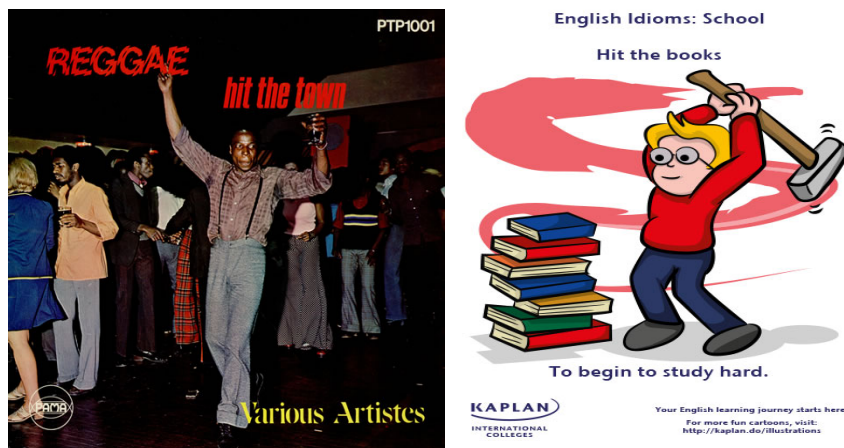


Figure 1. *Image annotations for the word hit.*

Statement of the Problem

Even though word treatment is known to aid L2 learners in attaining vocabulary during media-assisted language learning, there is a gap in the past research on word treatment.

Moreover, advanced language learners face a major dilemma in locating appropriate language-learning materials. The following describes these issues in more detail.

Past research on word treatments. Research in the past few decades on various word treatments, such as glosses, annotations, and online dictionaries shares various commonalities. First, the research has focused on comparing various levels of a treatment or one or two types of word treatment. For example, text-only annotations have been compared with text-plus-pictures annotations as well as with text-plus-video annotations. Likewise, no glossing has been compared with minimal glossing, and full glossing (Chun & Plass, 1996; Jones & Plass, 2002; Lomicka, 1998; Yeh & Wang, 2003). While these studies contain valuable insights, the results from these studies were not unanimous. Moreover, other than one recent study by Todd (2014), there has been no study that compared all the existing methods of word treatments— dictionary lookup, glossing, and annotation—at the same time. Todd’s study with advanced L2 Russian learners found that glosses were the most helpful in aiding Russian vocabulary

comprehension. These findings suggest that, the area of L2 vocabulary learning via word treatment merits more research effort.

Second, past studies on word treatment have focused on research with students who are acquiring languages similar to English, but have mostly neglected the investigation of situations in which learners are acquiring “*truly foreign languages*,” or languages that are more culturally distant and linguistically challenging for native English speakers (Jordan & Walton, 1987). Jordan and Walton define truly foreign languages as “non-Indo-European [languages]... spoken within societies that are culturally in marked contrast to our own” (Jordan & Walton, 1987, p. 111). For example, languages with unique orthography rather than alphabet systems have been neglected in word treatment studies conducted to date except that of Todd (2014).

Third, past research on word treatment and vocabulary learning has not taken into consideration the perspective of the individual learner. For example, no study has compared all three methods of word treatment from the learner’s perspective, which would evaluate preferences of individual learners for, or perceived helpfulness of the methods. Since the assessment of learner preferences is vital in order to create a learner-based curriculum and to aid in the development of learner autonomy, this is a major gap in the research in this area (Bada & Okan, 2000; Barkhuizen, 1998; Nunan, 1987).

The dilemma of advanced learners. Advanced language learners face a dilemma. Today’s media-rich world has plentiful materials available in many foreign languages. However, materials created in accordance with sound pedagogical principles that are suitable for language learners at the upper proficiency levels of the American Council on The Teaching of Foreign Languages (ACTFL) or the Interagency Language Roundtable (ILR) tend to be fewer. As a result, advanced language learners and their teachers are most often left to their own design to

locate and adapt adequate language materials. Dana Bourgerie and Matthew Christensen, former and current Chinese Flagship Program Directors at Brigham Young University claim that this is the case at least for the Chinese language (personal communication, 22 April, 2014). They also add that advanced language learners with a very distinctive focus or interest (e.g., law, computer engineering) are most likely to struggle to find appropriate language resources or may be forced to rely solely on being immersed in the second language speech community, though those opportunities are not always available to students.

The researcher performed a design-and-development project to create the materials (subtitles and word definitions) for this study. A needs analysis conducted as a part of the project demonstrated that finding materials that teach field- or discipline-specific vocabulary and terms are difficult for advanced Chinese students. According to the needs analysis, advanced Chinese students at BYU indeed rely on Chinese textbooks to learn field-specific terms or learn from language instructors who are equally unfamiliar with those terms. Therefore, the researcher concluded that various types of word treatments might assist the students to comprehend accurate and contextualized word definitions.

Significance of the Problem

As mentioned in the previous section, there has not been research effort on the effect of word treatment that compares all three word definition types—dictionary, glossed, and annotated definitions—in a *truly foreign language*. Therefore, an in-depth investigation of word treatment in the Chinese language merits effort in this research area. Additionally, scholars of Chinese as a foreign language found that Chinese vocabulary often poses a challenge for non-native learners in the reading process (Everson & Ke, 1997; Wang & Upton, 2012). Therefore, considering the

distinctive nature of Chinese orthography and the lack of a similar study that comprehensively compares all available definition types, this study makes a distinctive contribution to the field.

In addition to comparing the effect of various treatments of words on the comprehension of Chinese vocabulary, this study also investigated participants' preferences and use of word treatment. By comparing the objective data (use of word treatment) with the subjective preference for various word definition types, the current study contributes to the development of learner-friendly vocabulary learning methods.

Research Questions

The current study sought to complement the findings of past research on word treatment types and vocabulary comprehension and addressed some of the neglected aspects in previous research efforts. A key purpose of the current study is to enhance understanding of the relationship between use of various types of word treatment and the rate of vocabulary comprehension. Because Chinese is one of the *truly foreign languages* for native speakers of English (Christensen & Warnick, 2006; Jordan & Walton, 1987), this study compared the effect of various treatments of words (lexical items) on the comprehension of Chinese vocabulary for L2 Chinese learners.

In addition, this study investigated which word treatment method participants used more frequently or preferred to use. Software was used to keep track of word treatment use, and a survey was administered to the study participants to inquire their preference and perceived helpfulness of treatment types.

Therefore, the research questions that this study sought to answer are:

- What is the comparative effect of using dictionary, glossed, and annotated word definitions on L2 Chinese vocabulary comprehension?

- What is the comparative effect of using word definitions versus not using word definitions on L2 Chinese vocabulary comprehension?
- What type of word treatments among online dictionaries, glosses, and annotations are more frequently used, preferred, and deemed helpful by study participants?

The independent variables for this study are the levels of use of word treatment methods. The dependent variable of the study is the vocabulary comprehension test scores by advanced Chinese learners.

Definitions

In this study, the term *word treatment* has been introduced. Treating words instead of human subjects is an important and innovative aspect of this study. Treating words denotes providing definitions to enhance the learning of words. Three different word definition types, namely dictionary definition, glossed definition, and annotated definition, constitute word treatment in this study.

Limitations of the Current Study

A tool that is gaining popularity among L2 learners via multimedia is a pop-up dictionary (Wang, 2014). Pop-up dictionaries provide more rapid access than conventional online dictionary lookup by providing users word definitions as they hover their computer mouse over unfamiliar words. Even though the software used in this study allows the use of a popup dictionary, this function has been disabled to keep track of use data.

Vocabulary learning is considered to be an essential part of language acquisition, and this study incorporated the measurement of participants' vocabulary retention a week after treatment. However, the scope of this study is limited to vocabulary comprehension rather than vocabulary acquisition. Even though sound vocabulary comprehension could lead to acquisition, equating

long-term vocabulary retention with language acquisition is a tenuous assumption since many other elements besides vocabulary retention contribute to language acquisition.

The following chapter, Chapter 2, lays out the theoretical foundations for the study and describes previous research that provides useful insights for addressing the research questions.

Chapter 3 describes the methodology used.

Chapter 2: Review of Literature

The purpose of the current study is to shed light on the relative effectiveness, the frequency of use, and learner preferences for various word treatments for vocabulary comprehension. This chapter describes the important theories explored in previous research and lays the theoretical foundation for this study. This chapter also helps operationalize important concepts that are key to the study. Pertinent theories and studies included in the chapter also reveal the need for the current study. This chapter first introduces the concepts of language comprehension and non-incidental language acquisition and how they relate to the research questions explored in the study. Next will be a discussion on the effect of viewing subtitled videos on L2 learning and ways to reduce cognitive overload during multimedia learning. I will then provide an overview of the previous research on word treatment and L2 vocabulary learning. Lastly, I will discuss several concepts that pertain to this study, specifically word frequency, unique challenges of Chinese reading, and advanced language learners.

Language Comprehension

In her 1997 book, Gass lays out a model for the L2 acquisition process, which was re-conceptualized by Chapelle (1997) as shown in Figure 2. According to this model, language input is first noticed, termed “apperception,” and then comprehended. Next comes intake, the limited amount of L2 input retained by the learner. Intake then leads to integration into the learner’s linguistic system, finally resulting in L2 output becoming possible (Rast, 2008).



Figure 2. Second language acquisition process (Chapelle, 1997).

Readers should note two things from this language acquisition model. One is that the language acquisition process is quite complex and many factors inhibit acquisition. In fact, a great number of research studies reviewed by Larsen-Freeman and Long (1991) state that input must be comprehensible in order to lead to intake (Larsen-Freeman & Long, 1991). In other words, there is a need to provide L2 input in such a way to help L2 learners receive key elements of the input, thus increasing the chance they will comprehend it. In addition to incomprehensible input, a factor that may also interfere with language acquisition is attrition or memory loss. Watanabe (1997) and Huang (2003) both found that the second language learners' English vocabulary recall declined the sharpest between immediate post-treatment recall and the first delayed posttest. This result is similar to the forgetting curve by Ebbinghaus (1987), which shows a dramatic drop in acquired knowledge within days unless a learner consciously reviews the learned material. Pimsleur (1967) also asserted that learned words would fade from memory unless there is a review process.

Second, it is important for readers to note that language comprehension and acquisition are two different concepts and must be distinguished from each other. As shown in Figure 2 above, the whole six stages constitute language acquisition, while comprehension consists only of the first three steps. The key to readers' understanding of the current study is the observation that it is concerned with the first three steps of this model of language acquisition. Language intake, integration to the system, and L2 output may follow, but these steps in the process are

beyond the main scope of this study. However, one may assume that increased L2 comprehension, which is the focus of this study, precedes and will likely lead to enhanced L2 acquisition over time.

Non-Incidental Language Learning

Another important distinction the current study makes is that it does not address the incidental learning process, which amounts to the acquisition of a foreign language without exerting explicit effort while encountering the target language during activities such as reading or watching subtitled foreign films. While incidental learning has been widely investigated (Kuppens, 2010, Neuman & Koskinen, 1992), I am taking the position for this study that once learners are engaged in learning vocabulary by using word definitions while pausing videos, the learning is no longer incidental. Moreover, the presentation and processing of multimedia information demands a greater level of cognitive effort than what can be deemed incidental, as illustrated in Figure 3 from Mayer and Moreno (2003).

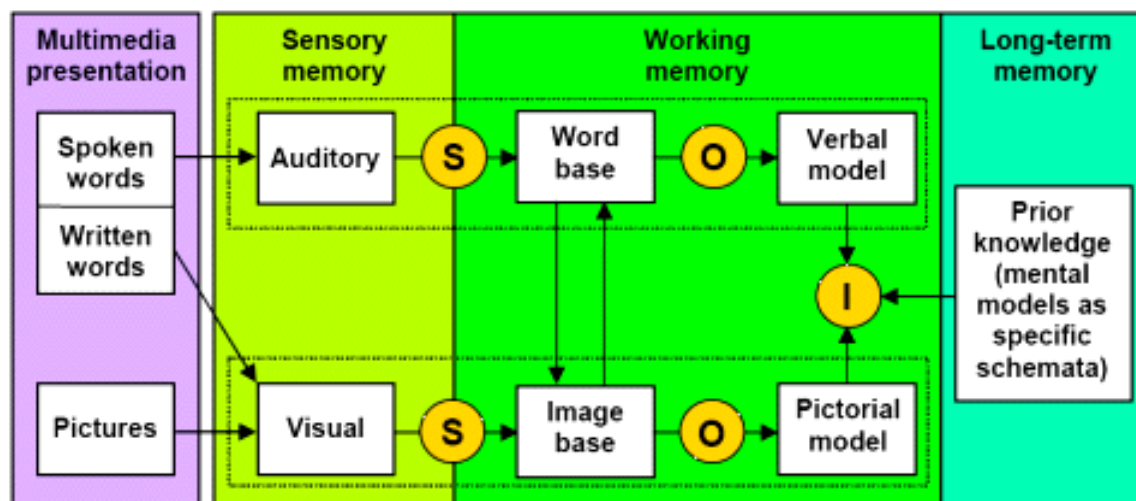


Figure 3. Multimedia-based language-learning process.

For example, creating meaning from words and images alone requires paying attention to many elements such as video images, vocabulary, grammar, sociolinguistic behavior, and

intonation. Therefore, the conclusion that L2 learners can incidentally yet successfully learn vocabulary from subtitled videos seems unlikely, given the video medium's rich nature.

Therefore, the focus of the current study is *deliberate* not *incidental* vocabulary learning. In other words, it assessed the impact on vocabulary comprehension of having users deliberately look up word definitions while watching videos, and not the impact of merely watching videos and observing vocabulary passively.

Subtitles and Language Learning

Some researchers who studied language learning via subtitled videos claimed that text representation confirms the audio input and thus enhance language learning (Danan, 2004; Holobow, Lambert, & Sayegh, 1984; Lambert & Holobow, 1984). Moreover, some aspects of L2 such as pronunciation, intonation, and sociolinguistic competence (how to use socially appropriate language) are difficult to acquire by the use of traditional textbooks alone. TV programs or videos not only model native speaker language, but also provides contexts for language use through video images. They also exemplify task performance in L2.

The means through which the current study participants learned vocabulary items was through subtitles. Therefore, the manner in which subtitles relate to L2 learning is crucial to understand how research on word treatment methods came about. The following section summarizes some noteworthy studies on subtitles and L2 learning.

