Spring 2016

Exploring depth of vocabulary knowledge among CFL learners of higher proficiency levels

Tingting Chen
University of Iowa

EXPLORING DEPTH OF VOCABULARY KNOWLEDGE AMONG CFL LEARNERS OF HIGHER PROFICIENCY LEVELS

by

Tingting Chen

A thesis submitted in partial fulfillment
of the requirements for the Doctor of Philosophy
degree in Teaching and Learning (Foreign Language & ESL Education) in the
Graduate College of
The University of Iowa

May 2016

Thesis Supervisor: Emeritus Associate Professor Michael E. Everson

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Tingting Chen

has been approved by the Examining Committee for
the thesis requirement for the Doctor of Philosophy degree
in Teaching and Learning (Foreign Language & ESL Education)
at the May 2016 graduation.

Thesis Committee: _____
Michael E. Everson, Thesis Supervisor

_____
David C. Johnson

_____
Lia M. Plakans

_____
Leslie L. Schrier

_____
Helen Shen

_____
Walter P. Vispoel

_____
Pamela M. Wesely

To my beloved husband and our three boys

ACKNOWLEDGEMENTS

I would like to express my very great appreciation to my academic advisor Prof. Michael Everson for his unfailing support during the entire process of my PhD study and thesis writing. He showed great understanding and patience in my initial proposal drafting stage, when I continued to change from one thesis topic to another in my indecisiveness. He worked very hard to ensure that my data collection could be conducted smoothly despite the many obstacles we encountered. His words of encouragement brought me through many moments of anxieties and frustrations, always reassuring me that I had the ability to carry my thesis to a successful completion. He is the person who not only gave the most helpful guidance at each step of my thesis development, but also carefully revised each of my grammatical or vocabulary errors whenever he noticed one. It is not exaggerating to say that without Prof. Everson's support, I would not have accomplished what I did in my thesis and PhD work.

My sincere gratitude also goes to each of my committee members—(in alphabetic order) Prof. David Johnson, Prof. Lia Plakans, Prof. Leslie Schrier, Prof. Helen Shen, Prof. Walter Vispoel, and Prof. Pamela Wesely. I was especially fortunate to have a large dissertation committee, as each of my committee members had helped me uniquely from his or her own special area of expertise—Prof. Johnson as an educational linguistics expert, Prof. Plakans from a second language assessment perspective, Prof. Schrier with her deep insights into second language education as a whole, Prof. Shen from her rich CFL research background, Prof. Vispoel from his strong educational measurement and statistics expertise, and Prof. Wesely with her profound knowledge in research methodology. Their feedback on my thesis, favorable or critical, was of tremendous help in refining my current work and will continue to greatly influence me in any of my future research endeavors.

A special acknowledgement also goes to all the professors who kindly offered me the opportunity to conduct data collection within their programs and all the students who participated in my study. I truly appreciate their help that made this thesis possible.

Finally, I would like to give my loving gratitude to my husband and our three sons, who sacrificed many hours of family time to make sure that I could work on my thesis in an undisturbed and timely way. Your love and support have been the strongest motivation for me to complete my thesis and PhD study.

ABSTRACT

L2 (second or foreign language) research indicates that vocabulary knowledge is not only the "single best predictor of text comprehension," but also a strong indicator of listening, speaking, and writing proficiency (Alderson, 2000, p. 35). Understanding the development of vocabulary knowledge, including both vocabulary size and vocabulary depth, or quality of vocabulary knowledge—is therefore essential to the building of an overall insight into L2 proficiency.

This study aims to explore the developmental status of vocabulary depth among postsecondary CFL (Chinese as a foreign language) learners of higher proficiency levels who have studied Chinese for over four years. In particular, it focuses on these learners' identification of two types of word association—synonym and collocational associations and how factors such as association type and target-word frequency impact association identification. The process and strategy use that are involved in the inference of word association are also explored.

For these purposes, this study employs a combination of quantitative and qualitative research methods. Cross-sectional data were collected through a paper-and-pencil test of Chinese word associates from seventeen learners within five postsecondary CFL programs in the U.S. Each participant worked on two test booklets assessing synonym and collocational associates respectively for the same 44 adjectives selected from the three word frequency levels of below 1000, between 1000 and 5000, and above 5000. A two-factor within-subjects ANOVA revealed both significant main effects for association type and word frequency on association identification and a significant interaction between the two. Simple effect analysis and pair-wise comparisons further revealed that association identification became increasingly stronger with the increase of word frequency for collocational association, yet remained non-impacted by frequency before reaching the mid- to high-frequency transition for synonym association.

Meanwhile, CFL learners' collocational knowledge was significantly higher than synonym knowledge at mainly the medium- and high-frequency levels. These results indicate that synonym knowledge seemed to lag behind in development as familiarity with words increased, but began to catch up at higher-frequency levels.

Interview data collected from six CFL learners show that they employed a wide variety of knowledge sources, such as radical knowledge, morphological knowledge, contextual clues, sound information, or L1 in inferring word association. Inference success seemed to be influenced not only by their preexisting word knowledge, but also an integrated and flexible use of linguistic and contextual information in the inference process.

Implications of these findings are discussed in relevance to curriculum and pedagogical development of CFL teaching and the understanding and definition of CFL proficiency in general. This study fills a gap in CFL vocabulary research by building a tentative measure of vocabulary depth and bringing greater insights into the developmental status of higher-level CFL learners in synonym and collocational association as well as the process that is involved in inference of word association.

PUBLIC ABSTRACT

This study investigates how Chinese vocabulary knowledge is developed among learners of Chinese as a foreign language at postsecondary level. In particular, it focuses on learners' ability to identify two types of association among words—synonym association or words that share the same or similar meanings and collocational association, or words that form a collocation or frequently appear together in a phrase.

A Chinese Word Associates Test was developed for this purpose. Quantitative data from seventeen participants' test performance and qualitative data from six participants' post-test interviews were collected.

Results of the study indicate that learners in the study were stronger in performance with synonym association than with collocational association. They were also stronger in performance with words that are more frequently used than those that are less frequently used. Moreover, the performance gap between the two association types depended on the level of word frequency, with a significant difference only at high and medium word frequency levels, but not at low frequency. The performance gap among the different levels of word frequency depended on the association type—for collocational association, the more frequently a word is used, the higher was learners' performance; for synonym association, a significant improvement in performance was only present when word frequency transitioned from medium to high frequency.

Due to the limitations in test design and the small number of participants, the performance differences observed in this study might not be truly representative of differences in association knowledge. However, results of the study do provide useful implications on how to develop association knowledge in a more effective way, and how to incorporate association knowledge as an essential aspect of CFL (Chinese as a foreign language) proficiency.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Despite its traditional status as one of the "less-commonly-taught" languages in the U.S., Chinese has been designated as a critical language for economic and strategic purposes with its learner population undergoing a dramatic increase over recent decades. A survey conducted by the Modern Language Association revealed that there had been a 51.0% increase in Chinese course enrollments in higher education in the U.S. from 2002 to 2006, and a 18.2% increase from 2006 to 2009 (Furman, Goldberg, & Lusin, 2010, p. 20). Similarly, advanced-level enrollments were also on the rise, going from 18.4% of the total undergraduate enrollments in 2006 up to 20.5% in 2009 (Furman, Goldberg, & Lusin, 2010, p. 27). To meet the pressing need for highly proficient Chinese users, Chinese Flagship programs have been established since 2002 with the specific goal of producing learners at the professional level of Chinese language proficiency.

The reality, however, is that this increasing popularity of the Chinese language is offset by the greater challenges involved in learning it, challenges not typically experienced by American students who learn many other foreign languages. In fact, Chinese was classified by the Foreign Service Institute and the Defense Language Institute as a Category 4 language together with Arabic, Japanese, and Korean (Walton, 1992, p. 163), requiring almost three times the contact hours needed to learn a Category 1 language in order to reach a similar proficiency level (Walton, 1992). Further complicating the situation is a relatively weak literature base in Chinese as a foreign language (CFL) learning and assessments, especially in understanding advanced language proficiency. One such issue is the development of vocabulary knowledge at higher levels of language proficiency as well as its assessment, given the complexity of the Chinese orthographic system. The following sections will discuss why this becomes the focus of the current study.

1.1 Vocabulary Knowledge and its Assessment

The important role of vocabulary in language learning and use is universally recognized. As put by Wilkins (1972), "without grammar, very little can be conveyed. Without vocabulary, nothing can be conveyed" (p. 111). In both first (L1) and second language (L2) acquisition, sophistication of vocabulary development is a key indicator or predictor of language proficiency. Particularly, in L2 reading, researchers have investigated the relationship between vocabulary competence and reading comprehension from at least the following three perspectives and findings: (1) Vocabulary knowledge, conceptualized as either vocabulary size or the quality of vocabulary knowledge, correlates highly with reading comprehension. Alderson (2000) recognized that vocabulary knowledge was the "single best predictor of text comprehension" (p. 35). Qian's study (2002) revealed that ESL learners' performance in both vocabulary size and depth (defined in a later section) explained over 50% of the variance of reading comprehension performance. (2) Vocabulary knowledge is considered a significant component of L2 reading. Based on a series of empirical studies across different L2 languages, Bernhardt's (2011) compensatory model of second language reading hypothesized that L2 language knowledge explains 30% of the variance in L2 reading comprehension. Even though this figure is relatively modest compared with many earlier studies (due to its consideration of L1 literacy), vocabulary knowledge was believed to be a major contributor. (3) A threshold level of vocabulary knowledge is considered essential for independent reading. In EFL, for instance, it was found that for "adequate" reading comprehension of academic materials to occur, readers need to have "between 6K and 8K word knowledge and about 98% of text coverage" (Laufer & Ravenhorst-Kalovski, 2010, p. 25). (4) Lexical measures, particularly those on word frequency and ambiguities of word meaning, provide useful indices on text readability or difficulty (Crossley, Greenfield, & McNamara, 2008). In addition to reading comprehension, research also has indicated that vocabulary knowledge is a significant factor in advanced

L2 listening comprehension and that a comparable lexical coverage (98%) was also essential for unaided understanding of listening texts as for written texts (Stæhr, 2009). Vocabulary knowledge also has strong explanatory power in speaking and writing performance, and rating of speaking and writing frequently relies on indices such as accuracy or appropriateness of word use and lexical richness or diversity (Milton, 2013).

1.1.1 Defining Vocabulary Knowledge

Given the importance of vocabulary knowledge in L2 performance, it is important to understand what constitutes learners' vocabulary knowledge and how it develops over time and across different proficiency levels. Earlier conceptualization of L2 vocabulary knowledge concerned mainly the number of words learners know in a language, alternatively termed as vocabulary size, width, or breadth. Researchers have long recognized, however, that knowing a word is not an all-or-nothing phenomenon. Even well-educated native speakers of a language may not have full knowledge of the meaning or use of a given word, let alone the fact that language development in itself may continue to redefine the semantic or linguistic boundaries of its vocabulary. For L2 learners, it is reasonable to assume that as the size of their vocabulary becomes larger, they are also likely to develop an expanded and deepened understanding of the words they know. A useful distinction within vocabulary knowledge is therefore between vocabulary breadth and depth, the latter being generally understood as the quality of vocabulary knowledge (e.g., Nation, 2001; Qian, 1999, 2000, 2002; Read, 1987, 2000, 2004; Wesche & Paribakht, 1996; Nassaji, 2004).

Nonetheless, while the concept of vocabulary depth may sound intuitively appealing, defining it in specific and unambiguous terms is not without complications. This is because there can be different ways as well as different degrees of knowing a word. For instance, a learner is likely to know how a word is pronounced without knowing its meaning, or she may understand what a word means without being able to

identify its grammatical functioning. Likewise, being able to perceive word meaning correctly does not necessarily mean that one can also use it accurately in specific contexts. There is also the difference between knowing a single word meaning and knowing its multiple senses as well as its sense relationships with other words. Knowledge of vocabulary depth, therefore, is a multi-dimensional construct.

1.1.2 Frameworks for Conceptualizing Vocabulary Knowledge

One of the most comprehensive frameworks that capture the various dimensions of vocabulary knowledge is Nation's (2001) framework of "what is involved in knowing a word." According to Nation, there are three general categories of vocabulary knowledge, namely, form (i.e., what a word look or sound like), meaning (i.e., what a word refers to or what concepts are involved), and use (i.e., how and in what situations a word is used). Each of these categories has its sub-dimensions, and each sub-dimension is further divided into receptive and productive knowledge and use. Taking "meaning" as an example, receptive word knowledge and use pertain to the following questions: (1) "What meaning does this word form signal?" (Form and meaning); (2) "What is included in the concept?" (Concept and referents); (3) "What other words does this make us think of?" (Associations) (Nation, 2001, p. 26). This multi-dimensional nature of vocabulary depth has also been reflected in other frameworks in somewhat different terms. Qian (1999), for instance, posited that vocabulary depth pertaining specifically to reading should include the aspects of pronunciation and spelling, morphological properties, syntactic properties, meaning, register, and word frequency. There are also those who believe that vocabulary depth indicates knowledge of different types of associations among words and the network of links they establish (Read, 2000). Summarizing these different conceptualizations, Read (2004) proposed that there were three major lines of development regarding research on the depth of vocabulary knowledge: precision of

meaning, comprehensive word knowledge, and network knowledge (p. 211). These variations lend considerable confusion over how the concept of vocabulary depth needs to be understood and operationalized in instruction, assessment, and research contexts. Its usefulness has also been questioned for the reason that vocabulary depth "does not seem to function entirely separately from breadth" (Milton, 2009, p. 169). Meanwhile, a convincing principle seems to be lacking that unifies the different elements of vocabulary depth in the various existing frameworks. (Milton, 2009, p. 150).

1.1.3 Assessment of Vocabulary Knowledge

Despite these controversies, investigating and assessing L2 learners' vocabulary depth are arguably of great value. Rather than a dichotomous status of learner performance—they either know or don't know a word—which conventional vocabulary size measures typically reveal, well-developed assessment of vocabulary depth has the potential of presenting a more comprehensive picture of L2 learners' vocabulary development.

There are, however, several major questions to be considered in developing such tests, such as:

(1) How many and which dimensions of vocabulary knowledge can it effectively measure? No existing assessments seem to be capable of measuring all possible aspects of word knowledge, such as those delineated in Nation's (2001) framework. In fact, any such attempts would be logistically infeasible due to the large amount of time required to complete such a test. Test developers may have to face the dilemma of either testing a small sample of words in greater detail or a large and more representative sample of words in a less-detailed way. Ultimately they have to make the decision on which dimensions of word knowledge should be prioritized.

(2) How do these dimensions interrelate with each other to form a holistic rather than fragmented view of learners' vocabulary development? This points to the

importance of an underlying theory of vocabulary acquisition that weaves together the different dimensions of vocabulary knowledge being assessed in a meaningful way.

(3) How can the selected dimensions be assessed in a valid, reliable as well as practical way? This pertains to the questions of whether the test is indeed assessing the dimensions of vocabulary knowledge of interest rather than something else, and to what extent the test produces consistent results about learners' performance in these specific dimensions. Meanwhile, practicality—especially time efficiency and ease of administration and grading—is an especially important concern when a large test-taking population is involved or when the test is intended to be used in large-scale assessment settings rather than for research purposes.

1.1.4 Pro's and Con's of Typical Vocabulary Depth Measures

Different available measures of vocabulary depth have their own strengths and limitations based on the above concerns. For instance, interviews, which have often been used in research situations to assess vocabulary depth, are generally unfit for large-scale standardized assessments due to its time requirement and potential complexities in administration and rating. The Vocabulary Knowledge Scale (Paribakht & Wesche, 1993a, 1993b; Wesche & Paribakht, 1996) attempts to identify learners' word knowledge on a receptive-productive continuum by asking learners to self-report whether they have encountered a word before, whether they understand its meaning, and whether they are able to use it correctly in a sentence context. One of the most obvious limitations is that it still focuses on a restricted range of vocabulary knowledge, i.e. primarily knowledge of a single word meaning. Moreover, the inclusion of constructive responses adds to the time requirement of this measure and reduces the possible number of words being tested. A third measure, the Word Associates Test (WAT) (Read, 1987, 1993, 1995, 2000), addresses the varying strength of learners' vocabulary network knowledge by asking

them to select words that have either meaning or collocational associations with the target words, known respectively as paradigmatic and syntagmatic associations. Compared with the former two measures, the WAT has several advantages: (1) Its validity as a measure of vocabulary depth is defensible from the perspective that vocabulary depth is an indication of "network building" of word relationships (Henriksen, 1999). Evidence of convergent validity (correlations being .76 and .85) was found in Read's comparison of WAT with two criterion measures of vocabulary depth (matching and interview) (Read, 1998). (2) The internal consistency of this measure was verified by multiple studies (Nassaji, 2004; Read, 1995; Qian, 1998, 1999). (3) In terms of time efficiency and potential representativeness of the target word sample, the selected-response format puts it at a clear advantage over interviews and Vocabulary Knowledge Scale. (4) Development of the measure and analysis of test results may benefit from the strong theoretical support of L1 and L2 word association studies as well as comparative studies of the two (e.g., Wolter, 2001; Namei, 2004; Nissen & Henriksen, 2006; Fitzpatrick & Izura, 2011). Among other things, these studies explore questions such as whether a paradigmatic-syntagmatic shift exists among L2 as well as L1 learners, whether L2 learners' association responses are more idiosyncratic and less stable than those of native speakers, and to what extent the developing mental lexicon (or the "mental dictionary" where word information is stored) of L2 learners resembles or differs from that of native speakers. Although a comparison between L1 and L2 learners' vocabulary development is not the focus of the current study, a better understanding of these questions will nonetheless give us a deeper insight into the developmental status of CFL learners in word association knowledge. For the above reasons, the WAT lends itself to a convenient and useful measure of vocabulary depth for the current study.

1.1.5 Assessment of CFL Vocabulary Depth

Since the WAT was developed to assess ESL/EFL vocabulary competence, adaptations would have to be made for its use in the CFL context. This does not mean, of course, that a simple translated version of the original test will be appropriate. Among other things, the special features of the Chinese writing system, Chinese vocabulary and the way CFL learners develop their vocabulary competence need to be considered. Firstly, the Chinese language employs a logographic writing system with a relatively weak correspondence between speech and its written symbols, Chinese characters. In other words, the pronunciation of a Chinese character is usually not directly perceivable by the way it is written. While word components termed phonetic radicals may provide some hints about a character's pronunciation, such information is often unreliable. Despite this, CFL researchers have still found that sound mediation is an essential strategy for accessing meaning in Chinese, much as it is in alphabetical L2 languages (Everson, 1998; Everson & Ke, 1997). This means that for a better understanding of learners' performance in word recognition and vocabulary knowledge tests, we also need to know to what extent learners rely on being able to know how to pronounce a word. Secondly, Chinese words are made up of one or more Chinese characters, with often semantic connections between a word and its component characters. Therefore, investigating learners' knowledge of word associations would help us understand how learners derive individual character meaning, as well as how they construct networks of semantically related characters. In addition, orthographic knowledge, particularly knowledge of semantic components, also contributes to an understanding of word meaning and how Chinese words associate with each other (Shu & Anderson, 1999; Shen & Ke, 2007). Thirdly, the way Chinese collocations are connected with each other is also different from English. For instance, adjectives and corresponding noun collocations are often connected by the particle "的" (DE). Adaption of the WAT would therefore have to take these characteristics into consideration.

1.2 Purpose Statement

This study investigates depth of vocabulary knowledge among postsecondary CFL learners of higher proficiency levels. In particular, it explores whether two different types of word association—synonym and collocational association—develop in a parallel way.

Due to the lack of effective measures in assessing CFL vocabulary depth, a test of Chinese word association was developed on the basis of Read's WAT, which was originally designed for ESL/EFL learners. Characteristics of Chinese morphology and word association patterns were given careful consideration in the development of this measure. However, the ultimate purpose of this study is more for vocabulary and language proficiency research rather than for the development and validation of a vocabulary depth measure.

