

August 2023

Insights into Chinese Second Language Acquisition: The Relationship between Glossing and Vocabulary Recall in Reading

Steven S. DeVellis
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/masters_theses_2



Part of the [Chinese Studies Commons](#), [Language Interpretation and Translation Commons](#), [Modern Languages Commons](#), and the [Reading and Language Commons](#)

Recommended Citation

DeVellis, Steven S., "Insights into Chinese Second Language Acquisition: The Relationship between Glossing and Vocabulary Recall in Reading" (2023). *Masters Theses*. 1318.
<https://doi.org/10.7275/35624368> https://scholarworks.umass.edu/masters_theses_2/1318

This Open Access Thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

INSIGHTS INTO CHINESE SECOND LANGUAGE ACQUISITION: THE RELATIONSHIP
BETWEEN GLOSSING AND VOCABULARY RECALL IN READING

A Thesis Presented

by

STEVEN DEVELLIS

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF ARTS

May 2023

East Asian Languages and Cultures
Chinese

INSIGHTS INTO CHINESE SECOND LANGUAGE ACQUISITION: THE RELATIONSHIP
BETWEEN GLOSSING AND VOCABULARY RECALL IN READING

A Thesis Presented

by

STEVEN DEVELLIS

Approved as to style and content by:

Zhijun Wang, Chair

Yuki Yoshimura, Member

Elena Chiu, Chinese Program Director
East Asian Languages and Cultures
Department of Languages, Literatures, and Cultures

Anthony Tuck, Chair
Department of Languages, Literatures, and Cultures

DEDICATION

For my Grandfather, who watched me write it.

ACKNOWLEDGEMENTS

I would like to first thank Professor Zhijun Wang for serving as my advisor and for providing me with a bountiful supply of related scholarly articles which helped me shape my research question, especially since I did not have any background in the field myself. Thank you for reviewing my work, and for teaching the Chinese courses which inspired it.

I would also like to thank Dr. Yuki Yoshimura for serving on my thesis committee. Her expertise in the field of second language acquisition was invaluable, and her insight and encouragement during the presentation was indispensable.

Thank you to Feng Laoshi for sharing your language students with me.

Thank you to the 10 students for participating in the study.

Thank you to Shengnan and Xiaoyan for helping me adapt the text for said students.

Thank you to Marc, Shen Laoshi, Chiu Laoshi, Zhang Laoshi, and the rest of the department.

Thank you to all my Belmont Hill and Colgate Chinese teachers who motivated me.

Thank you to my UMass friends for helping me find my sanity after I lost it.

Thank you to my grandparents, aunts, uncles, and cousins for believing in me.

Thank you to my family for raising a Chinese Master.

ABSTRACT

INSIGHTS INTO CHINESE SECOND LANGUAGE ACQUISITION: THE RELATIONSHIP BETWEEN GLOSSING AND VOCABULARY RECALL IN READING

MAY 2023

STEVEN DEVELLIS, B.A., COLGATE UNIVERSITY
M.A., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Zhijun Wang

Reading enhancement strategies are an important tool utilized by foreign language teachers. One of the most common types of during-reading assistance is the marginal gloss, which provides first language (L1) translations of selected foreign language (L2) terms in the margins of a text.

Glossing is an inherently individual approach to reading. It is predominantly motivation-driven, and provides as much assistance to the learner as he or she is willing to use it. Studies to this point have been largely inconsistent in regards to the exact usefulness of glossing, analyzing many variables such as the size, scope, and type of gloss, as well as the language read.

The present study uses glossing while reading a text to investigate whether glossing helps L2 learners learn Chinese and/or helps with their long-term memory. The text was adapted from a pedagogical website for Western learners of Chinese, and the gloss comprised six of the most meaningful terms in the text.

A total of 10 students learning Chinese as a foreign language participated in the present study. These beginner learners were divided into one of two groups, where the only variable was the presence or absence of a marginal vocabulary gloss.

The participants were tasked with reading a short passage in Chinese while responding to comprehension questions in English. Group A read the text without the assistance of a gloss, while Group B read while using a gloss. Participants had access to the text (and gloss, if applicable) while working on the text.

One week later, participants were tasked with remembering the meanings of the selected Key Terms which appeared in the passage. They were also asked to briefly summarize the text in as much detail as they remembered.

In analyzing the responses of the participants after both Part 1 and Part 2, it was found that glossing provided minimal assistance in regards to vocabulary recall in the future, but did much more to aid in comprehension at the time of reading. In addition, L1 translations which appear in the gloss are more likely to be remembered than the L2 terms to which they correspond.

Moreover, this study sheds light on various language acquisition theories, such as incidental vocabulary learning, involvement load, cognitive load, and Input Hypothesis. The results support the efficacy of glossing on during-reading comprehension through the aforementioned theories. Evidence of many of these theories of language processing are evident in the responses of each of the participants and will be analyzed on a case-by-case basis.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
ABSTRACT	v
LIST OF TABLES	ix
CHAPTER 1: INTRODUCTION	1
1.1 Background	1
1.2 Gloss studies	2
CHAPTER 2: THEORETICAL BACKGROUND	6
2.1 Incidental vocabulary learning	6
2.2 Processing theories	9
CHAPTER 3: METHOD	13
3.1 Aims and research question	13
3.2 Participants	14
3.3 Design	18
3.4 Materials	20
3.5 Procedure	26
CHAPTER 4: RESULTS AND OBSERVATIONS	30
4.1 Data collection	30
4.2 Data trends (as observed in Appendix Table 2.A)	33
4.3 Comprehension and retention	37
4.4 Providing given L1 translations	38
4.5 Circled characters during screening	39

4.6 Part 2 summary question.....	40
CHAPTER 5: ANALYSIS AND DISCUSSION	43
5.1 On glossing.....	43
5.2 On incidental vocabulary learning.....	45
5.3 On processing.....	48
CHAPTER 6: CONCLUSION.....	58
6.1 Summary.....	58
6.2 Implications.....	59
CHAPTER 7: LIMITATIONS.....	60
7.1 Study limitations.....	60
7.2 Screening concerns and justifications.....	62
BIBLIOGRAPHY.....	65
APPENDIX	
A Study part 1: reading (Group A, no gloss).....	69
B Study part 1: reading (Group B, gloss).....	70
C Study part 2: memory (both groups).....	71
Appendix tables.....	72

LIST OF TABLES

Table 1. Screening question results.....	17
Table 2. Key Term accuracy percentages in Part 1 and Part 2.....	32
Table 3.1. Part 2 response trends, Group A.....	34
Table 3.2. Part 2 response trends, Group B.....	35
Table 4. Simple t-test.....	36
Table 5. Comprehension and screening results.....	39
Table 6. Key Term, in-text translation output in the Part 2 summary question.....	41
Table 7. Supported theories in the present study.....	57
Appendix Table 2.A. Results of Part 2.....	72
Appendix Table 2.B. Results of Part 1.....	73

CHAPTER 1

INTRODUCTION

1.1 Background

In second language acquisition (SLA), reading enhancement strategies are of vital importance to vocabulary acquisition (Hulstijn, 1992; Nation, 2001; Taylor, 2013). Nation (1983) defines glosses as simply “short definitions.” Pak (1986) uses “explanations of the meanings of words” (Lomicka, 1998). Ko (2005) prefers “information on important words via definitions or synonyms.”

Marginal glossing serves as one of the most common vocabulary intervention techniques utilized to boost second language (L2) reading comprehension. The use of glosses in reading has been studied at length, often with inconclusive or inconsistent findings (Hulstijn, 1992; Jacobs, 1994; Taylor, 2014). The degree to which glosses provide aid to the learner often varies (Taylor, 2013). This is because glosses are a highly individualized tool; learners use the gloss when they need to, and ignore it when they don't (Taylor, 2014).

Parry (1991) shows that even though readers may comprehend unfamiliar words in a text, they may not be able to recall them later. Prince (1996) suggests that in order to comprehend a term, the term must be temporarily “isolated from context” in order to establish individual meaning of the term outside of the larger theme. Glosses serve to create this isolation, directly separating the individual word from the text and providing a translation to relate it to.

This paper will begin with an introduction to several previously performed gloss studies, and will go on to note some of the relevant theories of second language acquisition as they relate to glossing. Next, the results of the present study will be introduced and discussed. Finally, limitations and suggestions for future studies will be proposed.

1.2 Gloss studies

In many studies, it was found that glosses significantly improved L2 reading ability (Hulstijn, 1996; Watanabe, 1997; Bowles, 2004; Alessi & Dwyer, 2008; Taylor, 2013; Duan, 2018). Taylor (2013) found that glossing works in tandem with other variables, such as proficiency level. Taylor (2013) also found that learners with at least one year of instruction made best use of the glosses. In testing advanced learners, Joyce (1997) and Lomicka (1998) did not find those with glosses performed significantly better than those without, suggesting that learners of a lower level benefit more from word interventions while reading.

Ko (2005) notes four major advantages from glossing. First, and most obviously, glosses help L2 readers understand new words more precisely by preventing incorrect guessing. Especially with low-level students, whose vocabulary may not yet be fully optimized for natural reading conditions, glosses provide contextual translations that otherwise may be inaccessible to the learners (Hulstijn, 1992; Nation, 2001). Second, glossing reduces the need for a print dictionary, thus minimizing interruption of the reading process (Nation, 2001). Third, glosses help readers connect prior knowledge with new knowledge from the text. According to Ko, (2005) the reader's interactions with the text and gloss promotes comprehension and retention of the content of the text. Finally, glosses provide learners with a stronger sense of independence. Language learning, as previously mentioned, is a highly individualistic process. A gloss allows each learner to allocate his or her individual attention to words he or she does not know (Nation, 2001; Jacobs, 1994). Ko (2005) also found that about 94% of her participants preferred having access to a gloss while reading.

Taylor (2006) notes several advantages and disadvantages to glossing in the L1. As mentioned, glossing provides faster access to the L2 meanings. Rather than wasting time

searching through a dictionary or inferring from context, an L1 gloss can quickly provide the learner with necessary information. This can lead to an increase in motivation by the learner. Another advantage is glossing allows for more efficient allocation of attentional resources. Learners can devote attention only to words to which they choose. On the flip side, glossing requires minimal effort by the learner, which can result in lower retention. Providing a single translation for a given term will also trap learners into thinking that there is always one perfect translation for a lexical item, which is obviously not the case. Finally, Taylor (2006) notes that L1 should be avoided in L2 studies, as it interferes with smooth L2 reading.

There are several noteworthy types of commonly used glosses. The first and most common type is single-translation glosses, which can be simply explained by Taylor (2014): “first language (L1) translations in the margins of L2 texts.”. The primary benefit of using a single-translation gloss above the other types is its convenience (Hulstijn, 1992; Watanabe, 1997). A print dictionary is often time-consuming and requires the learner to break from the text, which can affect comprehension. A single-translation gloss; however, is readily accessible and can be accessed with a quick glance. Naturally, the downside to this convenience is that the single-translation gloss does not require the allocation of attention, and the glossed term is often skimmed over without proper acquisition (Hulstijn, 1996).

L2 glosses provide the meaning of selected terms in the L2. Owing to a heavy reliance on second language ability, these kinds of glosses are largely reserved for higher-proficiency learners (Ko, 2005). L2 glosses are generally preferred by higher-level readers because they establish a more specific context in the target language than do L1 glosses. Ko (2005) further notes that few studies have been done directly comparing L1 glosses to L2 glosses, citing Jacobs (1994) and Bell & LeBlanc (2000), and notes that findings in learner preference have been

inconsistent as well. Lomicka (1998) found that higher level learners preferred using an L2 gloss, and she theorized this was because the L2 gloss more effectively establishes context in the target language.

Multiple-choice glosses, proposed first by Hulstijn (1992), are another type that have been studied in detail. In a multiple-choice gloss, the correct translation is given along with several incorrect translations, requiring the reader to actively engage the text to infer the correct meaning. Watanabe's (1997) study demonstrated that both single-translation and multiple-choice glosses performed better than no-gloss groups on post-reading comprehension. Hulstijn's (1992) study demonstrated a stronger vocabulary recall by those participants using a multiple-choice gloss, which differed from Watanabe's (1997) study. Duan's (2018) study showed higher performances in incidental vocabulary learning by those with multiple-choice glosses than by those with single-translation glosses. Nagata's (1999) study demonstrates the importance of providing immediate feedback in glossing, noting that a student who learns a term incorrectly will remember it incorrectly. Thus, the major downside of multiple-choice glosses is the absence of immediate feedback and the propensity of learners to make and retain mistakes (Nagata, 1999).

Another kind of gloss that has become much-studied recently is the electronic CALL (computer-assisted language learning) gloss. Bowles (2004) found that learners with a CALL gloss did not perform as well as those with a traditional print gloss, though the result was not statistically significant. The primary difference between CALL and traditional glossing is the hiddenness of the CALL gloss in relation to the text. Whereas a print gloss is readily accessible and occupies the same visual space as the text, the CALL gloss remains hidden until accessed by the user with the cursor. The CALL gloss generally employs hypertext and can theoretically be

accessed for any word in the text, while a print gloss is generally only for a fixed list of key words. Alessi & Dwyer (2008), as well as Taylor (2013), note that CALL glossing is largely dependent on the learner's needs owing to the "invisibility" of the hypertext until manually activated.

The present study will focus on L1 marginal single-translation glosses, which provide direct translations of selected terms in the margin of the text. An L1 gloss was selected for its ease of use, general preference by learners, and for simplicity of the study. Duan (2018) notes as well that single-translation glosses are especially popular in China. A gloss placed directly in the margins of the text ensures that it is noticed by all readers with access to it.