Subtitles and vocabulary learning. Research on the relationship between subtitles and vocabulary learning reaped the most robust results. Koolstra and Beentjes (1999) studied Dutch fourth and sixth graders' vocabulary learning by randomly assigning them into three groups: (a) English TV with Dutch subtitles, (b) English TV without Dutch subtitles, (c) Dutch TV (control) and showing them assigned 15-minute videos twice in a single session. The English vocabulary

matching test results showed that students in the English TV with Dutch subtitles group did significantly better on vocabulary recognition tests than those students who watched English TV without Dutch subtitles. The English TV without Dutch subtitles group did significantly better than students who only watched Dutch TV.

Neuman and Koskinen (1992) found that watching a captioned TV program could be useful for content learning as well. They studied 192 bilingual seventh and ninth graders who spoke minority languages as a first language (L1). They divided the students into four different learning-type groups: (a) a captioned science production, (b) the same science production without captions, (c) read-along-and listen-to-text, and (d) textbook only. After the treatment, the students were assessed through word recognition, sentence anomaly, and word meaning tests. The results showed that bilingual students who watched the captioned science production did significantly better in acquiring science vocabulary and conceptual knowledge than the other treatment group students. The researchers also found that overall second language (L2) competence of students influenced vocabulary acquisition. The authors inferred that captioned television could be a tool to create environments for comprehensible input, which is beneficial for bilingual language acquisition.

Interlingual and intralingual subtitles. Zarei and Rashvand (2011) investigated the use of intralingual subtitles (the language of the subtitles is the same as the dialogue of the film. i.e. a Chinese movie with Chinese character subtitles) compared to the use of interlingual subtitles (the languages of the subtitles and the dialogue are different. i.e. a Chinese movie with English subtitles) on vocabulary production. They found that intralingual subtitles were more conducive to vocabulary production of 30 ESL learners than interlingual subtitles. In other words, verbatim

subtitles led to significantly better English vocabulary comprehension and production than non-verbatim subtitles.

However, the superiority of intralingual subtitles found by Zarei and Rashvand (2011) might well be dependent upon the proficiency level of the audience. Taylor's study (2005) compared first-year learners of Spanish as a second language with third-or fourth-year Spanish learners. The study showed that lower level learners felt positively toward the subtitles, but found they distracted from focusing on other forms of information (image, audio, etc.). Taylor (2005) commented that added exposure and practice might have helped the lower-level learners in the study to read subtitles and could potentially have led to different results.

Because the current study investigates vocabulary comprehension of *advanced* Chinese learners through watching subtitled videos, it is expected that intralingual or verbatim subtitles are more helpful for their L2 Chinese vocabulary learning.

Reducing Cognitive Overload

As discussed above, the number of elements presented during multimedia learning is likely to limit successful incidental language learning. As shown in Figure 3, multiple elements, such as spoken words (audio), written words (subtitles or captions), and images (video scenes), are presented and compete for cognitive processing during multimedia learning, which in turn may cause cognitive overload. Second language researchers and instructional designers alike suggest ways to minimize the learners' cognitive demands as much as possible.

One way to reduce cognitive overload during multimedia learning is to carefully select and adapt audiovisual materials to learners' language proficiency (Danan, 2004). This suggestion supports Krashen's theory of *comprehensible input*, or $i+1$, which denotes input just above the learner's current level.

Another way that learners reduce overload is to use various cognitive and metacognitive strategies. Thompson and Rubin (1996) recommend that jotting down unknown words during viewing and looking them up later is helpful. Other strategies include relying on known elements (known words or cognates) or predicting based on visual cues.

Word Treatment and Vocabulary Learning

The previous section provided an overview of strategies for lessening cognitive overload, one of which is looking up word definitions. The idea of providing word definitions or treating words with various definitions originally came from reading research, but has been increasingly attracting the attention of researchers involved in video-based language learning as a means to enhance vocabulary learning. The following describes the effect of three known word treatments—dictionary, glosses, and annotations—on L2 vocabulary learning. Sometimes these three methods have been termed interchangeably since the term gloss has a comprehensive meaning (Roby, 1999). The following summaries compare the terms from previous studies with the operationalized definitions for this study.

Dictionary use. Many second language acquisition researchers claim that dictionary use enhances vocabulary learning. Hulstijn (1993) asserts that “a sound reading pedagogy... sees to it that students verify their inferences by consulting an authority, such as a dictionary.” (p. 142). Smidt and Hegelheimer (2004) also argues that “. . . the use of online video in combination with . . . access to further resources in the form of a dictionary, for example, encourage learning beyond the content and help learners acquire language as well, making it a valid pedagogical choice.”

Studies on dictionary use have shown positive impact of referring to online dictionaries on L2 learning. One noteworthy study by Laufer (2000) investigated the vocabulary recall of 55

advanced English learners in Israel. She divided the subjects into two groups, one with online dictionary and one with marginal gloss on paper. The online dictionary group performed significantly better on vocabulary learning tests than the paper gloss group. This study also kept track of students' lookup behavior. Students could choose to view any options—L1 definition, L2 definition, or L2 example. The results showed that though most participants in the online dictionary group used the L1 definition only, those who viewed both L1 and L2 definitions or the L1 definition and the L2 definition along with the L2 example performed significantly better on a retention test.

Time and convenience seem to be important reasons to prefer online dictionaries (e-dictionary) to paper or downloadable dictionaries. Xie (2010) surveyed Chinese learners online on the types of Chinese dictionaries they use in their study, which included downloadable dictionaries, online dictionaries, pop-up dictionaries (accessible both online and offline), and online translators. The survey results showed that out of 303 survey respondents, most Chinese learners preferred online dictionaries due to convenience and the fact that they were free.

The study conducted by Wang (2012) is also one of the few that examined the effect of dictionary use on L2 Chinese learning. He compared 20 Chinese learners' use of an e-dictionary while reading a Chinese expository text online. Ten learners were intermediate Chinese learners and the other 10 were advanced learners. Lookup behavior, which consisted of using the e-dictionary while reading, was observed and recorded. After reading, learners were asked to write the main idea of the text in English. Results showed that the intermediate group looked up middle and high-frequency words while the advanced group did not. However, a retention test also showed that the intermediate group retained the words that had been looked up. The author

concluded that the “e-dictionary effectively closed the vocabulary knowledge gap between intermediate and advanced learners” (p. 483).

Glosses. Several studies have found that the use of glosses aids the learning of vocabulary. For example, Abraham (2008) found that hypertext glosses were more useful for L2 text comprehension and incidental vocabulary learning than not having access to the glosses.

Another study compared the use of glosses and dictionary for incidental vocabulary acquisition among advanced French learners at Dutch universities (Hulstijn, Hollander, & Greidanus, 1996). Testing 16 low-frequency words, the researchers divided the subjects into three groups: marginal gloss, conventional dictionary access, or control groups. Results showed better retention scores for the marginal gloss group than dictionary access group. Although only limited use of the dictionary was reported, when students did use it, the group’s retention was better than that of the marginal gloss group.

The method by which glosses are made available also seems to be of importance. A study by Lyman-Hager, Davis, Burnett, and Chennault (1993) showed superior vocabulary acquisition when instantaneous lookup of glosses is available. The authors compared two groups of L2 French learners, a treatment group with interactive reading software (which enables instantaneous lookup of glosses) and a control group without the interactive software, but with the same glosses available in traditional paper-and-picture format. The students participated in an immediate written recall protocol and an in-class vocabulary quiz a week later after a class discussion. The researchers found that the instantaneous lookup group performed significantly better on a vocabulary quiz of 20 words.

The types of glossing available have been found to be pertinent as well. Lomicka (1998) investigated the relationship between the number and type of glosses consulted and text

comprehension in a pilot study. Using a think aloud procedure, she compared 12 French learners' comprehension of a French poem. The students were divided into three conditions: those who had access to (a) full glossing: "definition, images, references, questions, pronunciation, and translation in English" (Lomicka, 1998, p. 46), (b) traditional glossing: English and French definitions, and (c) no glossing. Results from this pilot study showed that the participants used traditional glossing primarily for the goal of translation and comprehension even when the full annotation was available.

Todd (2014) compared three conditions: dictionary definition, glossed definition, and annotated definition in a study in which 27 Russian learners viewed a 1.5 hour-long Russian movie with Russian subtitles. A fourth of the target words were treated with dictionary definitions, another fourth with glosses, another fourth with annotations, and the other fourth showing blank. The participants did not know what definitions they would view until they clicked on the words. Results showed that the participants comprehended glossed words better than annotated words. Both of those treatments surpassed words for which there was a dictionary definition.

Annotations. Most previous research on annotations has focused on determining which format of annotation is most effective for vocabulary learning. Chun and Plass (1996) examined the effectiveness of a multimedia annotation application called *CyberBuch* created for learners of German. In their study 103 students read a short story on a computer with more difficult vocabulary items that were annotated in the following formats: (a) text only, (b) picture plus text, and (c) video only. Results show that students in the group that had the words with picture plus text annotations obtained higher scores on a vocabulary test than those in the group that had access to the text-only and video-only annotations.

Various other researchers also found that picture plus text was the more effective type of annotation than annotation in text alone (Jones & Plass, 2002; Kost, Foss, & Lenzini, 1999; Yeh & Wang, 2003). Some of these researchers explain that bimodal input of text and image is more effective than a single method of input. The varying types of annotations used in these studies, however, suggest the need for further investigation. Specifically, without comparing one method against another in additional comparison groups, it is hard to judge annotations in which one format (among text, picture, and video formats) is more helpful for L2 vocabulary learning than another.

Word Frequency

Vocabulary learning, the focus of the current study, may be affected by how many times a word appears in a text. Ellis (2002) asserts that the frequent exposure to an L2 word increases the likelihood of remembering that word. The previously mentioned study by Hulstijn, Hollander, and Greidanus (1996) additionally found a positive effect of word frequency on vocabulary retention. Out of 16 target words, 8 of them appeared three times in the text while the other only appeared once. While reading an L2 text with dictionary or glosses, the frequency of word occurrence positively affected word retention.

In the study by Wang (2012) cited above, frequency of Chinese words in Modern Chinese Frequency Dictionary (Language Research Institute, Beijing Language Institute, 1986) was used as a measure of word difficulty. This is similar to claims by Ellis (2002) that the difficulty of an L2 word may be estimated by its frequency of occurrence.

Preferences

Learner preference has for years been an important topic for language curriculum developers and teachers. Nunan (1987) queried immigrant ESL learners in Australia and their

teachers regarding their preferred learning tasks and activities. He found drastic differences in preferences for issues such as error correction and pair work. Nunan argues that learner preference should be an essential element of a learner-based curriculum (Nunan, 1996). Willing (1987) surveyed the preferences of 517 adult ESL learners in Australia. He tried to see if a correlation exists between biographical variables, such as age, ethnicity, level of education, or speaking proficiency level, and preferences for learning activities. Surprisingly, results from interviews and questionnaire showed no significant correlation between any of the biographical variables and learning preferences. However, learners varied in their preferences for a particular sensory mode for learning.

In regards to using learner preferences, Merrill (2000) asserts that learner preferences can be used to fine tune instructional strategies, determined based on content and instructional goals. Additionally, Bada and Okan (2000) claimed that letting learners assess and reflect on their preferences could help them build learner autonomy.

Advanced Language Learners

Yun (2011) investigated the effect of proficiency on how much advanced learners benefit from word treatment. He conducted a meta-analysis of 35 empirical research studies concerning the effect of hypertext glosses on L2 vocabulary learning, and found that studies with low-proficiency learners have significantly greater effect sizes than the studies with high-proficiency learners. In other words, hypertext glosses benefited learners with low L2 proficiency more than for those with high proficiency. However, this result contrasts with the finding by Plass, Chun, Mayer and Leutner (2003), which found that multimedia learning environments cause such high cognitive overload that low-proficiency learners are less likely to benefit from word treatment.

Past research studies also reveal important characteristics of advanced L2 learners that need to be taken into account. Research showed that advanced L2 learners learn most effectively in a learner-initiated learning environment (Byrnes, 2006). Byrnes asserts that this is especially true for adult learners whose volition serves as an important factor for successful learning. In addition to benefiting from autonomous learning settings, advanced L2 learners also effectively acquire language when they set their own goals or have specific learning objectives (Byrnes, 2006). Moreover, several research studies indicate that advanced learners are more motivated to learn when instruction focuses on fulfilling tasks and projects related to their goals (Beckett & Miller, 2006).

Reading Chinese

The Chinese language has a unique character-based, non-alphabetic orthography that English speakers can view as a significant challenge. First, each Chinese character is composed of one or any combination of the following three elements: graphemic component (shape), phonemic component (sound), or semantic component (meaning) (Christensen & Warnick, 2006; Ke, 1996). The end result of these features is that there is no immediate connection between written and spoken Chinese. Second, Chinese characters often combine with other characters and the “meanings of two-character words are related to the meanings of composite characters in complicated ways” (Chao, 1968). In other words, knowledge of a character does not necessarily lead to understanding a composite word in which that character appears. Moreover, written Chinese does not have word boundaries, which means that no space is used between Chinese words to signify word boundaries (Hoosain, 1991; Wang, 2012). In addition, the lack of inflection and existence of many function words is often challenging for learners of Chinese as a

second language (Li, 2005). The end result of these various characteristics is that Chinese presents unique challenges to learners who wish to become proficient in reading the language.

Summary

Language acquisition is a complex process, and language comprehension constitutes a part of it. Though media-aided vocabulary learning, especially through subtitled videos, is found to be helpful, the rich nature of media easily inhibits effective learning. One of the strategies to overcome such cognitive overload is providing definitions to target words in learning material or word treatment. There are three known word treatment methods discussed in second language acquisition research: dictionary definition, glossed definition, and annotated definition.

Though previous research studies reveal that word treatment aids vocabulary learning, the study results do not agree with each other, nor are they complete: except one previous study, there has not been a study that effectively compared all three existing word treatment methods, and the results lack consistency in indicating which word treatment method is most effective in aiding L2 vocabulary learning. Also, there has not been a research study that compared the effect of word treatment methods on L2 Chinese vocabulary learning. Given the difficulty of learning Chinese orthography and reading Chinese materials, this is a major gap in the research. Moreover, the need arose to understand and incorporate learners' preference and perspective on word treatment methods.