1.3 Research Questions

The primary focus of this study is on the investigation of depth of vocabulary knowledge among CFL learners of higher proficiency levels, especially their identification of synonym and collocational association. Research questions are:

1. Do CFL learners in this study identify synonym and collocational association equally well?
2. Do CFL learners identify synonym and collocational association equally well at different levels of word frequency?
3. What process is involved for CFL learners in inferring if two words are associated?

1.4 Significance of the Study

The significance of this study is closely tied to the need to more accurately define advanced- and higher levels of L2 proficiency. While language standards such as the ACTFL Proficiency Guidelines (1982, 1986, 2012) have made significant contributions

to understanding proficiency development amongst L2 learners, their validity is constantly questioned. A major criticism against the ACTFL Guidelines is its lack of empirical evidence, so much so that its level designation and proficiency descriptions are considered as "lucky approximations at best and mere fabrications at worst" (Liskin-Gasparro, 2003, p. 484). Refining (or revising when necessary) the proficiency standards by building a stronger empirical base would therefore be necessary. Since the variable of interest for this study is vocabulary knowledge, it is worthy of note that descriptions about vocabulary competence in the ACTFL Guidelines are minimal and vague. With Superior-level reading proficiency, for instance, it is only mentioned that learners' comprehension needs to be supported by "a broad vocabulary" and that they should be able to understand texts "that use precise, often specialized vocabulary" (ACTFL, 2012, p. 21). Empirical research is therefore needed to better understand vocabulary development at different levels of L2 proficiency and operationalize these differences in precise terms.

On a smaller scale, any potential contributions of the study are most relevant to vocabulary and language proficiency development in the CFL context. Vocabulary acquisition research in CFL has traditionally given greater emphasis to lower-level character and word recognition (e.g., Everson, 1998; Ke, 1996, 1998; Shen, 2005, 2010; Williams, 2013). This emphasis is defensible especially since word recognition is one of the greatest challenges for beginning Chinese learners due to the dense orthographic nature of the Chinese writing system. As their proficiency level increases, however, students' establishing one-on-one form-meaning correspondence is far from sufficient in coping with the more sophisticated language interpretation and use situations that proficient reading demands. Leaver & Shekhtman (2002) identified "precision of lexicon" as an important trait of superior-level language proficiency, which involves "an intuition of the multiple meanings of words and their correct (grammatical and cultural) usage" (Leaver & Shekhtman, p. 26). Similarly, when discussing CFL programs at higher

proficiency levels, Kubler (2002) emphasized that “the individual words themselves gain in importance” at these levels and that “the learner must know the exact meaning and usage of words – both common and rare, as well as synonyms and antonyms” (Kubler, p. 111). However, whether higher-level CFL learners do develop such knowledge is largely unknown and research in this direction is limited. It is important that the current study may fill this gap.

In order to explore how word knowledge is developed at higher proficiency levels, a valid way of assessing vocabulary depth in the context of CFL needs to be developed. Rather than building one from scratch, a more advisable approach would be to adopt an existing measure that has been theoretically and empirically validated and revise it for our own purposes. Compared with other measure, the WAT can effectively assess several important dimensions of vocabulary depth at the same time in an efficient manner— (1) Polysemy, or multiplicity of meaning in a word; (2) Synonymy, or equivalence in meaning between two or more words; (3) Collocation, or combination of words frequency used together. The WAT was also built upon rich L1 and L2 research foundations on word associations, which not only gave strong support to the validity of the instrument but also made it possible to interpret learner performance in a more theoretically meaningful way. Adapting the test for CFL also may have several significant benefits to our field: (1) It helps fill a gap in CFL where effective measures of vocabulary depth are lacking; (2) A test such as this may yield important insights about how vocabulary depth develops across proficiency levels, especially at higher proficiency levels given the learner sample that participated in this study. (3) A useful by-product of the instrument development and use is greater sensitivity to the cross-linguistic differences between English and Chinese, especially in the different ways the vocabulary networks are constructed.

1.5 Definitions of Terms

Less Commonly Taught Languages: Less Commonly Taught Languages in the U.S. has traditionally referred to languages other than the three most commonly taught foreign languages in U.S. public schools, namely, French, German, and Spanish. Examples of LCTLs include Mandarin Chinese, Japanese, Korean, Russian, Arabic, etc. There has been a rapid growth in enrollments at both secondary and postsecondary level in some of the LCTLs (such as Chinese), especially due to the government's effort to develop proficiency in languages critical to U.S. competitiveness and security.

The ACTFL Proficiency Scale: The ACTFL Proficiency Scale (1982, 1986, 2012) was developed from the Federal Government's ILR Scale by the American Council on the Teaching of Foreign Languages. It has five main levels: Novice, Intermediate, Advanced, Superior, and Distinguished. The first three levels are further divided into low, mid, and high sublevels. The scale has detailed performance descriptions for each level and its sublevel in the four skill areas of Speaking, Writing, Listening, and Reading.

Superior-level proficiency: Superior-level proficiency is the second highest proficiency level in the ACTFL Proficiency Scale. For instance, reading at a Superior level is defined by the ACTFL Proficiency Guidelines in the following way:

> At the Superior level, readers are able to understand texts from many genres dealing with a wide range of subjects, both familiar and unfamiliar. Comprehension is no longer limited to the reader's familiarity with subject matter, but also comes from a command of the language that is supported by a broad vocabulary, an understanding of complex structures and knowledge of the target culture. Readers at the Superior level can draw inferences from textual and extralinguistic clues.
>
> Superior-level readers understand texts that use precise, often specialized vocabulary and complex grammatical structures. These texts feature argumentation, supported opinion, and hypothesis, and use abstract

> linguistic formulations as encountered in academic and professional reading.
>
> Such texts are typically reasoned and/or analytic and may frequently contain cultural references. Superior-level readers are able to understand lengthy texts of a professional, academic, or literary nature. In addition, readers at the Superior level are generally aware of the aesthetic properties of language and of its literary styles, but may not fully understand texts in which cultural references and assumptions are deeply embedded. (ACTFL, 2012)

Radical knowledge: Some Chinese compound characters are composed of a semantic component and a phonetic component. Radical knowledge refers in general to Chinese learners' knowledge of the sound, shape, and meaning of a phonetic or semantic radical and the ability to use such knowledge in identifying character sound or meaning (Shen, 2000; Shen & Ke, 2007).

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

Challenges in learning a non-Roman-script writing system such as Chinese are well recognized by both CFL practitioners and researchers. This is also why a large number of CFL studies have focused on understanding the Chinese character and word recognition process, or how learners establish symbol-sound and symbol-meaning correspondence and how orthographical awareness impacts the process (e.g., Bassetti, 2005; Everson, 1998; Shen & Ke, 2007). Comparatively, there are far fewer studies devoted to understanding how learners' word knowledge develops in terms of both breadth and depth. Systematic efforts at developing valid and reliable measures of vocabulary knowledge in CFL are also rare. This is compounded by a lack of understanding of CFL language proficiency in general at the advanced and superior levels, with vocabulary development at these levels being an important missing piece. Filling these gaps is therefore necessary for deeper insights into what it means to be an advanced- or superior-level CFL learner.

This chapter will begin with a general overview of theoretical frameworks of vocabulary knowledge and related research and will be followed by a discussion of vocabulary knowledge development across proficiency levels, especially at higher levels of proficiency. Different measures of vocabulary knowledge will then be reviewed, with particular emphasis given to the WAT as a measure of vocabulary depth. Set in this general background of L2 vocabulary research and assessment, Sections 4 and 5 will be devoted to the discussion of the Chinese writing system, characteristics of Chinese vocabulary, and major vocabulary studies in the CFL context. Finally, Section 6 will be a summary of its previous sections and set a framework for understanding and investigating the research questions.

2.2 Vocabulary Knowledge

2.2.1 Dimensions of Vocabulary Knowledge

Researchers have long recognized that vocabulary knowledge is not a singular construct and that learners may know a word in different ways and to different degrees. Various attempts have therefore been made to exhaust the possible dimensions of word knowledge. An earlier work by Cronbach (1942) summarized various types of behavior indicating learners' knowledge of a word: (1) *generalization*, i.e. whether a learner is able to define a word; (2) *application*, or recognition of whether an "illustration" is "properly named" by a word; (3) *breadth* of meaning, or the recall of different word meanings in different contexts; (4) *precision* of meaning, or correct application of a word to "all possible situations"; (5) *availability*, or the ability to actually use a word in "thinking and discourse" (p. 207). A limitation of this framework is that it focused particularly on meaning and use, but neglected form—both how a word is pronounced and the orthographic structure of its written form. Among his assumptions about lexical competence, Richards (1976) included such dimensions as "probability of occurrence," "knowledge of collocation," "recognition of the constraints of function and situation on word choice," "syntactic behavior," underlying forms and derivations, "knowledge of the network of associations," semantic features, and multiple word meanings. Some obvious advantages of this framework over Cronbach's (1942) are that it emphasized such vocabulary features as frequency and register and that it also stressed the collocational relationship and meaning associations between one word and another.

Integrating the distinction of the receptive and productive use of words, Nation's (1990, 2001) framework further expanded on what was involved in knowing a word. Raising a total of 18 questions regarding "form," "meaning," and "use" from both receptive and productive perspectives, his framework (2001) was very comprehensive and widely employed in later vocabulary studies (as shown in Table 1).

Table 1

Nation's (2001) Framework of "What is Involved in Knowing a Word"

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express this meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use (Register, frequency…)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often would we expect to meet this word?

Note. R = Receptive knowledge, P = Productive knowledge

(Nation, 2001, p. 27)

Despite the variations among the above frameworks, a common underlying assumption is that knowing a word is not an all-or-nothing phenomenon. Consequently, the number of words one knows and the extent to which an individual word is known

should be two separate though related concepts. This is why a binary structure of vocabulary depth versus breadth was often used when addressing vocabulary knowledge (Chapelle, 1998; Haastrup & Henriksen, 2000; Meara, 1996; Qian, 1999, 2000, 2002; Read, 2000; Wesche & Paribakht, 1996). Qian (1999), for example, defined vocabulary breadth as vocabulary size, or "the number of words for which a learner has at least some minimum knowledge of meaning" (p. 283). Depth of vocabulary knowledge, on the other hand, was understood as "the quality of the learner's vocabulary knowledge" (Read, 1993, p. 357) or their "level of knowledge of various aspects of a given word" (Qian, 1999, p. 283). Alternatively, Haatrup and Henriksen (2000) defined depth of vocabulary knowledge as learners' knowledge of "a word's different sense relations to other words in the lexicon," which included paradigmatic relations such as antonymy, synonymy, hyponymy, and gradation and syntagmatic relations such as collocational restrictions (p. 222).

Arguably, the construct of vocabulary depth should involve multiple dimensions. For instance, depth of vocabulary knowledge in the reading process was considered to involve pronunciation and spelling, morphological and syntactic properties, meaning, register, and frequency (Qian, 1999). Similarly, Ordóñez et al. (2002) proposed that vocabulary depth should include knowledge of phonology, syntactic properties, morphological structure, semantic representation, and pragmatic rules for word use. Henriksen (1999) understood depth of vocabulary knowledge as mainly "network building" of both "extensional and intentional relations" among words (p. 312). In his review of works on vocabulary depth, Milton (2009) summarized that depth was generally used to refer to "a wide variety of word characteristics," including "the shades of meaning a word may carry, its connotations and collocations, the phrases and patterns of use it is likely to be found in, and the associations the word creates in the mind of the user" (p. 149). As can be seen, vocabulary depth may be interpreted from two major perspectives: First, knowledge of the multiple features of a word as a standalone unity;

Second, knowledge of a word's relationship to other words in learners' mental lexicon. When a word is understood as a standalone unity, we are concerned only about how it functions independently of other words—for instance, how a word is pronounced and what word parts it consists of seem to be the inherent properties of a word. However, considering words as isolated written symbols is fundamentally incomplete, since a word ultimately has to be used within the contexts of other words. Features of individual words also bring them into different types of relationships—words that share common phonological or morphological components, for example, are often connected in some way in learners' vocabulary network; and words that have similar meanings can often replace each other in language use, or have to be differentiated with care for more appropriate or precise use in context. Collocation is another type of word relationship, in which a word tends to appear together with some words but not with others, indicating that a word is defined through its relationships with other words. Vocabulary depth, therefore, has to take into account learners' network knowledge of words.

In addition to the binary division of vocabulary knowledge, there are also researchers who regard depth of knowledge as one of multiple dimensions along which learners' vocabulary competence develops. Henriksen (1999), for instance, criticized the "random grouping of knowledge components and learning processes" in earlier frameworks and emphasized the need to develop a "unified theoretical construct of lexical competence" (p. 304). He proposed three developmental continua along (a) partial to precise comprehension, (b) depth of knowledge, and (c) receptive to productive use ability. Read's (2000) model was very similar, with the three dimensions of partial and complete knowledge, receptive-productive knowledge, and depth of knowledge. Drawing upon the strengths of previous models, Qian (2002) adopted a four-dimensional framework, including vocabulary size, depth of vocabulary knowledge, lexical organization, and automaticity of receptive-productive knowledge. Daller, Milton, and Treffers-Daller (2007) understood learners' vocabulary knowledge as a three-dimensional

lexical space, involving breadth, depth, and fluency, with "fluency" defined as readiness or automaticity of word use. Kieffer and Lesaux (2012) categorized vocabulary knowledge into word-specific and word-general knowledge, with the former indicating "breadth and depth of linguistic knowledge of individual word meanings," and the latter emphasizing "metalinguistic knowledge about words and their meanings" (p. 351). Morphological awareness, for instance, is part of the word-general knowledge. While these different categorizations may appear confusing, they have at least two points in common: First, they unanimously recognized depth of knowledge as a significant dimension of vocabulary competence; Second, learners' vocabulary competence tends to develop along different continua, and depth of knowledge is just one of them. In other words, learners at different proficiency levels are likely to vary not only in terms of vocabulary size, but also in how well they know given words, the ease and automaticity they access and use words, and the precision of word perception and production. It is yet to be known, however, how these developmental continua (should they be present) interact with each other—whether they are parallel and whether there a driving force among them that move other dimensions along with it.

Despite its accepted position in vocabulary competence, there are also doubts whether the concept of vocabulary depth is a useful one, which are derived from the following perspectives. First, the current lack of consensus over the concept makes it hard to apply it in instructional and assessment contexts. According to Read (2004), depth of vocabulary knowledge had been applied to second language acquisition (SLA) in various ways, so much so that there was confusion over how the concept needed to be understood. He believed that there were "three distinct lines of development" in its application – (1) *precision of meaning*, or learners' "elaborated and specific" knowledge of word meaning; (2) *comprehensive word knowledge*, including knowledge of a word's semantic, orthographic, morphological, syntactic, collocational and pragmatic characteristics; (3) *network knowledge*, or the lexical network learners establish that

either connects different words or distinguishes them from one another in the mental lexicon. Second, the strong correlation between depth and breadth of vocabulary knowledge and their equivalent contribution to reading comprehension leads to the question of whether it is meaningful to separate the two concepts. Theoretically, it can be assumed that learners with a larger size of vocabulary are also likely to have a stronger vocabulary network due to the number of connections they have available. As Vermeer (2001) put it, “the more words one knows, the finer networks and the deeper the word knowledge” (p. 222). Empirical results such as those in Marzban and Hadipour (2012) also indicated that depth and breadth of vocabulary knowledge “cannot be considered separable” based on their finding of a very high Pearson product-moment correlation between the two ($r = .921$) (p. 5299). Third, the lack of correlation among its components as found by some studies calls into question whether vocabulary depth is a unitary construct. For instance, McGavigan found that correlations between idioms and collocation measures were “very close to zero” (as cited in Milton, 2009, p. 158). As Milton (2009) put it, “(i)t is hard to see what principle unifies collocational, associational knowledge, constraints on use, polysemy and the other qualities that are placed within this dimension” (p. 150). Learners may very likely be able to acquire knowledge in one area “without becoming commensurably able in other areas” of vocabulary depth (Milton, 2009, p. 169). A review of previous literature by Kieffer and Lesaux (2012) also indicated that depth and breadth might be indeed “indistinguishable” and that assessments of depth and breadth might be addressing “the same unitary construct of vocabulary” (p. 349). Due to these concerns, it is a question of whether it is more useful to investigate vocabulary depth as a holistic construct or to focus instead on each distinctive dimension or aspect as individual constructs. Read (2004) suggested that we needed to dispense with the term depth as well as any other substitutes which were intended as a cover term for all dimensions of vocabulary knowledge (p. 224). Any

dimension of word knowledge we intend to measure should be “carefully defined and not simply labeled with a catch-all term like depth” (Read, 2004, p. 224).

Even though no consensus has been reached on how to define vocabulary depth and its relationship with vocabulary breadth, the investigation of the quality of vocabulary knowledge at least compensates for the situation where only superficial knowledge of words is addressed. It is important, however, for any research on vocabulary depth, including the current study, to carefully define and delimit what we mean by vocabulary depth rather than taking its implications for granted. For the sake of efficiency and unity, it is also more meaningful to focus on a limited number of related dimensions of vocabulary depth instead of trying to cover all possible dimensions.

2.2.2 Development of Vocabulary Knowledge

Similar to L1 vocabulary acquisition, development in L2 vocabulary involves more than just the building of a larger vocabulary size. It also involves an increasing amount of knowledge regarding both individual words and connections among words. Haastrup and Henriksen (2000) believed that L2 vocabulary development was a process of “network building” in which the relational links became gradually expanded and strengthened (p. 222). In a similar vein, Milton (2009) also pointed out that the underlying assumption of vocabulary depth as networked word knowledge was that L1 learners had greater numbers of links as well as more links that were “correct and appropriate” than L2 learners (p. 149).

Different links in the lexicon network, however, seem to develop in an unparalleled fashion. For instance, L1 research on word associations seems to indicate that a syntagmatic-paradigmatic shift might be present, i.e., young children produce more syntagmatic associations, or associations among words that are “frequently found in continuity” in a “syntactic sequence,” whereas paradigmatic associations, or words with “the same grammatical form class,” tend to develop later in life (Nelson, 1977, p. 93).

Whether the same developmental patterns hold for L2 learners is an issue still being investigated. Greidnus, Beks, and Wakely's (2001) study indicated that adult L2 learners of French at an advanced level scored higher in paradigmatic associations than they did in syntagmatic associations. They also found that more advanced L2 learners performed better in all three dimensions of paradigmatic, syntagmatic, and analytic associations than less proficient learners. On the other hand, Wolter's (2001) comparison of L1 and L2 learners of English (including ESL teachers and those with postsecondary degrees in the U. S.) indicated that while both groups of learners produced a lot of "childlike" responses to the not-well-known words, their responses to the well-known words showed "almost a mirror-image effect" (Wolter, p. 58). The L2 learners demonstrated a preference for syntagmatic responses over paradigmatic responses, whereas the L1 learners were just the opposite. According to Wolter, this syntagmatic dominance among L2 learners did not necessarily represent a lower degree of vocabulary development. Rather, for both L1 and L2 learners, vocabulary development seemed to follow a shift from "semantically meaningless" associations to "semantically meaningful" associations (Wolter, 2001, p. 63).

Other than network building, there is also evidence that even for advanced L2 learners who have a large vocabulary size, their vocabulary knowledge may still be "hazy" when it comes to "polysemy, contextual and collocational restrictions, phrasal verb combinations, grammatical environment" (Lennon, 1996, p. 35). For one thing, they may still have difficulties correctly using simple high-frequency verbs due to their overreliance on "core meaning of polysemous verbs" and "treacherous translation equivalents in L1" (Lennon, 1996, p. 35).

Understandably, the development of vocabulary depth tends to lag behind the development of breadth. Greidanus, Beks, and Wakely (2005) found, for instance, that the advanced L2 learners in their study had greater depth of knowledge regarding "frequent, earlier-acquired words" than "less frequent, more recently acquired words,"

hence the conclusion that “vocabulary grows in depth more slowly than it grows in breadth” (p. 512).