The various aforementioned types of glosses were considered but not implemented into the present study. Due to the low proficiency level of the participants, L2 glosses were not considered. L2 glosses rely heavily on the L2 proficiency of the reader, as they require the reader to describe L2 terms using other L2 vocabulary they are familiar with. Multiple-choice glosses will not be studied either owing to the need for immediate feedback in order to fully optimize reading enhancement. In testing a group of participants, this immediate individual feedback is very difficult to achieve by one researcher without the use of computerized programs. Finally, CALL glosses will not be studied in the present study owing to the technical complexities of creating a computerized gloss.

As will be explained in greater detail in 3.1 Aims and research question, the ultimate goal of the present study is to measure comprehension abilities of Chinese L2 learners under the conditions of *presence* or *absence* of a vocabulary gloss. Single-translation glosses are generally the simplest and most widely-used form of gloss, and were ultimately decided for this reason above all else. Comparison between gloss types changes the nature of the experiment and eliminates the focus on Chinese language in favor of a more general SLA-based approach.

CHAPTER 2

THEORETICAL BACKGROUND

2.1 Incidental vocabulary learning

It is widely accepted that extensive foreign language reading results in improved vocabulary acquisition (Nation, 2001; Hulstijn et al., 1996). Reading with access to a gloss in particular draws the reader's attention to the selected target words which appear in the gloss; however, words in the text that do not appear in the gloss can also be learned by the reader in a process known as incidental vocabulary learning (IVL).

According to second language scholars such as Hulstijn (1989) and Schmidt (1994), incidental vocabulary learning is seen as the "accidental" absorption of non-target vocabulary words while reading a second language text. Read (2004) refers to IVL as acquisition of vocabulary knowledge during an activity not intended for vocabulary instruction. Reading in the L2 can passively boost a learner's vocabulary familiarity through IVL, especially if the acquired vocabulary is dominant in the text.

Incidental vocabulary learning is, by its nature, passive. Incidental learning as a theory is juxtaposed against traditional classroom learning, which heavily emphasizes deliberate studying, rather than casual passivity. Krashen (1982, 1989) equates this passive absorption of language through incidental vocabulary learning to tacit knowledge, or knowledge gained through experience rather than study. Incidental vocabulary learning, like all L2 learning, is driven primarily by learner motivation.

Hulstijn et al. (1996) propose several factors which may contribute to the inability of second language students to immediately acquire the meanings of previously unknown terms encountered in the text. First, learners often simply fail to notice new terms, or mistakenly believe they already know a term. Similarly, learners can ignore words they do not recognize. Context also plays a role in word familiarity. If a context is too redundant, a learner will not focus his or her attention on the individual words, but rather, on the content as a whole. Inversely, the context of the text may be too weak for a learner to properly infer the full meaning of the unknown term. Especially in a longer text, learners may not be properly motivated to frequent a dictionary, which can be time-consuming and difficult to navigate. In the present study, the idea to use Chinese-to-English dictionaries was rejected for this very reason. Finally, according to Nagy, Herman, and Anderson (1985), reading is a “process of small increments;” a single encounter with an unknown term will not necessarily guarantee its familiarity.

Hulstijn (1992) and Watanabe (1997) found that learners remember meanings slightly better when they are inferred by the reader, rather than given. A prominent issue with this theory, of course, arises when a learner infers an incorrect meaning for an unknown term. Knight (1994) found that readers with stronger verbal skills have stronger incidental reading skills as well. In addition, Knight (1994) found that dictionaries and glosses also aid incidental vocabulary acquisition. Finally, Saragi, Nation, & Meister’s (1978) famous *Clockwork Orange* study found that there is a correlation between word frequency and chance of acquisition, though this correlation is not absolute. The Hulstijn et al. (1996) study found evidence to support this claim; learners have a higher chance of acquiring a word incidentally if it appears frequently in the text and is glossed because it will reinforce form-meaning connections, which will be elaborated upon in 2.2 Processing theories.

Incidental vocabulary learning occurs most beneficially when a reader is familiar with an overwhelming majority of the text. According to Laufer's (1989) Lexical Threshold Theory, L2 readers should be able to recognize 90-95% of words in a text for optimal reading conditions to occur. Hu & Nation (2000) find that percentage should be around 98%. The difference of 2-5% should be made up through use of a vocabulary gloss.

As noted by studies such as that of Paribakht & Wesche (1997), a major drawback of incidental vocabulary learning is its uncontrollability; there is no way to account for which words will be learned incidentally, when they will be learned, and for how long they will be remembered. It is also difficult to measure an L2 learner's prior exposure to unfamiliar words. According to Duan (2018), incidental vocabulary learning is dependent on noticing and processing the meanings of unfamiliar characters, along with repetition of the form-meaning mapping.

Because IVL is dependent on noticing, processing, and form-repeating of lexical items in the text, it stands to reason that words with unique, identifying characteristics would be most likely to be noticed. Beginner Chinese learners would likely have difficulty in discerning identifying characteristics of Chinese characters, such as radicals, owing to their lack of experience in decoding the orthographically uniform structure of the Chinese written language compared to alphabet-based languages such as English.

Furthermore, IVL relates to words which are not considered target words, which would exclude any words given in a gloss from observation. Those with access to a gloss cannot learn the target words incidentally; as the meanings are already given, they cannot be passively absorbed from reading. Those without a gloss are unaware of the existence of target words and thus are able to learn them incidentally. A summary question in the present study asks

participants to recall the text, in English, one week after reading. A byproduct of this question is an indirect quasi-measurement of IVL, which will be elaborated upon further in Chapter 4:

Results and Observations.

In addition, IVL is largely dependent on word frequency (Rott, 2005). Terms which recur throughout the text have a higher likelihood of being recognized at a later point in the reading than words which appear once. In the present study, Key Terms are included only once, but other characters are repeated throughout. Because IVL is too unpredictable, comprehension questions are utilized to ensure participants actively understand each target word (referred to as a “Key Term” in the present study), rather than relying on random chance of acquisition. To accurately measure IVL of specific Key Terms, it must be the case that those Key Terms are the only unfamiliar words in the text, which is an unrealistic expectation of natural reading conditions, and they cannot be glossed, which is contradictory to the nature of the study.

Relatedly, Laufer’s (1989) Lexical Threshold theory will be difficult to quantify in the present study. Because Chinese is a character-based language which does not utilize spacing, it is difficult to determine precisely the number of words in a Chinese text. Thus, as in any Chinese text, it is difficult in the present study to establish the exact percentage of unfamiliar words in the text.

2.2 Processing theories

In order for the learner to transform L2 input into L1 output, language processing must occur. This section will analyze several different theories of language processing which have been proposed by various scholars as they relate to language acquisition. The theories will be applied in an attempt to justify the results in Chapter 4: Results and Observations.

Craik & Lockhart (1972) proposed the Depth of Processing Theory in the field of cognitive psychology, which claims the levels of processing depth at the time of learning directly affect memory and recall. Under this theory, the chance a new lexical item will be stored in the long-term memory is determined by the item's shallowness and depth, wherein shallowness refers to the sensory properties of the word and depth refers to semantic-associative features (Rott, 2005). Wittrock (1974) expanded on this idea by proposing a generative model which claimed that learning and retention will be improved if learners can create strong form-meaning connections (FMCs), or connections between old and new knowledge aided by the supplementation of new information; i.e. an L1 gloss (Rott, 2005). Noting the difficulties of quantifying "deep processing," Hulstijn & Laufer (2001) proposed the Task-Induced Involvement Load Hypothesis.

The focal point of the Task-Induced Involvement Load Hypothesis, naturally, is the involvement load. The involvement load describes the "motivational and cognitive dimensions of a task that simulates word processing to establish and retain FMCs" (Rott, 2005). The involvement load is observable and measurable; it employs a motivational-cognitive approach, comprising three main components: need, search, and evaluation (Hulstijn & Laufer, 2001). First, there must be a need for a word. In the L2, this corresponds to the need to translate an unfamiliar term, which leads to the second component: the search for its meaning. Readers can use the context to determine the meaning and syntactic function of an unknown word. Finally, readers must engage in evaluation to determine whether or not the meaning and function selected by the reader to represent the word are indeed correct. Reading tasks which utilize all three components of involvement load processing are more likely to strengthen FMCs. Words which are processed with higher levels of involvement loads are more likely to be retained by the reader

than words which are not (Hulstijn & Laufer, 2001). Glossing increases involvement load at the time of reading by creating the “need,” “search,” and “evaluation” aspects for certain terms.

Sweller’s (2005) Cognitive Load Theory is a related theory of processing. Cognitive Load Theory suggests that human processing is limited; as such, during-reading vocabulary assistance can be beneficial in reducing some of the language processing barriers and freeing the learner to devote the saved attention somewhere else. Alessi & Dwyer (2008) reference three types of cognitive loads which may affect a learner’s reading comprehension ability: intrinsic, germane, and extraneous cognitive load. First, intrinsic cognitive load is the unavoidable difficulty associated with the intrinsic complexity of the material. In the case of Chinese L2 reading, the text itself represents intrinsic cognitive load owing to the innate difficulty of foreign language reading. Germane cognitive load relates to the learning strategies or activities which facilitate learning, such as the act of translating the Chinese text with a gloss. Germane cognitive load is considered good because it benefits learning. The use of a gloss contributes to germane cognitive load because it facilitates the reading process for the learner. Finally, extraneous cognitive load refers to those activities which are not necessary for learning to occur, such as looking up an unknown term in a print dictionary. As mentioned, this process is time-consuming and breaks the reader’s concentration and processing. Extraneous cognitive load is detrimental to learning and should be avoided, when possible. L1 glossing in L2 reading reduces cognitive load, allowing the reader to devote his or her finite processing power to areas most in need of it.

On the topic of attentional allocation, several theories involving noticing the input are also relevant. Schmidt’s (1990) Noticing Hypothesis, claims that input must be noticed by the reader in order to be acquired. Furthermore, the input must be comprehensible by the reader. Input modification can be performed to draw the reader’s attention to particular words in the text

and can be achieved by highlighting or underlining target words or by isolating the word from the text, i.e. marginal glossing. Schmidt's (1990) Noticing Hypothesis is likewise consistent with Hulstijn's (1992) Mental Effort Hypothesis, which notes that information learned with more mental effort will be more likely to be recalled in the future. In the aforementioned theories of processing and noticing, attention is a finite resource to aid in acquisition which can be saved and/or expended at the discretion of the learner.

Taylor (2014) notes the importance of comparing the L2 to the L1 in language acquisition. Taylor also draws attention to Swain's (1995) "noticing the gap" principle during the comparing process, which occurs when a learner recognizes the difference between his or her own comprehension capabilities and what is expected at the global level, is relevant in comparing. Consulting a gloss is an example of a learner "noticing the gap;" if the learner is unfamiliar with a term (a gap in the learner's knowledge), he or she can consult a gloss to fill the gap with new information. Comparing occurs when using a gloss as well, as a marginal gloss is a direct link between the L2 and the L1, and generally occurs during or after noticing (Taylor, 2014). Acquisition occurs when the attended input, i.e. the glossed item, is properly integrated into the learner's output.

Finally, Krashen's (1982) Input Hypothesis, like Schmidt's (1990) later Noticing Hypothesis, functions on the central claim that "competence in spelling and vocabulary is most efficiently attained by comprehensible input in the form of reading" (Krashen, 1989). The Input Hypothesis assumes that a learner acquires language by understanding messages. In that respect, language is primarily subconsciously acquired; a learner will focus on the content of the L2 text rather than the form or structure. Krashen (1989) notes there are some aspects of language which can be consciously studied and learned, but this capacity is severely limited because it utilizes

parts of the brain not associated with language. The Input Hypothesis predicts that an increase in input will result in an increase in comprehension. Like many other theories of processing, the Input Hypothesis requires the learner to notice and comprehend the input in order for acquisition to occur. Glossing facilitates this act of noticing by highlighting important and/or unfamiliar items in the text.

In the present study, the Chinese text serves as input. Six Key Terms in the text are of significance to the study and are glossed. In terms of processing, Schmidt, Swain, Hulstijn, etc. stress the significance of noticing and comparing. Translation is, by nature, a comparison, as it involves noting similarities (and differences) between two languages, specifically in regards to establishing equivalent meanings. Krashen's (1982) Input Hypothesis will be well served by an increased attention to form while reading. Simply increasing input will likely not increase comprehension if the input is neglected by the reader.

CHAPTER 3

METHOD

3.1 Aims and research question

Given the unevenness of results in prior gloss studies, this study aims to add to the existing literature. Originally, the study was intended to be more statistical; however, the present study, owing to the small number of participants, will be largely observational. The purpose is thus to gain insight into the unique thought processes behind students who utilize glosses while reading versus those who do not, and to speculate which theories may have been responsible for the results. This paper will lean more into observing the individual nature of L2 reading rather

than establishing new rules or theories relating to the practice, which has already been studied at length.

While there exist several types of glosses, the present study will focus on the most straightforward of them, the marginal single-translation gloss. The goal is to analyze gloss use in relation to vocabulary acquisition and retention after one week, leading rise to the following research questions:

- 1. How does utilizing a gloss affect comprehension during and immediately after reading a Chinese L2 text?*
- 2. Does utilizing a gloss at the beginner level affect L2 translation memory after one week?*
- 3. Which of the aforementioned theories of language acquisition, if any, apply to beginner Chinese L2 readers? Does one theory have stronger explanatory power than others?*

3.2 Participants

a. Source

Twelve participants contributed data to this study, all of whom were students in their first year of studying Chinese at University of Massachusetts Amherst (four sourced from *Chinese 120, Non-Intensive Elementary Chinese II*; eight sourced from *Chinese 126, Intensive Elementary Chinese I*). A formal proficiency test was not given owing to the even level makeup of each language course. The students were all college-age and none had a native language background. The participants all possessed equal motivation but differed in confidence.