After reviewing important and pertinent research in the field of media-assisted vocabulary learning and word treatment, the need for the current study arose to examine the effect of word treatment on L2 Chinese learning. The following chapter describes the detailed study design, which was intended to close some existing gaps in the field, as mentioned previously.

Chapter 3: Research Design & Procedures

The current study sought to investigate whether the use of certain types of word treatments—online dictionary, glosses, and annotations—is related to Chinese vocabulary comprehension. This study also examined the use of word treatment versus no word treatment and whether certain word treatment type(s) were preferred or deemed helpful by advanced Chinese learners. The purpose of the present chapter is to provide an overview of the design and methodology to be used, which includes details on the experimental groups, participants, materials and instruments, and data analysis.

Overview of the Study Design

Participants in the study were advanced learners of Chinese as a second language. Each participant took a language background survey and a vocabulary comprehension pretest and then watched two 20-minute Chinese videos, one about a Chinese doctor and another about a Chinese government official. Each video had L2 subtitles available that contained a total of 80 target words treated with either dictionary definitions, glosses, annotations or no definition. The participants were able to click on the treated words in the subtitles to view dictionary, glossed, or annotated definitions. After watching the two videos, learners took a vocabulary comprehension posttest and a post-experience survey about their preference. One week later, the participants took a vocabulary retention test.

Participants

The main participants for this study were 53 advanced L2 Chinese learners in several U.S. universities or graduates of those institutions. Advanced proficiency was defined as being qualified to be currently taking or having taken upper level Chinese classes (300 level or above). Forty-nine of the participants were native English speakers while 4 of them were non-native

English speakers with English proficiency to pursue higher education in the US. The native languages of those four participants were: Japanese, Korean (two students), and Spanish.

Preliminary analysis of the study results between the 49 native English speakers and the 4 non-native English speakers were not significantly different. Therefore, the non-native English speakers' data have been included. There were 18 males and 35 females. All participants participated online in their chosen places of convenience during the pretest, treatment, and posttest. Participants received detailed instructions online.

In this study, the participants held two roles: first as raters of word treatment methods, and second as autonomous learners with preferences. In order to measure the effectiveness of word treatment in aiding Chinese learners' vocabulary comprehension, the treatment of words was compared with participants' performance on vocabulary comprehension tests. Another role of the participants was as autonomous learners with preferences for word treatment method, which was measured via a post-experience survey, which is included in Appendix B.

Word as Unit of Observation

The role of participants as raters is related to a distinctive characteristic of this study. A key difference with this approach and a conventional experimental design is that instead of having human subjects be divided into treatment and control groups, this study treated *words* as the unit of observation and had the target vocabulary divided into three treatment groups and a control group. Human participants' performance on vocabulary comprehension tests helped explain how word treatment use is related to word comprehension. The participants' use of word treatment methods, the frequency of clicks on word definitions, and viewing times of word definitions, all combine to help measure and rate the effectiveness of the treatment methods from

the users' point of view. This concept is novel and to the researcher's knowledge, the only language acquisition study to have employed this method is that of Todd (2014).

This approach was created in consultation with an experienced statistician in BYU's Department of Statistics and has the advantage of not requiring researchers to find and recruit as many human participants as would be required in a study that implements a conventional experimental design.

Materials and Instruments

The main materials and instruments used in this study were pretest, treatment, posttest, and delayed posttest. The internal reliability of the three tests was measured using Cronbach's alpha. The three tests exhibited high internal reliability: the alpha values of the pretest, posttest, and delayed posttest were 0.924, 0.941, and 0.976, respectively. The following includes more details about the three tests and the treatment.

Pretest and surveys. Prior to the experiment, a language background survey (Appendix A) and a multiple-choice vocabulary comprehension test was administered to assess participants' Chinese-learning experience and knowledge of the target words used in the videos. The pretest required the test takers to match a given definition of an English (or Chinese) word with the best Chinese (or English) word in a multiple-choice format.

Pretests have been found to have an effect on posttest results by exposing the test items or topic to participants (Willson & Putnam, 1982). Even though the pretest effect is considered to be more significant in behavioral science research, the current study acknowledged the bias that it might bring to the experiment. Thus, to control for possible pretest effects, a Solomon Four Group Design was used. This was possible by including only half of the target items in the pretest and all items in posttest (Solomon, 1949).

Treatment. Participants were asked to carefully view two 20-minute videos—the Doctor video and the Government Official video, which are part of *China Professionals Speak* video series. This series was created by the BYU Chinese Flagship Center to help flagship students understand the language and culture of seven different professions in China.

Given that the human participants were advanced L2 Chinese learners, the researcher based the design of the learning experience on the effectiveness of intralingual (verbatim) subtitles on language learning for this level of learners (Koolstra and Beentjes, 1999). The learning experience included the capability to view the videos with simplified Chinese character subtitles.

A total of 80 target words from the two videos have been chosen based on difficulty level set by the Chinese Proficiency Test (汉语水平考试 Hànyǔ Shuǐpíng Kǎoshì) (HSK), which is China's only standardized Chinese proficiency test for this purpose. Wang (2012) used the Modern Chinese Frequency Dictionary. However, since HSK wordlist also considers word frequency and is readily accessible, I decided to use HSK word level to measure difficulty of words. Half of the target words (40 words) chosen from the medium-difficulty level (HSK level 5) and the other half are from the high-difficulty level (HSK level 6) for advanced Chinese learners. The frequency of target words appearing in the video varied.

As mentioned previously, *words* were the unit of observation and analysis in this study rather than participants. Thus each word selected for use in the study received one of the following three treatments or no treatment:

- Dictionary (WordReference.com)
- Glosses (word definition in context)
- Annotations (explanatory notes)

The words were assigned to eight different group of words that contain a randomized selection of word treatments, in which the target words are assigned to one of the four situations (glosses, annotations, dictionary, or no treatment) in a randomized order. The eight randomized conditions were created on an Excel spreadsheet following Latin Square (“Latin square,” 2014). Each condition contained an equal amount of target words, half of them in HSK Level 5 and the other half in HSK Level 6. The randomization design is included in Appendix C.

When a study participant finished the pretest, he or she was assigned to one of those eight different conditions. Half of the treatment conditions involved viewing the doctor video then the government official video while the other half of the treatment conditions consisted of the government official video being shown first, followed by the doctor video.

Each time a user clicked on a treated word, which appeared in yellow, the word definition type was identified followed by display of the definition content. For example, “gloss: (glossed definition)” or “annotation: (annotated definition)”. For words in the no-treatment group, the phrase, “no definition available” appeared.

Posttests and use preference survey. After the online study sessions, the participants completed a vocabulary recall posttest on Qualtrics, which was in the same format as the pretest. Like the pretest, the multiple-choice questions assessed the participants’ vocabulary comprehension. The posttest was different from the pretest in that it assessed participants’ vocabulary knowledge of all of the target words. A short post-experience survey followed the posttest. The survey inquired about learner preferences and perceived helpfulness of the three, word treatment methods. The full survey is found in Appendix B.

After a week, students participated in a delayed posttest to measure their vocabulary retention over time, using the same target words in the same format as the posttest. The order of

the question items in all three vocabulary tests (pretest, posttest, and retention tests) was randomized.

Context: Ayamel

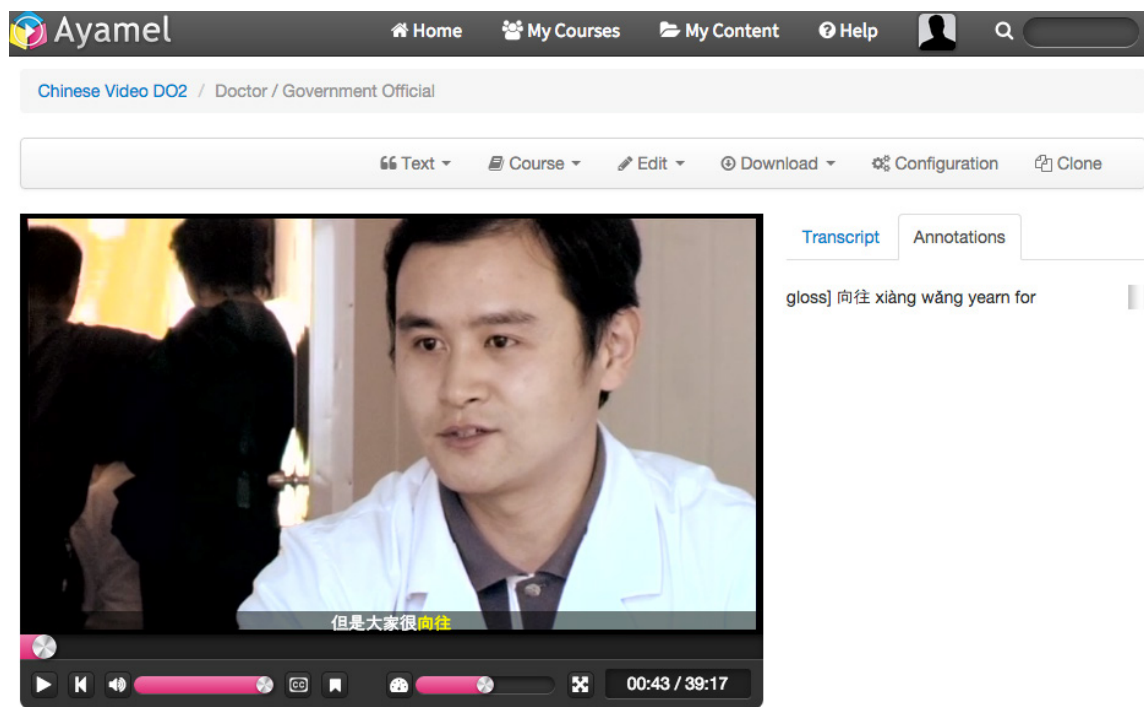
The Ayamel website provided the context for the treatment of the current study. The video playback and word definition technology enabled study participants to access the treatment, while the activity stream technology allowed me to access students' word definition lookups.

Video playback. The current study was conducted via video playback and data recording technology of the Ayamel website created by BYU's ARCLITE Lab (<http://ayamel.byu.edu>). Ayamel's video playback function provided access for learners to the word treatments used in this study: (a) dictionary definitions (WordReference.com), (b) glosses (word definition in context), and (c) annotations (explanatory notes). Since the study participants were not be familiar with the playback capabilities of the Ayamel system, which included word definition functions, they were encouraged to watch a brief video tutorial after taking the pretest to familiarize themselves with the capabilities that are available. A video tutorial transcript that explains the basic functions of Ayamel's video playback is included in Appendix D. As a user clicks on a treated word, which is highlighted in yellow, the video playback automatically stopped and the website led users to view the various versions of word treatment. An example of dictionary definition on Ayamel is captured in the following screenshot (see Figure 4).

The screenshot shows the Ayamel web interface. At the top, there is a navigation bar with the Ayamel logo and links for Home, My Courses, My Content, and Help. Below this is a breadcrumb trail: Chinese Video DO3 / Doctor / Government Official. A secondary navigation bar contains options for Text, Course, Edit, Download, Configuration, and Clone. The main content area features a video player on the left and a right-hand panel with two tabs: Transcript and Annotations. The Annotations tab is active, displaying a dictionary entry for the word '向往' (xiàng wǎng), which is highlighted in yellow in the video player's subtitle. The definition provided is 'dictionary] 向往 xiàng wǎng to yearn for / to look forward to'. The video player shows a man in a white lab coat, and the subtitle at the bottom reads '但是大家很向往'. The video progress bar indicates the current time is 00:43 out of 39:17.

Figure 4. Example of dictionary definition.

The following is an example of a target word treated with a glossed definition on Ayamel (see Figure 5). As was the case for dictionary and annotated definitions, the glossed definitions are shown under the Annotations tab on the right side of the screen.



The screenshot displays the Ayamel web interface. At the top, there is a navigation bar with the Ayamel logo and links for Home, My Courses, My Content, and Help. Below this, a breadcrumb trail reads "Chinese Video DO2 / Doctor / Government Official". A secondary navigation bar contains options for Text, Course, Edit, Download, Configuration, and Clone. The main content area features a video player on the left and a transcript/annotation panel on the right. The video player shows a man in a white lab coat with the subtitle "但是大家很向往". The transcript panel has tabs for "Transcript" and "Annotations", with the text "gloss] 向往 xiàng wǎng yearn for" displayed below.

Figure 5. Example of glossed definition.

Lastly, the following screen shot is an example of an annotated definition shown to the study participants as they watched the videos (see Figure 6).

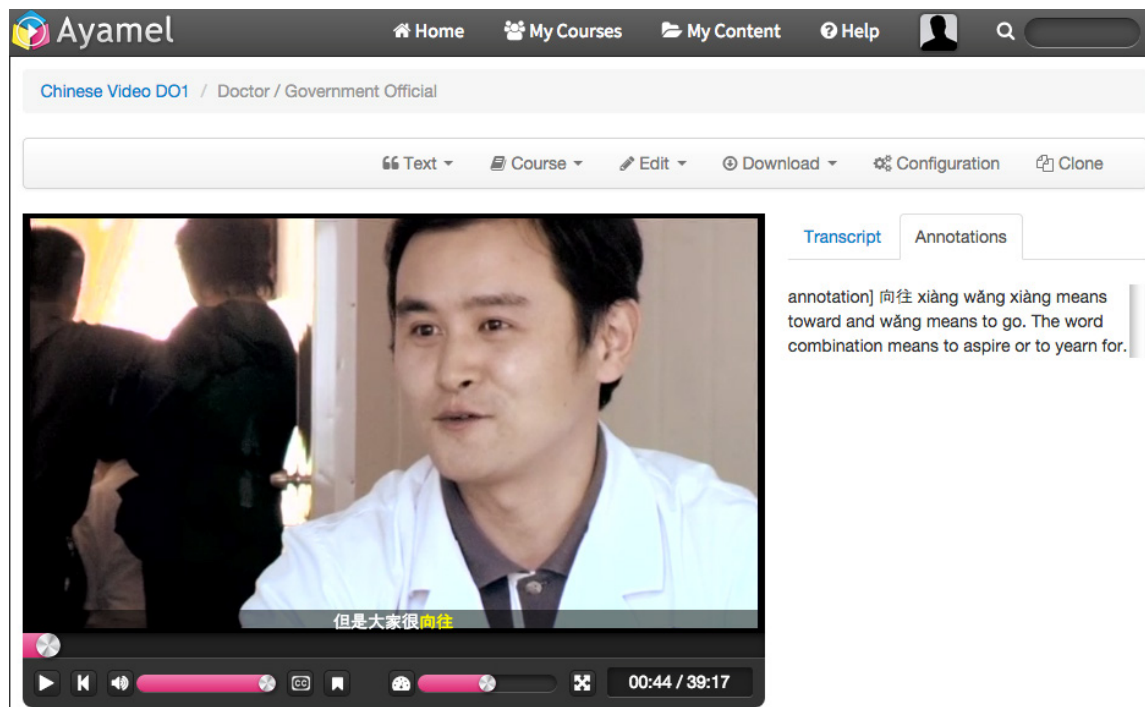


Figure 6. Example of annotation definition.