Lessard-Clouston’s study (2006) was one of the few that examined vocabulary development between L1 and L2 learners. It was a descriptive case study of how breadth and depth of specialized theological vocabulary developed among five non-native English speakers and seven native English speakers through a semester’s study in a theology course. A breadth test modeled after the Yes/No test and a depth test adapted from Vocabulary Knowledge Scale were employed, with the results indicating that the combined NES and NNES group made significant gains as measured by both tests. However, the NES group outperformed NNES in the pre-test in both breadth and depth; in the post-test, while the gap in breadth closed between the two groups, the NES group showed greater increase in depth than the NNES group. The author attributed this result to the better developed “lexical organization” among the NESs. In other words, it was likely that there were more links or associations in their overall lexical organization, so that it was easier for them to develop the vocabulary depth (Lessard-Clouston, p. 189). In addition, the study also revealed “the very individual nature” of vocabulary knowledge and learning, in that “no two participants respond(ed) in exactly the same fashion” (Lessard-Clouston, p. 187).

In summary, L2 vocabulary development is a complicated process which may involve the development of both vocabulary breadth and depth and the different dimensions of vocabulary depth. Whether and how the development differs from L1 vocabulary acquisition and across different L2 proficiency levels is an important issue to explore.

2.2.3 Vocabulary Development at Higher Proficiency Levels

Research over vocabulary development at higher (advanced and above) L2 proficiency levels is still very limited. Even though the ACTFL Proficiency Guidelines provides detailed descriptions of what learners can do with a language at different proficiency levels, these descriptions have been criticized as being "*a priori* definitions" (Alderson, 2000, p.279) and not empirically based (Alderson, 2000; Liskin-Gasparro, 2003). Mentioning of vocabulary competence in these guidelines is also scarce; in fact, the only place where vocabulary competence has been mentioned at the advanced level was in the description of texts appropriate for advanced-low learners, stating that "(t)hese texts predominantly contain high-frequency vocabulary…" (ACTFL, 2012, p. 22). On the other hand, texts attainable by superior-level readers use "precise, often specialized vocabulary" and learners' comprehension is supported by "a broad vocabulary" (ACTFL, 2012, p. 21). Other than the above vague mentioning of vocabulary size, word frequency and preciseness of vocabulary use, these descriptions are not helpful for our understanding of what kind of vocabulary competence is expected at each proficiency level, let alone learners' mastery of different dimensions of vocabulary knowledge.

For the current study, it is of particular interest to us how well vocabulary knowledge is developed at higher proficiency levels. Leaver and Shekhtman (2002) argued that success at higher levels of language proficiency may require different attributes than at lower levels and hence does not mean just "more of the same" for teaching and learning (Leaver & Shekhtman, 2002). Among other things, "precision of lexicon" is one major characteristic that distinguishes learners of higher L2 proficiency (Leaver & Shekhtman, p. 26). In particular, the ability to determine word meaning was emphasized, but "not by using context as much as by using an already-developed understanding of the linguistic framework of the language—a sensitivity to the morphemes and syntagms of the foreign language, as well as an intuition of the multiple

meanings of words and their correct (grammatical and cultural) usage" (Leaver & Shekhtman, p. 26). In other words, decontextualized understanding of polysemous word meanings and syntagmatic relations between words is expected and should be given higher priority. When discussing superior-level CFL learning, Kubler also emphasized that while vocabulary was only "secondary" at beginning and intermediate levels of Chinese study, deepened and precise knowledge of "the individual words themselves" became more important at the superior level. That is, learners at this level need to know the "exact meaning and usage" of words— "both common and rare, as well as synonyms and antonyms" (Kubler, 2002, p. 111).

We have very limited understanding, however, of how well CFL vocabulary knowledge is developed at advanced and higher levels of proficiency. As will be discussed in a later section, vocabulary studies in the context of CFL still focus largely on the word recognition process or the access to simple form-meaning correspondence. Studies on how learners gradually build more precise and sophisticated lexical knowledge are rather limited. Meanwhile, well-developed assessment procedures on vocabulary size and depth in CFL are virtually non-existent. This leads to the necessity of building a vocabulary measure to better understand CFL vocabulary development at higher levels of Chinese proficiency.

Placed in a larger framework of L2 language proficiency, such efforts will also help fulfill the gap of scholarly knowledge on "the nature of advancedness" (Maxim, 2009, p. 126). In other words, understanding the quality of advanced- and superior-level learners' vocabulary knowledge may bring a deeper insight into what it means to be an advanced or superior user of a language.

2.2.4 Conclusion

The above review of literature indicates that vocabulary depth has been a useful construct in conceptualizing learners' quality of word knowledge or sophistication of

their vocabulary network. Despite controversies over its relationship with vocabulary breadth and other concepts in vocabulary competence, the employment of vocabulary depth in investigating and assessing vocabulary knowledge has a clear advantage over traditional approaches of looking at word knowledge as an "all-or-nothing" phenomenon. The research base, however, is still weak over the developmental trajectories of vocabulary depth, especially its development at advanced and higher proficiency levels.

2.3 Assessment of Vocabulary Knowledge

2.3.1 Tests of Vocabulary Breadth

Assessment of vocabulary breadth has a longer history of development and is in more prevalent use in second language education and research. Most vocabulary breadth measures examine learners' single word-meaning (usually most prominent meaning) correspondence through tasks such as translation, paraphrase, multiple-choice, or matching. Target words may appear either in an isolated way or within sentences or texts.

The Yes/No test (Meara & Buxton, 1987; Meara, 1989, 1996; Meara & Milton, 2003) adopted a simple checklist format in which learners mark whether a word is known or not. Target words were sampled from different proficiency levels. To correct for guessing or false reporting, a combination of real words and pseudowords was used. Learners' vocabulary knowledge was estimated to be the number of real words known as adjusted by the rate of "false alarms" or "yes" responses to pseudowords. Another widely adopted vocabulary size test is Nation's (1990; 2001) Vocabulary Levels Test, in which learners matched a certain number of words with a smaller number of translations or paraphrases of word meaning. The sample item below is taken from Form B of the Vocabulary Levels Test at the 2,000 word level (Nation, 2001, p. 416). Part of the strengths of this test lies in the fact that it is "quick to take, easy to mark, and easy to interpret" (Nation, 2001, p. 21). Therefore, it is suitable to be used as a diagnostic test of vocabulary size at different proficiency levels.

1	copy	
2	event	_____end or highest point
3	motor	_____ this moves a car
4	pity	_____ thing made to be like another
5	profit	
6	tip	

Major concerns in building a breadth test include the representativeness of the target words for given frequency bands, time efficiency, and ease of administration and rating due to the large number of words that have to be assessed. The underlying assumption of a breadth test is that learners with greater vocabulary size should be able to know by sight or identify the meanings of a larger number of words at each word frequency level. Such tests are useful in getting a rough estimate of learners' general, superficial knowledge of words, either common or rare. However, how well learners know any given word is often not captured. Meaning rather than form and use remains the primary focus of most vocabulary breadth measures. Within the dimension of meaning, often only a singular form-meaning correspondence is addressed. While the meaning being tested may be the "primary" or most commonly-used meaning of a word, it is not necessarily the meaning with which a given learner is most familiar. Therefore, not being able to provide a correct answer does not necessarily indicate that a learner knows nothing about a word.

For the above reasons, it is fair to say that vocabulary breadth tests provide only a partial and often biased profile of learners' word knowledge. It may not even provide a faithful picture of how many words a learner knows depending on the match between word meanings being tested and those learners already know. This is where a test that addresses in-depth understanding of word meaning and use as well as meaning connections between words could be very useful.

2.3.2 Tests of Vocabulary Depth

Various attempts have been made at assessing depth of vocabulary knowledge. A useful qualitative approach is to interview learners and ask them to provide as much information as they know about a word. Such a procedure offers the flexibility to either test several dimensions of vocabulary knowledge simultaneously or explore one dimension in great depth. The disadvantages are that it tends to be very time-consuming and may involve difficulties in rating. It is therefore more appropriate for research rather than testing situations. Researchers also pointed out that learners may sometimes "pursue a meaning that is on the wrong track" and may not have an opportunity to discuss everything they know about a word unless clear guidance is provided (Nation, 2001, p. 357).

Considering time efficiency, the more widely-accepted measures take a selected-response approach or a combined format with both selected- and constructed-response questions. One such measure is the VKS, or Vocabulary Knowledge Scale (Paribakht & Wesche, 1993; Wesche & Paribakht, 1996), in which learners rate their own vocabulary knowledge on a five-level scale as below:

I. I have never seen this word

II. I have seen this word before, but I don't know what it means

III. I have seen this word before, and I <u>think</u> it means _____.
(Synonym or translation)

IV. I know this word. It means _____. (Synonym or translation)

V. I can use this word in a sentence _____.

One advantage of this scale is that it put learners' vocabulary knowledge on a receptive-productive continuum, with the assumption that learners with greater depth of word knowledge are closer to the productive end—they are able to not only accurately identify word meaning, but also use the word appropriately in a sentence context.

However, due to its reliance on learners' self-report, the results obtained via this measure may over- or under-estimate learners' true knowledge level. The first two levels, for instance, do not effectively differentiate learners since they do not require any evidence of learners' knowledge level, so learners may well give an inaccurate indication of knowledge for these two levels. The distinction between Levels III and IV also relies partly on learners' level of confidence with their word knowledge, even if they can provide correct synonym or translation in either case. For a receptive-productive continuum, the number of levels may seem crude, i.e., the scale does not take into consideration stages where learners may be able to conceptually understand word meaning, but not able to translate or paraphrase it in the right words. In addition, being able to recall word meaning tends to be a more difficult skill than being able to identify word meaning out of several different choices. The lack of context for the target word is likely to make the task of recall even harder for learners. Researchers (Milton, 2009; Wolter, 2005) also criticized that in practice the middle of the scale might end up being rarely used with scores tending to cluster at both ends. Read (2000, p. 137) also pointed out that Level 5 might allow learners to use the target word in a "semantically neutral context" without having to truly know its meaning. Likewise, Lessard-Clouston (2006) raised the point that providing a sentence example in Level 5 might serve better to "reveal syntactic or grammatical information" than to "verify semantic understanding" (pp. 190-191).

Lastly, this measure does not require or encourage learners to provide more than one word meaning, nor does it intend to tap into other dimensions of vocabulary knowledge. The scope of knowledge this measure is able to cover is therefore limited. In order for a second meaning of the target word to be tested, the scale would have to be extended and even more time would be needed for completion, especially considering the time needed for completing the constructed tasks in Levels 4 and 5 (Qian & Schedl, 2004).

Due to the limitations of the above measures, the WAT, as will be discussed in the next section, is considered a better option in assessing depth of vocabulary knowledge.

2.3.3 The Word Associates Test

The Word Associates Test (Read, 1987, 1993, 1995, 2000) has a solid theoretical base in word association studies in both L1 and L2. It is also closely related to word association tasks.

Word Association Task

When vocabulary depth is conceptualized as network knowledge, a viable approach to examining vocabulary depth would be to look at the number and quality of associations learners can establish among words. The word association task, for instance, presents learners with a set of stimulus words and asks them to spontaneously produce the first associated words that come to their mind. However, the validity of this measure is restricted by learners' ability to recall associates on the spot. Besides, how to sort through all the possible associates learners may have and understand what they imply about learners' vocabulary development can be very challenging.

Word Association Studies

It is believed that adult L1 users typically produce two main types of associations, i.e., syntagmatic and paradigmatic associations (Meara, 2009). Syntagmatic associations refer to associations that "complete a phrase (syntagm)" (e.g., "brush" and "teeth"), while paradigmatic associations are those in which "the stimulus word and the response that it evokes both belong to the same part of speech" and "usually share a large part of their meaning" (e.g., "tree" and "bush") (Meara, 2009, p. 6). Other less common associations, which are more prevalent among children, are "clang associates," or those that share

phonological similarities with stimulus words (Meara, 2009). Examples of the last type would include “night” and “knight,” and “blue” and “glue.”

The proportions of the different types of associations, especially syntagmatic and paradigmatic associates, are believed to represent the structure of a language user’s mental lexicon. With L1 users, there is general agreement that adults tend to produce more paradigmatic responses than syntagmatic ones, while syntagmatic and clang responses seem to be dominant among children, hence the so-called “syntagmatic-paradigmatic shift.” Whether the same shift is present among L2 learners and to what extent L2 learners’ mental lexicon resemble those of native speakers are topics of ongoing investigation. Read (1993) pointed out that word association patterns among native speakers were “remarkably stable,” whereas those of L2 learners tended to be “much more diverse and unstable,” but show trends of gradually resembling “native speaker norms” as their proficiency levels increased (Read, p. 358). Wolter’s (2001) study indicated that while L1 and L2 learners’ response patterns closely resembled each other for words not well known—both groups produced a lot of “childlike” responses, their response patterns for well-known words were almost completely opposite, with native speakers producing more paradigmatic responses and non-natives producing more syntagmatic responses. Meara (2009) argued that L2 learners’ association responses tended to be “more varied and less homogeneous” than those of “a comparable group of native speakers” (p. 22). One explanation for this difference is that L2 learners were more likely to produce clang responses like children (Meara, 2009). Another possibility is that L2 learners might not have correctly identified the stimulus words, so that the associations they came up with appeared odd or out of place (Meara, 2009).

The Word Associates Test

While word association tasks as described above may provide useful implications on how L2 learners’ mental lexicon developed, they are more appropriate for research

than for assessment purposes. A much more practical measure based on essentially the same principle is the Word Associates Test (Read, 1987, 1993, 1995, 2000), which requires learners to "*select* responses rather than supplying their own" (Read, 2000, p. 181). The format of the test underwent several revisions, and continues to be validated and modified. The earliest exploration of this measure started with an interview procedure and an analysis of interview results (Read, 1987). The initial version (Read, 1993) presented learners with a stimulus word and asked them to identify out of a group of other words those related to the stimulus word in some way. Three major types of word associations were addressed: paradigmatic, or relations of synonymy or near synonymy; syntagmatic, or collocational relationships; and analytic, with the associates representing a component meaning of the stimulus word. The associates were carefully selected to not only represent the three major association types, but also "cover the main meaning of polysemous words" (Read, 1993, p. 360).

One potential drawback of this early version as recognized by Read (1993) was that learners, especially those of higher proficiency levels, were sometimes able to correctly guess the associates by examining "the semantic relationships among them" without having to refer to the stimulus word (Read, 1993, p. 368). Read (1993) posited, therefore, that this measure would be better considered as "assessing the test-takers' ability to identify the lexical network formed by the stimulus plus the four associates" rather than simply knowing the target word (Read, 1993, p. 360). Another concern was the heterogeneity of the stimulus words, which was likely to reconcile the consistency of the measure from one item to another (Read, 2000). For one thing, nouns, verbs, and adjectives as stimulus words tended to have different kinds of associations, and a word might function as more than one part of speech, both of which added complications to the association task. Moreover, for those of the same part of speech, some stimulus words may have a broader range of meaning and use than others. Some of these concerns were taken into consideration in the revised version of WAT, so that only adjectives with more

than one meaning and a broader range of uses were used as stimulus words. In addition, the eight associates in the original format were replaced by two groups of four associates, with the former group representing paradigmatic relationships with the stimulus word, and the latter representing syntagmatic relationships. Learners were asked to select four associates out of the two boxes, with the combinations of correct associates varying from 2+2, 1+3, to 3+1 for the two boxes. This revision was made to reduce the impact of the guessing factor. An example of this test format is as follows (Read, 2000, p. 184):

sudden

beautiful	quick	surprising	thirsty	change	doctor	noise	school

Various validation studies were conducted on the WAT or its revised versions. In general, reliability for the WAT was found to be fairly high—.92 (KR-20) and .93 (Rasch) according to Read (Read, 1993, 1995), .92 (split-half) for a revised version by Qian (Qian, 1998, 1999), and .89 (split-half) in Nassaji's study (2004). Qian & Schedl (2004) compared the Depth of Vocabulary Knowledge measure (DVK, or the revised WAT) and existing TOEFL vocabulary measures and concluded that both had "a similar difficulty level" and "provide(d) a similar amount of prediction of test-takers' reading performance" (p. 28). Qian and Schedl (2004) summarized several strengths of DVK: (1) the two aspects of meaning and collocation reflect "the primary aspects of lexical semantic association"; (2) the dimensions of synonymy, polysemy, and collocation lend themselves conveniently to descriptions of proficiency profiles in vocabulary knowledge; (3) the sole focus on receptive vocabulary knowledge makes it an appropriate test to address the kind of knowledge needed for reading comprehension (p. 33). However, there are also practical concerns that reduce the potential usefulness of the WAT in especially large-scale standardized assessment. The difficulty to "develop uncontroversial keys"—especially for collocations—is one of them (Qian & Schedl, 2004, p. 46). In addition, the fact that only adjectives were used as stimulus words also greatly limited the extent of

vocabulary coverage, even though Qian (2002) argued that nouns were also "indirectly tested" in the collocation section. In terms of the validity of its use as a vocabulary depth measure, there was also the concern that it "might be regarded as another measure of vocabulary size rather than a measure of vocabulary depth" due to its high correlation (r = .88 according to Qian, 2002) with Vocabulary Levels Test (Mochizuki, 2012, p. 48).

Another validation attempt was directed at understanding the construct of the WAT. Batty (2012), for instance, proposed three possible dimensional structures of the WAT—unidimensional, two-factor and bi-factor models. Results of confirmatory structural equation modeling indicated that the bi-factor model was most applicable to WAT. Specifically, the test items loaded on "a single vocabulary g-factor" (general factor) and additionally on two separate sub-skill factors of synonym and collocational knowledge (Batty, p. 74). According to Batty, the vocabulary knowledge g-factor underlying both test sections "muddles" the interpretability of synonym and collocation items; meanwhile, the two additional sub-skill factors also call into question the unidimensionality of the WAT as put forth by previous researchers (p. 76).

Based on the above evidence, the Word Associates Test is overall a test with high score reliability. It has a solid theoretical basis in word association studies and a clear focus on the key dimensions of vocabulary depth. However, to what extent it provides a valid representation of synonym and collocational knowledge instead of vocabulary size or general vocabulary knowledge is still being questioned. Its usefulness and practicality in large-scale standardized assessment is another issue of concern.

2.3.4 Critique of Vocabulary Depth Measures

There are several major controversies regarding the development and use of vocabulary depth measures, the first being questions over their feasibility. Milton (2009) argued that the quality of vocabulary depth was "not an idea of vocabulary measurement where we have an accepted methodology, still less a generally accepted test to make

these measurements" (p. 150). Considering the variations in how vocabulary depth is conceptualized, this may be true. It does not necessarily mean, however, that we need to wait till a consensus is reached over the construct before making any attempts to develop such a test. Meara (2009) raised an even more pessimistic view about vocabulary depth measures, arguing that "this enterprise is fundamentally doomed," since by trying to test minute details of individual word knowledge, it is very likely that we will lose "the larger characteristics of whole vocabularies"—"a classic example of not being able to see the wood for looking at the trees" (p. 74). Testing a large number of words in depth is logistically infeasible, whereas testing a few words in infinite details may nevertheless greatly reconcile the generalizability of the test (Meara, 2009). Again, rather than dismissing the very possibility of developing a useful vocabulary depth measure, his comment could be taken to imply the importance of assessing a unified construct of vocabulary depth rather than attempting to assess every minute detail of it.