None of the participants were required to participate; all volunteered their time and effort to the study willingly. The researcher worked in tandem with the professor to advertise the study and students were compensated for their help with snacks. It was important that the participants be entirely willing, as a mandatory reading experiment outside of class time would undoubtedly demonstrate a lack of motivation in many of the participants, the consequence of which would be rushed performances and tarnished results. A willing participant would be more likely to read and translate the way he or she naturally would, using whichever reading strategies and as much time as he or she would normally require. To this end, the participants were able to schedule and perform the experiment at a time convenient to them, resulting in four groups in different time slots over the course of two days.

Most importantly, participants were not told about the existence of a Part 2 to the study in order to avoid any intentional or unintentional studying of the text. As a result, two participants dropped out of the study without completing Part 2, bringing the total number of full participants to 10. Both withdrawn participants were from the no-gloss group, resulting in the uneven distribution of participants between groups after Part 2. Both withdrew as a result of poor self-confidence, likely after being in the no-gloss group.

Beginner students were selected to be the target level of this study for several reasons. First, “elementary” students were decided upon owing largely to the researcher’s own limitations in Chinese foreign language proficiency. As a non-native speaker of Chinese, it is logical to observe learners of a lower level to minimize risk of error in translation and increase the legitimacy of the observer and study in the eyes of the participants. Limitations in Chinese course offerings narrowed the prospective participant pool further. It was originally considered to include Chinese language students who may not have been registered in a Chinese language

course at the time of the study, but this was decided against due to the unpredictability and unreliability of comparing these students' language levels to those in an established, traceable level.

It was necessary to draw from both Chinese 120 and 126 courses owing to the limited number of students available to participate. Ideally, all participants would have come from one single course, but the language level required for each course was similar enough that it was possible to find a text with vocabulary that could apply to both groups. Notably, this study does not include enough participants to produce statistically significant data.

b. Screening

Crucial to this study was that the Key Terms be unfamiliar to the participants. As a method of screening, Question 9 of the comprehension questions asked the participants: "*Please circle at least 6 characters in the text you did not recognize.*" The word "*recognize*" was chosen due to its ambivalence, encompassing such definitions as, "*seen before and remember seeing;*" "*studied comprehensively;*" "*remember the translation of;*" or some other interpretation. An additional purpose of this vagueness was to hide the true nature of the study from the participants. Furthermore, the question specifically asked for "*characters*" rather than "*words*" to avoid the vague concepts surrounding the Chinese word. This allowed for a fast and simple, though far from foolproof, method of mass screening without revealing the nature of the study to the participants. None of the participants failed the screening test, as justified in 7.2 Screening concerns and justifications.

As will be elaborated upon further in Chapter 4: Results and Observations, most participants circled at least one of the characters of all or most of the six Key Terms; an ideal

participant would circle at least one of each character comprising the six Key Terms. A summary of the screening question results is displayed in the table below. Note that each Key Term appeared only once in the text, with the exception of 建立, “to build,” which appeared subsequently in the modified form 建. Table 1 below displays the results of the screening question.

Table 1. Screening question results

Participant	1. 养	2. 建立	3. 后悔	4. 应该	5. 成语	6. 补救	Total Unknown
#1	一个人 养 了很多羊	给羊 建立 一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	4/6 7/11
#2	一个人 养 了很多羊	给羊 建立 一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	6/6 9/11
#3	一个人 养 了很多羊	给羊 建立 *一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	4/6* 5/11*
#4	一个人 养 了很多羊	给羊 建立 一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	6/6 8/11
#5	一个人 养 了很多羊	给羊 建立 一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	6/6 10/11
#6	一个人 养 了很多羊	给羊 建立 一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	6/6 7/11
#7	一个人 养 了很多羊	给羊 建立 一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	6/6 7/11
#8	一个人 养 了很多羊	给羊 建立 一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	1/6 2/11
#9	一个人 养 了很多羊	给羊 建立 *一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	1/6* 1/11*
#10	一个人 养 了很多羊	给羊 建立 一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	3/6 5/11
* #W1	一个人 养 了很多羊	给羊 建立 *一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	4/6* 6/11
* #W2	一个人 养 了很多羊	给羊 建立 *一个羊圈	那个人很 后悔	他 应该 给羊	这个 成语 的意思是	才可以 补救	4/6* 4/11
Average:							4.25/6 6.08/11

Table 1: Results of the screening question. The table displays the Key Term, bolded, in the context in which it appears in the text. The highlights represent the character(s) circled by the participant. Only the Key Terms are emphasized in the table. The “Total Unknown” column represents data from both total Key Terms (6) wherein at least one character is unfamiliar and total characters comprising the Key Terms (11). Sequential characters circled in one stroke are highlighted together; sequential characters circled separately are highlighted in different colors to represent the separation in language processing. An asterisk (*) represents a Key Term that was not circled in its first appearance, but was circled in a subsequent, modified appearance (only 建立).

As observed in the table, the average number of Key Terms identified as unfamiliar by each participant was 4.25 out of 6, or just over 70%. Casual verbal confirmation upon completion of the study was also implemented to ensure the participants had truly not

encountered the Key Terms in reading before. Concerns with screening will be elaborated upon further in 7.2: Screening concerns and justifications.

3.3 Design

This study was designed to test whether or not the presence of a single-translation L1 gloss while reading would impact the language acquisition and/or retention abilities of beginner Chinese language students. Participants were randomly divided between two groups (Group A: no gloss; Group B: gloss), which read identical texts and answered identical comprehension questions, which they possessed at the time of reading. The only altered variable between the two groups was the presence or absence of a Key Term gloss.

Participants were handed a single sheet of paper, upon which was included the text and nine comprehension questions. For those in Group B, a marginal gloss comprising six Key Terms was included directly below the text, before the questions. This placement was to ensure the participants with access to the gloss would actually notice and use the gloss. As previously noted, the participants were not told they were divided between two groups; but rather, believed they all possessed an identical test. The reason for this omission was to maintain even motivation amongst all the participants. The comprehension questions were implemented to test the participants' familiarity with the Key Terms during and immediately after reading the text.

Part 2 of the study occurred exactly one week later, when participants were asked to reconvene in order to test their Key Term memory. Part 2 comprised a simple six questions, each one requesting to recall an English translation of a Chinese Key Term. A final prompt asked the participants to provide, in as much detail as they could remember, a summary of the story described in the text. This prompt was included primarily to discern whether the participants'

retention of the Key Term translations had any correlation to the participants' recollection of the text, and resulted in some interesting trends which will be elaborated upon in detail later.

Preliminary designs of this study would have tested the participants' memory much later in the future. Initially, Part 2 was planned to occur one month after Part 1, but, owing to complications with scheduling and the realization that college participants would likely find difficulty recalling language they hadn't used in a month, one week was chosen to better reflect the language level and the environment. This shorter time would hopefully ensure fresher memory of the text and the process as a whole.

Several aspects of the experiment were intentionally left vague, in order to prompt participants to produce as natural output as they understand it. As mentioned previously, vagueness should be tolerated and encouraged in L2 learning because it engages the learner more deeply with the text. Phrasing in several questions is also vague, as detailed in 3.2.b. Screening and 7.2 Screening concerns and justifications, and participants who asked questions during the study about the exact translation of a particular term or interpretation of a comprehension question were provided nebulous responses.

The idea for this study came from the observation that, as a Chinese L2 learner with English L1, it is easier to read out loud other alphabet-based languages than it is to read pictographic languages such as Chinese. To a beginner Chinese L2 learner, it is impossible to read unfamiliar characters. A beginning reader in Spanish, with an English L1 for example, could read Spanish, however crudely, owing to the presence of a nearly identical alphabet and spacing system. Even though the direct translation may be elusive, a learner can still sound out the written words. As much exposure to Chinese as an L2 learner, or even an L1 speaker, may

have, it is impossible to fully be able to recognize every written unit of language in the lexicon, which cannot be said about alphabet-based languages with a finite number of graphemes.

3.4 Materials

a. Text and in-text translations

The text was adapted from a pedagogical website intended for Western learners of standard Mandarin. The website follows the HSK (Hanyu Shuiping Kaoshi, “Chinese Level Exam”) Chinese reading system, and the text selected for this study came from the low HSK 3 level (there are six HSK levels). The title of the text is 亡羊补牢 (Literally, “to mend the pen after the sheep or lost.” Or, an idiom meaning “better late than never.”) The content of the text is very simple: it is a short story in which a man who raised sheep did not build them a pen, which resulted in a wolf eating ten of the sheep. The man then built a stable pen to prevent future loss but regrets losing the unprotected sheep. The text ends with the moral of the story: One can only remedy a situation after something bad has already happened.

This particular story was chosen due to its simplicity, as knowing the content of a story can help in identifying the meanings of individual terms. For this text, certain changes were made to the language to make it more accessible for beginners. For example, certain non-essential complements were omitted or altered, description words were added to ensure differentiation between different events and themes, and sentences deemed too complicated were abridged or removed entirely. With the exception of some of the more abstract grammar markers, the text was intended to comprise mostly familiar characters to significantly differentiate the six unfamiliar Key Terms. As the results will show, this was not necessarily the case.

The text was intentionally short. Initially, a longer text was planned, but a simpler text was chosen to ensure steady motivation and confidence of the readers. A participant who was volunteering his or her time would be less likely to return for an unannounced second part if he or she felt that he or she did not perform adequately on the first part, as was the case with two of the initial participants. Furthermore, if the text was too long, the participant would likely start to rush at some point, and his or her noticing power would deplete by the end of the text.

To further assist in combating a potential deterioration of motivation, certain terms which were critical to understanding the meaning of the story were given to both groups as in-text translations. These five terms comprised some of the most important words in the story, such as “sheep” and “pen,” which, if the reader was unfamiliar with, would be extremely difficult to infer from context and would derail the reading process. With these five terms given, it was possible for those in the no-gloss group to easier understand and infer the basic meaning of the story, as not understanding the basics of the story would fail to result in acquisition of the Key Terms.

The text was selected by the researcher based on his own second language reading experience. Modifications to the text were approved by the students’ teachers, who are all native speakers, to ensure authenticity. The text was deemed level-appropriate by the teachers, with the acknowledgement that some of the more complex grammar structures, such as the passive tense marker 被 and the directional complement 来 had yet to be studied in detail. These complex structures remained minor and supplementary context surrounding them minimized interference with sentence reading and text comprehension.

b. Key Terms

As previously mentioned, the gloss in this study comprised six lexical items (Key Terms) anticipated to be unfamiliar to the reader. Most of the Key Terms were selected directly as they appeared in the text, as they were already deemed level-appropriate for the reader and required no further modification. These included 1. 养, “to raise;” 3. 后悔, “to regret;” 5. 成语, “idiom;” and 6. 补救, “to remedy.” One term, 2. 建立, “to build,” was modified from its original single-character transcription 建 in order to give the term a more concrete meaning and differentiate it from the two subsequent 建’s. One term, 4. 应该, “should,” was included as part of a sentence inserted by the researcher to establish a stronger conclusion which would concisely sum up the events of the story.

The Key Terms were selected by the researcher and, like the text, were screened with the help of the students’ teachers. The teachers, native speakers, ensured that those Key Terms had not made a centralized appearance in the curriculum to that point. In addition, the teachers were more familiar with the abilities of their individual students, and could confirm the general language ability of each student as it related to character recognition. The Key Terms were selected largely for their relative straightforwardness of translation in context.

The Key Terms increased in difficulty of form as the text progressed. 养 is a fairly simple character to write and can be encountered in beginner texts, whereas 补救 is a disyllabic item which has more complicated strokes and is not studied in beginner classrooms. The Key Terms also vary greatly in appearance, as Key Terms that look too similar would be too complicated for beginner learners to distinguish, and the design aimed to test character acquisition rather than differentiation. The Key Terms all possess concrete meanings rather than abstract ones to increase likelihood of comprehension.

With the exception of the first term, 养, “to raise,” all of the Key Terms are disyllabic, representing the most common structure of word formation in Chinese. Only 成语, “idiom,” is a noun; the others are all verbs. The part of speech of each Key Term was intended to be easily discerned from context, to make the comprehension questions slightly less overwhelming for the beginner students.

The reason for the low number of Key Terms was due to the brevity of the text. The text consists of 96 characters total, 11 of which were directly associated with a Key Term. 建立, “to build,” appeared in the modified form 建 two subsequent times following its initial introduction, while other Key Terms only appeared once. Assuming the 11 characters of the Key Terms were the only unfamiliar characters, this would result in a comprehension rate of about 88.5%, a bit lower than Laufer’s (1989) preferred 90-95% text familiarity rate. Naturally, the ongoing debate regarding classification of word versus character in the Chinese language creates some difficulty in pinpointing the exact number of “words” in any Chinese text. To combat this vagueness surrounding the familiarity rate of “words” in the text, a simple question was added at the end which asked participants: “*What percentage of the text did you understand?*”

The Key Terms were “randomly” distributed throughout the text to simulate semi-natural L2 reading, an intricate task given the low L2 reading level of the participants, and to avoid inadvertent recognition of their special status by pattern-recognizers. However, they do appear in equal distributions within the layout of the text; each of the three paragraphs contains two Key Terms, to ensure that comprehension would be tested at the beginning, middle, and end of the reading process.

c. Gloss

For this study, a single-translation gloss was chosen to represent the reading enhancement strategy. A single-translation gloss displays the term in the L2, as it appears in the text, followed by a direct translation in the L1. This format was chosen owing to its similarity to a pedagogical L2 textbook vocabulary list, a format which would likely be intimately familiar to the beginner language students serving as the participants. This is also the simplest means of displaying the information. As mentioned, other gloss types, such as multiple-choice glosses and L2 glosses, were not chosen due to their relative complexity and the small sample size of learners.