While using the dictionary definition function on Ayamel, Chinese learners may need to keep the idiosyncrasies of Chinese orthography in mind. Since each Chinese character carries a meaning and Chinese words are often character compounds, recognizing word boundaries can be challenging for learners. The online dictionary function helped recognition of word boundaries by not bringing up non-word definitions or definitions of each character in a compound. The function only brought up the definition of that specific character-compound word.

Activity stream. The activity stream function on Ayamel tracked each student's use of the various word treatments (e.g., when, how long, and how many times a student looked at a word treatment). Activity stream is "list of recent activities performed by an individual, typically on a single website." ("Activity stream," 2014) The incorporation of software to track a learner's Activity Stream in multimedia learning experiences can provide a new tool and valuable insights into L2 learning (Noblitt & Bland, 1991, Cubillos, 1998). For example, as a

learner clicks on a treated word, one can assume that the learner paid attention to that word treatment, however brief that attention might have been. The time that lapses between the click on the word definition until the next click to replay the video can be calculated and assumed to represent the amount of time spent on studying that word definition.

Data Collection and Analysis

Study participants' scores on the pretest, posttest, and retention test items were recorded. Each participant's word treatment use data was tracked through Ayamel's Activity Stream recording capabilities. The Activity Stream data was used to calculate the number of clicks a word definition received and the total lapsed time spent on viewing that word definition for all target words. Students' post-survey responses on preference of word definition and attitudes towards word treatments generated additional subjective data to complement the Activity Stream data.

The dependent variable for this study, the level of recall of target vocabulary definitions, was measured by the participants' performance on the posttest and retention test. The vocabulary test scores for the treatment groups (dictionary, gloss, annotation) were combined for each target word, then the average vocabulary comprehension rate for the groups was compared to each other and to the no definition group to determine any significant differences in word treatment methods as well as the effect of time on recall. Students' post survey responses on self-assessed preference for and helpfulness of the various types of definitions were compared with objective Activity Stream data as well.

The main methods of data analysis were mixed-model ANOVA and regression analysis blocking on word. All comparisons of means were done using post-hoc pairwise comparisons with Tukey-Kramer adjustments for multiple comparisons. In each case the independent

variables were various word treatment methods and the dependent variable is the rate of vocabulary recall. Regression analysis allowed the researchers to assess the effect of the various independent variables that may have affected vocabulary comprehension. The following are some of those variables:

- Frequency of target words appearance: how many times the target words appeared in the video
- Difficulty of target words: HSK levels to which the target words were assigned to
- Order of target video: which of the two videos the participants watched first

Chapter 4: Analysis of Data

Taking the word as the unit of analysis, the study compared the comprehension level of the words that received different treatments or no treatment in pre-, post-, and delayed post-vocabulary tests. In other words, the research questions this study investigated are:

- What is the comparative effect of using dictionary, glossed, and annotated word definitions on L2 Chinese vocabulary comprehension?
- What is the comparative effect of using word definitions versus not using word definitions on L2 Chinese vocabulary comprehension?
- What type of word treatments among online dictionaries, glosses, and annotations are more frequently used, preferred, and deemed helpful by study participants?

In doing so, I addressed the effects of several independent variables such as how many times each target word appeared in the treatment videos, the difficulty level (either medium or hard) of target words, and the order of the two video presentations, on vocabulary comprehension and acquisition.

The following sections show the results of the key inquiries based on the above research questions. After describing the study participants, the sections describe the results for the first, second, and third questions. Answers to the first and second research questions are given by comparing the effects of the three word treatment types and then comparing treatment versus non-treatment. The sections then describe the effects of the independent variables—frequency of target word appearance, difficulty of target words, and the order of target video presentations.

Comparative Effect of Dictionary, Glosses, and Annotations

This section compares the effect of three treatment conditions. Each of the 80 target words received one of the following three treatments or a no-treatment condition: dictionary

definition (D), glosses (G), annotations (A), or No definition available (N). During the data analysis, a statistics faculty advisor and I realized the need to differentiate the act of clicking a word and finding “No Definition Available” from the act of purposely not looking up a word definition. More specifically, *No definition available* is when participants clicked a yellow-highlighted word, but it showed *No definition available* for the study purposes while *no click* means participants did not look up the word definition. Thus, we decided to call the no click action a control action (C).

The following line graph marks the average comprehension level for the five conditions at pretest, posttest, and delayed posttest periods. As shown in the graph, words in the gloss (G) group and the annotation (A) group exhibit very similar patterns of mean comprehension rate rising during the posttest and then dropping in the delayed posttest. Even though words in the dictionary (D) group show a slightly lower level of mean comprehension rate in the pretest, the gain between pre- to posttest is very similar to that of the G and A groups. D group also exhibits a decreased level of comprehension between post-to delayed posttest, except that the decrease is not as severe as those for G and A groups. The no-definition (N) group and no-click (C) group words exhibit a somewhat different pattern from the D, G, and A word groups. The average pretest comprehension rate of the N group words is very similar to that of the D, G, and A groups, while the C group’s mean pretest score is the higher than the rest of the word groups. However, the line that connects the N group’s pretest score to posttest score is less steep than those of the G, A, and D groups and the line for the C group is even less steep than that of the N group. Unlike the other four groups, the C group’s delayed posttest score is lower than its pretest score.

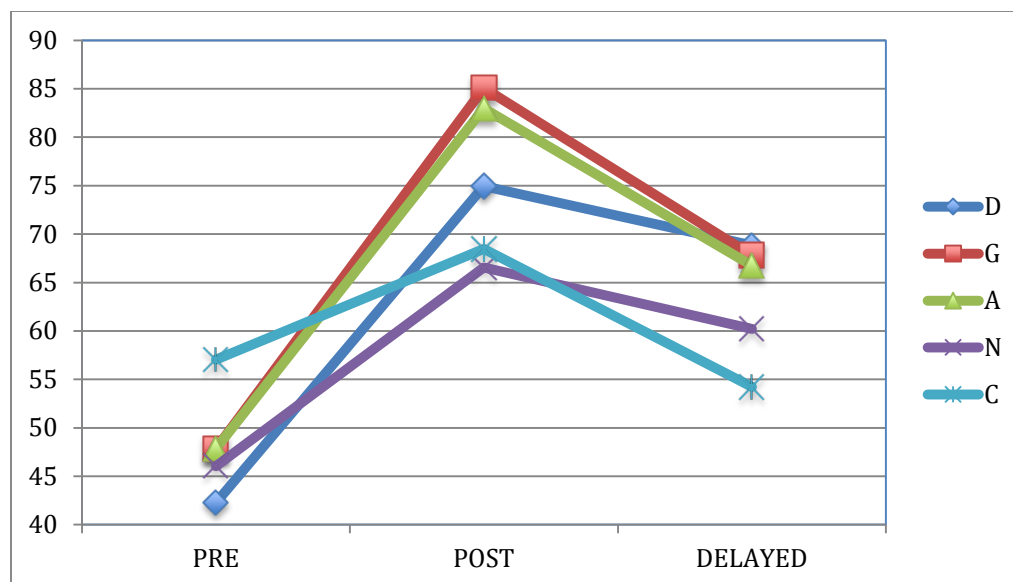


Figure 7. Line graph showing vocabulary test means (D, G, A, N, C words).

Table 1 shows the descriptive statistics—the mean, standard error, and statistical significance of the words—in the D, G, A, N, C groups in pre-, post-, and delayed posttest administrations, organized in D, G, A, N, C order. The letters in the fifth column are used to denote statistical significance: treatment-by-time combinations that share letters are not statistically significantly different from each other ($p < .05$ level). As shown below, the words in the D, G, A, and N groups during pretest all share the letter U, and thus, statistically speaking, the comprehension level of the D, G, A, and N group words during the pretest were not significantly different from each other. The words in the D, G, and A groups during the posttest all share the letter Z, and thus the mean comprehension rates of the three word treatment groups during the posttest were not significantly different from each other. When comparing the pretest letters with the posttest letters, the D, G, A, and N groups did not share letters, but the C group did (VWX and XY). Statistical significance data coupled with the means data showed that the D, G, A, and N groups had significantly higher posttest vocabulary comprehension than pretest

comprehension. On the other hand, the comprehension of the C group words during the posttest was higher than that of the pretest, but the difference was not statistically significant.

During the delayed posttest, D, G, and A words showed significantly different levels of mean comprehension (letters XY) than that of C words (UVW). However, the comprehension of N group words (WX) did not vary significantly from those of D, G, and A words, nor did it vary from that of C group words.

Comparing posttest with delayed posttest administrations, the D groups' delayed posttest score was lower than its posttest score, but was *not* significantly different from it (letters XY and YZ). Likewise, the N group's posttest score was not significantly different from its delayed posttest score (letters XY and WX). On the other hand, the C group's delayed posttest score was significantly lower than its posttest score (letters UVW and XY). The G and A groups' posttest scores were also significantly different from delayed posttest scores (letters Z and XY).

Table 1

Descriptive Data on the Time-by-Treatment Effect (D, G, A, N, C)

Effect	Time	Treatment	Mean	StdErr	Letters*
Time*treatment	Pre	D	42.28227	3.497868	U
Time*treatment	Pre	G	47.86025	3.510772	UVW
Time*treatment	Pre	A	47.8634	3.574122	UVW
Time*treatment	Pre	N	46.05429	3.486768	UV
Time*treatment	Pre	C	57.03785	3.383743	VWX
Time*treatment	Post	D	74.95536	2.724104	YZ
Time*treatment	Post	G	85.09998	2.72162	Z
Time*treatment	Post	A	82.96031	2.721289	Z
Time*treatment	Post	N	66.55826	2.681617	XY
Time*treatment	Post	C	68.46285	2.622041	XY
Time*treatment	Delayed	D	68.83118	2.724104	XY
Time*treatment	Delayed	G	67.81117	2.72162	XY
Time*treatment	Delayed	A	66.74817	2.721289	XY
Time*treatment	Delayed	N	60.19816	2.681617	WX
Time*treatment	Delayed	C	54.23094	2.622041	UVW

Note. Treatment-by-time combinations that share letters are not statistically significantly different from each other ($p < .05$ level).

Comparative Effect of Treatment and No Treatment

In order to answer the second research question of this study, the various treatments and no-treatment effects have been narrowed down to two categories:

- Dictionary, glosses, and annotations = Treatment Group (T)
- No definition available, no click = Control Group (C)

The following table shows the comparative effect of treatment versus no treatment or control at the three test administrations. The mean is the average comprehension rate for the target words. As shown in the table, the comprehension rate of the treatment group in the pretest is lower than the control group, but not significantly (Letters YZ and Z). However, the treatment

group's comprehension rate in the post- and delayed posttests are both significantly higher than those of the control group.

For the C group, the mean posttest score was significantly higher than the pretest score and the delayed posttest score. However, its pretest and delayed posttest scores were not significantly different from each other, sharing the letter Y. For the T group, however, the pretest mean was significantly lower than both the posttest and delayed posttest means.

Table 2

Descriptive Data on the Time-by-Treatment Effect (T, C)

Effect	Time	Treatment	Mean	StdErr	Letters
Time*treatment	Pre	C	51.97	2.74	YZ
Time*treatment	Pre	T	46.09	2.51	Z
Time*treatment	Post	C	67.64	2.18	X
Time*treatment	Post	T	81.04	2.05	W
Time*treatment	Delayed	C	57.14	2.18	Y
Time*treatment	Delayed	T	67.84	2.05	X

The following line chart depicts the means for the T and C conditions during pre-, post-, and delayed posttests. The average comprehension level of target words in the treatment and no treatment groups both increased from pretest to posttest, then decreased in delayed posttest. However, the increase from pre to posttest was greater for the T group.

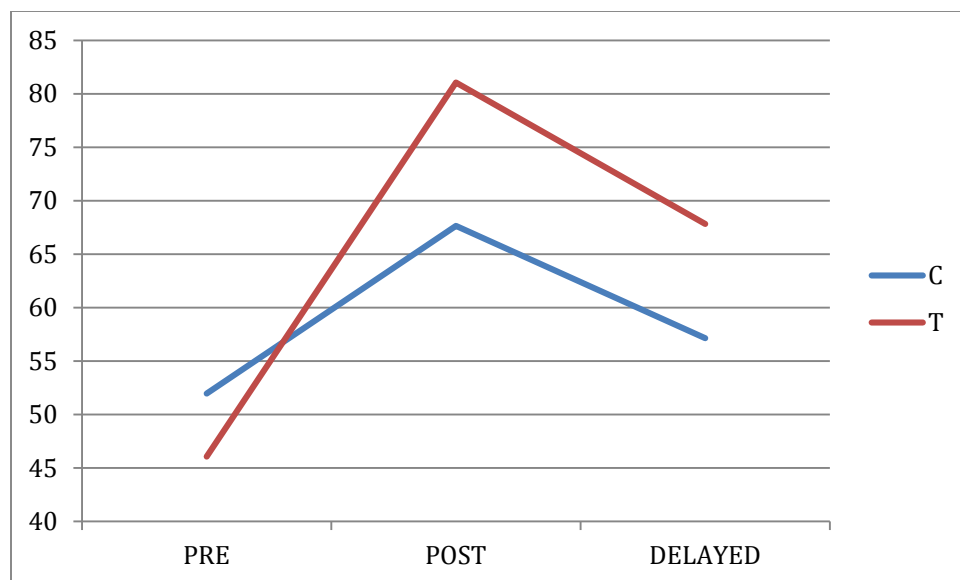


Figure 8. Line graph showing vocabulary test means (T, C words).