Another controversy is over the relationship between tests of vocabulary breadth and those of vocabulary depth. Some argue that vocabulary breadth tests are superior to the latter in terms of representativeness. This is because when multiple dimensions of vocabulary knowledge are addressed in the same test, the number of vocabulary items that can be covered is by necessity limited, whereas vocabulary size tests are capable of assessing a large sample of words at different frequency levels, thus providing a more representative profile of learners' vocabulary knowledge status (Read, 2000; Laufer et al., 2004). There are also those who suggested that tests of vocabulary breadth and depth were roughly equivalent in ranking learners' vocabulary abilities, thus a separate depth test might not be necessary. Based on her findings of high correlations between depth and breadth measures for both monolingual and bilingual children, Vermeer (2001) argued that a breadth test could well serve the same purpose as a depth test in assessing language proficiency in vocabulary as long as it "include(d) words from various domains and frequency levels" (p. 231). Rather than supporting one measure in preference to another,

a third line of thought is to decide whether and how tests of vocabulary breadth and depth can be integrated in practice, as according to Ishii and Schmitt (2009), “using depth measures by themselves is unlikely to be completely satisfactory” (p. 7). When learners perform poorly in a depth measure, it may be due as much to a small vocabulary size as to their insufficient knowledge about given words (Ishii and Schmitt, 2009). They therefore suggested evaluating breadth and depth simultaneously in a single test battery. However, the test battery they developed for this purpose was able to cover only three aspects of vocabulary knowledge—multiple meaning senses, derivative word forms, and near-synonyms, though they argued that the three aspects were related to the most frequent problems among L2 learners in their study. Besides, the four measures they employed were relatively independent from one another. Whether it is possible to integrate breadth and multiple dimensions of depth within one test in a unified and systematic way remains a question, let alone how useful it is in practice considering time and rating issues.

In addition, there is also the controversy of whether to test vocabulary in or without context. Read and Chapelle (2001) argued that discrete and decontextualized vocabulary measures were “out of step with current thinking in educational measurement and applied linguistics” (p. 21). Such tests are also likely to have a “negative educational impact” since they fail to meet the needs of developing and assessing communicative language use. Contextualized vocabulary measures should therefore be encouraged to bring more positive washback effects on vocabulary and language instruction in general. However, others may argue that such contextualized measures are not “pure” measures of vocabulary knowledge and that variables such as learners’ language proficiency may contaminate the assessment results. For instance, when words are tested in context, learners without preexisting knowledge of a word may be able to guess word meaning correctly using contextual clues. As a result, “it is not always clear whether it is word knowledge, or inferencing skills, which is being tested” (Laufer et al., 2004, p. 204).

Nation (2001) also pointed out that vocabulary items "with opportunities for inferencing" might measure "other things besides previous vocabulary knowledge" and could therefore be misleading (p. 353). He also discussed other disadvantages of using sentence contexts for vocabulary assessment, which included "the extra time required to make an item, and the fewer items that can be tested within the same time" (p. 354). When sentence contexts are not well designed, learners are likely to select a correct answer not by knowing the meaning of the target word, but by "using substitution within the context sentence to eliminate the distractors" (Nation, 2001, p. 354).

Yet another criticism on existing assessments of vocabulary knowledge in general is the strong focus on declarative knowledge and the neglect of procedural knowledge (Read, 2004). The latter can be more important in L2 learning since "ultimately the question is not what learners know about a word but what they can do with it" (Read, 2004). Therefore, in addition to tests of declarative knowledge of vocabulary in isolation, particular attention needs to be made to testing vocabulary in use (Read, 2004). Regarding this criticism, it may be argued that vocabulary measures may have different focus depending on the mode of language use. When receptive language use is emphasized (in the contexts of reading and writing), assessment of declarative knowledge may be more appropriate, whereas procedural knowledge is better suited when learners need to communicate in the language either orally or in writing. While an "ideal" test would probably need to involve the integration of the two, practical needs and constraints often allow us to focus on just one of them.

2.3.5 Conclusion

In order to obtain a more comprehensive profile of learners' vocabulary development, tests of both breadth and depth are indispensable. In assessing vocabulary depth, the major concerns are how to balance the representativeness of the vocabulary sample with time requirement, and how to integrate various dimensions of vocabulary

knowledge in one measure in a unified way. Compared with other existing measures, the WAT is better able to fulfill the criteria for representativeness, time-efficiency, and construct unity, hence a better candidate for assessing vocabulary depth. This study will therefore employ an adapted WAT as a tentative measure to assess depth of vocabulary knowledge in CFL, while validating in the meantime its usefulness for related future assessment.

2.4 The Chinese Writing System

As mentioned in Chapter One, the Chinese language was classified as a Category 4 language by the Foreign Service Institute and the Defense Language Institute, on the basis that it requires nearly three times as many "contact hours" as a Category 1 language in order to develop a comparable level of proficiency (Walton, 1992). One of the greatest challenges in Chinese L2 learning comes from its writing system. Compared with the English language that employs a phonetic writing system, the Chinese writing systems is logographic, "in which the basic grapheme is a character, a symbol that represents a morpheme" (Wang & Yang, 2008). Meanwhile, according to the orthographic depth hypothesis (Katz & Feldman, 1983; Frost et al., 1987; Katz & Frost, 1992), Chinese employs a deep orthography, in which there is little direct or reliable mapping between written symbols and sound (symbol-sound correspondence). While learners of English may often tell with considerable certainty what an English word sounds like through its graphemic structure, Chinese characters seldom give accurate clues about their sounds. Learners of Chinese, therefore, have to learn a separate phonetic system to represent how a new character is pronounced. The phonetic system they use (pinyin and bopomofo being the most commonly employed) has not been an integral component of the writing system, but was developed a few decades ago (1950s and 1910s for pinyin and bopomofo respectively) to facilitate the learning of the writing system. This is not to say, however, that there is no principle to follow in knowing or deducing the sound of a Chinese

character or that experienced readers of Chinese will not develop any sense of symbol-sound correspondence. In fact, approximately 90% of Chinese characters are compound characters with a phonetic component and a semantic component (Zhu, 1987). Compound characters with these components may take on either a similar sound or meaning with them. Thus, learners who recognize these components within a Chinese character have a better chance of knowing what the character sounds like. However, the sound or meaning correspondence thus established is often unreliable. For instance, while sharing the same phonetic component 同 (tóng), the characters 铜 and 酮 have the same sound of tóng, whereas the character 筒 is pronounced as tǒng and the character 洞 is pronounced as dòng. The facts that the Chinese is a tonal language and that two sounds may differ not only in syllables but also in tones further complicate the situation. Since learners are not able to get reliable sound clues from Chinese characters, they may end up having to rote memorize the pronunciation of each character on an individual basis.

2.4.1 Challenges in Chinese Word Recognition

Partly for the above reasons, word recognition is a laborious process for beginning CFL learners. For one thing, access to word meaning is believed to be mediated through sound (Perfetti et al., 1992), and since the sound information is not always readily available, learners may have a difficult time recalling word meaning. Admittedly, the semantic components do provide some hints about the meaning of a word. However, rather than provide accurate meaning information, these components often serve only to set loose meaning boundaries for words. For instance, characters like 腿 (leg), 脚 (foot), 肘 (elbow), 脑 (brain), 肝 (liver) share a common semantic component 月, which means “flesh.” However, knowing the meaning of this semantic component does not help one accurately determine what a given character means. In many cases, the meaning connection between a character and a meaning component is so loose that the latter is not at all helpful in word recognition.

In addition to the special feature of the Chinese characters, the way Chinese words are constructed constitutes another challenge to L2 learners. Packard (2000) commented that the concept of a "word" in Chinese was not a clear and intuitive notion. While it is easy for English learners to segment a text into words using spaces between them, this does apply to Chinese, since "Chinese orthography segments written texts into characters, which generally represent morphemes rather than 'words'" (Packard, 2000, p. 8). In other words, what constitutes a word is not directly perceivable in Chinese. This is true particularly because the Chinese language has a large number of compound words, consisting of two or more Chinese characters. In fact, approximately 74% of commonly used Chinese words are two-character words (北京语言学院语言教学研究所 [Language Education and Research Center of Beijing Language Institute], 1986). The fact that there is no interword spacing in Chinese further adds to the difficulty in determining word boundaries. As will be detailed in the following section, English CFL learners tended to mark shorter words than native speakers of Chinese due partly to the influence of the former's L1 writing system (Bassetti, 2005). Their success rates in determining the correctness of word boundaries were also very low (Shen, 2008).

2.4.2 Differences between Reading in Chinese and English

As summarized by Wang and Yang (2008), there are four major distinctions between learning to read in Chinese and English: First, the holistic morpho-syllabic system of Chinese orthography contrasts with the linear structure of words in English; Second, characters in Chinese map primarily onto morphemes or words rather than onto phonemes; Third, Chinese phonology is characterized by the use of tones and simpler syllable structures; Fourth, compared with English, Chinese employs much less grammatical affixation (inflectional and derivational processes), while semantic compounding plays a much more prominent role (Wang & Yang, pp. 135-136).

These differences have several implications for the acquisition and assessment of word knowledge in Chinese L2 learning. For one thing, the extent to which learners know the sound of a word may have an impact on their identification of word meaning, and part of the knowledge of sound may come from their familiarity with phonetic components of given characters. For another, the extent to which learners know the meanings of individual characters composing a word or their semantic components may impact their identification or inference of word meaning. Finally, their understanding of meaning associations between words may come partially from how their component characters connect with each other. For instance, learners may be able to infer that 酷暑 (intense summer heat) and 炎热 (scorching hot) are related in some way when they recognize the meaning association between the characters 暑 (summer) and 热 (heat), even if they are not familiar with those two words as a whole. Bearing these points in mind may help us better understand learners' lexicon network and interpret their performance in the WAT.

2.5 L1 and L2 Studies in Chinese Vocabulary Learning

2.5.1 L1 Studies of Chinese Word Recognition

There are at least two major sources of L1 research that inform us of how Chinese characters and words are recognized and how their meanings are accessed. One pertains to the role of phonological access to Chinese word recognition and the other to the importance of radical knowledge.

Perfetti & Tan's (1998) study addressed a common misconception of Chinese character reading which states that since there is no direct mapping between speech and writing, Chinese readers read words by "going directly from writing units (characters) to meanings" (p. 101). This assumption is contradictory to the universal phonological principle, which states that phonology is "an essential constituent" of word identification (Perfetti & Tan, p. 101) regardless of orthography. Their study found that it took significantly shorter time overall for Chinese speakers to identify a target character when

a prime and a target character (shown successively within a short time interval) were homophonic than if they were semantically (but not phonologically) related. According to the authors, this indicated that phonological information of Chinese characters was activated prior to their semantic information, verifying that the universal phonological principle does apply to Chinese word recognition. Zhou & Marslen-Wilson's Study (2000), however, produced somewhat inconsistent results with Perfetti & Tan's findings. They found that phonological effects were either absent, smaller than, or equal with semantic effects in character or word identification in different experimental situations, whereas semantic effects remained consistently strong. These results indicated that semantic information was "activated at least as early and just as strongly as phonological information" (Zhou & Marslen-Wilson, p. 1245).

Other than phonological access, radical knowledge is another important factor that influences Chinese word recognition. L1 studies such as that of McBride-Chang et al. (2003) found that morphological awareness uniquely predicted variance in Chinese character recognition among lower-grade Chinese children after variables such as age, phonological awareness and vocabulary are controlled for. An earlier study by Shu and Anderson (1999) also found that many Chinese children had "a functional awareness" of how radicals relate to character meaning (p. 13). However, higher ability children could make better use of morphology in recognizing new Chinese characters, but this advantage was only observed with morphologically transparent characters with familiar radicals. As compared, lower-ability children either had "not discovered the basic morphological features of Chinese" or were not able to "spontaneously" apply morphological knowledge to the processing of new characters (Shu & Anderson, p. 13).

2.5.2 L2 Studies of Chinese Vocabulary Learning

Some CFL studies followed the same line of research as the L1 studies above and yielded interesting results about L2 learners' word recognition. Revisiting the issue of

phonological priming in the CFL context, Everson (1998) investigated the relationship between speech and meaning access in Chinese word recognition among beginning CFL learners. The study found that when learners were able to correctly pronounce a word, they also had a very high probability (mean probability of 90.7%) of being able to correctly identifying its meaning. Everson posited that learning Chinese characters was a "package deal" where spoken language and meaning were closely connected and that the "primacy of speech" that applied to most phonetic writing systems was also applicable to Chinese orthography (Everson, p. 200). In Everson & Ke's (1997) study of reading strategies among intermediate and advanced CFL learners, abundant instances of overt vocalization were also observed when learners made various attempts to sound out a word while trying to retrieve word meaning or construct textual information. Similarly, vocalization was a prevalent strategy used by the intermediate-level CFL learners in Lee-Thompson's (2008) study. These findings were consistent with L1 studies which indicated that phonological access was at least an essential component in Chinese word recognition (Perfetti & Tan, 1998).

Regarding radical knowledge, Shen's (2000) study on CFL learners confirmed its facilitating effect on word recognition as found among Chinese children (McBride-Chang et al., 2003; Shu & Anderson, 1999). The study also revealed that knowledge of semantic radicals was helpful with CFL learners' acquisition of morphologically transparent characters rather than morphologically opaque characters. A recent study by Williams (2013) yielded similar results, i.e., learners' reading time and accuracy were facilitated when the semantic component was "a clear indicator of the semantic category of the whole character," and were impaired when the meanings of the radical and the character were not congruent (p. 304).

Shen and Ke (2007) further investigated the development of three levels of knowing and using Chinese radical knowledge and their relationships with character acquisition. The three levels were operationalized in the study as "radical perception," or

the ability to decompose Chinese characters into radicals, "radical knowledge," or learners' knowledge of the sound, shape, and meaning information of semantic radicals, and "radical knowledge application," or the ability to use semantic radicals in learning "morphologically transparent new compound characters" (Shen & Ke, pp. 101-103). Results of the study indicated that each of these three levels had its own developmental trend and that radical knowledge application had a moderate positive correlation with Chinese word acquisition, which was possibly due to the fact that "not all compound characters are morphologically transparent" (Shen & Ke, p. 109).

A third branch of study in CFL word recognition is in word segmentation, or the determination of word boundaries. As was mentioned earlier, this is a challenging task for L2 learners of Chinese. In fact, even native Chinese speakers may not necessarily agree on how a text is segmented into words. Bassetti's study (2005) revealed that Chinese native speakers had lower level of intragroup agreement than English-speaking CFL learners on word segmentation. According to the author, this was because Chinese native speakers employed a larger number of and a greater variety of word segmentation strategies, while CFL learners seemed to rely frequently on translation into English. Another significant difference between the two groups was that CFL learners tended to mark shorter words than native speakers did. This was attributed partly to the influence of the L1 (English) writing system. However, since there was also evidence that CFL learners were beginning to develop a sense of words similar to the native Chinese, Bassetti argued that the Chinese word awareness of English CFL learners was "not simply a consequence of cross-orthographic influence," but also showed influence of characteristics of the Chinese writing system (p. 349). In a similar vein, Bassetti (2009) used a picture-sentence verification task and asked native speakers of Chinese and English CSL users to determine the match between a picture and a pinyin or Chinese character sentence written with either interword spacing or intersyllable spacing. The participants' reading rate and correctness of response were recorded. It was found that

while interword spacing facilitated pinyin reading among CSL learners (but not in native Chinese), it had no effect on Chinese character reading in either group.

However, the absence of effect of interword spacing on Chinese character reading might be partially explained by the use of visual cues and only frequent words in the above study. In fact, Shen's (2008) study indicated that making word decisions, or determining word boundaries, could be a very challenging task for CFL learners due to the lack of interword spacing in Chinese. Beginning and advanced CFL learners were asked in this study to make word judgments within ten sentences (i.e., determining whether an underlined component is a word), supply meanings for what they regard as words and provide reasons for their decisions. Results of the study indicated that the overall accuracy rates on word decisions were only 50% and 54% respectively for beginning and advanced CFL learners. The study also suggested that word decision was a complex process that involved the activation of various linguistic knowledge such as "phonology, orthography, semantics, and grammar for either individual words or entire sentences" (Shen, 2008, p. 516). Three strategies, in particular, were found to contribute significantly to word decision success: (1) "Deriving word meaning beyond just adding the meaning of constituent characters"; (2) "Trying to identify parts of speech for the target item and then making a word judgment"; and (3) "Using contextual cues" (Shen, 2008).

Other than the recognition of "known" words, another line of research is on the inferencing of unknown word meaning within or without context. In Mori's study (2003) of L2 learning of Japanese, which also employed Chinese characters, or *kanji*, semantic information from word morphology and contextual clues were found to have cumulative effectives on word meaning inference. While context provides more syntactic information, word morphology "provides semantic information that context does not provide" (pp. 410-411). Knowledge of word morphology for compound *kanji* (or

Chinese) words in this case would involve the understanding of how a word is constructed and how word components contribute to the meaning of a word.

Even though the above studies did not directly address CFL learners' word knowledge, they nevertheless brought a better understanding of what factors may mediate learners' retrieval of word meaning. Their results also indicate that the different dimensions of word knowledge as those in Nation's (2001) framework are likely to be interrelated and interdependent in word acquisition. Knowledge of word meaning, in particular, may rely to a certain degree on knowledge of both spoken and written forms of a word. Therefore, when investigating CFL learners' knowledge of Chinese words, a worthwhile effort would be to examine how much they know about the sound and written form of a word and how such knowledge facilitates or hinders their access to knowledge of word meaning and use.

Among the few CFL studies that investigated depth of vocabulary knowledge are Shen's (2009) study on the relationship between vocabulary size and strength among advanced CFL learners and Lü's (2010) study of CFL learners' word association patterns and development of mental lexicon. Vocabulary strength in Shen's study was delimited by the number of words that "students are able to use correctly in a linguistic context" (Shen, p. 75). The Vocabulary Knowledge Scale was used to measure both vocabulary size and strength, sampling a one-percent of the 8,500 high frequency words in Chinese. The study found that the participants' active vocabulary (words they can use correctly in a sentence context) constituted about 59% of their total vocabulary size by the end their third year of study and that there was a moderate yet significant correlation between their vocabulary size and strength ($r = .579$). The author concluded that the development of accurate word use, while related to that of accurate word recognition, tends to lag behind the latter (Shen, 2009). However, limitations of the Vocabulary Knowledge Scale as discussed in an earlier section have to be taken into consideration while evaluating the validity of these findings.

Lü's (2010) study employed a word association task in which the participants were asked to listen to a stimulus word stored in a computer program and respond with the first associated word(s) that came up to their mind within ten seconds. These data were used to examine how the different dimensions of word knowledge were acquired among CSL learners and verify whether a syntagmatic-paradigmatic shift would occur as their proficiency level increased, i.e., whether there would be a decrease in syntagmatic response and a corresponding increase in paradigmatic response as learners became more proficient. Such a shift, however, was not observed in Lü's study. Both syntagmatic and paradigmatic responses became greater in number among intermediate-level learners as compared with beginning-level learners. There was instead a decrease in "unrelated" response. Meanwhile, there were also a considerable number of phonological responses observed at both beginning and intermediate level, even though its increase was not significant. The author argued that learners' mental lexicon had "evolved from sound- to meaning-oriented" and that words in learners' mental lexicon began to be "connected in a more organized and meaning-related way" as readers became more proficient (Lü, p. 68). The study also found that there were more syntagmatic and paradigmatic responses associated with more familiar stimuli words, whereas unfamiliar words elicited more unrelated responses. In terms of parts of speech, nouns were found to be associated among themselves through a larger number of paradigmatic responses than verbs and adjectives. Based on previous research indicating stronger connections among words linked through paradigmatic and syntagmatic relationships than those through other relationships, the author argued that nouns might have stronger connections among themselves than verbs or adjectives did.