The gloss appears immediately following the text, preempting the comprehension questions. There is nothing on the test suggesting that the terms appearing in the gloss are of any distinct significance to the study. Since only one group had access to the gloss, and none of the participants knew about Part 2, the gloss provided as much assistance as the participant was willing to provide attention to it.

Initial designs of this study also allowed for use of an electronic dictionary by the participant. The intention was to maintain high motivation throughout, especially by those participants without access to a gloss, and to mimic natural beginner L2 reading. The electronic dictionary was eventually decided against due to the propensity of college students to abuse translation services; in addition, the electronic dictionary would have also added an extra variable to account for in the analysis, as there is a high likelihood the amount of time devoted to looking up individual characters would have impacted participants' allocation of attentional resources. A standard print dictionary was not considered owing to the complexity of structure of Chinese—English language dictionaries.

d. Comprehension questions

The comprehension questions were designed to test the participants' literal comprehension of the text. In noting the different types of comprehension a reader may demonstrate, Day & Park (2005) refer to literal comprehension as an "understanding of the straightforward meaning of the text." For the present study, participants were tasked with reading a short passage in Chinese and answering several comprehension questions about the content of the passage.

Nine comprehension questions were included, to be completed during or immediately following reading. Both groups received identical comprehension questions. The comprehension questions were both written and answered in English. Doing so prevented the participants from simply copying down the characters from the text they believed corresponded to the correct answer.

Six of the questions are directly related to the meanings of the Key Terms; wherein a correct "desired" response would elicit the intended Key Term. This would establish that the participant interacted with the Key Term in the text. Two questions are about the plot and simply serve to gauge each participant's translation and comprehension abilities. Incorrect answers to these questions would demonstrate too great an unfamiliarity with the language, and would call the results into question. The final question was included as a screening question and to get a stronger sense of each participant's perceived Chinese reading ability.

Importantly, the comprehension questions could not directly ask for the meanings of these characters because those with access to the gloss could simply transcribe the given terms, and would have very little incentive to actively engage in the text. The comprehension questions and the text were adjusted simultaneously while in development to ensure that the answers could

be deduced clearly from context. Day & Park (2005) note comprehension questions should not intentionally trick the reader if their ultimate goal is understanding of the text.

e. Part 2 exam

The questions in Part 2 were more straightforward. Participants were once more given a single sheet of paper, each numbered question simply displaying a Key Term and a subsequent blank space to provide the English translation. The text, along with any Part 1 materials, was not provided this time to ensure that the participants relied solely on memory, rather than on context. Participants were not required to provide a response to every question; as a result, many are left blank.

A seventh question asked participants to “*Please briefly summarize the story.*” This was intended to provide the researcher with insight into the individual’s recall process or long-term memory. Furthermore, it would demonstrate that a participant was not simply guessing a meaning if it was related to what the participant wrote in the summary. Asking participants to summarize the story also served to provide data to analyze peripheral reading trends. Responses to this question will be elaborated upon in later sections.

3.5 Procedure

The procedure of this experiment was largely derived from similar studies in second language acquisition through reading enhancement techniques, such as Watanabe’s (1997) and Nation’s (2001) classic gloss studies, as well as more modern ones such as Duan (2018) and Rott (2005). Parts of the procedure have already been alluded to.

a. Step 1: Scheduling

Students from multiple sections of Chinese 120 and Chinese 126 were forwarded an email requesting volunteers for a short reading study. To increase motivation and thus participant turnout, it was stressed that the time commitment was short, though as much time could be taken as would be needed. Students were also told they would be rewarded for their help. Attached to the email was a digital survey regarding scheduling availability in the upcoming week. Interested participants simply had to respond to this survey with time availability and the researcher would reach out and schedule the sessions. This way, participants would not have to email the researcher privately, the extra step of which might have decreased motivation and lessened volunteer turnout.

Those who responded to the survey were emailed privately to arrange a time, resulting in four separate testing sessions. During the scheduling process, participants were told this would be the one and only in-person testing requirement, which could be completed in their own time one-on-one with the researcher or in one of the scheduled sessions. All reasonable flexibility in regards to scheduling was offered in order to ensure volunteer turnout. Knowledge of Part 2 was deliberately withheld to ensure natural reading conditions. 12 of the 15 respondents to the survey showed up to participate at their scheduled time.

b. Step 2: Reading and comprehension

At each of the four sessions, the test began and ended at staggered intervals. The sessions met in an empty language classroom void of outside disturbances. Instructions sent via prior email communication and transcribed on the sheet itself were clear enough to merit minimal further verbal explanation of the process; furthermore, because there was no time constraint for

the reading and comprehension part of the study, participants were given the text immediately upon arrival and were secretly divided between Group A (no gloss) and Group B (gloss). They were told to read as they normally would, and it was reiterated that they could leave as soon as they were finished. Adopting a think-aloud procedure, as in Lomicka (1998), to allow for individual insight into each participant's learning method was considered, but dropped to allow for simultaneous group testing.

As previously stated, participants received the text, comprehension questions, and gloss (if applicable) as one single one-sided sheet of paper. Participants were told to do their best, even if it seemed there were many characters they did not recognize. Questions directed to the researcher about the meanings of individual characters were rejected, while questions regarding the nature of the comprehension questions were met with answers suggesting it be up to the participant's own interpretation. The researcher remained in the room for the duration of the process in order to ensure adherence to the procedure.

Though untimed, participants generally required between 15 and 25 minutes of reading and answering time. They were requested to remain silent throughout the process and to not discuss it outside of the room, as many of them were classmates. Upon completion, participants were informed they would be contacted exactly one later for a follow-up survey. The researcher spoke informally with each participant on his or her way out, to gauge external factors such as confidence, skill, time, background, etc. and to assure each participant that he or she was doing a service to the study.

For accommodation purposes, participants were told they did not have to return to a formal sit-down setting to complete Part 2, and the researcher could seek out participants wherever they happened to be at the corresponding time. This ensured that all participants would

be tested exactly one week later, since many of their schedules did not allow for them to return to a formal classroom setting at the same time as the week prior. Two of the 12 participants did not continue on to Part 2, likely owing to the clandestine nature of the follow-up test and the rigidity of class schedules.

c. Step 3: Memory test

Participants were found individually and in small groups one week after their initial testing date. Part 2 requested the participants simply list any translations they could remember regarding the six Key Terms. Most remarked that they remembered the contents of the story and, even more specifically, having seen the Key Terms before, though they could not remember the exact translations. The final question of Part 2 asked the participants to recall, in as much detail as possible, the events of the text, and was added to gauge the participants' L1 recall, and, as will be demonstrated in Chapter 4: Results and Observations, most participants recalled specific details of the text in the L1, even if they could not remember the exact corresponding characters.

Participants each spent around five minutes on Part 2, which was promised to be considerably shorter owing to the lack of new reading. All participants spent time thinking about each term; none simply abandoned the task immediately due to lack of recall. Following completion of Part 2, which was conducted under much more casual conditions, participants were thanked for their time, and the true nature of the experiment was revealed. The researcher compiled all 22 submitted forms to compare and analyze trends in the data. Owing to the small sample size, the following results will be analyzed more in terms of anecdotal observation than statistical quantification.

CHAPTER 4

RESULTS AND OBSERVATIONS

4.1 Data collection

As previously mentioned, this study will focus more on observed trends than on statistical data, owing to the limited data stemming from shortage of participants. In analyzing the written responses of each participant from both Part 1 and Part 2, several notable trends emerge, and different theories will be applied to explain these observations. This chapter will briefly introduce several discernable patterns in the results along various stages of the acquisition (Part 1) and memory (Part 2) processes, which will be elaborated upon further in Chapter 5: Analysis and Discussion. It is important to remember that there were not enough participants to establish a statistical report, but some general trends do emerge which could be explored in a larger study. Statistical values in this and subsequent sections, such as averages and modes, should be interpreted loosely.

10 participants completed this study in its entirety. Owing to the withdrawal of two participants, there were six participants in Group B (gloss), and four participants in Group A (no gloss) who completed both parts of the study.

In this section, “desired response” refers to a response by a participant which produces an L1 translation of an L2 Key Term exactly as it appears in the gloss. A non-desired response is not necessarily an incorrect response; but rather, any response which differs in wording from the gloss to a noticeable degree. “Incorrect response” refers to a non-desired response which produces an L1 translation which differs in meaning from the translation provided by the gloss. “Acceptable response” refers to a non-desired response which contains similar form and/or meaning to the desired response, and/or a Part 2 response which mimics an incorrect or

acceptable response provided by that same individual in the Part 1 comprehension section. The category of each response was decided at the discretion of the researcher.

In recognizing retention, acquisition must have occurred first; thus, it is important to analyze the output produced as part of responses to the comprehension questions in Part 1 as well. For those in Group A (no gloss), it was less likely they would produce the exact desired response to each comprehension question in Part 1. For example, Participant #2 translated 补救, “to remedy,” as “*deal with*.” This response contains the same meaning as the desired response yet differs in form. Those in Group B (gloss) would be more likely to produce responses analogous to the glossed items, as they all had them listed out in front of them. Each participant’s Part 2 response must line up with the Part 1 response to demonstrate comprehension of the term. A full breakdown of Part 1 responses exists in Appendix Table 2.B and a full breakdown of Part 2 responses exists in Appendix Table 2.A.

Table 2 below shows the percentage of accuracy of the six Key Term translations produced by each participant to each corresponding comprehension question in Part 1 and Part 2. An ideal response in Part 1 would include a direct translation of the Key Term being tested in that question, but some questions do encourage inferential responses, such as “*He is a shepherd*” instead of “*He raises sheep*,” which, for the purpose of acquisition, cannot be considered correct as they do not display knowledge of the individual term being tested (in this example, 养, “to raise”). An ideal response in Part 2 would be a desired response (word-for-word recall) or acceptable response (slight variability in word choice). In Part 1, *acceptable responses were considered accurate for those in Group A* due to the increased variability from lack of gloss. In this manner, for those in Group A, a Part 2 acceptable response which mirrors a Part 1 acceptable response will be considered accurate. For those in Group B, an acceptable response is not

considered accurate in Table 2. Three Part 2 acceptable responses emerge, and are shown in Appendix Table 2.B.

Table 2. Key Term accuracy percentages in Part 1 and Part 2

Group	Participant	Part 1 Accuracy	Part 2 Accuracy
A	#2	4/6	0
A	#5	1/6	0
A	#6	3/6	0
A	#7	2/6	0
B	#1	6/6	0
B	#3	5/6	0
B	#4	5/6	0
B	#8	3/6	0
B	#9	4/6	3/6
B	#10	5/6	0
	Group A Average	2.5/6=42%	0
	Group B Average	4.67/6=78%	0.5/6=8%
	Total Average	6.3/10=63%	0.5/10=5%

*Table 2: Accuracy percentages for L1 output corresponding to the six Key Terms.
Participants #W1 and #W2 scored 4/6% and 2/6%, respectively, in Part 1.

From Table 2, it is clear most participants had a solid general understanding of the Key Terms from context during Part 1. Many responses indicated familiarity with the concept, but did not directly connect meaning to the Key Term, and were thus not included in the accuracy calculations. During Part 1, the average Key Term accuracy was 42% for those in Group A and 78% for those in Group B. Total average accuracy was 63%.

Notably, those without the gloss were less likely to produce responses to the comprehension questions than those with the gloss. None of the responses provided to the comprehension questions demonstrated a complete misunderstanding of the desired character while reading; responses generally contained a similar enough meaning to the Key Term, with

those in Group B naturally providing more uniform responses. It was clear from the comprehension questions that, with a few exceptions, most of the participants had been exposed to the Key Terms in the reading and could elicit their meanings during reading.

4.2 Data trends (as observed in Appendix Table 2.A)

Before breaking down the memory table, let us briefly attend the research question. Regarding Part 2 memory, the data in Appendix Table 2.A reveals three correct “desired” responses, and three semi-correct “acceptable responses.” Other responses were incorrect or left blank. All of the desired and acceptable responses came from members of Group B (gloss). In a larger, more precise study which could focus on statistics, these trends would slightly endorse the hypothesis as to the efficacy of the single term gloss as it relates to vocabulary retention.

Though there were few desired responses, there are some notable trends worth observing. The most notable trend is the propensity of participants to respond to the Part 2 questions with L1 terms which had been given to them as translations in Part 1, and the inability to correctly associate recalled L1 input to the corresponding L2 term.

a. Group A results

Four of the initial six participants from Group A (no gloss) completed Part 2 of the study. Unfortunately, owing to issues of scheduling and, more influentially, confidence, two members of the no-gloss group, Participant #W1 and Participant #W2, became immediately unreachable upon completion of Part 1. Both remarked on the difficulty of the text and their lack of extensive language learning background, though both did objectively well in Part 1 and would have been

exceptionally interesting candidates for Part 2 owing to their understanding of the Key Terms demonstrated in their responses to the Part 1 comprehension questions.