Gain over Time

In order to determine the level of vocabulary gain that study participants achieved during their learning experience, I compared the vocabulary comprehension data for treatment and no-treatment group words between the three test administrations (shown in Time). More specifically, pretest word comprehension was subtracted from posttest comprehension for both T and C groups, and they are compared with each other. Likewise, posttest data was subtracted from delayed posttest, and pretest from delayed posttest and then the difference in gains was compared between T and C groups.

The following table summarizes the descriptive data for the difference in gains. The values in the second column (Time Treatment I) were subtracted by the values in the third column (Time Treatment II). As shown below, the gain from the pretest to posttest for the T group words averaged 34.96, while the gain for the C group was less than half of that, at 15.67. As a result, the pre- to post gain for the T group words was 19.29 points higher than the C group words on average and the difference in gain for the two groups was statistically significant ($p <$

.0001). From pre- to delayed posttest, the comprehension level increased for both T and C group words. However, the increase was even greater for the T group than the C group and the difference between the two groups was statistically significant ($p < .0001$). From posttest to delayed posttest, the comprehension level decreased significantly for both groups, but the decrease was not statistically different between the T and the C group words ($p = 0.3038$).

Table 3

Descriptive Data on the Gains over Time (T, C)

Effect	Time Treatment I	Time Treatment II	Mean	StdErr	<i>p</i> -values
Time*treatment	Post T	Pre T	34.96	2.30	<.0001
Time*treatment	Post C	Pre C	15.67	2.66	<.0001
Differences in gains	Post-pre T	Post-pre C	19.29	3.33	<.0001
Time*treatment	Delayed T	Pre T	21.76	2.30	<.0001
Time*treatment	Delayed C	Pre C	15.87	2.66	<.0001
Differences in gains	Delayed-Pre T	Delayed-Pre C	16.58	3.33	<.0001
Time*treatment	Delayed T	Post T	-13.20	1.79	<.0001
Time*treatment	Delayed C	Post C	-10.50	2.07	<.0001
Differences in gains	Delayed-Post T	Delayed-Post C	-2.70	2.63	0.3038

Effect of Word Difficulty

In order to appropriately investigate the comparative effect of word treatment on word comprehension for this study, it was imperative to look at the effect of several independent variables. The first of those variables is the effect of word difficulty on the comprehension rate. As mentioned in Chapter 3, half of the target words were HSK level 5 words and the other half, HSK level 6, which were termed medium and high difficulty words for advanced Chinese learners.

We analyzed the effect of word difficulty alone and its interaction effect with time (over the three test administrations) and the three-way interaction of time-by-treatment-by-word difficulty. The following table summarizes the three effects. As shown below, the effect of

word difficulty by itself was not significant. However, the interaction effect of word difficulty-by-treatment was significant at the $p=0.0169$ level. The effect of difficulty level interacting with three test periods was not significant ($p=0.7697$). However, the interaction effect of word difficulty-by-time-by-treatment was significant ($p=0.0044$). Since the interaction effect of time-by-difficulty is not significant, the descriptive data is not reported.

Table 4

Summary Data of the Word Difficulty Effect

Effect	Number DF	Den DF	<i>F</i> Value	<i>p</i>
Difficulty	1	77	0.14	0.7070
Treatment*Difficulty	5	389	2.80	0.0169
Time* Difficulty	2	156	0.26	0.7693
Time*Treatm*diffic	10	1935	2.56	0.0044

The following table shows descriptive data for the interaction between treatment and word difficulty. For the words in the no-treatment (C) group, the mean comprehension level for high-difficulty (H) words and medium-difficulty (M) words did not vary significantly from each other. Likewise, word difficulty did not have a significant effect on the words in the T group either. However, the average comprehension level of the H words in the C group (mean= 55.72) was significantly lower than that of the H words in the T group (mean= 64.50). Nonetheless, for the M words, the mean comprehension levels for the C group and T group did not differ significantly (letters XY for both).

Table 5

Descriptive Data on the Treatment-by-Difficulty Effect

Effect	Treatment	Difficulty	Mean	StdErr	Letters
treatment*difficulty	C	H	55.72	2.70	X
treatment*difficulty	C	M	63.12	2.71	XY
treatment*difficulty	T	H	64.50	2.60	Y
treatment*difficulty	T	M	65.48	2.62	XY

Figure 9 shows the average comprehension rate of the words for the three-way interaction of treatment-by-difficulty-by-time. As shown in the line graph, the H difficulty words in the treatment group (or T_H words) scored the lowest in the pretest, lower than the H words in no-treatment group (or C_H group words). However, during the posttest, the comprehension of the T_H words increased more drastically than the comprehension of the C_H words. The same pattern appeared for the M difficulty words except the pattern exhibited a more mild difference between T_M and C_M words. In other words, the T_M words showed lesser amount of increase from pre- to posttest than the T_H words, but still greater amount of gain than the C_M words.

In general, all the words, regardless of difficulty or treatment, showed increased comprehension from pre to posttest, then decreased in the delayed posttest. However, the decreases from posttest to delayed posttest in the four groups were almost parallel to each other.

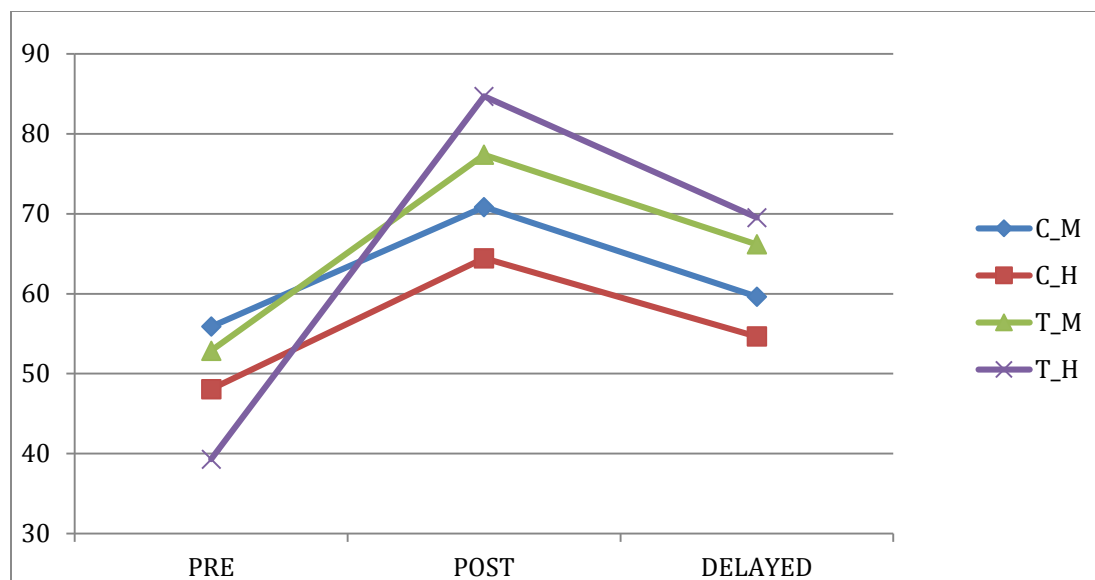


Figure 9. Line graph showing vocabulary test means (Time-by-treatment-by-word difficulty).

Table 6 provides more descriptive data for the interaction between difficulty, treatment, and time, organized by time, treatment, and difficulty. All the pretest mean scores for the four comparison groups were not significantly different from each other, sharing the letter V for statistical significance. For the posttest and delayed posttest, the T_H words' comprehension scores were significantly different from those of the C_H words (letters SUZ and W; TU and SXY). In other words, the comprehension of the treated high-difficulty words was significantly higher than that of non-treated high-difficulty words during post- and delayed posttests. The comprehension of the T_H words during the posttest was significantly different from and higher than those of pre- and delayed posttest. In fact, the treated, high-difficulty (T_H) words during the posttest showed the highest level of comprehension. Likewise, the posttest comprehension of T_M words was also significantly higher than those of pretest and delayed posttest. The comprehension of the C_M words also significantly increased from pre-to posttest, but its delayed posttest comprehension was not significantly different from the pretest comprehension.

Table 6

Descriptive Data on the Time-by-Treatment-by-Difficulty Effect

Effect	Time	Treatment	Difficulty	Mean	StdErr	Letters
Time*treatm*diffic	Pre	C	H	48.06	3.87	UV
Time*treatm*diffic	Pre	C	M	55.89	3.90	STUV
Time*treatm*diffic	Pre	T	H	39.30	3.53	V
Time*treatm*diffic	Pre	T	M	52.87	3.58	UV
Time*treatm*diffic	Post	C	H	64.44	3.08	SUZ
Time*treatm*diffic	Post	C	M	70.84	3.10	WXY
Time*treatm*diffic	Post	T	H	84.69	2.89	W
Time*treatm*diffic	Post	T	M	77.39	2.93	WXZ
Time*treatm*diffic	Delayed	C	H	54.65	3.08	TU
Time*treatm*diffic	Delayed	C	M	59.63	3.10	STU
Time*treatm*diffic	Delayed	T	H	69.51	2.89	SXY
Time*treatm*diffic	Delayed	T	M	66.17	2.93	STY

Effect of Word Occurrence

The effect of how many times target words occurred in the treatment videos was the second independent variable to investigate. The word occurrence for the 80 target words varied, ranging anywhere between 1 and 8. The fixed effects of the frequency of the target word occurrence on comprehension were analyzed. Table 7 sums up the descriptive data from the analysis including the slope and p value. As a continuous variable, the slope of 2.98 was reported. In other words, if one increases the word occurrence by one, the comprehension score is projected to go up by 2.98%. Word occurrence had a statistically significant effect on word comprehension ($p=0.0093$).

Table 7

Descriptive Data on the Word Occurrence Effect

Effect	Slope	StdErr	DF	t Value	<i>p</i>
Word Occurrence	2.98	1.144	1475	2.60	0.0093

Effect of Video Presentation Order

The last independent variable to consider in this experiment was the order of the two videos presented during the treatment. The participants were randomly assigned to either of the two video presentation orders, either doctor video first then the government official video (DO) or the official video then the doctor video (OD). Since the order of the video presentation affected how participants accessed the video subtitles, target words, and various word definitions, the effect of video presentation order was analyzed as an interaction between treatment and time. As the following table summarizes, the video presentation order itself did not have a significant effect on the dependent variable, the comprehension rate of target vocabulary ($p=0.1207$). The two-way interaction between treatment (treatment or no treatment) and video order, as well as the three-way interaction of time-by-treatment-by-video order were significant ($p=0.0309$; $p=0.0070$).

Table 8

Summary Data of Video Order Effect

Effect	Number DF	Den DF	F Value	<i>p</i>
video order	1	79	2.46	0.1207
treatm*video order	1	1475	4.66	0.0309
time*treatm*video order	4	1475	3.54	0.0070

Table 9 contains the descriptive data for the significant treatment-by-video order interaction. The C words viewed in DO order did not vary significantly from those viewed in OD order in comprehension level. However, the C words and T words in DO order had significantly different comprehension rates.

Table 9

Descriptive Data on the Treatment-by-Video Order Effect

Effect	Treatment	Video Order	Mean	StdErr	Letters
treatment*video_	C	DO	58.59	2.12	X
treatment*video_	C	OD	59.24	2.15	X
treatment*video_	T	DO	67.35	1.98	Y
treatment*video_	T	OD	62.63	2.04	X

The following line graph emphasizes the difference in means for the C and T group words interacting with the DO and OD video presentation orders. As shown below, the treated words were comprehended significantly better than the no-treatment words for the DO video order. However, for the OD video order, the treatment or no treatment did not make much difference on participants' comprehension.

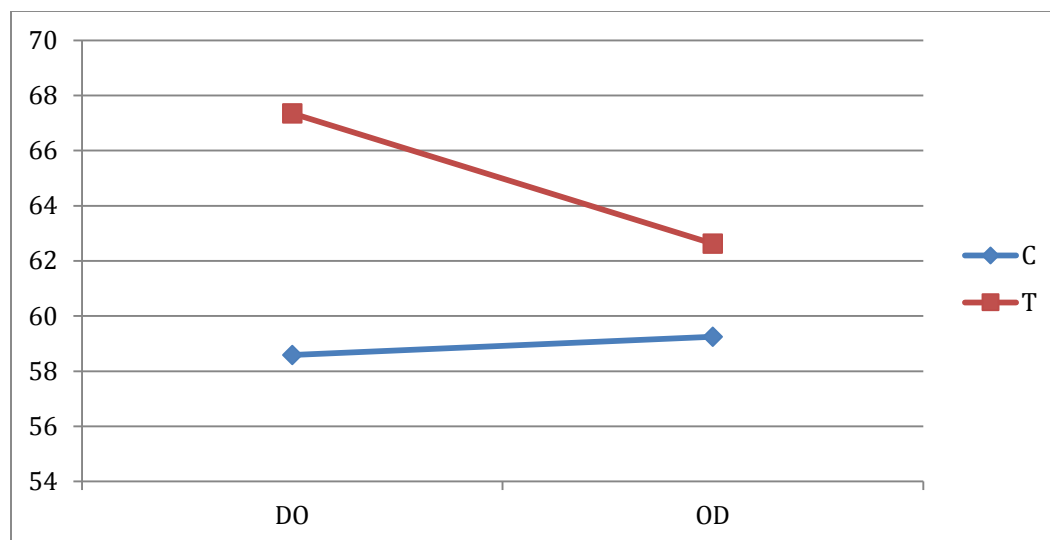


Figure 10. Line graph showing vocabulary test means (treatment-by-video order).

The following line graph shows the three-way interactions of time-by-treatment-by-video order. Treated words in both video orders, DO_T and OD_T, display slightly lower level of comprehension during pretest than not treated words, or DO_C and OD_C. During the posttest, however, the DO_T and OD_T words display greater average comprehension than the DO_C and OD_C words. The loss from posttest to delayed posttest is greatest in the OD_C group words and the smallest in the DO_C words.