Lü's findings provided useful insights into how depth of knowledge developed among CSL learners. However, as the word association task in this study required learners to produce word associates on their own within a limited time frame, its validity as a vocabulary depth measure is questionable. Depending on the participants' ability to

recall associates spontaneously, their true knowledge of word associations may or may not be truthfully represented. Researchers also found that it was not uncommon for learners to produce “bizarre associations” in such tasks (Meara, 2009, p. 63). It is not clear, for instance, how useful phonological and unrelated responses in this study are in understanding learners’ lexical network, as they do not reveal meaning or grammatical connections among words. Rating associations in a reliable way can be “particularly problematical” for this methodology, which is why identification of word associations was considered advantageous over generating them (Meara, 2009, p. 63). Besides, as the study involved only two proficiency levels—beginning and intermediate—it is yet to be known whether the developmental patterns as observed in the study would hold across the entire continuum of CFL learning.

2.6 Summary

The main purpose of this study is to investigate the developmental status of vocabulary depth among advanced and higher-level CFL postsecondary learners. For this purpose, several major sources of research have been reviewed in this chapter: (1) the construct of vocabulary depth and its role in the overall vocabulary knowledge framework; (2) comparative strengths and limitations of existing vocabulary measures, particularly measure on vocabulary depth; (3) the unique features of the Chinese writing system and Chinese vocabulary that need to be considered in developing and using a vocabulary depth measure for the current study; (4) previous CFL studies on word recognition and vocabulary knowledge development.

These studies indicate to us that investigation into vocabulary depth is much needed for a more comprehensive understanding of second language learners’ vocabulary development. CFL vocabulary research, in particular, has focused mainly on the recognition of individual Chinese characters and words, calling for a need to explore into how connections of words are established among CFL learners. Based upon previous L2

research, two of the most important connections are synonym and collocational associations, targeting both the meaning and syntactical dimensions of word association. As CFL vocabulary depth research is still at its beginning stage, limiting our scope to these two types of association is also more approachable than trying to obtain an all-round picture of all vocabulary depth dimensions. In addition, a comparison of existing vocabulary measures shows to us that the Word Associates Test, while still needing further refinement, may lend itself to effectively assessing the above two types of CFL word association if properly adapted.

CHAPTER THREE

METHODOLOGY

3.1 Overview

This study explored the developmental status of vocabulary depth among postsecondary CFL learners at higher levels of CFL proficiency. In particular, it examined their identification of two types of word association—synonym and collocational association as related to word frequency, as well as the process that was involved in the inference of word association. For this purpose, this study employed quantitative research methods as supplemented by interview data. Cross-sectional quantitative data were collected through a paper-and-pencil test of Chinese word association from several postsecondary CFL programs in the U.S. The major variables investigated in this study include knowledge of synonym association, knowledge of collocational association, and word frequency. A pilot study was conducted for mainly instrument development purposes before the main study.

3.2 Sampling and Participants

The participants for the pilot study include both native speakers of Chinese and CFL learners from a postsecondary CFL program. The main study includes CFL learners from several different CFL postsecondary programs in the U.S. Convenience sampling was used among those who met the selection criteria as detailed below:

3.2.1 Chinese Native Speaker Sample

A total of six Chinese native speakers were selected for the pilot study using convenience sampling. The native speaker sample was used exclusively for instrument development purposes. Their performance data were not compared with those of CFL learners in any way. While there were no requirements for individual native speakers' Chinese vocabulary and reading competence, all participants had completed their K-12

and undergraduate education in China prior to participation, indicating that their Chinese literacy level should all be very high.

3.2.2 CFL Samples

All CFL participants for the study (including both pilot and main study) satisfied at least two criteria: (1) They needed to be enrolled at one or more advanced-level CFL courses or demonstrate advanced-level or higher CFL proficiency in particularly reading at the time of their participation; (2) They needed to be non-heritage Chinese learners, i.e., learners who grew up in a family where no family members spoke Chinese as their native language. Whether a participant met the above criteria was determined through the following two sources: (1) Instructors' report of the participant's proficiency level, level of courses in which they are enrolled, and the heritage/non-heritage status. These criteria applied mainly to the pilot study; (2) Participants' self-report of standardized proficiency test scores, length of CFL study, and heritage/non-heritage status. These criteria applied to the main study. For the latter, since standardized proficiency test scores were not available for many of the main study participants, length of CFL study was the major criterion used for determining the participants' proficiency level. Only those who reported having studied Chinese for at least four years at the time of participation were included for the final data analyses.

Normally, a student who participated in the pilot study was not expected to be involved in the main study due to previous exposures to similar test items. Because of recruitment difficulties, there was one student who participated in both parts of the study. However, this student's test data were later excluded from the final analysis due to the heritage-learner status she reported in the main study. What follows are details for the pilot and main study CFL samples.

For the pilot study, a convenience sample of four CFL learners was selected from a CFL program in a large Midwestern university. The students were enrolled in either a

fourth or fifth year CFL course. According to their instructors, none of them was a heritage learner. However, one of the participants (as mentioned above) did indicate in a later demographic questionnaire that one or more of her family members with whom she grew up spoke Chinese as his or her native language. Due to the very small sample sizes, her data were included in the pilot study but excluded from the later main study in which she also participated.

Convenience sampling was employed also for the main study, and a total of twenty-five CFL learners were initially recruited from five postsecondary CFL programs. Eight learners were later excluded from the data analysis, among whom three learners were heritage learners, four learners had a length of study of only 2-3 years, while another learner did not indicate his length of study. As a result, the actual sample for the study involved a total of seventeen CFL learners who came from non-heritage backgrounds and had studied Chinese for at least four years in formal classroom settings. Six among the seventeen participants were involved in interview data collection.

Due to the possible differences among the curriculum and assessment systems within these five CFL programs, obtaining information on the participants' CFL proficiency level was a challenge. The participants were asked to provide their proficiency information through responses to two questions in a demographic questionnaire (introduced in a later section)—(1) Their most recent score or level passed in a standardized Chinese proficiency test, with first priority given to scores in a comprehensive proficiency test and then to reading test scores. Whenever possible, test scores which used a different proficiency scale were transformed to its equivalent level on the ACTFL scale. (2) The number of years they had spent studying Chinese in a formal classroom setting. Unfortunately, very few participants were able to provide proficiency test scores, and the scores the participants did provide were from different proficiency tests and were often not comparable. The actual sample used for the current

study consisted of the seventeen participants mentioned above who had studied Chinese for either 4-5 years or over 5 years by the time of their participation.

Understandably, there could be significant heterogeneity among the CFL participants in this study, due not only to the different Chinese programs in which they were enrolled, but also to their diverse language learning background (either L1 or L2), language exposure, capacities of learning, and time investment. Therefore, it was likely there would be variations among individual learners' Chinese proficiency in general as well as in the four skill areas of listening, speaking, reading, and writing. These variations will be addressed during the data analysis and interpretation process.

3.2.3 Sampling

Convenience sampling was adopted for both the pilot study and main study due to the policy restrictions within most CFL programs on participant recruitment and the low availability of CFL learners who would serve as research subjects on a voluntary basis. While larger sample sizes were of course preferable in order to increase statistical power during the data analysis phase, we were only able to recruit a total of twenty-five participants from five CFL programs for the main study.

CFL program directors were contacted for initial site access with a brief introduction of the study. Once the access was approved, a recruitment email detailing the purpose of the study and participation eligibility was sent to the program director along with an electronic sign-up sheet, who could then share the email either directly with students in his or her program or with instructors who would help distribute the information. Students who were interested in participation would fill out the sign-up sheet and email it back to the investigator.

3.3 Variables

Variables investigated for each research question related to the quantitative research design are summarized as follow:

Table 2

Variables for Quantitative Research Questions

Research questions	Variables
Do CFL learners in this study identify synonym and collocational association equally well?	Independent: Type of association Dependent: Identification of word association
Do CFL learners identify synonym and collocational association equally well at different levels of word frequency?	Independent: Type of association Target-word frequency level Dependent: Identification of word association Identification of synonym association Identification of collocational association

3.4 Instrumentation and Pilot Study

Three major instruments were employed in this study: a Chinese Word Associates Test, a Demographic Information Sheet, and a guided interview protocol. A pilot study was conducted among both Chinese native speakers and CFL learners for mainly test development and refinement purposes.

3.4.1 Chinese Word Associates Test

Development of Initial Test Items

A total of forty-five stimulus words—all adjectives—were used in the study. Three major sources were used in creating the initial word associate test items: Commonly-Used Chinese Words Frequency List (汉语常用词词频表), Chinese Synonym Dictionary (同义词大词典), and CCL Corpus. The Commonly-Used Chinese Words Frequency List was developed by Beijing Language and Culture University based on a Modern Chinese corpus of 500 million Chinese characters. The frequency of each

word was derived by the number of occurrence of this word in the corpus. The Chinese Synonym Dictionary was produced by 上海辞书出版社 (Shanghai Lexicographic Publishing House). It includes close to 15,000 pairs of Chinese synonyms or near synonyms. CCL Corpus, or the Peking University CCL Online Corpus was developed by Center for Chinese Linguistics PKU. The corpus includes a total of 783,463,175 characters, of which there are a total of 581,794,456 characters for Modern Chinese, and 201,668,719 characters for Classical Chinese. Data for the Modern Chinese corpus came from a wide range of oral and written linguistic materials in modern-day use, including modern literary works, translation of literary works from other languages, applied writing, historical works, transcripts of TV talk-shows and TV plays, government and legal documents, business writing, newspaper and journal articles, movie scripts, blogs, tweets, and web pages.

Target words for the initial 70 items were selected based on the following guidelines:

1. They were all adjectives or used predominantly as adjectives instead of any other part of speech.
2. For purposes of consistency, only two-character compound words were selected. This selection criterion was also based on the finding that approximately 74% of commonly used Chinese words are two-character words (北京语言学院语言教学研究所 [Language Education and Research Center of Beijing Language Institute], 1986).
3. They came from three levels of word frequency—below 1000, between 1000 and 5000, and above 5000.
4. Words that have synonyms were given greater priority to those with only near synonyms. Words with closer meaning association with their synonyms were given greater priority to those with less, so that when students selected synonyms for a target word, their responses were less likely to be interfered with by

whatever distinctions in meaning that might lie between the target word and the keyed synonyms. This would reduce any possible ambiguity with the items.

5. Words that have more than one synonym were usually given greater priority to those with only one, though a small number of words with single synonyms were also included. With a larger number of synonym associations, it is easier to see students' differences in association knowledge, with the more proficient students presumably knowing more associations.

The creation of options for the items followed the following guidelines:

1. Each item had four options for synonym association and four options for collocational association.
2. Whenever possible, all word options would be of roughly equivalent or higher word frequency than that of the target words, so that students' response to an item would not be impacted by the frequency of word options to a great extent.
3. Collocational associations needed to be selected from those that appear more frequently in the CCL corpus.
4. When synonym associates (keyed responses) had shared characters with the target words, distracters should also have shared characters with the target words, although the characters shared in the two situations did not need to be the same. This was done to make both associates and distracters appear equally attractive to students.
5. When synonym associates (or keyed responses) did not have shared characters with the target words, distracters usually did not have shared characters with the target words, either.
6. Distracters for synonym associations were selected in such a way that they should not bear close meaning similarities with the target words while still appearing attractive to students less proficient in association knowledge.

7. Distracters for collocational associations were selected in such a way that their combinations with the target words clearly did not appear in the CCL Corpus.
8. Selection of distracters was based partly on their effectiveness in differentiating the more proficient learners from those less proficient. Distracters were to present themselves as plausible alternatives to the correct answer(s) when learners were not proficient enough and should not be so obviously incorrect that even the least proficient learners would exclude them from possible options. Greidanus and Nienhuis (2001) found that distracters semantically related to stimulus words were better suited to "assess the quality of word knowledge of advanced learners" in the WAT than semantically nonrelated distractors. This indicated that the paradigmatic distractors used in this study should ideally be words that have some semantic relationship with the target stimulus words and that the syntagmatic distractors should also have seemingly workable collocational relationships with the target words.
9. The number and position of associates and distracters within an item should have some variations within the test booklet, especially between adjacent items.

In addition to these criteria on target word and item option selection, a new item format was adopted that was much different from Read's WAT. The earlier version of the WAT had exactly two associates for each type of word association. Due to its susceptibility to guessing, it was later changed to a format where there were still a total of four associates for each item, yet the number of associates for each type of word association ranged from 1 to 3. In other words, the combinations of the number of keyed responses for the two association types can be 1 + 3, 2 + 2, and 3 + 1. With this revised format, it was believed that students had a lower chance of guessing the correct responses. Nonetheless, guessing is still a great concern with the varying number of keyed responses. For instance, if a student is certain that only one of the four options for synonym association is associated with the target word, then he would be at least able to

deduce that three (rather one or two) out of the four options for collocational association should be selected. Should the student happen to know that one of the collocational options is definitely incorrect, she would not need to know any of the three other options to decide that all of them should be marked as associates. In other words, students' knowledge of one type of association may affect their performance in the other, and their knowledge of some options may affect their performance on other options within the same association type. To ensure that their performance on each type of association and on each individual option is independent, a multiple true/false item format was adopted for the current study. Using this format, students would make individual decisions based upon whether each word option is an associate, and the decision to select one option does not provide any clues that influence selection of another. Students can also focus on one word option at a time rather than trying to balance their responses on different word options. Additionally, with this format, an item could have between zero and four associates for each association type rather than the 1-3 for the revised WAT, giving greater flexibility to item development.

This multiple true/false test format was employed for the initial batch of 70 items with a sample item appearing below. The top portion of the item assesses synonym associations and the bottom portion assesses collocational associations. Written directions were provided at the beginning of the test, which explained the two types of associations with an example illustrating what students were expected to do to complete an item.

Table 3

Initial Sample Item

严重（的）		
表情	Y	N
错误	Y	N
伤害	Y	N
方案	Y	N

Table 3—continued

严肃	Y	N
严密	Y	N
严峻	Y	N
严格	Y	N

Pilot Study among Chinese Native Speakers

An initial version of the Chinese Associates Test was administered to five Chinese native speakers. There were altogether 70 items with 70 target words tested. Each item tested two types of word association simultaneously using a multiple true/false item format. The participants were instructed to decide whether each word option was or was not an association of the given target word by marking T (for "True") or F (for "False").

The participants were asked to work on the test in a quiet and uninterrupted setting and to spend as much time as they wanted. However, they were not allowed to consult any reference materials or ask another native speaker for help.

The completed test booklets were not rated on an individual basis. In other words, the total score of each participant and his or her accuracy rate were not concerns for this part of the pilot study. Rather, data analysis focused on the level of agreement among all participants and how their responses taken collectively compared with the keyed response. Below is an item that shows the participants' agreement level for each word option. In this case, the participants had unanimous agreement that "应用" (application) and "兴趣" (interest) were collocational associates of "广泛" (wide, extensive), and that "胸怀" (mind) and "天空" (sky) were not. However, they had less agreement on the synonym word options, except for "空泛" (vague), which they all correctly decided was a non-associate. For the associate "普遍" (widespread), one participant made a wrong

choice by marking “N,” whereas for the non-associate “广阔” (spacious), four out of the five participants made a wrong choice by marking “Y.” They had even less agreement with the associate “宽泛” (broad), with participants selecting two Y’s and three N’s.

Table 4

Sample Agreement Pattern among Chinese Native Speakers

广泛（的）	Number of participants who selected “Y”	Number of participants who selected “N”
应用	5	0
胸怀	0	5
兴趣	5	0
天空	0	5
宽泛	2	3
广阔	4	1
普遍	4	1
空泛	0	5

Below is a summary of Chinese native-speaker participants’ agreement rate on all items as indicated by the ratio of Y’s and N’s. There was a total of 70 items or 560 sub-items in this first pilot test, with each true/false question on one of the eight word options for each target word regarded as a sub-item. Five items were excluded from the analysis, since only four participants gave a response for them. For the rest of the items, the participants were able to have full agreement in 322 situations. In other words, five out of five participants were able to agree on a response for 322 sub-items or 57.5% of the entire test. Comparatively, four out of five people agreed on a response for 153 sub-items or 27.3% of the test, and three out of five people agreed on a response for 80 sub-items or 14.3% of the test. However, the responses that all or the majority of the participants were able to agree upon were not always the keyed responses. For instance, there were 3

situations when the participants had full agreement on the incorrect response, and 19 situations when they had 80% agreement on the incorrect response. One of the examples was the word option “管理” (management) for the target word “呆板” (rigid); while the two words do form a collocational association, all five participants had selected “N” as a response.

Table 5

Native Speaker Agreement Rates on Initial Items

100% agreement	80% agreement	60% agreement
322	153	80
57.5%	27.3%	14.3%
100% agreement on keyed response	80% agreement on keyed response	60% agreement on keyed response
319	134	52
57.0%	23.9%	9.3%

In addition to participant agreement, the pilot study focused on obtaining and examining feedback from the Chinese native-speaker participants (represented by Participant NS1 through NS5 below) on the test items so that the test could be improved. Below is a summary of their feedback in four respects—

1. Clarity of directions

Participant NS4 was not completely clear about the test directions and asked for clarification.

2. Item format

Participant NS1 suggested that the two types of word associations should be tested separately. He found that for any given item, after he had selected collocational associates and moved on to synonym associates, he would test if a word option was a

synonym by putting it in a collocation, i.e., trying to find out if a word option could form a collocational relationship with his selected associates in the collocational section. Due to this feedback, responses in the two sections were deemed to interfere with each other, with one section potentially providing additional cues for the other section.

3. Item difficulty

Three participants voiced concerns that CFL learners might find it hard to complete the test, since they themselves had difficulties answering some of the questions. For instance, Participant NS2 pointed out that she was not familiar with some of low-frequency words, such as “乖僻” (eccentric), and she doubted that non-native speakers would know these words. The participants suggested that there should be a greater number of easier words.

4. Item ambiguity

Participant NS2 pointed out that due to increasing communication on the internet, language use had become less standardized, so that collocations that had been traditionally unacceptable were beginning to be accepted. This had brought ambiguities to many of the collocations. She also indicated that age might be a factor that affects people’s understanding of what collocations were correct and what were not, i.e., younger people might have different responses to some of the questions from older people. Participant NS4 indicated that it was harder to make a judgment on synonym associations than on collocational associations, since it was often difficult to determine whether two words were synonyms due to the different degrees of meaning association.

Based on this feedback, some adjustments were made to item and test format, test directions, and item difficulty. Rather than having synonym and collocational association tested in the same item, two separate booklets were created, each testing a different association type. Another revision to the item format was the addition of a “U” box (for “unknown”) at the end of each word option, giving CFL learners the option to mark “unknown” if they decided they had no knowledge of a word. Detailed directions and a

sample item were included at the beginning of each booklet, demonstrating what test takers were expected to do, and explaining how the correct responses were derived for each association type.

The synonym booklet was presented to students prior to the collocational booklet and this order was kept for the main study. This decision was based on the investigator's observation that when CFL learners did not know a target word, the collocation section seemed to provide them with more clues on guessing the target word meaning than did the synonym section, so that there would be less of a carry-over effect by presenting the synonym section first . However, this assumption could be invalid so that the lack of counterbalancing could become a possible limitation of the study.

In terms of item difficulty, some of the less-familiar target words as judged by native speakers were removed from further consideration, as they were likely to be even more difficult for CFL learners. Sample items in the two booklets are provided below.

Table 6

Version 2 Sample Items

严重			
严肃	Y	N	U
严密	Y	N	U
严峻	Y	N	U
严格	Y	N	U

严重（的）			
表情	Y	N	U
错误	Y	N	U
伤害	Y	N	U
方案	Y	N	U

A new list of 40 items for each booklet was created out of the 70 items evaluated by the native speakers with agreement rates being used as one of the criteria for item selection. Although there could be multiple reasons why the participants selected different responses on the same words, target words and word options with higher levels of agreement were given greater priority. Usually, only words with 100% or 80% agreement on keyed responses were adopted. When an item with low-agreement options needed to be kept, those options were replaced with potentially less controversial choices. When a synonym associate was found to be problematic, however, often the whole item was removed due to the difficulty in finding a replacement.

Regarding the breakout of word frequency, there were at this stage a total of 13 items for each booklet at a frequency of 5000 or above, 13 items between the frequency of 1000 and 5000, and 14 items below a frequency of 1000.