The remaining participants were each asked to recall L1 translations for six Key Terms, resulting in a total of 24 individual responses by those in Group A. Of the 24 potential responses, 13 were left blank, whereas 11 received written responses. Of the 11 written responses, eight were initially encountered as in-text translations during the reading. Only three written responses (herder, consequence, count) had no relation to L1 words that appeared directly in Part 1, and were recalled from incorrect participant responses to Part 1 comprehension questions. One response, “*later on*,” was derived with the knowledge that 后 generally means “later.” None of the responses produced were desired or acceptable. Table 3.1 below summarizes the response trends of those in Group A during Part 2.

Table 3.1. Part 2 Response Trends, Group A

Participant	Number of responses left blank	Number of responses which originated as in-text translations	Number of responses not given as L1 in Part 1	Accuracy (%)
#2	0/6	4/6	2/6	0
#5	4/6	2/6	0/6	0
#6	3/6	2/6	1/6	0
#7	6/6	0/6	0/6	0
Average	3.25/6=54%	2/6=33%	0.75/6=12.5%	0

Table 3.1: Part 2 response trends in Group A. Much of the recalled L1 originated as a given in-text translation during Part 1. Each row should add up to 6/6.

b. Group B results

There were six participants in Group B who completed Part 2 for a total of 36 individual translation responses recorded. Of these 36 responses, 25 received written output, whereas 11 were left blank. Three of the responses were entirely correct, producing the exact desired response (“*to establish/build*,” “*regret*,” “*idiom*”). All three responses belong to Participant #9.

Two more responses are similar enough to the glossed terms in meaning and structure to be considered correct, and are deemed acceptable responses as such. These belong to Participant #10 (“*supposed*” instead of “should” and “*fable/story*” instead of “idiom”). One response from Participant #1 was very close in form to the desired response, but with just a slightly different meaning (“*repent*” instead of “to regret”). This will be considered an acceptable response owing to the flexibility of the study and the similarity in L1 form.

Of the 25 written responses provided by the six members of Group B, nine of the terms written by those in Group B were Key Terms, eight were given as in-text translations (sheep (shepherd was included as well), wolf, pen), and only five had no relation at all to the given terms (farmer, again, solution, problem, daily). Only Participant #4 left a majority of his responses blank from Group B. Though most of the responses originated from the Part 1 gloss and in-text translation words, it is important to note that these responses were not written next to the correct characters. Table 3.2 below summarizes the response trends of those in Group B during Part 2.

Table 3.2. Part 2 Response Trends, Group B

Participant	Number of responses left blank	Number of responses which originated as in-text translations	Number of responses which originated in the gloss	Number of acceptable responses (still considered incorrect)	Number of responses not given as L1 in Part 1	Accuracy (%)
#1	1/6	2/6	2/6	1/6	0/6	0
#3	1/6	2/6	2/6	0/6	1/6	0
#4	5/6	1/6	0/6	0/6	0/6	0
#8	0/6	0/6	2/6	0/6	4/6	0
#9	2/6	1/6	3/6	0/6	0/6	50
#10	2/6	1/6	0/6	2/6	1/6	0
Average	1.83/6=31%	1.17/6=19%	1.5/6=25%	.5/6=8%	1/6=17%	.5/6=8%

Table 3.2: Part 2 response trends in Group B. Much of the recalled L1 originated as a given in-text translation or gloss during Part 1. Each row should add up to 6/6.

c. Differences

Table 4. Simple T-Test

Group	Part 1 Average Score	Part 2 Average Score	t-Value	p-Value
A	0.42	0		
B	0.78	0.08		
Total	0.63	0.05		
			5.297	.003

Table 4: Simple T-test comparing Group A and B average response accuracy between Part 1 and Part 2. In this study, $t=5.297$. With 2 degrees of freedom, the p-value is .003. The data supports the null hypothesis.

The T-Test results show that there is a significant difference between Group A and Group B ($p < 0.05$). Participants in Group B performed better than those in Group A, and they were more likely than those in Group A to propose written responses to the Part 2 questions, thus indicating that glossing may help with learners' reading comprehension. Those in Group B did not provide as many in-text translations as responses to Part 2 as did those in Group A, likely owing to the existence and perceived significance of the gloss in Part 1 being the differentiating variable between the two groups.

It is also very clear from the data that participants were able to remember L1 input more than they were L2. Only nine of the 36 written responses from participants did not appear exactly as written in L1 in the Part 1 reading, which means that 75% of all produced written output in Part 2 was directly remembered from Part 1, albeit with the incorrect corresponding L2 term. This suggests that it is something in the form or structure of the Chinese L2 which caused difficulty in recall.

Notably, all participants were able to retell the story from the text. In analyzing the written summaries of the text produced by each participant at the conclusion of Part 2, it was

clear everyone understood and remembered the main points of the story, with several participants recalling such specific details as the number of sheep eaten by the wolf. In recalling the summary of the text in their own words one week later, many participants used either the Key Term or in-text translations and used them in their own responses in a manner similar to the Key Term translation questions. The summary of the text by each participant can be seen in *Appendix A* and will be elaborated more in Chapter 5: Analysis and Discussion.

4.3 Comprehension and retention

As mentioned, it was predicted that those in Group B would provide more consistent answers to the Part 2 memory section, owing to having seen the exact desired responses during Part 1. As a result of not providing the L1 translations to the Key Terms in Part 1, it was expected that the responses provided by those in Group A would be more subject to individual interpretation. For example, in analyzing Table 2.B the Key Term 建立, “to build,” received the direct translation desired response by all participants in Group B, while inviting some more flexibility by those in Group A. 建立 was translated as “*build*” by all in the gloss group in comprehension questions #4 and #6, while it was translated as “*give*,” “*set up*,” “*improves*,” and, incorrectly, “*count*” by those in the no-gloss group. Group A was noticeably less rigid in their L1 output production.

Incorrect responses to the comprehension questions were also often repeated in the Part 2 memory portion. For example, Participant #6 remembered that she had associated one of the terms with “*count*” in Part 1, and her Part 2 responses include the translation “*count*” as well. Similarly, Participant #8 remembered the translation “*solution*.” Though these participants

remembered the L1 they had used a week prior, they did not remember the correct L2 term to which they associated the translation.

4.4 Providing given L1 translations

The first notable pattern is the tendency of all participants to respond with English terms in the Part 2 retention test which were given as gloss words or in-text translations. Of the 36 memory responses in Part 2 which were not blank, 22 were directly given to the participant in English in Part 1, while 6 more responses (“*herder*,” “*farmer*,” “*shepherd*,” “*repent*,” “*supposed*,” “*fable/story*”) were very similar both in form and meaning to the given terms. It seems the participants could remember many of the English translations of the gloss terms and in-text translations, but could not recall accurately to which character each belonged.

Participants generally favored the simpler terms which were more crucial to the story. Of the nine participants who provided written responses (Participant #7 left her sheet blank and apologized for not remembering any of them. From her summary response, it is likely she remembered these terms as well but her visible lack of confidence led her to withhold her attempt), six included the word “*sheep*” or “*shepherd*” as one of translations, five included “*pen*,” four included “*wolf*” or “*wolves*.” These terms were most central to the story, were given as in-text translations to all participants, and repeated several times throughout the text. Naturally, those in Group A remembered L1 in-text translations only because they did not have access to the gloss. Those in Group B remembered in-text translations and gloss terms in relative equal quantity. Since none of the Key Terms were given as in-text translations, none of the in-text translations received as responses to Part 2 could be correct.

4.5 Circled characters during screening

As noted, the final comprehension question to be completed alongside the initial reading asked the participant to estimate what percentage of the text he or she understood. In addition, it asked each participant to please circle at least six unfamiliar characters. Several notable trends emerged in the screening process. Table 5 below shows the results of the question #9.

Importantly, the data below was self-reported by the participants and may or may not accurately reflect performance.

Table 5. Comprehension and Screening Results

Group	Participant	Percentage of text understood	# of circles made	Total # of unknown characters
A	#2	80	15	18
A	#5	75	16	21
A	#6	50	13	15
A	#7	-	11	14
B	#1	90	6	10
B	#3	65	6	6
B	#4	80	12	15
B	#8	70	6	6
B	#9	95	2	3
B	#10	80	4	6
	Group A Average	68	13.75	17
	Group B Average	80	6	7.67
	Total Average	74	9.88	12.33

Table 5: The results of the questions, “What percentage of the text did you understand?” and “Please circle at least 6 characters in the text you did not recognize.” Some participants circled multiple characters in one stroke and some circled individual characters only.

Due to the phrasing of the question, some participants circled six individual characters, and some circled two-character items. Those in Group B generally circled more two-character items than those in Group A. All participants in both groups, except Participant #9, circled in-text translations and/or glossed terms, suggesting that participants did not consider newly-learned

information to be familiarity, which is a positive. With the exception of Participant #4, participants in Group B generally circled fewer terms than those in Group A.

In circling characters, three participants circled two non-related characters together as one item. Participant #1 circled 羊建 as one item, interestingly including the glossed term 建, which he would later recall correctly in Part 2, and 羊, which had previously been given as an in-text translation. Participant #7 and Participant #9, who were the extremes on both ends of the Part 2 memory test, each incorrectly grouped 经发 from the phrase 已经发生 as one lexical item.

4.6 Part 2 summary question

Elements of the summaries of the text provided by the participants at the conclusion of Part 2 follow very similar trends to the vocabulary recall. Like their Key Term translation responses, most participants used specific L1 terms provided to them during Part 1 in their summary responses. Naturally, the in-text translation words “*sheep*,” “*pen*,” and “*wolf*” were all vital to the retelling of the story, and had to be used in all of the responses. Those in Group A all used these terms; in addition, Participant #2 and Participant #10 were both able to produce the given in-text translation 才, “only then” in their summaries.

Those in Group B were generally able to recount the text using more of the Key Terms in their summaries than did those in Group A (33% versus 8% average). Participants in Group B also elicited more vocabulary that would have been deemed acceptable responses had they been provided as Part 2 translation (for example, Participant #8 eliciting “*could*” instead of “*should*” and Participant #9 eliciting “*moral*” instead of “*idiom*.” Again, many related terms which demonstrated too much inferencing (for example, “*herder*” or “*shepherd*” instead of “*raise sheep*”) were not counted as accurate owing to the lack of emphasis on the individual Key Term

translation. The only Key Term desired responses used in a Group A summary was “*build*” used by Participant #2 and “*should*” used by Participant #6. Those in Group A responded with more in-text translation terms on average than did those in Group B, likely owing to a heavier reliance on the fewer number of given L1 terms. In contrast, those in Group B were able to provide more specific descriptions, which more often included the Key Terms “*raise*,” “*build*,” and “*regret*” or “*regretful*.” Not all of the Key Term translations were equally likely to be remembered, as they were not all necessary for the summary of the main content. Table 6 below summarizes the summary question trends.

Table 6. Key Term, in-text translation output in the Part 2 summary question

Group	Participant	Key Terms used in summary	In-text translations used in summary
A	#2	1/6	4/6
A	#5	0/6	3/6
A	#6	1/6	4/6
A	#7	0/6	4/6
B	#1	2/6	3/6
B	#3	3/6	3/6
B	#4	3/6	3/6
B	#8	2/6	3/6
B	#9	1/6	4/6
B	#10	1/6	4/6
	Group A Average	.5/6=8%	3.75/6=63%
	Group B Average	2/6=33%	3.33/6=56%
	Total Average	1.25/6=20%	3.54/6=59%

Table 6: The number of Key Terms and in-text translations used by each participant in recalling the events of the text after one week.

The L1 terminology used by each participant in recalling the story typically mirrored the L1 terms they used in responding to the Key Terms. L1 words used by participants in the summary tended to mirror L1 words used to answer the Part 1 comprehension questions as well.

Participant #4 translated 成语 as “*moral of the story*” both in Part 1 and recalled the “*moral of the story*” in his Part 2 summary. Participant #5 translated 建立 as “*get*” in both Parts. From the summaries only, it is impossible to tell which, if any, L2 characters from the text the participant associated each written L1 word with; more likely, the participants remembered the L1 words from Part 1 only in the context of the story and were remembering them without a specific associated character in mind.

CHAPTER 5

ANALYSIS AND DISCUSSION

5.1 On glossing

In the present study, glossing improved comprehension significantly, but did very little to help with retention over one week. As evidenced by Table 2, those with the gloss performed much better, on average, at providing accurate translations to the Key Terms in Part 1 (78%) than those without the gloss (42%). Those with the gloss performed slightly better (8%) than those without the gloss (0%), a very minor difference. Furthermore, those in Group B were really only better at accessing the previous L1 input, and not at accurately placing it. Hulstijn's (1992) study supports that single-translation glosses are not as effective as multiple-choice glosses, owing to the lack of necessity for engagement between the reader and the text.

In regards to the sentence translations (Part 1 comprehension question #8), those with the gloss were able to provide much more specific detail. The marginality of the gloss ensured that it was clearly visible, increasing the likelihood of repeated exposure to the selected L1 output. Though most of the responses used broken English to some degree, those with the gloss all included the glossed term "to remedy" in their response, while none of those without the gloss

did. As a result, those in the gloss group were able to piece together much more of the sentence, resulting in more complete translations than those in the no-gloss group. Coupled with their higher reported comprehension scores and applying this sentence theory to the rest of the text, it is clear how reliant those in Group B were on the gloss in the comprehension and acquisition stage. The poor sentence translations clearly foreshadow future recall difficulties.