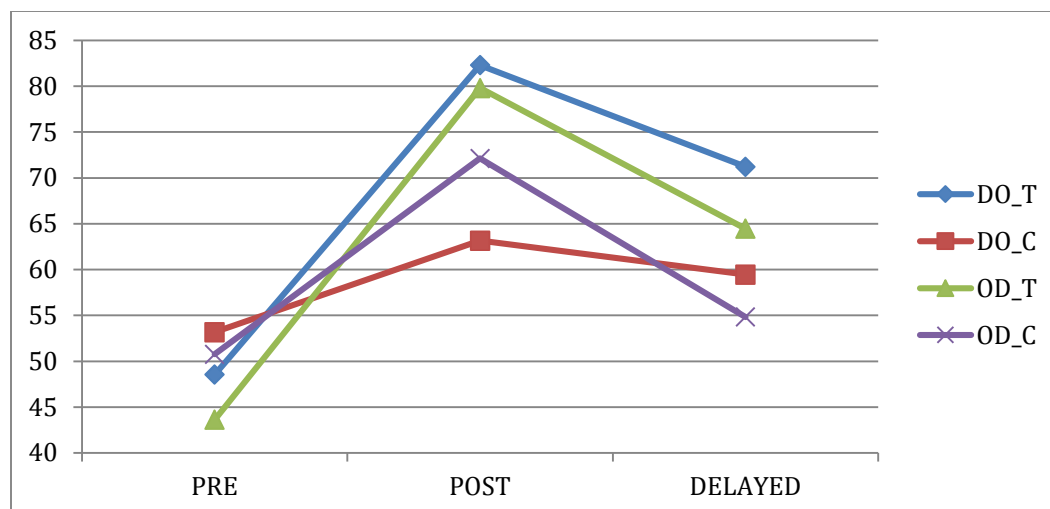


Figure 11. Line graph showing vocabulary test means (Time-by-treatment-by-video order).

The following table includes more descriptive data on the three-way interaction between time, treatment and video order. As the table data suggests, the treated words presented with DO video order during posttest marked the highest level of comprehension, averaging 82.29% of the words. During pretest, no significant difference was found between the four groups (sharing letter V). During the posttest and delayed posttest, the comprehension levels of the non-treated words in both video orders (C_DO and C_OD) were not significantly different from each other (letters RSTZ and XYZ; letters RSTU and STUV). The same pattern appeared for treated words in both video orders, T_DO and T_OD, during the posttest and delayed posttest. The comprehension rate of the C_DO words was *not* significantly different from pre- to post- to delayed posttests. Contrastingly, the comprehension rates of the T_DO as well as T_OD words differed significantly among all three test administrations. The posttest comprehension for the C_OD words was significantly higher than pretest and delayed posttest comprehensions, but the pretest and delayed posttest comprehensions did not differ significantly from each other.

Table 10

Descriptive Data on the Time-by-Treatment-by-Video Order Effect

Effect	Time	Treatment	Video order	Mean	StdErr	Letters
Time*treatm*video	Pre	C	DO	53.17	3.36	TUV
Time*treatm*video	Pre	C	OD	50.77	3.45	UV
Time*treatm*video	Pre	T	DO	48.54	2.96	UV
Time*treatm*video	Pre	T	OD	43.63	3.12	V
Time*treatm*video	Post	C	DO	63.15	2.58	RSTZ
Time*treatm*video	Post	C	OD	72.13	2.63	XYZ
Time*treatm*video	Post	T	DO	82.29	2.32	W
Time*treatm*video	Post	T	OD	79.80	2.44	WX
Time*treatm*video	Delayed	C	DO	59.45	2.58	RSTU
Time*treatm*video	Delayed	C	OD	54.84	2.63	STUV
Time*treatm*video	Delayed	T	DO	71.21	2.32	YZ
Time*treatm*video	Delayed	T	OD	64.47	2.44	RZ

Frequency of Usage

One aspect of the second research question in this study concerns which of the three word treatment types were more frequently used. In order to calculate the frequency of clicks, all the clicks greater than one were calculated from the activity data. When participants first click on a word, the type of word definition he or she will view is purely determined by randomized treatment; thus, the first click was eliminated from the data analysis. As shown in the following table, the most frequently clicked word definition type among the three definition types was gloss, being additionally clicked a total of 194 times. The raw frequency of clicks on dictionary definitions and annotations were very close to each other (179 and 177 additional clicks).

Table 11

Frequency of Clicks on Three Word Definition Types

	Dictionary	Glosses	Annotations
Additional Clicks	179	194	177

Furthermore, I also looked at the relationship between the number of additional clicks on each target word and the change in the comprehension of that word during pretest and posttests. Additionally, I investigated the relationship between the time spent on viewing word definitions and the change in comprehension. In order to do this, only the words that received treatment have been analyzed. In other words, the words that contained no definition (N) or the words that never got clicked (C) were removed from the analysis. I also used the pretest score and word occurrences as covariates. The results showed that the interaction between treatment and number of clicks was not significant, so it has been removed from analysis.

As shown in the following table, the number of clicks on the target word was not significantly correlated to its comprehension. Moreover, the total time spent on viewing word definition and comprehension rate did not have a significant relationship with each other.

Table 12

The Effect of Frequency and Length of Word Definition View

Effect	Number DF	Den DF	F Value	<i>p</i>
Frequency of Definition View	1	1721	0.63	0.4265
Length of Definition View	1	1721	1.49	0.2227

Preference and Helpfulness

Another aspect of the second research area of the current study was to determine participants' perceived preference for and helpfulness of the three word definition types via the post-experience survey. After the treatment and post vocabulary test, the participants were asked to rank the three different types of word definitions, the most, medium, or least preferred or helpful. The following table shows the total number of responses for preference and helpfulness ratings. As shown below, the participants perceived the glossed definition to be the least preferred and helpful.

Table 13

Preference and Helpfulness Ratings

Rating	Prefer D	Prefer G	Prefer A	Helpful D	Helpful G	Helpful A
Most	21	9	24	23	5	26
Medium	19	20	15	18	22	14
Least	14	25	15	13	27	14
Total	54	54	54	54	54	54

An analysis of the correlation between preference and helpfulness ratings using Pearson product-moment correlation coefficient (Pearson's r) revealed that preference and helpfulness ratings for dictionary definitions were significantly correlated with each other ($r=.985$). The preference and helpfulness ratings for glossed definition were significantly correlated ($r=.978$). The same pattern appeared for the preference and helpfulness ratings for annotated definition types ($r=.986$).

Table 14

Correlation Data on Preference and Helpfulness Ratings

		PreferD	HelpD	PreferG	HelpG	PreferA	HelpA
PreferD	Pearson Correlation	1	.985**	.787**	.724**	.915**	.903**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
HelpD	Pearson Correlation	.985**	1	.747**	.663**	.931**	.915**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
PreferG	Pearson Correlation	.787**	.747**	1	.978**	.735**	.699**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
HelpG	Pearson Correlation	.724**	.663**	.978**	1	.648**	.592**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
PreferA	Pearson Correlation	.915**	.931**	.735**	.648**	1	.986**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
HelpA	Pearson Correlation	.903**	.915**	.699**	.592**	.986**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	57	57	57	57	57	57

** . Correlation is significant at the 0.01 level (2-tailed).

Pearson's chi-square goodness of fit test was used to compare the frequency of occurrence of preference rankings for the three definition types. A significant difference was found among the preference rankings for dictionary, gloss, and annotated definitions ($\chi^2(4) = 11.89, p = .018$). Another chi-square test was conducted comparing preference rankings for dictionary and annotated definition types. The result was not significant ($\chi^2(2) = .71, p = .703$) and thus confirmed that any difference between the preference rankings for the dictionary and annotated definition types was by chance. Since the preference and helpfulness rankings were significantly correlated, no further analysis was conducted to compare helpfulness rankings of the three word definition types.

Helpfulness of Subtitles

At the end of the post-experience survey, I also asked participants to provide ratings on the helpfulness of the simplified Chinese subtitles they viewed during treatment. As shown in

Appendix B, it was a 7-point Likert scale ranging from “not helpful at all” to “very helpful.” The average rating on the helpfulness of Chinese subtitles was 6.1 out of 7.

Chapter 5: Discussion and Conclusions

Based on the data reported in Chapter 4, this chapter provides a discussion concerning vocabulary learning via word treatment. In addition to the discussion of the reported data, I also relate these findings to previous research and then draw several conclusions.

The key inquiry of this study was to determine the relationship between use of three types of word definitions and L2 Chinese vocabulary comprehension, as well as to learn whether the treatment affected vocabulary comprehension, and which of those three were more frequently used and perceived as preferred and helpful by study participants. This chapter is structured based on the three research questions of this study:

- What is the comparative effect of using dictionary, glossed, and annotated word definitions on L2 Chinese vocabulary comprehension?
- What is the comparative effect of using word definitions versus not using word definitions on L2 Chinese vocabulary comprehension?
- What type of word treatments among online dictionaries, glosses, and annotations are more frequently used, preferred, and deemed helpful by study participants?

Comparative Effect of Three Word Treatment Methods

As for the comparative effect of the three types of word definitions, dictionary, glossed, and annotated definitions; all had a similar impact on vocabulary comprehension between three test administrations. In other words, there was no significant difference among the three word treatment types on vocabulary comprehension during the pre-, post-, and delayed posttests. All three groups gained in vocabulary comprehension from pre- to posttests, and the gain was statistically significant.

Post- and delayed posttest comparison data, however, shows a different pattern for the dictionary definition group from those of gloss and annotation groups. For the gloss and annotation groups, the delayed posttest comprehension was significantly lower than that of posttest. This phenomenon can probably be attributed to the attrition effect or memory loss due to the passage of time. However, the dictionary group's post- to delayed post attrition was *not* statistically significant. In other words, there was no significant attrition effect for the dictionary group. This finding is consistent with Hulstijn, Hollander, and Greidanus (1996)'s study comparing gloss and dictionary use. When students used dictionary definitions, the vocabulary retention rate was higher for the dictionary group than for the marginal gloss group. As Todd (2014) asserted, dictionary definitions often include other possible meanings of target words besides the meanings used in a specific context (glossed definition). Therefore, comprehending a dictionary definition requires critical thinking in order to be able to extract one specific definition among other possible options of meanings. This critical thinking might have helped participants to retain vocabulary viewed with a dictionary definition better than with a glossed definition or annotated definition, which often provide meanings of relatively smaller scope.

The results from the current study were somewhat different from the findings of the study by Todd (2014). Todd found that glossed definitions were better comprehended than annotated definitions, but the difference in their mean comprehension rates was not significant. However, her findings showed that both the gloss and annotation groups' comprehension levels far surpassed that of the dictionary group, which actually decreased after treatment and was not significantly different from the control group's comprehension level. The differences in the two studies' findings, namely the effect of dictionary definition, may be due to the nature of the online dictionary services that the two studies employed. The difference between the Russian

and Chinese languages also may have contributed to the result. For example, Russian is a highly inflective language while Chinese is not. Thus, dictionary definitions may be much longer for Russian words than Chinese words. Additionally, all the Chinese target words for this study were two-character combinations, which reflect the majority of Chinese words. Therefore, unlike the Russian language, understanding each target word required at least fair knowledge of both of the component characters.

As for the two control conditions, there was no significant difference found between the no-definition group and no-click group words during pre-, post-, and delayed posttest administrations. While the comprehension of the no-click words and the no-definition words both increased from pre- to posttest, the pre- to post gain was significant for the no-definition group, but not significant for the no-click group. While the attrition effect measured by the posttest to delayed posttest was not significant for the no-definition group, it *was* significant for the no-click group.

The pre- to delayed post- comparison data showed that the no-definition group had a significantly higher delayed posttest score than its pretest score. This could mean that wanting to look up definitions of certain words, and thus paying increased attention to those words, still could have reinforced comprehension even though no help came. The research on retrieval-based learning by Grimaldi, and Karpicke (2012) supports this assumption. In their experiment, they compared three conditions: (a) students guessing word pairs before they study them, (b) students not guessing, but just studying the word pairs, and (c) students generating word pairs from fragments of words. The results showed that even attempting to guess or retrieve significantly enhanced learning. Thus, they argued that inference making is an active retrieval process, which promotes learning.

On the contrary, the no-click group had lower delayed posttest scores than the pretest scores, but the decrease was not significant. In other words, when participants did not access word definitions, those words were comprehended less well some time after the experiment than the before the experiment. Thus, participating in the study experience did not benefit the participants when they chose not to look up words.

Comparative Effect of Treatment vs. No Treatment

When D, G, and A conditions were combined into treatment and N and C were combined into control, the effects of treatment and no treatment were manifest more strongly. The treatment and no-treatment group words were comprehended similarly before the treatment. However, the treated words (those for which the participants accessed one of the three definitions) were much better comprehended in the post-treatment test than non-treated words (those for which the definitions have not been accessed or shown). The treatment versus no treatment comparison data supports the previous research findings on the use of word treatment in aiding vocabulary learning (Abraham, 2008; Hulstijn, 1993; Hulstijn, Hollander, & Greidanus, 1996; Lyman-Hager, Davis, Burnett, & Chennault, 1993).

In addition, the words viewed with definitions were much better retained than non-treated words during the delayed posttest. The data showed that the comprehension rate of the treated words was significantly higher than the pretest one, showing that the target words were effectively retained. For the non-treated words, however, delayed posttest scores were not significantly higher than their pretest scores. In other words, when participants did not take advantage of word treatment, those words were not effectively retained.

Effect of Independent Variables on Word Comprehension

The following summarizes the effects of three independent variables of this study: word difficulty, word occurrence, and video presentation order.

Word difficulty. Even though word difficulty alone did not affect the results of the experiment, overall, the interaction between treatment and word difficulty was significant. The study data showed that word difficulty did not have significant effects within the treatment or no-treatment groups. In other words, among the target words whose definitions had been accessed, the difficulty levels, either medium or high difficulty, did not have a differentiating effect. Likewise, the comprehension rate of medium-difficulty words and high-difficulty words did not vary significantly from each other if the word definitions were not shown or accessed. The reason why the word difficulty did not have a significant effect on comprehension within treatment and no-treatment groups could well be because the HSK level 5 (medium-difficulty) and level 6 (high-difficulty) words did not vary enough from each other. In other words, for advanced Chinese learners, HSK 5 and HSK 6 words may not have been different enough with respect to actual difficulty.

Among the high-difficulty word groups, however, the mean score across all treatment groups was significantly higher than that of the no-treatment group. In other words, clicking on and viewing word definitions made a significant difference on the comprehension of words in the high-difficulty group. Whether a word definition was viewed or not (treatment or no treatment), however, did not make a difference for medium-difficulty words. Therefore, one may conclude that looking up a word definition only helped high-difficulty words, but not medium-difficulty words in this experiment. The exact reason why the word definition was more helpful in comprehending the high-difficulty words than the medium-difficulty words is not clear.