Pilot Study among CFL Learners

The new 80-item test (40 items in each booklet) was administered to four CFL learners (represented by Participant NNS1 through NNS4) who participated in the pilot study. All four students were in their fourth or fifth year of study in a postsecondary CFL program. Each student was given a total of 50 minutes to complete both booklets in a quiet and uninterrupted setting. They were encouraged to ask questions if they had difficulties understanding the test directions. Each test was followed by a 10-minute interview which asked for their general impression of the test as well as their impression of the test-taking experience.

As shown in the table below, students had an accuracy rate of between 26.3% and 38.8% for the booklet on synonym association and between 40% and 59.4% for the booklet on collocational association. The rate of correctly-identified associates ranged from 23.2% to 39.0% for synonym association, and 38.3% to 66.0% for collocational association. The best-performing student was able to achieve approximately 39% correct

for Booklet 1 (synonym association), and around 60% correct for Booklet 2, and able to identify 66% of the collocational associates. From a test development perspective, the overall difficulty of this test was very high for this group of learners. For CFL learners who would be involved in the main study with comparable length of study in Chinese, achieving full accuracy on this test seemed highly unlikely.

Table 7

Pilot Study Results among CFL Learners

		Student 1	Student 2	Student 3	Student 4
Booklet 1	Overall accuracy rate	30.0%	26.3%	34.4%	38.8%
	Rate of correctly identified associates	23.2%	24.4%	30.5%	39.0%
	Rate of correctly identified distracters	37.2%	28.2%	38.5%	38.5%
Booklet 2	Overall accuracy rate	45.6%	40.0%	53.1%	59.4%
	Rate of correctly identified associates	38.3%	38.3%	55.3%	66.0%
	Rate of correctly identified distracters	56.1%	42.4%	50.0%	50.0%

Another purpose of the pilot study was to determine whether the time allotment for the test administration was appropriate, whether test directions were clear, and how learners felt about the test as expressed through their own words. Regarding time allotment, it typically took students less than 20 minutes to complete each booklet, a total of less than 40 minutes for both booklets. One participant spent 17 minutes completing each booklet, and a total of 34 minutes for both. The participants did indicate that they might take longer if they did know every word. The 50-minute time allotment seemed to be more than sufficient for all CFL learners involved in this part of the pilot study. All

four participants felt that overall the test directions were clear but also offered suggestions for improvement: Participant NNS3, for instance, mentioned that when she did not know a target word, she naturally wouldn't know if it was associated with any word option, so she would just mark "unknown" for all four word options. She suggested that it would be helpful to have an "unknown" option for the target word, as well. These suggestions were incorporated in further revisions of item format and directions.

Creation of Additional Test Items

After the pilot study was administered to this group of learners, it was decided that an additional 10 items needed to be created for each booklet to provide exactly 15 items for each frequency level in each booklet. One target word "繁荣" (prosperous) was removed from the test and so required a replacement, as none of the four CFL pilot study participants had any knowledge of this word. Therefore, a second list of 11 items for each booklet was created and pilot tested among five native speakers, four of whom also participated in the first part of the pilot study. These items followed the same item format as appeared in the pilot study on CFL learners, with synonym and collocational associates tested separately. Based on suggestions from the CFL pilot study participants, a "U" (for "unknown") box was added for each target word as well as for each of the word options.

Particular attention was given to keeping the difficulty level of these additional items under control. Words of higher frequency were selected in preference to those of lower frequency as options, and ambiguous associations were avoided whenever possible. As shown in Table 8 below, the level of agreement among native-speaker participants seemed to improve. That is, the participants achieved full agreement for 71.6% of the items, as compared to the 57.5% in the earlier test. Out of all situations when the participants had 100% agreement, they were able to select the correct responses in all but one instance. Put another way, there was only a 1% difference between full agreement and full agreement on keyed response. This difference was 2.3% for the 80% agreement.

In other words, when the participants agreed, they typically agreed on the correct responses.

There was a higher occurrence of full agreement and full agreement on the keyed responses in Booklet 2 than Booklet 1 for this test. Such a difference was hardly observable for 80% agreement. These results seem to indicate that native speakers in this test had greater agreement for collocational associations than for synonym associations.

Table 8

Agreement Rates among Chinese Native Speakers on Additional Items

	100% agreement	80% agreement	60% agreement	Total
Test	63	16	9	88
	(71.6%)	(18.2%)	(10.2%)	
Booklet 1	29	8	7	
Booklet 2	34	8	2	
	100% agreement on keyed response	80% agreement on keyed response	60% agreement on keyed response	Total
Test	62	14	5	88
	(70.5%)	(15.9%)	(5.7%)	
Booklet 1	29	6	4	
Booklet 2	33	8	1	

Final Review and Revision of Test Items

The test booklets used for the pilot study, together with the 11 additional items for each booklet, were submitted to a Chinese professor (Professor A as a pseudonym) at a postsecondary CFL program for final review. Four items in Booklet 1 (“严重, ” “丰富, ” “诙谐, ” and “含蓄”) were found to be problematic, due mainly to the fact that the keyed responses could not be considered as true synonyms of the target words. For instance, even though both “诙谐” (the target word) and “滑稽” (one of the word options) have the

underlying meaning of “playful” or “funny,” the former is more often understood as “humorous” and the latter as “comical.” Consequently, Professor A believed that these two words were not usually replaceable, and items related to all four target words were subsequently removed from both booklets.

In Booklet 2, only one target word (“圆滑(的)”) was found to be problematic, since the collocations it made with the keyed responses appeared awkward and were rarely used. It was decided that replacement options would be created for this target word.

In addition, since target words such as “鲁莽” and “繁荣” were marked as “unknown” by all or nearly all pilot study participants, all related items were removed from both booklets.

On the other hand, certain items where native speakers did not have total agreement were considered appropriate and allowable. For instance, even though “老成” did not have a high agreement among native speakers, Professor A believed that it was a legitimate synonym of “成熟,” and therefore should be retained.

Professor A also suggested that 45 items for each booklet should be sufficient for this test, so that there would be exactly fifteen items for each of the proposed frequency levels—above 5000, between 1000 and 5000, and below 5000. Two items (“舒畅” and “明显”) from the original 70 items were selected as replacements for the dropped items, and a new item was created. Two other items were dropped from the middle frequency level in order to reduce the number of items at this level from 17 to 15. “充足” was one of the items removed as its response options shared considerable similarities with those of “充沛” in the low-frequency level.

Based on the results from the pilot study, it was estimated that CFL learners at advanced or even superior level might be able to reach a 70% accuracy rate. However, other than a very few exceptional students, it was thought to be very unlikely that they would be able to reach full or nearly full accuracy, thus making it difficult to discriminate

among the high-performing students. In other words, a ceiling effect would not be a concern for the current test.

3.4.2 Demographic Information Sheet and Interview Protocol

A short Demographic Information Sheet was created for use in the main study. The Information Sheet collected the participants' information on their language learning background, in-country Chinese learning experience, proficiency test scores, as well as their consent to participate in a Skype interview after the test. The format of this information sheet consisted of a combination of multiple-choice and short-answer questions.

The interview protocol contained a list of questions for the guided interview, focusing on the participants' general impression on the test items and test administration, and the processes and strategies employed by the participants in completing the items. These were intended to serve as general guiding questions in the interviews, and were not always strictly followed due to time constraints. Both the Demographic Information Sheet and the interview protocol are included in the Appendices.

3.5 Data Collection Procedures

Due to the different locations where the participants lived, data collection was done either on site or remotely through mail. The majority of the participants took the test in their classroom with a proctor. The time allotment was 50 minutes for completing both booklets of the Chinese Word Associates Test. The proctor started the test by handing all participants Booklet One. After they had completed Booklet 1, they handed the booklet back to the proctor, took Booklet 2 from him, and spent the remaining time working on Booklet 2. Once they had completed the test, they were asked to fill out the Demographic Information Sheet and turn them in to the proctor.

The last question on the Demographic Information Sheet asked participants whether they would like to take a Skype interview with the investigator and leave an email for the investigator to contact them. After receiving all test materials back, the investigator contacted all interested participants to schedule an interview.

Prior to the Skype interview, the investigator sent scanned images of all booklet pages to the participants and asked them to briefly review the questions and their responses. In the interviews, the investigator went through all interview questions with the participant, and asked them to elaborate on their responses using examples from the test booklets. Each Skype interview took approximately 10-15 minutes, and the conversations in their entireties were recorded and later transcribed with the participants' consent.

In situations when a student was the only participant within his or her own program, a packet of test materials was mailed to them instead. Test materials were carefully labeled, and a detailed instruction sheet on how to complete the test and mail back the materials was enclosed in the packet. The participants were instructed to complete the test within the same time allotment at one sitting without using any type of reference materials or seeking help from other people. They were also asked to complete the booklets and Demographic Information Sheet in the same order as the participants in who completed the test in a proctored environment. Considering that these students volunteered to participate and there were no consequences for how well they performed on the tests, their chances of cheating were very low, and their test data were considered to be as trustworthy as those in a proctored test. In fact, an informal analysis revealed that the test scores of participants who completed the test in an unproctored environment showed no significant scoring increases over participants who took the test in the proctored environment.

All participants were informed that their performance on the test would be used only for this particular study, would in no way affect grades for any courses they may

currently be taking, and would not be shared with any third party, including their instructors or program directors. The participants did have the option of having a brief score report and the overall findings of the study reported to them so as to motivate them to complete the tests independently to the best of their ability.

3.6 Data Analysis

3.6.1 Scoring of the Chinese Word Associates Test

The participants received one point when they correctly identified an associate or distracter for an item. The score for any given item therefore ranged from 0 to 4. The score a participant received for an item in this test would be the number of matches between the selected responses (either "Y" or "N") and the keyed responses. Participants were not penalized for incorrect selections, and "unknowns" were recorded but not considered in the scoring.

3.6.2 Data Analysis Procedures

Both quantitative test data and interview data were collected for this study. Descriptive and inferential statistics were employed to answer the first two research questions, i.e. the comparison of CFL learners' performance across association types, and across association types by different word-frequency levels. Three types of indices were used to understand any performance differences in the test: total score or total accuracy rate, rate of correctly-identified associates, and rate of correctly-identified distracters. The reason why identification of associates and distracters were evaluated separately was the assumption that it would be likely that learners might be stronger at identifying associates for a target word in some situations, and stronger at excluding distracters from their selections in other situations.

Descriptive statistics for the indices above included mean, standard deviation, minimum, and maximum. Inferential statistical analysis focused on the differences in

performance across association types or across both association types and word frequency levels.

For comparisons across association types, tests for normality were conducted initially for the presence of any serious departure from normal distribution. As non-normality was not found to be a concern for any of the comparisons, paired-sample *t*-tests were conducted on the differences in CFL learners' scores, rates of correctly-identified associates, and rates of correctly-identified distracters across the two association types. Results of non-parametric alternatives to *t*-tests, such as sign tests and Wilcoxon signed-rank tests were also presented so as to provide a comparison with the *t*-test results, thus taking into consideration the small sample size used in the study.

In order to investigate any possible interaction between word association type and word frequency on learner performance, a two-factor within-subjects ANOVA was employed, with both main and interaction effects examined, and post-hoc simple effect analyses and pair-wise comparisons conducted. Effect sizes for comparisons were derived whenever needed to provide a better estimate of the magnitude of the differences between means. Cohen's *d* (Cohen, 1988) using pooled standard deviations were employed in computing the effect sizes. A combined use of the SAS software (Statistical Analysis System) and the Microsoft Excel was employed for all required data analyses. An α level of .05 was used in interpreting the statistical testing results.

All interview data of six participants (represented as Participant A through Participant F) were transcribed and proofed, and subsequently coded for any key words related to the following questions: (1) sources of knowledge or strategies in their self-report of the test-taking process, particularly in the examples they provided; (2) comments on the test instrument, including time allotments, clarity of directions, general difficulty of the test, comparison between different sections, etc.; (3) past study background that might have influenced their performance in the test, particularly their

study-abroad experiences. These codes were further examined to find any emerging themes.

CHAPTER FOUR

RESULTS

This study aimed to investigate whether CFL learners of higher proficiency levels knew synonym and collocational associations equally well and whether their knowledge of word associations varied by word frequency level. Both descriptive and inferential statistics were employed in answering these research questions. Test data were analyzed from the perspectives of both vocabulary breadth, i.e. the number of known or unknown target words or word options, and vocabulary depth, or the identification accuracy of both types of word association. Several different indices were used as indicators of participant performance: (1) Number or proportion of target words unknown within a booklet or a given frequency level of a booklet; (2) Number or proportion of word options unknown within a booklet or a given frequency level of a booklet; (3) Total score or accuracy rate for the test, a booklet, or a given frequency level of a booklet; (4) Number or rate of correctly-identified associates or distracters for a booklet or a given frequency level of a booklet. Comparisons across association types were made using *t*-tests. The potential effect of both association type and word frequency was examined using ANOVA as well as post-hoc analyses. Corresponding effect sizes were also computed.

A second purpose of the study was to explore how word associations were inferred and what general strategies were employed in the inference process. The participants' interview data were used in answering these questions.

This chapter will report the data analysis results through three sections: identification of synonym and collocational associations, identification of synonym and collocational associations as related to word frequency, and process of word association inference. The first two sections focused primarily on quantitative data analysis results, followed by supporting interview data. The third section was addressed exclusively by interview data.

4.1 Identification of Synonym and Collocational Associations

4.1.1 Vocabulary Breadth

Even though target words in this study were not randomly selected to represent each Chinese frequency level, vocabulary breadth was operationalized as the number of known target words in any given booklet and at each word frequency level of a booklet.

A total of forty-five target words were tested for both association types. However, it was found in the data collection process that one target word "简单" (Item 12 in both booklets) had identical word options for the two booklets. That is, instead of collocational associates or distracters for Booklet 2, synonym associates or distracters were entered instead due to an error in the test development stage. As a result, Item 12 was removed from the data analysis and only forty-four items were included.

As shown in Table 9 below, CFL learners in this study had on average a total of 16 unknown target words for Booklet 1 (synonym association) and 16.29 unknown target words for Booklet 2 (collocational association). In reality, these two numbers should be exactly the same, because the same set of target numbers had been tested in both booklets. The slight difference was probably due to the fact that some participants, instead of marking each "U" box for an unknown target word, neglected doing so when they decided they could not answer an item. This was very common towards the end of each booklet, or the low-frequency sections. As a result, it is likely that the actual average number of target words unknown was larger than 16 or 16.29.

These numbers indicate that CFL learners in the study could recognize a total of approximately 28 target words out of 44, or 63.6% of all the target words tested. However, the number of target words recognized varied widely among the learners, with some recognizing as few as 17 (38.6%) or 18 (40.9%) target words, and others recognizing as many as 39 (88.6%) or 41 (93.2%) target words. The standard deviations

for the number of unknown target words were 6.79 and 7.15 respectively for synonym and collocational association. Assuming that participant performance was normally distributed, this means that approximately 68% of the participants had between 9 and 23 unknown target words (one standard deviation below or above the mean), or were able to recognize between 21 and 35 target words out of the 44 words total.

Table 9

Unknown Target Words in Each Booklet

Number of unknown target words	M	SD	Min	Max
Booklet 1	16.00	6.79	5	26
Booklet 2	16.29	7.15	3	27

Note. Min = Minimum; Max = Maximum.

As shown in Table 10 below, the average number of unknown word options is 34.29 for Booklet 1 and 21.65 for Booklet 2, indicating that on average the participants had approximately twelve or thirteen fewer unknown word options in Booklet 2 than in Booklet 1. In percentages, Booklet 1 and Booklet 2 had an average of 19.5% and 12.3% unknown word options respectively for the participants, and a difference of 7.2% between the two. Again, the actual percentages could be higher, because it had been observed from the marked booklets that when a target word was unknown, the participants often marked only the unknown box for the target word but not any of the unknown boxes for the word options. Consequently, what is represented in the following table may have underestimated the participants' actual number of word options unknown to them.

The maximum number of unknown word options for the participants was 77 for Booklet One and 97 for Booklet Two, which both constituted large proportions (43.8% and 55.1% respectively) of the total of 176 word options for each test booklet.

Table 10

Unknown Word Options in Each Booklet

Number of unknown word options	M	SD	Min	Max
Booklet 1	34.29	18.63	0	77
Booklet 2	21.65	23.98	0	97

Note. Min = Minimum; Max = Maximum.

4.1.2 Association Identification

In this study, vocabulary depth was operationalized as CFL learners' accuracy regarding the identification of two important types of word association: synonym and collocational association. Specifically, this study investigated the number of correct decisions over whether a word option was associated with a target word through either type of association, the percentage of correct decisions, and the frequency and percentage of correctly-identified associates or distracters.

Descriptive statistics such as mean, standard deviation, minimum, and maximum were presented first on CFL learners' performance on the two association types. To determine whether any differences we observed applied also to the CFL population these learners represented, inferential statistical testing was used.

As shown in Table 11 below, CFL learners in the study had an average score of 55.12 (or 31% accurate) on synonym association and 65.94 (or 37% accurate) on collocational association, with over a 10-point difference between the two. As each point

represents a correct response to one word option, this difference indicates that overall the participants were able to make 10 more correct association decisions for collocational association than for synonym association. There is, however, quite a large gap in the scores the different participants received in each booklet, ranging from 24 to 105 for synonym association and from 35 to 134 for collocational association. There also seems to be a slightly larger standard deviation among participant scores in Booklet 2 than in Booklet 1, indicating that in this participant sample, performance in collocational association was slightly more variable than in synonym association.

Table 11

Total Score and Accuracy Rate of Each Booklet

Variable	M	SD	Min	Max
Booklet 1 Total	55.12	24.16	24	105
Booklet 2 Total	65.94	29.01	35	134
Booklet 1 accuracy rate	0.31	0.14	0.14	0.60
Booklet 2 accuracy rate	0.37	0.16	0.20	0.76

Note. Min = Minimum; Max = Maximum.

On average the participants were able to correctly identify 31% (as shown in Table 12 below) of the synonym associates in Booklet One and 41% of the collocational associates in Booklet Two. They were able to correctly identify 30% of the distracters (word options that are not synonym associates of the target words) in Booklet One and 31% of the distracters (word options that are not collocational associates of the target words) in Booklet Two. Clearly, for this particular group of participants, performance in Booklet Two was higher than that in Booklet One regarding the identification of associates, even though there was hardly any difference for the identification of distracters. Overall, the participants in the sample were almost equally competent in

identifying associates and distracters for Booklet One or the synonym section, but had an easier time identifying associates than they did distracters for Booklet Two or the collocational section.

At the individual learner level, the gaps were very large between the least competent and most competent CFL learners in this test, with the former identifying only 12% of Booklet One associates and 23% of Booklet Two associates, and the latter identifying 66% of Booklet One associates and 78% of Booklet Two associates.

Table 12

Rate of Correctly-Identified Associates/Distracters by Booklet

Variable	M	SD	Min	Max
		Booklet 1		
Rate of correctly-identified associates	0.31	0.15	0.12	0.66
Rate of correctly-identified distracters	0.30	0.13	0.11	0.49
		Booklet 2		
Rate of correctly-identified associates	0.41	0.17	0.23	0.78
Rate of correctly-identified distracters	0.31	0.17	0.08	0.71

Note. Min = Minimum; Max = Maximum.

In order to determine whether the above observed differences in the sample were statistically significant, inferential statistical testing such as the *t*-test needed to be conducted. An initial step was to check if the assumption of normal distribution had been met. Tests for normality were conducted using the SAS UNIVARIATE procedure on the variables of interest. While the distribution of two individual variables (Booklet 2 total score and Booklet 2 rate of correctly-identified associates) were found to be non-normal,

there was insufficient evidence to show that any of the “difference” variables had a significant departure from normal distribution.