Ko (2005) noted that students without access to a gloss used far more lower-level reading strategies, such as making inferences, than those with access to a gloss. This is in line with Hulstijn's (1995) observations that learners focus more on the words than on the form of the language. In the same paper, Hulstijn (1995) advocates for prioritizing learning vocabulary over grammar, even advocating for avoiding teaching "explicit grammar rules" to beginner learners with low educational background. In the present study, there is a trend of those with the gloss making more inferences and predictions in the Part 1 comprehension questions, in line with Ko (2005). For example, in response to the question "*What does the man do for work?*" Participant #8 (Group B, gloss) responded "*farmer, presumably.*" The question was intended to provoke a translation of the key term 养, "to raise," but instead it is clear from his response that he included other relevant details from the text to put forth a more specific answer than "he raises sheep." Conversely, Participant #7 and Participant #5 (both Group A, no gloss) did not even put forth a guess when asked for a direct translation of 成语, "idiom" in the comprehension questions, demonstrating an inability or unwillingness to infer from context. Participant #5 similarly responded to a subsequent question ("*How does the man feel after the events of the story?*") by circling a term in the text and drawing an arrow to the question, understanding the importance of the term but unable or unwilling to infer meaning from context without the gloss. Participant #7 responded with "*content/sad,*" which essentially equates to a non-answer.

As in Ko (2005), students who read with the gloss reported an easier time in comprehension and were more outwardly confident and motivated. Those without the gloss reported higher difficulty reading and answering the comprehension questions, and many apologized for their seemingly low language level. The two participants who withdrew from the process seemed to have done so owing to the difficulty of reading an unknown text without vocabulary assistance.

The present study also sheds light on some of the existing theories of language processing, which have been described previously in Chapter 2: Theoretical Background. The subsequent sections will elaborate in more detail on these aforementioned theories. Table 7 at the end of Chapter 5 concisely lists the theories supported by the data.

5.2 On incidental vocabulary learning

In the present study, learners focus entirely on reading and understanding the text, so the vocabulary learning, including the Key Term acquisition, is incidental. To recall, there are several factors involved in incidental vocabulary learning. As described in Duan (2018), IVL is primarily dependent on noticing unfamiliar characters. From there, readers must process meaning in order to acquire the vocabulary. Finally, repeated form-mapping ensures that the new vocabulary is ingrained in memory. Comprehension questions served to promote meaning processing, as well as to draw attention back to those Key Terms in the text which would serve as responses. Key Terms were not repeated (to mimic natural reading conditions), which may explain why Key Term recall is only 8% by those with gloss and 0% accurate by those without.

It was expected that the participants in Group A would struggle to recall the terms with as much accuracy as those in Group B, since Group B secretly received the exact desired response

(form-mapping) while reading in the form of the gloss. As noted by Watanabe (1997), context has a direct role in incidental vocabulary learning. If the context is too redundant, readers may not focus on the individual word forms. This phenomenon likely happened during this particular study. Especially in a language like Chinese, which does not utilize spacing or other “word” differentiators, it can be difficult or impossible (or just impractical) for learners to divert attention to each individual character when reading a longer text. The participants likely understood the meaning of the story by piecing together parts they did understand and filling in the blanks with appropriate guessing from context. It is very unlikely any of the participants paid specific attention to the individual characters in the text, outside of interacting with the gloss and/or in-text translations.

Hulstijn (1993) found that readers tend to pay more attention to terms they deem relevant to the text. In the present study, the glossed words and the words with given in-text translations were the only terms with any kind of special indication. This likely ensured they were noticed by participants more than non-differentiated characters. The Key Terms and in-text translations were also the most meaningful content words in the story, leading to a higher likelihood of L1 recall.

Schmidt’s (1990) Noticing Hypothesis suggests that readers must be paying attention to L2 input in order for learning to occur. Coupled with the complexity of learning to read and differentiate Chinese characters as a beginner owing to the similar shapes, it is likely most of the participants were unable to differentiate between the L2 terms during their Part 2 test. They could, however, access the previous L1 terms, and largely guessed with which Chinese character each English word was associated. The data suggests it is easier to recall information in the L1 than L2. Similarly, Guszak’s (1967) finding that language students performed best when

answering questions of factual recall noted by Day & Park (2005) could explain why the participants performed so much better on the text summary recall than on the individual direct translations.

Xu (2010) also notes that word concreteness may play a role in incidental vocabulary learning, concluding that words that are easier to visualize will be easier to recall. This is notable in the results, as the wide majority of the 36 written L1 responses in Part 2 were nouns (23) or action verbs (10). Very few were adverbs (3). (Noted in Appendix Table 2.B). Though the recalled L1 output was generally incorrect in matching the L1 term, L1 translations that were more concrete in meaning were put forth more often by the participants.

Duan (2018) notes that incidental vocabulary learning is contingent on three factors: noticing of unfamiliar words, processing of their meanings, and repetition of the form-meaning mapping. As it relates to the present study, the lack of repeated Key Term form-mapping in the text likely contributed to the low memory performance of the participants (5% average). As will be mentioned in 5.3 On processing, it is also very likely the participants, particularly those without access to the gloss, failed to adequately notice the individual Key Terms in the text.

According to Laufer (1989) and Hu & Nation (2000), readers should understand about 95-98% of an L2 text for optimal reading conditions to occur. The 2-5% of unknown text should be inferable from context. In the present study, it is apparent that the text slightly exceeded the students' abilities. In reviewing the self-reported percentage of text understood by each participant, only Participant #9 hits the 95% comprehension mark, with Participant #1 coming close at 90%. Notably, those with access to the gloss reported higher overall comprehension, which speaks to an unfamiliarity with the Key Terms prior to reading and the fundamental role of those vocabulary words in establishing meaning of the story. Both were goals in selecting text

and Key Terms, however; it was expected/hoped that they would recognize most of the other characters in the text. Group A (including the two who withdrew before Part 2) reported 68% text comprehension on average; Group B reported 80% on average. The average of all 11 participants who reported was 74%, significantly lower than the desired percentage. Naturally, the difficulty of the text further calls the results into question, as there were far too many unknown terms to differentiate the Key Terms in recall.

A major problem with incidental vocabulary learning, of course, is its unpredictability. Because incidental vocabulary learning is essentially “accidental,” there is no way to account for which words will be learned, how often a reader will pick up a word, or for how long the word will be retained (Paribakht & Wesche, 1997). Participants in the present study were likely able to recall the English terms but not the Chinese characters in Part 2 owing to the vast differences in structure and form between written Chinese and written English language.

5.3 On processing

a. On involvement load

Returning to Craik and Lockhart’s (1972) Depth of Processing Theory, which states that the chance of a new word being stored in the long-term memory is determined by the word’s sensory properties and semantic-associative features. In the case of Chinese, it is very likely the “words” appeared too similar for the participants to distinguish individual characters out of a crowded text. The most vivid example supporting this idea lies in the most common Part 2 memory response error: identical mistranslations of the key term 养, “to raise.” Of the 10 participants, three translated 养 as “sheep,” while three more translated it as “pen.” In addition, several of the participants, particularly Participant #4, were extremely confident about this

response. The most plausible explanation for this outcome is that the character 养 looks orthographically similar to the character 羊, “sheep,” sharing a very similar top part. The character 羊 was also given as an in-text translation to all participants, as was the word 羊圈, “pen,” allowing for extra attention to be drawn to the character 羊 during reading. Because the character 羊 was important to the story, it was repeated many times throughout the text, causing many of the participants to associate an incorrect form-meaning connection after confusing the characters 羊 and 养 (which, unrelatedly, are also pronounced the same way). Beginner Chinese L2 learners have less experience differentiating similar characters by orthographic structures such as radicals, so it is likely they simply mistook one character for another. Wittrock’s (1974) model also suggests retention will be improved if learners create strong FMCs. In this example, Wittrock’s model is correct, but the FMCs were often incorrect. Similarly, some of the more complex Key Terms, such as 补救, were likely too complicated in appearance for participants to notice at length, supported by the fact that only four of the 10 participants even attempted a response of this particular item in Part 2.

One example of Schema Theory (Cook, 1989) presents itself in the Part 2 responses. Participant #2 (Group A, no gloss) responded that 后悔 means “*later on.*” 后悔 means “to regret,” so Participant #2’s response was not acceptable; however, she did relate the character 后 back to its individual meaning, “later” or “after.” In this case, Participant #2 demonstrated a clear understanding of the first character, which was discernibly already in her long-term memory. Because Participant #2 did not have the gloss, it is likely she associated an unfamiliar term with a familiar character to erect her response.

The involvement load in this study was observable as well. In the Part 1 comprehension section, the division between those in Group A and those in Group B in terms of desired responses directly paralleled the absence or presence of a gloss. As observed in Appendix Table 2.B, those with access to the gloss clearly utilized it; seldom did a Group B participant provide a non-desired response in Part 1. For example, none of those in Group A properly translated 养, “to raise,” or 补救, “to remedy,” whereas most of those in Group B transcribed the exact gloss input as their own, except in places where inferences occurred, such as translating 养羊 (literally, “to raise sheep”) as “herder.” The three main components of the involvement load (need for a word, search for its meaning, evaluation of correctness) (Hulstijn & Laufer, 2001) are clearly present in the comprehension question responses of those with access to the gloss, though Hulstijn notes multiple-choice glosses better encourage and quantify evaluation efforts.

As reinforced by Taylor (2013), learners place more importance on the content of the words than on the structure of the language when reading in the L2. They do this because words have much higher communicative value than grammatical items. This is directly observable in the sentence translations at the end of the Part 1 comprehension questions (#8). Most responses included the given words “*things*,” “*only then*,” and (for those with the gloss) “*to remedy*.” Participant #4, who had the gloss, translated the sentence as, “*Bad things... but it’s not too late to remedy*.” Many other translations followed this skeletal pattern, where only the main content words were translated and other unknown characters were simply ignored. Participant #6 (Group A) responded with, “*Don’t validate things until you lose it*.” Without the gloss, she had to determine the meaning of the word 补救 from context. Though she did not translate the term correctly, it is clear she recognized it as an important content word in the latter clause (才可以补救). She also circled both characters together, recognizing it as one linguistic unit in a

five-character clause, lending credibility to the idea that her chosen translation “*validate*” directly corresponded to the term 补救.

b. On cognitive load

Sweller’s (2005) Cognitive Load Theory suggests that during-reading assistance reduces the processing load, allowing the reader to devote his or her attention where he or she sees fit. This is evident by the use of higher reading strategies by those with the gloss, particularly inferencing. For example, the first comprehension question asks, “*What does the man do (for work)?*” A desired response would be: “*He raises sheep.*” Many in Group B took it one step further, however, by inferring the man’s job title based off the information received about him in the text, indicating that the man was perhaps a shepherd or a farmer. Interestingly, several of their responses directly note that they are speculating. Participant #1 responded with, “*He’s a sheep herder (raises sheep).*” The part in parentheses suggests that she understood the desired response; at the very least, she was making it clear she saw the correct answer in the text and knew she had access to it. Similarly, Participant #8 responded with, “*farmer, presumably,*” revealing that he was inferencing based on the information he acquired from the text. Participant #9 and Participant #10 also noted the man was a “*shepherd.*” Participant #3 and Participant #4 were the only two from Group B who directly copied the gloss on that response. Participant #2 (and Participant #W2) was the only participant in Group A who ventured a job title, while the other participants tried to directly translate the simple verb-noun phrase 养羊.

Reading the Chinese text represents intrinsic cognitive load, as it was a high-involvement task for all those involved. The act of translating the text using the L1 gloss represents germane cognitive load, as the gloss reduced the natural difficulty of translation while facilitating reading

comprehension, evidenced by the greater quantity of desired responses in Part 1 by those with access to the gloss. Finally, extraneous cognitive load was represented by non-text activities, such as potentially becoming distracted by sounds from outside the space or by interacting with the researcher during the reading process. Because dictionaries were not allowed during this study, a major source of extraneous cognitive load was eliminated; however, the time it took for participants to guess words not vital to the text could have been a source of unnecessary load as well. Glossing greatly reduced intrinsic cognitive load during reading, likely explaining why those with the gloss performed considerably better at Key Term comprehension and acquisition than those without (78% versus 42% average), while those with access to the gloss did not perform considerably better at Key Term recall than those without (8% versus 0% average).

c. On noticing

Schmidt's (1990) Noticing Hypothesis claims that input must be noticed in order to be acquired. It is very likely that readers did not attend individual characters while reading the text, instead understanding the content as a whole. According to Schmidt, isolating a word, as in a gloss, allows the reader to allocate specific attention to the physical appearance of the word. In the case of Chinese, where all characters possess the same basic square-form structure, it is very difficult to distinguish between characters at a quick glance, and even more difficult to remember the minor stroke differences among similar characters.

Rott (2005) noted as a result of her study that bolding the Target Words in the text triggered the readers' need for meaning, which resulted in a consulting of the gloss and an increased focus on word form. In the present study, Key Terms were not differentiated from other characters in the text, and were likely not recognized as important until after consulting the

gloss. They likely did not know to use the gloss until after attending the sentence, because Key Terms were not otherwise indicated as being such, resulting in selective attending of the input and thus a lower probability of recall. Those without the gloss had no way of noticing the significance of the Key Terms, other than their being answers to the comprehension questions, because the gloss serves as a marginal list of all Key Terms. The Key Terms were not bolded in the present study in order to more adequately simulate natural reading conditions.

Schmidt (1994) further claims that isolation of a target word, as in a gloss, allows the reader to focus on the form and structure of the word. This focus strengthens the likelihood of retention. In the case of Chinese, especially at the beginner level, it is possible that even with the gloss, the words still look too similar for learners to adequately differentiate amongst them. As evidenced by the Part 2 responses, the Noticing Hypothesis did result in those with access to the gloss encoding the glossed language into their mental lexicons; however, it only encoded the English L1 translations, likely owing to the complexity of the orthography of Chinese characters compared to the more familiar English alphabet structure.