However, one guess is that generally speaking, lower difficulty words are usually composed of higher frequency characters and higher difficulty words are made up of lower-frequency characters. As a result, understanding lower-difficulty words may be easier, simply because one could guess the meaning by looking at the component characters without accessing word definition for help. On the other hand, guessing the meaning of high-difficulty words may not be so easy, especially when the component characters are low-frequency characters. In the current study, it is possible that the treatment affected high-difficulty words, but not medium-difficulty words since the composing characters of medium-difficulty words were at the level that participants could guess the meaning of target words and that guessing was just as effective as accessing word definitions.

The three way interaction of time by treatment by word difficulty showed that the non-treated medium-difficulty word group, and two high-difficulty word groups (treated and not treated) all increased in comprehension after treatment and showed a sound level of word retention. The non-treated high-difficulty group words, however, did not gain in comprehension after the treatment, and retention was not significant either. The two treated word groups, treated high-difficulty and treated medium-difficulty groups, also showed that participants comprehended those words more effectively a week after the treatment than before participation. In other words, viewing word definitions aided retaining both medium and high-difficulty target words when compared before and delayed post measurements.

Word occurrence. It was reported that the frequency of word occurrence had a significant effect on vocabulary comprehension. If one increases the appearance of a target word from this study by one, there will be 2.98% increase in its comprehension. This finding confirms the study by Ellis (2002) that frequency of exposure increases vocabulary learning. This finding

also supports the results of Hulstijn, Hollander, and Greidanus (1996) study, which revealed a positive relationship between word frequency and vocabulary retention.

Video presentation order. The order of the two-video presentation did not have a significant effect itself. However, the effects of the two-way interaction of treatment by video order, and the three-way interaction of time by treatment by video order were significant. More specifically, the treatment effect was seen more obviously in the DO video than in the OD video: the treated and non-treated words were significantly different in comprehension when viewed in DO order, while this was not the case when viewed in OD order. Moreover, the video order did not make a significant difference in the comprehension of non-treated words, but for the treated words.

The reason why the treatment effect was not significant in the OD video presentation order is not clear. The participants all viewed both of the videos and each video contained equal numbers of H and M words and words with treatment conditions. It may be due to a fatigue effect or varying interest levels of the two videos. Without much change in the scenes and having native Chinese speakers speaking advanced-level speech almost the whole time, and adding the time to look up word definitions on top of that, the participants might have felt fatigue, which could have adversely affected students' performance. It could be that the doctor video was harder to comprehend and watching it at the end was more challenging than watching it first.

Other than the varying interest levels of the two videos, which is dependent upon personal interest, one other difference was the combined target word frequencies in the two videos. For the doctor video, high-difficulty (HSK level 6) words occurred 33 times and the medium-difficulty ones (HSK level 5) occurred 49 times. For the government official video,

high-difficulty words occurred 28 times and medium ones, 32 times. The relatively lower target word occurrence in the government official video than the doctor video might have resulted in less effective priming of those words, making it difficult to watch the official video first.

Another possible explanation is perhaps a few more high performing students were assigned to the DO group than to the OD group. Post analyses of participants' language background survey responses and their pretest scores revealed that there were about five more people in the DO group than the OD group at the end of the randomized assignment of the groups and experiment. Unfortunately, the exact cause of the observed difference is unclear.

Learner Perspective for Word Treatment Methods

The second research question of this study concerns which of the three word definitions investigated in this study was more frequently used and perceived preferable and helpful by the study participants. Using the Activity Stream data, the frequency of additional clicks on word definitions and total length of the definition view time were calculated. Results from the statistical analyses showed that neither the frequency nor the length of definition view had a significant effect on word comprehension. It is hard to infer the meaning of a greater number of clicks or viewing for a longer period of time. With the type of word definition predetermined for each target word, however, additional clicks could mean a participant did not understand the definition very well on the first click. Likewise, when a participant spends a great length of time viewing a definition, it could mean that the processing of that definition required a longer time, perhaps because the definition was longer or was not very clear or straightforward.

Participants' responses on their post-experience survey showed that they perceived annotated definitions to be the most preferable and helpful. Dictionary definitions closely followed annotation in preference and helpfulness ratings. Glossed definitions were significantly

less preferred and perceived to be less helpful than the dictionary and annotated definitions. Correlation data showed that the participants' preference and helpfulness ratings were significantly correlated. This result is not surprising because language learners would likely prefer language-learning resources that they perceive to be helpful.

It should be noted, however, that the preference and helpfulness were measured by ranking the three types of word definitions. Therefore, the measurement was relative. Had I employed a more objective measurement, such as rating the effectiveness of a definition type in a one to ten scale, the results might have been different.

In the post-treatment survey, however, participants were also asked to briefly provide reasons for their preference and helpfulness rankings. The reasons and opinions varied, but those people who liked dictionary definitions preferred seeing "... the full range of what the word means according to a dictionary definition" (taken from a participant response). Those who liked glossed definitions the most said the glosses were the "simplest and most accurate" (taken from a participant response). Those who preferred annotated definitions the most said that they were "able to see what characters mean independently, which... help... remember the word later" (taken from a participant response). A respondent who gave the lowest ranking for dictionary definitions, however, said that "... I feel seeing all the different plausible meanings is sometimes annoying and make it harder for me to remember the meaning of the word." Some reported they did not prefer glossed definitions as they were too short or simple; there were some who felt annotated definitions were too much information or "[t]wo parts of a word don't necessarily create their meaning together" (taken from a participant response). Reasons for helpfulness rankings were very similar to reasons for preference rankings.

Implications

Accessing the dictionary, glossed, and annotated definitions was equally effective in helping participants to comprehend advanced-level Chinese vocabulary. Moreover, viewing dictionary definitions was positively related to vocabulary retention while viewing glossed and annotated definitions was related to significant attrition. The fact that the dictionary definition was as effective as glossed and annotated definitions has an important implication. An electronic dictionary is usually free of charge and automatic while glosses and annotations are human-developed and thus more expensive time- and resource-wise. If the three word treatment methods are similar in effect, language learners and teachers can consider using more dictionary definitions to learn and teach vocabulary during media-assisted language learning. The survey data by Xie (2010) showed that Chinese learners preferred online dictionaries over paper dictionaries, downloadable dictionaries, and pop-up dictionaries for the cost and convenience.

As a whole, accessing word definitions resulted in a significantly higher comprehension rate of the target words. On the other hand, not viewing word definitions because of no click resulted in a non-significant gain in word comprehension immediately after the treatment and as more time went by. Therefore, the results from this study suggest that word treatment is an effective way to comprehend vocabulary while viewing target language video, which can lead to vocabulary retention. If learners choose to reinforce the retention, the learned materials can integrate into the learner's system and this can lead to vocabulary acquisition, though this pattern is not necessarily guaranteed (Chapelle, 1997, Ebbinghaus, 1987, Pimsleur, 1967).

Some independent variables of this study had significant effects on the word comprehension. High-difficulty words with lookup of definitions resulted in a significantly more effective comprehension while medium-difficulty words did not. Therefore, language teachers

may consider using word treatment to help students learn high-difficulty words. The number of times a target word occurred in the treatment videos significantly affected word comprehension. If a language teacher can control how many times a target word appears in an instruction material, it might be beneficial to expose learners to target words ample amount of time. The order of the two treatment videos that participants watched did not have a significant effect on target vocabulary comprehension itself, but for video order interacting with time and treatment, the treatment did not have a discriminating effect on the treated and non-treated target words viewed in doctor then government official order. Lastly, the number of additional clicks or time spent on viewing word definition time did not affect vocabulary learning. In other words, the frequency and length of definition view did not yield a significant result on the study data. Therefore, instructors should keep in mind that while the frequency and length of clicks on word definition can be an indicator of students' pre knowledge, the lookup behavior is not necessarily related to how well students comprehend.

Participants preferred and found dictionary and annotated definitions more helpful than glossed definitions. The reader should keep in mind, however, that viewing glossed definitions yielded a comparable result in aiding vocabulary comprehension as viewing dictionary and annotated definition types did in the experiment.

Suggestions for Future Research

Given that the study by Todd (2014) and the current study both used the same design to use individual words as the observational unit, yet they yielded slightly different results, further studies should be conducted to compare the effects of dictionary, glossed, and annotated definitions. Others may consider replicating this study with more distinctive word difficulty

groups or with beginning and intermediate Chinese learners instead of advanced language learners.

The treatment videos for this study were interviews with Chinese professionals. Since our videos mostly contained interviews, not many changes of scenes occurred. This could have helped the participants focus more on the language input, but also might have led to an increased level of fatigue or lack of interest. In the future, other studies can use more interesting, story-like videos to gauge Chinese learners' vocabulary comprehension while accessing word definitions. For these reasons, future studies with various video types could also help answer why the interaction effect of video presentation order by treatment had a significant effect on word comprehension.

This study compared three different *types* of word definitions, but not different *mediums* for word definitions. Therefore, another useful study might be one that compares various modes of input, such as text, picture, and video formats. While there are several studies that revealed word definitions with combined mediums were more effective than the definitions with merely a text medium alone, no study exists that compared all modes of input or all possible combinations among various modes of inputs (Jones & Plass, 2002; Kost, Foss, & Lenzini, 1999; Yeh & Wang, 2003).

Conclusions

Through this study, I sought to find whether word definitions help advanced Chinese learners to comprehend words and if so, which of the three definition types performed better than the others. The Chinese language has a complicated orthography for native English speakers, which makes vocabulary learning quite challenging. The findings of the current study confirm that accessing word definitions while watching online video aids vocabulary learning for

advanced Chinese learners. As Smidt and Hegelheimer (2004) argued, “. . . the use of online video in combination with . . . access to further resources in the form of a dictionary, for example, encourage learning beyond the content and help learners acquire language as well, making it a valid pedagogical choice.” Hulstijn (1993) confirms that, “. . .students verify[ing] their inferences by consulting an authority, such as a dictionary could be a sound pedagogical choice” (p. 142). In the future, I hope more language videos will incorporate language-learning resources, not just the videos and subtitles. Glosses and annotations are usually developed by human labor, but online dictionaries are readily available. If owners of video content for language learning cannot afford developing glossed or annotated definitions, making available dictionary definitions from an existing dictionary service could be significantly beneficial to users who wish to comprehend and acquire the target language.

For advanced language learners, learner autonomy is crucial in successful learning. As Bada and Okan (2000) asserted, helping students reflect on their preferences could help them become more autonomous learners. Hopefully, this study could inspire language content developers and teachers to consider learners’ perspective for language resources. Language materials that carefully reflect learner perspective and solid research findings will equip language learners with tools to achieve language proficiency and successful autonomous learning.

References

- Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, 21(3), 199–226.
- Activity stream. (n.d.). In *Wikipedia*. Retrieved January 5, 2014 from http://en.wikipedia.org/wiki/Activity_stream
- Bada, E., & Okan, Z. (2000). Students' language learning preferences. *The Electronic Journal for English as a Second Language (TESL-EJ)*, 4(3), 1-15.
- Barkhuizen, G.P. (1998). Discovering learners' perceptions of ESL classroom teaching/learning activities in a South African context. *TESOL Quarterly*, 32(1), 85-108.
- Beckett, G. H., & Miller, P. C. (2006). *Project-based second and foreign language education: Past, present, and future*. Greenwich, CN: Information Age Publishing.
- Byrnes, H. (2006). *Advanced language learning: The contribution of Halliday and Vygotsky*. London, England/New York, NY: Continuum.
- Chao, Y.R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.
- Chapelle, C. (1997). *Multimedia CALL: Lessons to be learned from research on instructed SLA*. *Language Learning & Technology*, 2(1), 21-39.
- Christensen, M. and Warnick, J. P. (2006). *Performed culture: An approach to East Asian language pedagogy*. Columbus, OH: National East Asian language resource center, Ohio State University.
- Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The modern language journal*, 80(2), 183-198.

- Coady, J. (1993). Research on ESL/EFL vocabulary acquisition: Putting it in context. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 3-23). Norwood, NJ: Ablex Publishing.
- Cubillos, J. H. (1998). Technology: A step forward in the teaching of foreign languages. In J. Harper, M. Lively, & M. Williams (Eds.), *The coming of age of the profession: Issues and emerging ideas for the teaching of foreign languages* (pp. 37-52). Boston, MA: Heinle & Heinle Publishers.
- Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta*, 49(1), 67-77.
- d'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, 28, 227-244.
- d'Ydewalle, G., & Van Rensbergen, J. (1987). Reading a message when the same message is available auditorily in another language: The case of subtitling. In J. K. O'Regan & A. Lévy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 313-321). Amsterdam, Netherlands: Elsevier Science.
- Ebbinghaus, H. (1987) *Translation of Memory: A Contribution to Experimental Psychology*. New York, NY: Dover Publications.
- Ellis, N. C. (2002). Frequency effects in language processing. *Studies in second language acquisition*, 24(2), 143-188.
- Everson, M. E., & Ke, C. (1997). An inquiry into the reading strategies of intermediate and advanced learners of Chinese as a foreign language. *Journal of the Chinese Language Teachers Association*, 32(1), 1-20.

- Gass, S. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Garza, T. J. (1991). Evaluating the use of captioned video material in advanced foreign language learning. *Foreign Language Annals*, 24(3), 239-258.
- Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, 40(4), 505–513.
- Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology*, 42(1), 181-192.
- Holobow, N., Lambert, W. E., Sayegh, L. (1984). Pairing script and dialogue: Combinations that show promise for second or foreign language acquisition. *Language Learning* 34(4), 59–74.
- Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, NJ: Erlbaum.
- Huang, Y. C. (2003). The effects of vocabulary glosses and example sentences on junior high school EFL students' reading comprehension and vocabulary learning. *Unpublished master thesis*. National Cheng-Kung University.
- Hulstijn, J. H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *The Modern Language Journal*, 77(2), 139-147.
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80(3), 327-339.

- Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal*, 86(4), 546-561.
- Jorden, E. H., & Walton, A. R. (1987). Truly foreign languages: Instructional challenges. *The Annals of the American Academy of Political and Social Science*, 490(1), 110-124.
- Ke, C. (1996). An empirical study on the relationship between Chinese character recognition and production. *The Modern Language Journal*, 80(3), 340-349.
- Kellogg, G. S., & Howe, M. J. A. (1971). Using words and pictures in foreign language learning. *Alberta Journal of Educational Research*, 17(2), 87-94.
- Koolstra, C. M., & Beentjes, J. W. 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.
- Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals*, 32(1), 89-97.
- Krashen, S.D. (1985). *The input hypothesis: Issues and implications*, New York, NY: Longman.
- Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, 35(1), 65-85.
- Lambert, W. E., & Holobow, N. E. (1984). Combinations of printed script and spoken dialogues that show promise for beginning students of a foreign language. *Canadian Journal of Behavioural Science* 16(1), 1-11.
- Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. New York, NY: Longman.

- Latin square. (n.d.). In *Wikipedia*. Retrieved April 24, 2014, from http://en.wikipedia.org/wiki/Latin_square
- Laufer, B. (2000). Task effect on instructed vocabulary learning: The hypothesis of 'involvement'. *Selected papers from AILA '99* (pp. 47–62). Tokyo, Japan: Waseda University Press.
- Li, X. Q. (2005). 现代汉语虚词讲义 [On function words in modern Chinese]. Beijing, China: Beijing University Press.
- Lomicka, L. (1998). "To gloss or not to gloss": An investigation of reading comprehension online. *Language Learning & Technology*, 1(2), 41.
- Lommel, S., Laenen, A., & d'Ydewalle, G. (2006). Foreign-grammar acquisition while watching subtitled television programmes. *British Journal of Educational Psychology*, 76(2), 243–258.
- Lyman-Hager, M., Davis, J. N., Burnett, J., & Chennault, R. (1993, March). *Une vie de boy: Interactive reading in French*. Paper presented at the meeting of CALICO, Durham, NC.
- Martinez-Lage, A. (1997). Hypermedia technology for teaching reading. In M. Bush, & R. Terry (Eds.), *Technology-enhanced language learning* (pp. 121–163). Lincolnwood, IL: National Textbook Company.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43–52.
- Merrill, M. D. (2000). Instructional transaction theory (ITT): Instructional design based on knowledge objects. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Nation, I. S. P. (1983). *Teaching and learning vocabulary*. Wellington: English Language Institute, Victoria University.
- 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*(1), 94-106.
- Noblitt, J. S., & Bland, S. K. (1991). Tracking the learner in computer-aided language learning. In B. F. Freed (Ed.), *Foreign language acquisition research and the classroom* (pp. 120-132). Lexington: Heath.
- Nunan, D. (1987). The teacher as curriculum developer. An investigation of curriculum processes within the AMEP. Sydney, Australia: NCELTR.
- Nunan, D. (1996). Towards autonomous learning: Some theoretical, empirical and practical issues. In R. Pemberton, E.S. Li, W.W. Or, & H. D. Pierson (Eds.), *Taking control: Autonomy in language learning* (pp. 13-26). Hong Kong, China: Hong Kong University Press.
- Pimsleur, P. (1967). A memory schedule. *Modern Language Journal, 51*(2), 73-75.
- Plass, J. L., Chun, D. M., Mayer, R.E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*(2), 221–243.
- Rast, R. (2008). *Foreign language input: Initial processing*. United Kingdom: Cromwell Press limited.
- Roby, W. (1999). What is in a gloss? *Language Learning & Technology, 2*(2), 94–101.

- Smidt, E., & Hegelheimer, V. (2004). Effects of online academic lectures on ESL listening comprehension, incidental vocabulary acquisition, and strategy use. *Computer Assisted Language Learning*, 17(5), 517-556.
- Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, 46(2), 137-150.
- Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals*, 38(3), 422-427.
- Terrell, T. (1986). Acquisition in the natural approach: The binding/access framework. *The Modern Language Journal*, 70(3), 213-227.
- Todd, E. C. (2014). *Facilitating vocabulary acquisition through annotations and definitions for 12 learners of Russian- a comparison of the effectiveness of dictionaries, glosses, and annotations*. (Unpublished master's thesis). Brigham Young University, Provo, UT.
- Thompson, I., & J. Rubin (1996). Can strategy instruction improve listening comprehension? *Foreign Language Annals*, 29(3), 331-342.
- Vanderplank, R. (1999). Global medium – global resource? Perspectives and paradoxes in using authentic broadcast material for teaching and learning English. In C. Gnutzmann (Ed.), *Teaching and learning English as a global language: Native and non-native perspectives* (pp. 253–266). Tübingen, Germany: Stauffenberg.
- Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, 43(1), 1-37.
- Wang, J. (2012). The use of e-dictionary to read e-text by intermediate and advanced learners of Chinese. *Computer Assisted Language Learning*, 25(5), 475-487.

- Wang, J. (2014). Reading strategies and reading comprehension retention with and without a pop-up dictionary by beginning learners of Chinese. *CALICO Journal*, 31(2), 244-260.
- Wang, J., & Upton, T. A. (2012). The impact of using a pop-up dictionary on the reading processes of beginning learners of Chinese. *Journal of the Chinese Language Teachers Association*, 47(1), 23-41.
- Watanabe, Y. (1997). Input, intake, and retention. *Studies in second language acquisition*, 19(3), 287-307.
- Willson, V. L., & Putnam, R. R. (1982). A meta-analysis of pretest sensitization effects in experimental design. *American Educational Research Journal*, 19(2), 249-258.
- Xie, T. W. (2010). 电脑词典在中文教学中的应用[The application of electronic dictionaries in the teaching of the Chinese language]. *Journal of the Chinese Teachers Association*, 45(3), 53-65.
- Yeh, Y., & Wang, C. W. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *CALICO Journal*, 21(1), 131-144.
- Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning*, 24(1), 39-58.
- Zarei, A. A., & Rashvand, Z. (2011). The effect of interlingual and intralingual, verbatim and nonverbatim subtitles on L2 vocabulary comprehension and production. *Journal of Language Teaching and Research*, 2(3), 618-625.

APPENDIX A: LANGUAGE BACKGROUND SURVEY

1. What is your native language?
2. How many semesters have you studied Chinese in your home country in high school?
3. How many semesters have you studied Chinese in your home country in college?
4. How many months have you lived in a Chinese-speaking country, speaking and listening to Chinese extensively?
5. How many months have you studied Chinese in an immersive Chinese-learning setting in your home country?
6. Please specify other Chinese learning experience, if any. (Purpose: _____ , Length: _____ months)
7. Please list other foreign languages you speak, if any.

APPENDIX B: POST STUDY SURVEY

1. The following shows the same three word definition types in the videos you watched. Please click and drag them to match the order that you **preferred**. (1: Most preference 2: Medium preference 3: Least preference)
2. Please provide reasons for your preference ranking above. Why did you find one type of word definition most preferable? Why did you find another type medium preferable and the other least preferable?
3. The following shows the same three word definition types in the videos you watched. Please click and drag them to match the order of **helpfulness**. (1: Most helpfulness 2: Medium helpfulness 3: Least helpfulness)
4. Please provide reasons for the helpfulness ranking above. Why did you find one type of word definition most helpful? Why did you find another type medium helpful and the other least helpful?
5. In a 7-point scale, please rate the helpfulness of the Chinese subtitles in understanding the Doctor and Government Official videos
(1: Not helpful at all, 2: Not helpful, 3: Not really helpful, 4: Neutral, 5: Somewhat helpful, 6: Helpful, 7: Very helpful)

APPENDIX C: RANDOMIZATION DESIGN

Word ID	Group	DO1	DO2	DO3	DO4	OD1	OD2	OD3	OD4
H01	H	A	G	D	N	G	A	D	N
H02	H	G	A	N	D	D	N	G	A
H03	H	G	A	N	D	G	A	D	N
H04	H	D	N	A	G	N	D	A	G
H05	H	A	G	D	N	A	G	N	D
H06	H	G	A	N	D	N	D	A	G
H07	H	N	D	G	A	D	N	G	A
H08	H	A	G	D	N	A	G	N	D
H09	H	A	G	D	N	N	D	A	G
H10	H	A	G	D	N	N	D	A	G
H11	H	N	D	G	A	N	D	A	G
H12	H	N	D	G	A	D	N	G	A
H13	H	N	D	G	A	G	A	D	N
H14	H	D	N	A	G	D	N	G	A
H15	H	G	A	N	D	D	N	G	A
H16	H	D	N	A	G	N	D	A	G
H17	H	A	G	D	N	N	D	A	G
H18	H	G	A	N	D	D	N	G	A
H19	H	N	D	G	A	A	G	N	D
H20	H	G	A	N	D	N	D	A	G
H21	H	A	G	D	N	A	G	N	D
H22	H	N	D	G	A	G	A	D	N
H23	H	N	D	G	A	G	A	D	N
H24	H	G	A	N	D	G	A	D	N
H25	H	A	G	D	N	D	N	G	A
H26	H	D	N	A	G	A	G	N	D
H27	H	D	N	A	G	G	A	D	N
H28	H	D	N	A	G	D	N	G	A
H29	H	N	D	G	A	G	A	D	N
H30	H	G	A	N	D	G	A	D	N
H31	H	D	N	A	G	A	G	N	D
H32	H	G	A	N	D	D	N	G	A
H33	H	N	D	G	A	G	A	D	N
H34	H	D	N	A	G	N	D	A	G
H35	H	D	N	A	G	D	N	G	A
H36	H	G	A	N	D	A	G	N	D
H37	H	A	G	D	N	A	G	N	D
H38	H	A	G	D	N	A	G	N	D
H39	H	D	N	A	G	A	G	N	D
H40	H	N	D	G	A	N	D	A	G
M01	M	D	A	N	G	N	G	A	D
M02	M	A	D	G	N	N	G	A	D
M03	M	A	D	G	N	G	N	D	A

M04	M	G	N	A	D	D	A	G	N
M05	M	N	G	D	A	G	N	D	A
M06	M	N	G	D	A	N	G	A	D
M07	M	A	D	G	N	D	A	G	N
M08	M	N	G	D	A	A	D	N	G
M09	M	D	A	N	G	N	G	A	D
M10	M	A	D	G	N	G	N	D	A
M11	M	N	G	D	A	D	A	G	N
M12	M	A	D	G	N	N	G	A	D
M13	M	D	A	N	G	D	A	G	N
M14	M	D	A	N	G	G	N	D	A
M15	M	G	N	A	D	D	A	G	N
M16	M	G	N	A	D	G	N	D	A
M17	M	N	G	D	A	N	G	A	D
M18	M	A	D	G	N	D	A	G	N
M19	M	G	N	A	D	A	D	N	G
M20	M	N	G	D	A	A	D	N	G
M21	M	G	N	A	D	N	G	A	D
M22	M	A	D	G	N	D	A	G	N
M23	M	D	A	N	G	N	G	A	D
M24	M	A	D	G	N	G	N	D	A
M25	M	A	D	G	N	A	D	N	G
M26	M	D	A	N	G	G	N	D	A
M27	M	N	G	D	A	A	D	N	G
M28	M	D	A	N	G	A	D	N	G
M29	M	G	N	A	D	G	N	D	A
M30	M	G	N	A	D	A	D	N	G
M31	M	G	N	A	D	D	A	G	N
M32	M	N	G	D	A	G	N	D	A
M33	M	G	N	A	D	A	D	N	G
M34	M	A	D	G	N	G	N	D	A
M35	M	N	G	D	A	D	A	G	N
M36	M	N	G	D	A	A	D	N	G
M37	M	D	A	N	G	A	D	N	G
M38	M	D	A	N	G	D	A	G	N
M39	M	G	N	A	D	N	G	A	D
M40	M	D	A	N	G	N	G	A	D

Note: In the second column of the above table, H stands for high difficulty, and M stands for medium difficulty. In the first row, DO stands for the condition to view Doctor video first, then Government Official video while OD stands for the opposite.

In the body of the table, G stands for Glosses, A stands for Annotations, and D stands for Dictionary definitions while N stands for no definition available.

APPENDIX D: TRANSCRIPT FOR AYAMEL VIDEO TUTORIAL

Welcome to Ayamel. In this video, you will learn basic, but crucial information on how to use Ayamel to learn your target language.

You can load your subtitles by clicking the cc button. Simplified Chinese subtitles are available for this video.

To help you learn unfamiliar words, some words are treated with either a dictionary definition or glosses (short definitions) or annotations (more extensive notes on linguistic or cultural matters). The words treated with definitions appear in yellow. As you click a yellow-colored word, it will bring either of those three definition types under the annotations tab. Sometimes, the phrase, “no definition available” will show. This does not mean a system error, but appears just for the purpose of this study, so do not worry when you see this. Just move on with learning. Notice that as you view any type of word definitions, video playback will stop automatically and you will be able to view the word definition under the annotations tab. If you want to continue with the video, then click the play button again.

Let me introduce you to some other video play functions on Ayamel here.

The left most button is the play/pause button. The second button from the left is the cool one: play-a-subtitle-line-before button. Let’s say you *just* heard a line of speech, but you didn’t really get it. You can click this button and it will take you back to that exact point of the video where that subtitles line started. You can control the volume of the video here, and do you remember what the cc button was for? Yes, loading subtitles.

Notice that you can control the video playback speed using this button on the bottom right side of the video screen. You can go full screen with this button. The numbers on the right

show how much time has lapsed since the beginning of the video out of the total length of the video.

On the right, you may also choose in what language you will view the transcript. Transcripts will allow you to view more subtitle lines. If you click any subtitle line on the transcript, the video play will take you to that exact spot of the video that the speech occurs.

Now, enjoy learning Chinese words using the Ayamel!

APPENDIX E: VOCABULARY TEST ITEMS SAMPLE

The following is a screen shot of the vocabulary test used to measure vocabulary comprehension for the study. All three vocabulary tests (pre-, post-, and delayed posttests) used the same format with items appearing in a randomized order.

This test is designed to gauge your comprehension of some of the vocabulary items in the videos you watched a week ago. Please *turn off* any electronic popup dictionary functions on your browser in order to accurately measure your knowledge.

For a given **Chinese** word, choose the *best English* equivalent. Likewise, find the best **Chinese** equivalent for a given **English** word (80 items).

to negotiate

访谈



谈论



面谈



谈判



pass through/spend

过度



过于



超过



渡过



to lead to

指导



导演



辅导



导致



差距

bias



messenger



gap



inaccuracy



开放

to depart



to publish



to open up



to launch



to put into practice

实用



实力



实验



实践



值班

to get off work



to work overtime



to take over a job



to be on duty



previously

曾经



经过



未经



历经