Table 13

Normality: Total Score Difference

Tests for Normality				
Test	Statistic		*p*	
Shapiro-Wilk	W	0.97	Pr < W	.77
Kolmogorov-Smirnov	D	0.14	Pr > D	>.15
Cramer-von Mises	W-Sq	0.05	Pr > W-Sq	>.25
Anderson-Darling	A-Sq	0.30	Pr > A-Sq	>.25

Table 14

Normality: Difference in Rate of Correctly-Identified Associates

Tests for Normality				
Test	Statistic		*p*	
Shapiro-Wilk	W	0.93	Pr < W	.22
Kolmogorov-Smirnov	D	0.20	Pr > D	.07
Cramer-von Mises	W-Sq	0.09	Pr > W-Sq	.14
Anderson-Darling	A-Sq	0.51	Pr > A-Sq	.18

Table 15

Normality: Difference in Rate of Correctly-Identified Distracters

Tests for Normality				
Test	Statistic		*p*	
Shapiro-Wilk	W	0.94	Pr < W	.34
Kolmogorov-Smirnov	D	0.14	Pr > D	>.15
Cramer-von Mises	W-Sq	0.06	Pr > W-Sq	>.25
Anderson-Darling	A-Sq	0.39	Pr > A-Sq	>.25

As shown in Tables 13-15, the Shapiro-Wilk test for normality indicated that the null hypothesis of normality needed to be retained for the difference between Booklet 1 and Booklet 2 total scores (p = .77), the difference between rate of correctly-identified associates (p = .22) and the difference between rate of correctly-identified distracters (p = .34) in the two booklets. The other three tests for normality also yielded non-significant results. In other words, the distributions of the above three differences did not show marked deviation from normality. A paired-sample *t*-test was therefore appropriate for this study.

Using also the UNIVARIATE procedure in SAS, the tests for location below (Tables 16-18) list results from both a paired-sample two-tailed *t*-test and two non-parametric tests—the sign test and the Wilcoxon signed-rank test. *T*-test result in Table 16 showed that the difference between Booklet 1 and Booklet 2 total scores was significantly different from zero (p = .02). The sign test (p = .004) and Wilcoxon signed-rank test (p = .02) also showed that the difference was statistically significant. Based on the observed total scores of the two booklets, we can conclude that the total score or accuracy rate of collocational association is higher than that of synonym association for the CFL learner population this particular sample represents.

Table 16

Tests for Location: Total Score

Test	Statistic		*p*	
Student's t	t	2.67	Pr > |t|	.02
Sign	M	6	Pr >= |M|	.004
Signed Rank	S	45.5	Pr >= |S|	.02

Table 17

Tests for Location: Rate of Correctly-Identified Associates

Test	Statistic		*p*	
Student's t	t	3.88	Pr > |t|	.001
Sign	M	5.5	Pr >= |M|	.01
Signed Rank	S	65.5	Pr >= |S|	<.001

As shown in Table 17 above, there was also a significant difference in the rate of correctly-identified associates between Booklet 1 and Booklet 2 indicated by the *t*-test ($p = .001$), sign test ($p = .01$), and signed rank test ($p < .001$). Based on the observed directionality, this indicates that CFL learners represented by this particular sample are stronger in identifying collocational associates than they are in identifying synonym associates.

Table 18

Tests for Location: Rate of Correctly-Identified Distracters

Test	Statistic		*p*	
Student's t	t	0.52	Pr > |t|	.61
Sign	M	2.5	Pr >= |M|	.33
Signed Rank	S	20.5	Pr >= |S|	.35

Table 18 above shows consistently non-significant results for the comparison between rates of correctly-identified distracters in the two booklets. This indicates that there is insufficient evidence to show that CFL learners' ability to identify non-associates differed across the two association types.

Table 19 below shows that the effect size in Cohen's *d* using pooled standard deviation for the difference in Booklet score or accuracy rate was close to medium at 0.41, and the effect size for the difference in rate of correctly-identified associates was medium-to-large at 0.62.

Table 19

Booklet 1 vs. Booklet 2: Effect Sizes

Booklet 1 vs. Booklet 2	Effect Size (Cohen's *d*)
Total Score/Accuracy Rate	0.41
Rates of correctly-identified associates	0.62
Rates of correctly-identified distracters	0.07

4.1.3 Interview Findings

The six participants who participated in the after-test interviews had divided opinions over which test booklet was harder for them, and for different reasons, as well.

Participant E believed that Booklet 1 was harder due to the number of unknown words. While Booklet 1 and Booklet 2 had the same target words, he was less familiar with word options in Booklet 1 than those in Booklet 2. He indicated that "the synonym portion of the test is hard because it's …I just don't know a lot of the characters … the synonyms for those … for those words that aren't as commonly used I don't know any of those." For Booklet Two or the collocational association part, he commented that "Booklet Two, I … I've seen a lot more of the options, so it's easier to … to at least guess or infer. But for Book One, I haven't seen very many of the options at all."

Participant A believed that Booklet 2 was definitely easier than Booklet 1 due to the context Booklet 2 provided—"I was able to kind of recognize more words by context with other words rather than compared against other synonyms." He further explained

that Booklet Two provided more words with which the target word could be used, thus giving clues to help identify the part of speech or meaning of the target word.

> I think it's because of seeing – instead of just synonyms, more words that might be used with, so if I … if I see … a lot of a certain kind of category of objects I might know—it might give me some kind of guesses of what this—if it's an adjective or something, what it might mean, whereas in the first one, I did have to go by radicals, which could sometimes—if I couldn't recognize the radical to begin with, then I have a lot less to go on.

Participant C who felt that Booklet Two was harder explained that in Booklet 1 he could "pick from the answers which had the same characters," whereas in Booklet 2, there were not many choices which had the same individual characters as those in the target words. In other words, he felt the shared characters provided clues for word meanings and facilitated his selection of the associates.

According to Participant B, which booklet was more difficult depended on one's personal strength in learning. For himself, he believed he was stronger at remembering synonyms but found it harder to remember "when it's appropriate to say a certain word or phrase and when it's not." Therefore, he found Booklet 2 harder than Booklet 1, but he did not think that it was a common experience for all participants:

> I think a lot of other people are better than I am in identifying the patterns while I kind of learn the patterns based on trial and error … I need to kind of get them wrong and then I realized I got them wrong. (Participant B)

4.2 Identification of Word associations as Related to Word Frequency

Three word frequency levels were addressed in this study: below 1000, between 1000 and 5000, and above 5000. Originally, the Chinese Word Associates Test was designed to have exactly fifteen items for each word frequency level in each booklet.

However, due to the removal of Item 12, there were only fourteen items for the high-frequency level as compared with fifteen items for each of the other levels.

In the following sections, vocabulary breadth or the number of target words and word options CFL learners were able to recognize at each word frequency level was first examined using descriptive statistics such as mean, standard deviation, minimum, and maximum. Subsequently, learners' accuracy rate at each frequency level as well as their rate of correctly-identified associates or distracters were also presented through the same descriptive statistics. Lastly, a two-way ANOVA was employed in examining the effects of both association type and word frequency level on learners' association identification.

4.2.1 Vocabulary Breadth at Each Frequency Level

As shown in Table 20 below, within this participant sample, there was an increasingly larger number of unknown target words as word frequency decreased. Given that the average number of unknown target words should be exactly the same across booklets (if the participants gave accurate reports), the participants had an estimated average of 1.5 unknown words at high-frequency level, 6 unknown words at mid-frequency level and 9 at low-frequency level. Taking into account the fact that the high-frequency level had just 14 target words, the percentages of unknown words amounted to 11%, 40%, and 60% respectively for the high-, mid-, and low-frequency levels.

Overall, the standard deviation of unknown target words became increasingly higher as word frequency decreased, which was more evident in the Booklet 1 data. There was a standard deviation increase from 1.50 at high-frequency level to 2.77 at mid-frequency level and 3.48 at low-frequency level for synonym association. For the collocational association, however, the standard deviation increased as word frequency dropped from high to medium, but remained nearly constant between medium and low frequency levels. Again, standard deviations of unknown target words should also be

exactly the same across booklets, and the observed difference may be due to the participants' failure to mark each unknown word.

The above results indicate that in this particular sample, CFL learners recognized fewer target words or had a smaller vocabulary breadth as word frequency dropped. Their vocabulary breadth was also overall more variable for lower-frequency words than for higher-frequency words.

Table 20

Unknown Target Words by Frequency

Number of unknown target words	M	SD	Min	Max
	Booklet 1			
High frequency	1.59	1.50	0	5
Mid frequency	5.76	2.77	1	10
Low frequency	8.65	3.48	3	14
	Booklet 2			
High frequency	1.47	1.23	0	4
Mid frequency	5.88	3.35	1	11
Low frequency	8.94	3.27	2	13

Note. Min = Minimum; Max = Maximum.

As shown in Table 21 below, the number of unknown word options for the high-, mid-, and low-frequency section was 11.41 (20% of all word options at this level), 13.18 (22%), and 9.71 (16%) respectively for Booklet One and 7.47 (13%), 7.06 (12%), and 7.12 (12%) respectively for Booklet Two. The fact that there was not a systematic pattern among these numbers for different frequency levels is not surprising. First, this test was not designed in a way that the frequency levels of word options decreased in the same manner of the target word frequency levels, hence the participants did not necessarily

know fewer word options toward the end of each booklet when target word frequencies were lower. Secondly, the participants might have failed to mark each unknown word option as “U” or “unknown,” especially towards the end of the test where items became more difficult. Again, what is shown in Table 21 may not accurately represent the actual number of unknown word options at each target-word frequency level.

Table 21

Unknown Word Options by Target-Word Frequency

Number of unknown word options	M	SD	Min	Max
	Booklet 1			
High-frequency items	11.41	8.16	0	30
Mid-frequency items	13.18	7.99	0	30
Low-frequency items	9.71	8.45	0	30
	Booklet 2			
High-frequency items	7.47	6.86	0	22
Mid-frequency items	7.06	8.84	0	35
Low-frequency items	7.12	11.48	0	49

Note. Min = Minimum; Max = Maximum.

4.2.2 Association Identification by Frequency

As shown in Table 22 below, for both Booklet 1 and Booklet 2, the participants’ average accuracy rate became increasing lower as word frequency dropped. Within the synonym section, the average accuracy decreased from 50% at high-frequency level to 24% at mid-frequency level and 21% at low-frequency level. Correspondingly, there was a decrease from 56% to 36% and 21% across the three levels of the collocational section. Again, there were large gaps at each frequency level within each booklet between the lowest and highest accuracy rate the participants obtained. For instance, at the high word-

frequency level, the least competent CFL learner was able to achieve only 23% accuracy in synonym association as compared to 79% for the most competent learner; for the collocational association the least competent learners was able to achieve only 34% accuracy as compared to 91% for the most competent learner. However, the six standard deviations (from three frequency levels of the two booklets) were very close, ranging from 0.14 (or 14%) to 0.18 (or 18%).

Table 22

Accuracy Rate by Target-Word Frequency

Accuracy rate	M	SD	Min	Max
	Booklet 1			
High-frequency items	0.50	0.15	0.23	0.79
Mid-frequency items	0.24	0.14	0.10	0.48
Low-frequency items	0.21	0.16	0	0.53
	Booklet 2			
High-frequency items	0.56	0.16	0.34	0.91
Mid-frequency items	0.36	0.18	0.13	0.72
Low-frequency items	0.21	0.18	0.03	0.67

Note. Min = Minimum; Max = Maximum.

The average percentages of correctly-identified associates (as shown in Table 23 below) also showed a decrease in the sample as the target word frequency dropped—from 51% at high-frequency level to 27% at mid-frequency level and 20% at low-frequency level for synonym association, and from 59% to 38% and 23% for collocational association. Similarly, there was also an observed decrease in the rate of correctly-identified distracters as word frequency became lower.

Again, the large gaps between the least and the most competent learners were observable at each word frequency level of both association types. At the high-frequency

level, for instance, the most competent learner was able to correctly identify 89% of the synonym associates and 97% of the collocational associates, whereas the least competent learner could only correctly identify 22% and 42% respectively for the two types of association.

Table 23

Percentage of Correctly-Identified Associates/Distracters by Frequency

Variable	Frequency	M	SD	Min	Max
		Booklet 1			
Percentage of correctly-identified associates	High	0.51	0.15	0.22	0.89
	Mid	0.27	0.15	0.09	0.64
	Low	0.20	0.17	0	0.51
Percentage of correctly-identified distracters	High	0.53	0.17	0.26	0.81
	Mid	0.17	0.11	0	0.39
	Low	0.14	0.11	0	0.35
		Booklet 2			
Percentage of correctly-identified associates	High	0.59	0.15	0.42	0.97
	Mid	0.38	0.18	0.14	0.74
	Low	0.23	0.20	0.03	0.74
Percentage of correctly-identified distracters	High	0.28	0.12	0.08	0.53
	Mid	0.24	0.14	0.06	0.54
	Low	0.19	0.15	0.03	0.55

Note. Min = Minimum; Max = Maximum.

In order to investigate whether there was a significant difference in CFL learners' performance across frequency levels, and whether the effect of frequency varied across word association types, a two-factor within-subjects ANOVA was employed. The two

factors were (1) word association or association type with two levels, and (2) target word frequency with three levels. The dependent variable was accuracy rate in association identification. Because the same 17 participants took the same test with all six association-frequency combinations, both factors were within-subject or repeated-measure factors, with all six cells, or combinations of association type and word frequency level, having roughly equivalent sample variances (ranging from 0.14 to 0.18). Given that the F statistic of ANOVA is robust to violations of normality (if any) when the same sample is used for all conditions, assumptions for using an ANOVA test were generally met for this study. The PROC ANOVA procedure in SAS was used for the analysis with a repeated-measures design on both factors, and selected tables from the SAS output have been included.

Tables 24 and 25 below list descriptive statistics for each level and each cell respectively of the two-way ANOVA on accuracy rate. Results in these tables are consistent with those in preceding analysis, with a larger sample mean for the collocation section than for the synonym section, and a larger sample mean for higher word frequency levels than for lower word frequency levels. Within each association type, the sample mean increased as the word frequency became higher. Within both high- and mid-frequency levels, the sample mean was higher for the collocation section than for the synonym section. However, there was no difference in sample means for the two association types at the low-frequency level.

Table 24

ANOVA: Descriptive Statistics for Each Level

Level	N	M	SD
Level of association			
Synonym	51	0.32	0.20
Collocation	51	0.38	0.22

Table 24—continued

Level	N	M	SD
Level of frequency			
High-Frequency	34	0.53	0.15
Mid-Frequency	34	0.30	0.17
Low-Frequency	34	0.21	0.17

Table 25

ANOVA: Descriptive Statistics for Each Cell

Level of association	Level of frequency	N	M	SD
Synonym	High-Frequency	17	0.50	0.15
Synonym	Mid-Frequency	17	0.24	0.14
Synonym	Low-Frequency	17	0.21	0.16
Collocation	High-Frequency	17	0.56	0.16
Collocation	Mid-Frequency	17	0.36	0.18
Collocation	Low-Frequency	17	0.21	0.18

Table 26

Two-Way ANOVA on Accuracy Rate: Association by Frequency

Source	*df*	SS	MS	F	*p*
Using MS for Subject*Association as an Error Term					
Association	1	0.10	0.10	7.04	.02
Using MS for Subject*Frequency as an Error Term					
Frequency	2	1.86	0.93	183.50	<.001
Using MS for Subject*Association*Frequency as an Error Term					
Association*Frequency	2	0.05	0.03	6.61	.004

As shown in Table 26 above, the main effects of both association type ($F(1, 16) = 7.04$, $p = .02$) and word frequency ($F(2, 32)$, $p < .001$) were statistically significant. The interaction between association type and word frequency was also statistically significant ($F(2, 32) = 6.61$, $p = .004$). These results indicate that CFL learners' accuracy rate in word association was significantly different across association types and across word frequency levels. However, the effect of frequency on accuracy rate was not consistent at each level of word association type (synonym vs. collocational association). Similarly, the effect of association type on accuracy rate was also not consistent at each level of word frequency (high, mid, and low).

For a better understanding of the interaction effect, a simple-effect analysis was conducted using the PROC MIXED procedure in SAS. Tables 25 and 26 represent respectively the effect of association at the three levels of word frequency and the effect of word frequency at the two levels of word association.

Table 27

Simple Effect Analysis: Sliced by Word Frequency

Tests of Effect Slices					
Effect	Frequency	Num *df*	Den *df*	F	*p*
Association*Frequency	High-Frequency	1	33.9	4.12	.05
Association*Frequency	Mid-Frequency	1	33.9	16.47	<.001
Association*Frequency	Low-Frequency	1	33.9	0.07	.79

Note. Num *df* = *df* of the numerator for F statistic; Den *df* = *df* of the dominator for F statistic

The simple effect analysis in Table 27 above indicates that the effect of association type on CFL learners' accuracy rate differed from one word frequency level to another. Association type had a statistically significant effect on association

identification at both high-frequency level ($p = .05$) and mid-frequency level ($p < .001$); whereas no significant effect was detected at the low-frequency level ($p = .79$). This indicates that the identification of synonym and collocational association differed mainly at the mid-word-frequency level, barely at the high-frequency level, but not at the low-frequency level. Specifically, CFL learners had significantly higher accuracy rate in the identification of collocational association than they did in synonym association for the high- and mid-frequency levels, but no significant difference was present at the low-frequency level.

The simple effect of word frequency for each association type is represented in Table 28 below. Word frequency had a significant effect ($p <.001$) on CFL learners' accuracy rate at both levels of word association. In other words, there was an overall significant difference among CFL learners' performance at the three word frequency levels for both synonym and collocational association.

Table 28

Simple Effect Analysis: Sliced by Association

Tests of Effect Slices					
Effect	Association	Num *df*	Den *df*	F	*p*
Association*Frequency	Synonym	2	62.9	98.53	<.001
Association*Frequency	Collocation	2	62.9	115.21	<.001

Note. Num *df* = *df* of the numerator for F statistic; Den *df* = *df* of the dominator for F statistic

Post-hoc pair-wise comparisons were conducted using Tukey-Kramer adjustment between accuracy rates at different word-frequency levels for each association type. Results from Table 29 below indicate that with synonym association, CFL learners' performance was significantly better at the high-frequency than both mid- and low-

frequency level ($p < .001$). However, the difference between mid- and low-frequency levels was not significant ($p = .56$). With collocational association, however, CFL learners' performance at all three frequency levels was significantly different ($p < .001$) from each other. This shows that the effect of word frequency on learner performance was dependent on association types. Learner performance in collocational association became increasingly higher with the increase of word frequency. However, there was insufficient evidence to indicate that learner performance in synonym association was impacted by frequency before the transition from mid- to high-frequency, a transition which seemed to bring improvement.