Those with access to the gloss had the advantage of comprehensibility. According to Schmidt (1990), input must be noticed and comprehensible in order to be acquired by the reader. In the present study, the gloss served this isolation role, separating individual characters from the dense block of text and providing L1 translations. Thus, those with the gloss were more likely to notice the individual Key Terms and could spend their attention translating non-given words, focusing on the more intricate details of the text while those in the no-gloss group spent all their effort searching for the main meaning of the text. This is evident in the dichotomous degrees of specificity to which those in Group A and those in Group B responded to the Part 1 comprehension questions and Part 2 text summary. Those without the gloss were less likely to

notice the Key Terms, likely owing to the similar blocky structure. This is evidenced by the uneven distribution of character circling outlined in 3.2 Participants.

According to Hulstijn's (1992) Mental Effort Hypothesis, increasing effort at the time of learning will result in a higher probability of recall. Hulstijn endorsed this hypothesis in relation to multiple-choice glosses, as described in 1.2 Gloss studies. In the present study, which utilized single-translation glosses over multiple-choice glosses, participants were able to recall many of the gloss L1 translations, but were unable to match them to the associated L2 Key Term. Hulstijn notes that an important feature of multiple-choice glosses is their consciousness-raising ability, allowing readers to become more aware of what they are reading. Furthermore, multiple-choice glosses require more back-and-forth interaction with the text, while single-translation glosses require only a single glance. Since there is less involvement in processing while using a single-translation gloss, it stands to reason that readers would be less likely to remember the meanings after some time has passed, which is exactly what happened in the present study. The present study did not measure multiple-choice glosses, so the previous observation is largely speculative. If multiple-choice glosses were studied as well, Hulstijn (1992) would predict stronger recall ability by those with the multiple-choice gloss.

On the subject of noticing and comparison, there were several examples of Swain's (1995) "noticing the gap" principle. Unsurprisingly, the participants, especially those in Group A, remarked on the difficulty of the text. Based on the self-reported comprehension scores, it is evident the participants were aware the text was above their level. In this case, those with access to the gloss had the advantage of reading enhancement, which is evidenced by their generally higher reported comprehension scores and utilization of glossed terms in answering the comprehension questions. Of course, as mentioned, the gloss is only as helpful as the amount of

attention allocated to it. Single-translation glosses are an example of bottom-up processing, wherein the desired response is given to the participant for free. According to Taylor (2013, 2014), participants who spent more time looking at the gloss and “noticing the gap” were more likely to recall the information later.

Taylor (2014) notes that learners simply may not be developmentally ready to process linguistic items present in the gloss. Comparing occurs when learners look at an L1 gloss and are able to see both languages side-by-side. In this context, Taylor notes that processing abstract terms is difficult in context, owing to the strain of trying to relate abstract concepts to prior knowledge. In the case of Chinese, beginner participants very likely are not ready to relate new characters back to ones they know, as they are working from much smaller individual lexicons and are still largely unfamiliar with the structuring system of Chinese characters; i.e. radical systems.

Taylor further notes, citing the study of Ellis (1997), that an L1 glossed item which is above the learner’s level may be partially learned but not fully integrated into what he calls the interlanguage, the jumbled intermediary language of a learner as he or she integrates new aspects of the L2 into the individual lexicon. This is evidenced by the trend of remembering the translations of the Key Terms but failing to remember the corresponding character. Krashen (1989) notes a 5-20% chance for learners to acquire a new word after one encounter in a text.

d. On input

Krashen’s (1982) Input Hypothesis suggests that learners primarily focus on the content of the text, rather than on the form or structure of the language. Based on the Part 2 text recall summaries, it is clear participants all understood the content of the text with enough strength to

recall significant portions of the main story at a later point. No participant misremembered the central story; all participants remembered that a man raised sheep and took better care of them after some were eaten. As mentioned, Group B generally recalled the story in more specific detail than Group A.

The most important content of the text in the present study was the Key Terms, and these were commonly present in both the Part 1 and Part 2 responses of the Group B participants (including the text content recall), and were entirely absent from the responses of the Group A participants. This suggests that glossing served as effective input enhancement, allowing those in Group B to access more of the content with less effort, leading to stronger immediate comprehension and later L1 translation retention than those in Group A. It is clear the gloss in the present study resulted in increased comprehension, and, to a degree, aided retention, seemingly confirming the Input Hypothesis' central claim that increasing input will increase comprehension.

Returning to the subject of excessive input, Nagle & Sanders (1986) note that the primary resource in L2 reading is attention. Attention “involves the application of mental energy to processing tasks and may range from focusing on specific features of input to controlled processing for retrieval” (Nagle & Sanders, 1986). This type of attention, as previously noted, is limited and can be impeded or hindered. Too much new input will result in the allocation of attentional resources into smaller processing units. This process is tedious, and many participants during the reading stage of the present study likely gave up on trying to translate each individual word.

Because Chinese text is very overwhelming to beginner learners, due to its uniform structure and lack of distinct spacing, it is likely there was too much new input for the

participants to focus on, resulting in a reliance on the provided translations to deduce content and evoke answers to the comprehension questions. As a result, Krashen’s (1982) observation that too much attention causes breakdown in processing is likely evident in this study. Furthermore, since the gloss provided the most important content words, it is likely other input in the text, i.e. grammatical markers or auxiliary characters, were ignored in favor of the known, useful words.

Table 7. Supported theories in the present study

Theory	Supported by the data?
Incidental vocabulary learning	Yes
Depth of Processing Theory	Yes
Task-Induced Involvement Load Hypothesis	Yes
Cognitive Load Theory	Yes
Noticing Hypothesis	Yes
Mental Effort Hypothesis	Yes
“noticing the gap” principle	Yes
Input Hypothesis	Yes

CHAPTER 6

CONCLUSION

6.1 Summary

From the data, it is clear the gloss slightly impacted language acquisition. While reading, those with access to the gloss were able to comprehend 78% of the Key Terms on average, while those without access to the gloss were able to comprehend 42% of the Key Terms on Average. In testing translation memory after one week; however, the gloss did not seem to affect translation accuracy. Those with the gloss correctly translated 5% of terms on average, while those without the gloss did not translate any terms accurately.

In the present study, the primary benefit of the gloss is seen during Part 1 reading comprehension. As mentioned, those with the gloss produced more desired responses to the comprehension questions, supporting the idea that glossing aids in comprehension during reading. However, it is clear glossing provides minimal aid in retention, with both groups unable to provide more than three desired responses total. It is probable that glossed terms went undifferentiated from other terms owing to the homogenous makeup of Chinese characters and, as a result, did not receive adequate attention from the readers.

Notably, many of the responses in Part 2 came from L1 terms which had been given either as in-text translations or in the gloss. On average, 33% of written responses from those in Group A originated in this fashion, while 44% of written responses from those in Group B did.

Those without access to the gloss may have never learned the meanings of the Key Terms in the first place. However, in trying to remember the meaning they associated with each Key Term from the Part 1 comprehension questions, it is clear they could not accurately recall any L1 translation they may have produced during the Part 1 reading.

In summarizing the events of the text one week later, a similar trend emerges. Those in Group A, on average, used 8% of Key Terms in their summaries, and 63% of in-text translations. In contrast, those in Group B used 33% of Key Terms and 56% of in-text translations, on average.

6.2 Implications

The present study slightly reinforces the advantages of glossing while reading in the L2. In the present study, those with access to the gloss were able to recall the English L1 terms at a later date. Though they could not accurately assign the correct meaning to the associated L2 Key Term, the impact of the gloss on their processing ability is clear. With more Key Term repetition in the text, it is likely more of those in Group B would have been able to recall terms more accurately. Single-translation glosses, at the very least, aid in character recognition, and can prompt general L1 recall.

Those in Group A did not accurately translate any Key Terms after one week, instead remembering the English terms given as in-text translations. The trend of both groups remembering the L1 they were given suggests that lack of repeated exposure was a major variable responsible for impeded form-mapping and processing. For language teachers interested in teaching vocabulary through natural reading with glossing and incidental vocabulary learning, it will be more effective to repeat the desired terms throughout the text in order to increase exposure and likelihood of acquisition.

Another variable that was mentioned but not specifically accounted for in the study is the uniformity of the written Chinese language relative to English. A language learner should have enough experience in recognizing the nuances of his or her language structure and form in order to fully utilize reading enhancement strategies. Enhancement works by strengthening existing skills; as a result, those with more language learning experience will be better served by such techniques.

CHAPTER 7

LIMITATIONS

7.1 Study limitations

As mentioned repeatedly, this study would have been much better served by a larger pool of participants. The present study, as a result of lower participation, shifted in premise from a statistical analysis to informal observations regarding correlation.

As outlined in 3.2 Participants, participants were sourced from two different Chinese language classes, 120 and 126. These two classes are roughly the same level, as evidenced by the data, which does not show any major discrepancy between classes. However, a more reliable study would source them from one class only, even though accounting for individual language background is nearly impossible regardless (Rott, 2005; Lomicka, 1998; Alessi & Dwyer, 2008).

Laufer's (1989) and Hu & Nation's (2000) principle of 95% text familiarity for optimized comprehension is a widely accepted standard of second language acquisition. Given that the average self-reported comprehension rate was around 80%, it can be safely assumed this particular text was too difficult for the learners, the result of which being an overly substantial reliance on the gloss by those with access to it and minimal comprehension by those without it. Ideally, the Key Terms would have been the only unfamiliar characters in the text, which is an unrealistic expectation owing to the diverse language backgrounds of the individual participants.

Though the formal think-aloud procedure did not occur during this particular study, several metacomments emerged from the participants throughout the process which confirmed that many of the intendedly familiar characters were, in fact, unknown. In the present study, the mixed results as to the accuracy of the translation responses during the comprehension test suggest that the L2 vocabulary was not as accessible as had been planned.

The present study included each term in the text only once, with the exception of 建立, “to build,” which appeared twice more in the modified form 建, reducing likelihood of repeated form-mapping. The brevity of this particular text was largely responsible for the infrequency of the Key Terms. The study aimed as well to mimic natural L2 reading as closely as possible, which is why the Key Terms were not repeated often within the text.

The present study did not analyze other dependent variables related to the gloss, such as speed of reading. The study also did not compare different gloss types against each other, owing to the shortage of participants and widespread agreement that glossing increases comprehension compared to no glossing. The present study did not analyze any variables besides the absence or presence of a gloss.

The present study also focused solely on Chinese reading ability, which is difficult to measure in isolation owing to the pictographic nature of the written characters. A future study desiring to emphasize the uniqueness of reading Chinese characters might analyze speaking and listening comprehension as well to incorporate pinyin, such as in Xu (2010), who suggests that learning is more likely to occur when learners can build meaningful connections between visual and verbal stimuli. If Chinese characters were to be analyzed in terms of visual stimuli in a manner similar to photographs or videos, a study using multimedia glosses would be interesting.

Several prior studies, such as Taylor (2014), note a tipping point, wherein glossing too much of the text will result in decreased comprehension, owing to an overreliance on L1 and less interaction with the text itself. Perhaps glossing too much increases extraneous cognitive load in a manner similar to looking up a term in a dictionary. The present study did not analyze this trend, instead opting to provide glosses of equal size to the participants with them. Different sized glosses were briefly considered in an earlier design, but were decided against owing to the

small pool of participants. A variable which may be accounted for in a future study could be at which point too much glossing becomes a distraction for the reader.

7.2 Screening concerns and justifications

The screening process was not perfect, of course. Owing to the beginner-learner status of the participants and the simplicity of the experiment design, some flexibility with screening was allowed. Notably, several participants did not circle one or more of the Key Terms, implying that they had come across the character in their studies before. Fortunately, this was not an issue, and the screening process itself did not result in the dismissal of any of the participants from the study. There are several reasons which may explain this trend of under-circling and justifications for disregarding it.

First, though a participant may have implied familiarity with a character, this translates only to a recognition of the written form. Recognizing a character's written form does not equate to knowledge of the character's meaning in the learner's L1. For example, Participant #9 did not circle 应该, "should" during the initial reading; however, he did not correctly recall the meaning of this character in Part 2, prompting the possibility that he recognized the character from having read it before but did not know its meaning without the gloss. Conversely, he may have seen the term's meaning in the gloss and recognized it from spoken language without having seen the written character, and thus disregarded it in his circling process. As evident in the results, though the participants may have indicated familiarity with a written character, it was clear from the overwhelmingly incorrect answers in Part 2 that the exact English meaning of the characters eluded them.

Similarly, several participants, such as Participant #3, circled only six unfamiliar characters, the minimum asked. She estimated that she understood only 65% of the text as well, suggesting that there were unfamiliar characters which she did not circle. The question, unfortunately, did not specifically ask to identify all unfamiliar characters in the text. Doing so would have provided more precise insight into the screening process, with the additional effect of the reader's devoting more attention to the text, specifically the characters comprising it.

Second, participants may have simply failed to notice an unfamiliar character while reading, in a manner similarly described in studies such as Hulstijn (1996). This can likely be attributed to the relative uniformity of size and structure of Chinese characters compared to the diversity of length of words using the Roman alphabet; an unfamiliar character would be more likely to avoid a novice L2 learner's attention than might an unfamiliar letter-based word. For example, Participant #3 did not circle the Key Term 建立, "to build," the first time it appeared, implying that she had seen both characters before while reading. She did not circle 建 the second time it appeared in the text. However, she did circle 建 the third time it appeared in the text, suggesting either that she had not noticed the character while reading it before, that she did not recognize it outside of the familiar context of 建立, or that she mistakenly thought it was a different character.

Interestingly, unlike other participants, Participant #3, when recounting the translations in Part 2, listed out pinyin in several places. Her written pinyin was not correct: for the 立 in 建立, she wrote out "wei," when the actual pinyin is "li." However, the similar-looking character 位 is pronounced "wei," suggesting that she had confused the two characters, or had never actually seen 立 in reading and her brain was falsely equating it with 位. Though phonetic

pronunciation of a character does not equate to familiarity with meaning, the implication is simply that L2 learners are not perfect speakers and often misperceive their own abilities.

Third, several participants may have been impacted by their own L2 abilities. This is most demonstrable where participants erroneously circled unconnected sequential characters as one linguistic item, as detailed in Chapter 4: Results and Observations. For example, in the phrase 已经发生了, both Participant #7 and Participant #9 incorrectly circled 经发 in one stroke, implying she thought those two characters formed one term, when, in actuality, 已经 is one unit which means “already” and 发生 is another which means “to happen.” Similarly, in context, it is easy to incorrectly assume familiarity with the second character in a phrase following a noticing of a familiar first character. The sharing of radicals amongst characters in the Chinese lexicon leads to many characters possessing similar forms, and learners do not have perfect recognition abilities.

Finally, several participants with access to the gloss may not have circled Key Terms that were presented to them below the text, such as Participant #9. The instructions did not indicate whether or not to circle glossed terms, owing to the individual nature of reading. It was also likely that a participant would, intentionally or not, mistakenly equate seeing a word in the gloss with prior familiarity, which would result in fewer Key Terms circled. Following the example of Participant #3 circling 建 the third time it appeared in the text, it is possible she saw it first in the gloss and then tricked herself into believing she recognized the character the first time it appeared in the text since it was in her most recent working memory, but then failed to recognize it later without properly accessing the gloss immediately before again.

BIBLIOGRAPHY

Alessi, S., & Dwyer, A. (2008). Vocabulary assistance before and during reading. *Reading in Foreign Language*, 20(2), 246-263. doi: 10125/66815

Bell, F. L., & LeBlanc, L. B. (2000). The Language of Glosses in L2 Reading on Computer: Learners' Preferences. *Hispania*, 83(2), 274–285. <https://doi.org/10.2307/346199>

Bowles, M. A. (2004). L2 Glossing: To CALL or Not to CALL. *Hispania*, 87(3), 541–552. <https://doi.org/10.2307/20063060>

Cook, G. (1989). *Discourse*. Oxford: Oxford University Press.

Craik, F. & Lockhart, R. (1972). Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671-684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X).

Day, R. R., & Park, J.-s. (2005). Developing reading comprehension questions. *Reading in a Foreign Language*, 17(1), 60–73.

Duan, S. (2018). Effects of Enhancement Techniques on L2 Incidental Vocabulary Learning. *English Language Teaching*, 11, 88-101. <http://doi.org/10.5539/elt.v11n3p88>.

Hulstijn, J. H. (1989). Implicit and incidental second language learning: experiments in the processing of natural and partly artificial input. In H. W. Dechert, & M. Raupach (Eds.), *Interlingual processes*, 49-73. (Language in performance; No. 1). Narr.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: experiments in incidental learning. In P. J. L. Arnaud, & H. Béjoint (Eds.), *Vocabulary and applied linguistics*, 113-125. Macmillan.

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. <https://doi.org/10.2307/328937>

Hulstijn, J. H. (1995). Not all grammar rules are equal: giving grammar instruction its proper place in foreign language teaching. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning*, 359-386. (Technical report; No. 9). Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.

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. <https://doi.org/10.2307/329439>

- Hu, M. & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430. <https://doi.org/10.26686/wgtn.12560354>
- Jacobs, G. (1994). What lurks in the margin: Use of vocabulary glosses as a strategy in second language learning. *Issues in Applied Linguistics*, 5, 115-137.
- Joyce, E. E. (1997). Which words should be glossed in L2 reading materials? A study of first, second and third semester French students' recall (report number FL 024 770). *Pennsylvania Language Forum* (ERIC Document Reproduction Service No. ED 427 508).
- Knight, S. (1994). Dictionary Use While Reading: The Effects on Comprehension and Vocabulary Acquisition for Students of Different Verbal Abilities. *The Modern Language Journal*, 78(3), 285–299. <https://doi.org/10.2307/330108>
- Ko, H. M. (2005). Glosses, comprehension, and strategy use. *Reading in a Foreign Language*, 17(2), 125-143.
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon Press Inc.
- Krashen, S. (1989). We Acquire Vocabulary and Spelling by Reading: Additional Evidence for the Input Hypothesis. *The Modern Language Journal*, 73(4), 440–464. <https://doi.org/10.2307/326879>
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension. In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines*, 316–323. Multilingual Matters.
- Laufer, B & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: the construct of Task-Induced Involvement. *Applied Linguistics*, 22, 1-26. doi: 10.1093/applin/22.1.1.
- Lomicka, L. L. (1998). “To gloss or not to gloss”: An investigation of reading comprehension online. *Language Learning & Technology*, 1(2), 41-50. <http://dx.doi.org/10125/25020>
- Nagata, N. (1999), The Effectiveness of Computer-Assisted Interactive Glosses. *Foreign Language Annals*, 32, 469-479. <https://doi.org/10.1111/j.1944-9720.1999.tb00876.x>
- Nagle, S. J., & Sanders, S. L. (1986). Comprehension Theory and Second Language Pedagogy. *TESOL Quarterly*, 20(1), 9–26. <https://doi.org/10.2307/3586386>
- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. *Reading Research Quarterly*, 20(2), 233–253. <https://doi.org/10.2307/747758>
- Nation, I. (2001). *Learning Vocabulary in Another Language* (Cambridge Applied Linguistics). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524759

- Paribakht, T.S. and Wesche, M. (1997) Vocabulary Enhancement Activities and Reading for Meaning in Second Language Vocabulary Acquisition. *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, 55, 174-200. <https://doi.org/10.1017/CBO9781139524643.013>
- Parry, K. (1991). Building a Vocabulary through Academic Reading. *TESOL Quarterly*, 25(4), 629–653. <https://doi.org/10.2307/3587080>
- Prince, P. (1996). Second Language Vocabulary Learning: The Role of Context versus Translations as a Function of Proficiency. *The Modern Language Journal*, 80(4), 478–493. <https://doi.org/10.2307/329727>
- Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24, 146-161. doi:10.1017/S0267190504000078
- Rott, S. (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language*, 17, 95-124.
- Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, 6, 72-78. [https://doi.org/10.1016/0346-251X\(78\)90027-1](https://doi.org/10.1016/0346-251X(78)90027-1)
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 1, 129-158. doi: 10.1093/applin/1.2.129
- Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages*, 165-209. London: Academic Press.
- Sweller, J. (2005). Implications of Cognitive Load Theory for Multimedia Learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning*, 19-30. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511816819.003
- Taylor, A. M. (2006). The Effects of CALL Versus Traditional L1 Glosses on L2 Reading Comprehension. *CALICO Journal*, 23(2), 309–318. <http://www.jstor.org/stable/24156249>
- Taylor, A. M. (2013). CALL versus Paper: In Which Context Are L1 Glosses More Effective? *CALICO Journal*, 30(1), 63–81. <http://www.jstor.org/stable/calicojournal.30.1.63>
- Taylor, A. M. (2014). Glossing Frequency and L2 Reading Comprehension: The Influence of CALL Glossing. *CALICO Journal*, 31(3), 374–389. <http://www.jstor.org/stable/calicojournal.31.3.374>
- Watanabe, Y. (1997). Input, Intake, and Retention: Effects of Increased Processing on Incidental Learning of Foreign Language Vocabulary. *Studies in Second Language Acquisition*, 19(3), 287–307. <http://www.jstor.org/stable/44487970>

Xu, J. (2010). Using Multimedia Vocabulary Annotations in L2 Reading and Listening Activities. *The CALICO Journal*, 27, 311-327. <http://dx.doi.org/10.11139/cj.27.2.311-327>

APPENDIX A
STUDY PART 1: READING, GROUP A (NO GLOSS)

Name:

Class #:

Date:

Group A

Text:

以前有一个人养了很多羊 (sheep)，可是他没给羊建立一个羊圈 (pen)。

有一天晚上，跑来很多狼 (wolves)，吃了十只羊。那个人很后悔，他应该给羊建一个好的羊圈。在羊被吃了以后，他才 (only then) 给羊建了一个羊圈。

这个成语的意思是：不好的事情 (things) 已经发生了，才可以补救。

Comprehension Questions

Please answer as specifically as possible in English.

1. When does this story take place?
2. What does the man do (for work)?
3. What happens to the sheep?
4. What does the man do after this happens?
5. How does the man feel after the events of the story?
6. What should the man have done, according to the text?
7. What is the meaning of 成语?
8. Translate the final sentence as best you can, starting after the “:”
9. Please circle at least 6 characters in the text you did not recognize
What percentage of the text did you understand? Please estimate.

APPENDIX B
STUDY PART 1: READING, GROUP B (GLOSS)

Name: _____ Class #: _____ Date: _____ Group B

Text:

以前有一个人养了很多羊 (sheep)，可是他没给羊建立一个羊圈 (pen)。

有一天晚上，跑来很多狼(wolves)，吃了十只羊。那个人很后悔，他应该给羊建一个好的羊圈。在羊被吃了以后，他才 (only then) 给羊建了一个羊圈。

这个成语的意思是：不好的事情 (things) 已经发生了，才可以补救。

Gloss:

1. 养 : To Raise
2. 建立 : To Build
3. 后悔 : To Regret
4. 应该 : Should
5. 成语 : Idiom
6. 补救 : To Remedy

Comprehension Questions

Please answer as specifically as possible in English.

1. When does this story take place?
2. What does the man do (for work)?
3. What happens to the sheep?
4. What does the man do after this happens?
5. How does the man feel after the events of the story?
6. What should the man have done, according to the text?
7. What is the meaning of 成语?
8. Translate the final sentence as best you can, starting after the “:”
9. Please circle at least 6 characters in the text you did not recognize
What percentage of the text did you understand? Please estimate.

APPENDIX C
STUDY PART 2: MEMORY (BOTH GROUPS)

Part 2

The following characters appeared in the text. Please recall the translations, based on the text, as best you can.

1. 养 :
2. 建立 :
3. 后悔 :
4. 应该 :
5. 成语 :
6. 补救 :

Please briefly summarize the story:

APPENDIX TABLES

Appendix Table 2.A. Written results of Part 2

Group	Participant	1. 养 “to raise”	2. 建立 “to build”	3. 后悔 “to regret”	4. 应该 “should”	5. 成语 “idiom”	6. 补救 “to remedy”
A	Participant #2	Sheep	Pen	Later on	Wolf	Herder (of sheep)	Consequence
A	Participant #5	Pen?	-	-	Sheep?	-	-
A	Participant #6	-	Sheep	Wolf	Count	-	-
A	Participant #7	-	-	-	-	-	-
B	Participant #1	Sheep	Idiom	<i>To Repent</i>	-	Wolf	To raise
B	Participant #3	Farmer	Wolves	Pen	Regrets	-	Should have
B	Participant #4	Sheep	-	-	-	-	-
B	Participant #8	To build	Shepherd	Again(?)	Solution	Regret	Problem
B	Participant #9	Pen	To establish/ build	Regret	-	Idiom	-
B	Participant #10	pen	-	daily	<i>supposed</i>	<i>Fable/ story</i>	-

Appendix Table 2.A: Appendix Table 2.A: Results of Part 2 recall portion, as recorded by the participants. A dash (-) refers to an answer that a participant left blank. A bolded answer represents a desired response; an italicized answer represents an acceptable (or otherwise notable) response

Appendix Table 2.B. Written results of Part 1

Group	Participant	1. 养 “to raise” Q #2	2. 建立 “to build” Q #4, 6	3. 后悔 “to regret” Q #5	4. 应该 “should” Q #6	5. 成语 “idiom” Q #7	6. 补救 “to remedy” Q #8
A	Participant #2	Shepherd	Built Built	Regretful	Should have	-	Deal with
A	Participant #5	Herd	Gives Given	-	-	-	-
A	Participant #6	Take care of	- Count	Regret?	Should have	Lesson?	Validate
A	Participant #7	Owns	Gave Built	Content/sad	-	-	-
B	Participant #1	Raises	Builds Building	Regrets	Should have	Idiom	Remedy
B	Participant #3	Raises	- Built	Regrets	Should've	Idiom	Be remedied
B	Participant #4	Raises	Builds Built	Regrets	Should have	Moral	To remedy
B	Participant #8	Farmer	Builds Built	Regretful	-	Idiom	Coming up with a solution
B	Participant #9	Shepherd	Built Built	Regrets	-	Idiom	Remedy
B	Participant #10	Shepherd	Builds Built	Regretful	Should've	Idiom	Remedy
A	*Participant #W1	Takes care of	Sets up Set up	Sad	-	Story	Learn your lesson
A	*Participant #W2	Shepherd	Improves Given treatment	Sad	-	Story, moral, narration	Solve

Appendix Table 2.B: Results of the Part 1 reading comprehension questions. This table displays the L1 word or phrase produced by each participant in answering the corresponding comprehension question of each Key Term. A dash (-) represents an irrelevant answer provided by a participant.