Table 29

Post-Hoc Pair-Wise Comparisons: Adjusted by Tukey-Kramer

Group 1	Group 2	*df*	*t*	*p*	Adjustment	Adj *p*
Synonym_HF	Synonym_LF	62.9	12.90	<.001	Tukey-Kramer	<.001
Synonym_HF	Synonym_MF	62.9	11.24	<.001	Tukey-Kramer	<.001
Synonym_MF	Syonym_LF	62.9	1.67	.10	Tukey-Kramer	.56
Collocation_HF	Collocation_LF	62.9	15.12	<.001	Tukey-Kramer	<.001
Collocation_HF	Collocation_MF	62.9	8.68	<.001	Tukey-Kramer	<.001
Collocation_MF	Collocation_LF	62.9	6.45	<.001	Tukey-Kramer	<.001

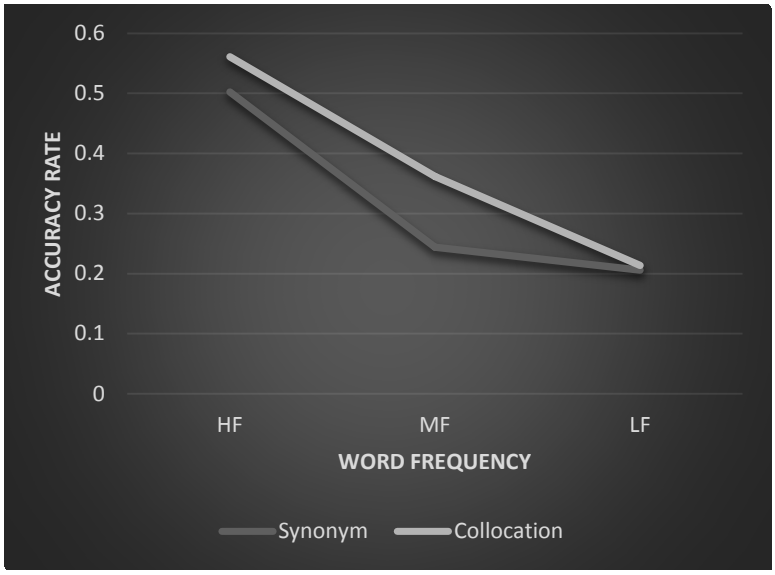
Note. Group 1 and Group 2 represent the two groups for a pair-wise comparison (i.e., Group 1 – Group 2); HF = High-frequency; MF = Mid-frequency; LF = Low-frequency

The line graph below (Figure 1) represents the interaction between the two factors of word frequency and word association type on CFL learners' accuracy rate on association identification. As the preceding analysis has indicated, the difference in accuracy rate across association types was most marked at the mid-frequency level, and less so at the high-frequency level. At the low-frequency level, however, hardly any

difference could be observed. The figure also shows that improvement in accuracy rate with frequency levels also differs with association type. That is, while accuracy rate seemed to increase as the target word frequency increased for collocational association, there seemed to be a "plateau" period from low- to mid-frequency where the learners' performance remained almost stable, with a sudden increase in performance when word frequency transitioned from mid- to high-frequency.

Figure 1

Two-Way ANOVA on Accuracy Rate: Line Graph



Note. HF = High-frequency; MF = Mid-frequency; LF = Low-frequency.

Effect sizes in Cohen's *d* for all pair-wise comparisons of interest were computed using pooled standard deviations. As is shown in Table 30 below, the comparison of high- and mid-frequency had a large effect size (Cohen's $d = 1.81$) for synonym association, whereas the effect size of the comparison of mid- and low-frequency was small (0.26). For collocational association, both comparisons of high- versus mid-frequency and of mid- versus low-frequency had large effect sizes (1.17 and 0.83

respectively). These effect size results coincided with the presence (or lack thereof) of significant differences, i.e., performance differences across word frequency levels which were found to be significant also had large effect sizes, whereas the one non-significant difference had a small effect size.

Table 30

Pair-Wise Comparisons for Each Association Type: Effect Sizes

Level	Comparison	Effect Size (Cohen's *d*)
Synonym	High-Frequency - Mid-Frequency	1.81
	Mid-Frequency - Low-Frequency	0.26
	High-Frequency - Low-Frequency	1.94
Collocation	High-Frequency - Mid-Frequency	1.17
	Mid- Frequency - Low-Frequency	0.83
	High- Frequency - Low-Frequency	2.04

Effects sizes for all comparisons of synonym and collocation (as shown in Table 31) were small at the three levels of word frequency. This indicates that while two out of the three comparisons were found to be statistically significant, the magnitude of these differences was nonetheless small.

Table 31

Pair-Wise Comparisons at Each Frequency Level: Effect Sizes

Level	Comparison	Effect Size (Cohen's *d*)
High-Frequency	Collocation - Synonym	0.09
Mid-Frequency	Collocation - Synonym	0.26
Low-Frequency	Collocation - Synonym	0.13

4.2.3 Interview Findings

The post-test interviews revealed that the participants were well aware that the target words became less familiar to them as the test progressed. They also felt that the familiarity level of both the target words and word options impacted their performance.

The participants found that it was easier for them to work on the beginning part of each booklet than the latter part. Participant E, for instance, said that "I knew most of the target words at the beginning but by the time I get to the end, I don't know a lot of them." Participant F also mentioned that words towards the end of each booklet were more difficult and that "some of the choices" were "really difficult," whereas the beginning was "really smooth." Likewise, Participant C commented that "generally it got harder as it went on."

Word frequency of not only the target words but also the word options seemed to have affected how well the participants performed on the test. Participant E indicated that his major difficulty was that the words in the test, particularly the options in the synonym part, were not as commonly used as those they had learned in the classroom—

> … there are a lot of words that are commonly used, and those are the ones that were – that we are taught as … as Chinese learners, so we know the commonly-used ones. And then the synonyms for those … for those words that aren't as commonly used I don't know any of those. So a lot of words in the test I just had … no … no clue.

Word frequency also appeared to be one major difference why some participants found Booklet 1 to be harder than Booklet 2. Participant E indicated, for instance, that he had seen "a lot more of the options" in Booklet 2 than he did in Booklet 1, which made the synonym association items harder for him.

4.3 Process of Word Association Inference

While CFL learners may be able to recall from their memory whether two words are associated either in meaning or in collocational use, oftentimes determination of word association is not a straight-forward or automatic process. This appeared to be especially true for words (either target words or word options) CFL learners were less familiar with or those they had not encountered earlier. The interview data indicated that they used a variety of strategies in inferring the relationship between two words when they were not able to instantly identify their relationship.

4.3.1 Use of Radical or Character Knowledge

In order to determine whether two words are associated, one usually has to know the meaning of each word. In situations when word meaning was unknown or partially known, many participants indicated relying on radicals or constituent characters in making an educated guess.

> "… I just made the best guess based on the … based on if I knew one of the characters, or if I knew a radical, the other part of the characters …" (Participant E)
>
> I think maybe one thing I try to do is just identify—especially if I didn't know the word as a whole, I would try to identify the individual components, like I said earlier how with "久" ((long) time) versus "远" (far; distant) … (Participant B)

Because each of the words in the test consisted of two characters, often the participants would synthesize what they knew about both component characters to infer the meaning of the whole word. Their knowledge of the individual characters would sometimes come from other words these characters could form with different characters. A case in point would be how Participant E guessed the meaning of "平常" (ordinary; common)—

I know both characters—I know both characters which “平时” is like “often” or “common” and “常常” is like … “often.” So something is “平常”—probably “oftenly seen” or “really common” or something like that.

The use of radicals was mentioned less frequently in participant interviews. The same participant—Participant E—indicated that compared with the use of “separate characters,” he did not “use radicals as often.” However, there were a very few instances when they tried to infer character meaning on the basis of radical knowledge. Below is an example of how knowledge of both sound and meaning radicals had brought a synthesized understanding of the character “疲. ”

Participant F: I remember this by something like “skin,” right. It also means “skin.”

Researcher: Right, yes. It does. It’s that inner part.

Participant F: Yes, the inner characters. So that’s how I vaguely remember “oh, this is pí” and then the … the outer part of pí—it means—it’s same as “病”—the outer part of … “sickness.” So I thought “oh, this must be like ‘tire,’ ‘fatigue.’” And then …

Rather than focusing on guessing the individual meanings of the target words or the word options, some participants tried to compare the target words and corresponding word options in order to infer their meaning connections. Participant F, who was a Korean CFL learner, described her guessing process as follows:

Participant F: So I guess I tried to look if just all the words were two-letter words. And if I could understand exactly at least one letter with this word, then I tried to—I guess tried to—so it’s like finding a root, the root, and then with the given word, then I tried to look at the other half of that, the word, the letter in—I think a lot of times it was like that—I know one letter, but I don’t’ know the other one. Even if I know the other one, sometimes I wasn’t sure what the exact meaning of the word, like the

> composition—I know the composition, but I don't know the exact meaning of that particular word. So that's how I guess—like try to …
>
> Researcher: So you start by looking at the shared character between the target word and the options …
>
> Participant F: Exactly, right.
>
> Researcher: … then you try to find out whether the other part of that word shares something… something (Participant F (simultaneously): similar) similar to the other part of the target word.
>
> Participant F: Yes, exactly.

The shared character between the target word and its word options was also used by other participants in deducing any possible meaning connections. An example of this is shown below:

> …for my educated guesses I would see if I knew one of the characters of the word and if I knew what it meant and then if I did know I would look for either a similar meaning in the four choices or even the exact same character… (Participant C)

In some situations, the participants used sophisticated strategies of radical and character knowledge in making an inference. Below is an example of how Participant F utilized a combination of radical and character knowledge of both the target word and the word options to infer the meaning of "圆滑" (slick and sly) and select its synonym associations:

> … when I saw this word (滑), it was really … it was really hard but at first I thought this wasn't "滑, " I thought it was "骨" (bone). So I was really, really confused, so at first I was thinking what is "圆骨," and then later I remembered this word by the water part, and then I thought "Oh I saw it when I was reading, reading something about skating or something." So I remember like a story book and then I said "Oh, this must be '圆滑,'" and

> then at that point I knew this has to do something with—so I don't know the meaning of the word, but it has to—I mean—there are two words with "滑"—"油滑" (slippery; cunning), "光滑" (smooth; glossy), but I think this is not relevant because "光" is either—it means "complete" or "light," so I just got … but I just wasn't—I wasn't quite sure so I put "unknown." But then "油滑"—"油" is oil, oily, so I thought "Oh, this must be similar," so that's how I 猜 (guess), and then "圆满" (satisfactory; perfect)—I wasn't quite sure of the meaning of the word, but then since this is like 团圆的圆 (the character "圆" in "团圆" (reunion)), and then it could also be—so I felt like something is full. So that's how I guessed these two."

As can be seen from the example, the participant used her knowledge of the water radical as well as her knowledge of seeing the character "滑" in earlier context ("skating" and the story book) to successfully infer the meaning of "滑." Once she understood the meaning of this shared character, she went on to find out the meaning of the second part of each word, and tried to guess how the two characters combined in each word would possibly mean. There was also an ongoing process of examining the target word and a word option in comparison, specifically the comparison of the two non-shared characters and determining whether there were any similarities in meaning. Although such complexed inference process was often well-grounded and able to bring the participants closer to the correct answer, it was not always reliable. In this case, for instance, the participant was able to correctly identify "油滑" (slippery; cunning) as an associate of "圆滑" (slick and sly), but mistakenly took "圆满" (satisfactory; perfect) as an associate based on the same strategic process.

Effectiveness of radical or character knowledge seemed to be impacted particularly when CFL learners had uncertainties about radical or character meaning. For instance, the participants might sometimes be torn between different possible meanings when he or she knew just the individual meanings of characters but not that of the word

as a whole. Participant C talked about how he could not decide whether the word “珍贵” (valuable; precious; rare) had more of the meaning of “precious” (the first character “珍”) or that of “valuable” (the second character “贵”). Due to this, he was unable to decide which word could go with it in a collocation—

> I had thought it might have meant “precious,” but I wasn’t 100% sure. And then I looked at the individual characters like “珍” (precious), and then “贵” (expensive; valuable) like has some sort of “valuable” meaning to it and then I looked at like the second choice like “礼物” (gift; present), like “gift,” like probably would make sense so I got … I didn’t know for sure the English definition of “珍贵” (valuable; precious; rare), but I—based on the individual characters, I thought—you know, it could make sense …

Often a word is not the simple addition of character meanings, but may take on additional or different meanings from those of its components. Participant A talked about the fact that he knew the shared character “名” (which he believed meant “name”) of all five words in an item, and that he would simply need to identify “the other half of the word that (he) might not have known.” He decided that the option “名贵” (famous and precious) was a close synonym to “知名” (well-known; noted), as the character “贵” has the meaning of “noble” in addition to being “expensive.” However, the combination of “出,” which means “out,” and “名” didn’t make any sense to him, so he made the wrong decision that “出名” (famous; well-known) was not an associate of “知名.” Therefore, combining character meaning to infer word meaning was not a very useful strategy in this latter case.

In addition, the incorrect response on “出名” was also due to the participant’s imprecise understanding of the meaning of the character “名,” which could mean in itself “fame” (as in the word “名气” (reputation; fame)) in addition to the more commonly-

used meaning of “name” (as in the word “名字” (name)). This misconception of word meaning also impacted his response to the corresponding item in Booklet Two:

> So like “作品” (works)—that doesn’t seem like something that I think would be famous in the way like “知名” (well-known) kind of indicated—like a famous name. I think that would be like, for like a brand maybe would be like “名牌” (famous brand). So I… that’s why I put a “no” on that one. And like I’ve had for the first one (“人物” (figure; personage)) that's similarly why I chose yes because I think “知名”—“名” obviously can have a … like a … like famous in the sense of like a name—might be something only a person is applicable with. (Participant A)

The above example represents a situation of how CFL learners’ understanding of a word as whole was impacted by imprecise knowledge of individual component characters, hence erroneous inference of word association. Such instances were not uncommon among the interview participants.

In fact, some participants did have the awareness that guessing on the basis of constituent characters was not always reliable—“… sometimes even if you know both characters, the combined meaning is not what … what you expected” (Participant E). Likewise, they didn’t think that radicals were a reliable source of knowledge for word inference, especially when they were not sure if a radical was “a meaning part or sound part of a word” (Participant A).

In addition, the participants’ use of radicals or characters for word inference was probably not equivalent between the two test booklets. Participant A indicated that he did not try to rely on radicals for Booklet 2 (collocational association) because he “didn’t think it would be as relevant,” as Booklet 2 was mainly about “finding usage between different words.”

4.3.2 The Role of L1

As compared with the use of radical or character knowledge, the use of L1 was brought up less often in participant interviews. However, there were some instances when the participants consciously or subconsciously drew upon their L1 word knowledge to facilitate their identification of word associations.

Participant E believed that the researcher had purposefully tried to "trick" them by including things that are English colloquialisms. He cited, for instance, the association of "mature fruit," which he believed was legitimate in English but might not be so in Chinese. He also cited the association of "clear eyes," which he believed was used in both Chinese and English. However, his judgment in both situations turned out to be incorrect and he was also aware of the fact that English and Chinese word associations may or may not coincide. He concluded that the fact that he was "still making associations in English while using Chinese" was "probably bad." He took a cautious attitude toward the use of L1 in making judgment on L2 associations—"I'm kind of wary of just guessing, because things that I … things that I might think make sense in English probably don't in Chinese."

Similarly, Participant A showed sensitivity to L1-L2 differences in making association decisions. When asked whether he relied upon his native language at all in the test, the participant indicated that he "did have to think about that a lot," but "it was mostly for differences." He believed that, for instance, the English word "famous" was applicable to "almost everything really," whereas the character "名" or the word "知名" (well-known; noted) was probably associated with a person, but not "inanimate objects." While his understanding of the Chinese word was somewhat inaccurate, this did indicate that he was aware of the subtle differences between the so-called counterpart words between different languages and that a Chinese word should not be understood merely as its translated counterpart.

More often than not, learners' use of L1 was a subconscious process. As Participant D put it, they "don't really think about that (their use of L1) very much," although he agreed that when speaking Chinese, he would try to "think of what (he) might say in English, and then put it into Chinese," and that in reading, as well, he would often "translate it into English internally."

Another instance of the subconscious use of L1 knowledge was in the determination of a collocational relationship. Participant C could not decide on the meaning of "珍贵" (valuable; precious), thinking that it could mean either "precious" or "valuable." He therefore put the English word "precious" or "valuable" in a collocation with a word option and tried to find out if it made sense.

> I looked at the individual characters like "珍" (precious) and then "贵" (expensive; valuable) like has some sort of "valuable" meaning to it and then I looked at like the second choice like "礼物" (gift; present) like "gift" like probably would make sense… So like "资料" (data; material) I think I wasn't sure if it was if "珍贵" was closer to "precious" or "valuable." And in that case if it meant "precious," like "precious information" wouldn't make sense. But if it was "valuable," and like "valuable information" would make a lot of sense. (Participant C)

Although it could be argued that the participant was evaluating if two concepts – i.e., "precious" and "information"; "valuable" and "information"—would make sense when used in combination, he was undoubtedly using his own native language (English) as the criterion of how two words fit together. He neglected the fact that while concepts may be transferrable across languages, a word and its translation in a different language do not always represent the same concept. For instance, the Chinese "资料" (which could mean data, material, or information) and the English "information" do not mean exactly the same thing.

When asked initially whether she relied on L1 in any situations to make a guess, Participant F, whose native language was Korean , at first negated, saying “I don’t even—I found myself not even thinking in Korean language.” Later, however, she did recall an instance when her native language actually facilitated her meaning inference of the character “惫” as in “疲惫” (weary; exhausted).

> Participant F: … so the second (character) – So I wasn’t quite sure when I saw this—it feels “Oh, something is … tire?” But then I wasn’t sure about this but then later I saw without “忄” (one component of the character) this could be “备” (bèi), and then that’s when I actually use … oh, this is when I actually use the Korean word.
>
> Researcher: Oh. Really?
>
> Participant F: It has the same word called “pipei.” So I was thinking oh maybe this is “pí … píbèi.” It means …
>
> Researcher: You have … like a word that is pronounced almost the same… like this one?
>
> Participant F: It’s the same character.
>
> Researcher: Same character, but …
>
> Participant F: But I don’t know the …
>
> Researcher: The meaning is the same?
>
> Participant F: Yes, the meaning is the same.

Therefore, trying to “find the association between Korean and Chinese language” was probably one strategy she applied subconsciously when encountering words unknown. The participant did indicate, however, that she did not grow up her entire life in Korea, so she did not know as many traditional Chinese characters (which, instead of simplified, are used in Korean) as other Koreans do. Nonetheless, this was one of the few instances when she did use her Korean language as a knowledge source for word inference.

To what extent learners relied on their L1 seemed to be related to their language environment and their proficiency level. Participant C talked about how he used to "focus on Chinese and would never translate in (his) head into English" when studying in China, but a year after he came back began to switch back into a state when he constantly asked himself what something meant in English and whether it would make sense in Chinese, etc. He believed that this was a sign that his Chinese had become worse.

Familiarity with a word, especially past encounters with it, may also impact learners' reliance on L1. The same participant as above talked about how he would have benefited from a past context where "珍贵的礼物" (precious present) was used, but without such a context, he found himself having to translate in English to determine if this was a legitimate collocation—

> … in this case maybe I would have heard someone said "珍贵的礼物," but even if I didn't know what it really meant, like I would have remembered the context, and now at this point I'm more translating in English. (Participant C)

4.3.3 The Role of Context

There was some disagreement among the participants on whether or not the test (particularly Booklet 2) provided some kind of a context that facilitated their identification of word meaning or word associations. Participant A believed that the context within Booklet Two helped him with word recognition – "I was able to kind of recognize more words by context with other words rather than compared against other synonyms." Participant E believed that "(t)here's no context really" and that the lack of context is one reason why it "makes it more difficult." He talked about how context could potentially affect his performance in this way:

> "I mean if I saw some of these words in … in context, I would remember what they mean probably or I could ensure their meaning much better. Without any context, it's really difficult." (Participant E)

When asked whether he would agree with some other participants that Booklet 2 provided some limited context of word use (as in collocations), he reluctantly agreed "I suppose a little bit," but went on to say that such context would not be very helpful since "unless I've studied or heard of it before, I don't know if a collocation exists or not in Chinese."

Although the test itself provided no or very limited context, the participants might draw upon past context they had encountered where the word was used in making an inference. Participant E mentioned that other than radical or character knowledge, he would make the best guess based on "if I'd seen it before but … but couldn't quite remember the meaning but I could remember the context in which I'd seen it."

Participant C talked about how he guessed the meaning of "显著" (remarkable; pronounced) — "I know the individual characters like '显'and '著'—maybe I've seen them together like a couple of times – like just in reading things, but I've never looked up the English definition, but I would remember based on context—like '显著' … like … could maybe mean something like 'clear' …" To be more exact, his inference on the word meaning was done in the past when he encountered this word in context, and in the test he's merely recalling this inference rather than making a guess on the spot.

At a later point in the conversation, the same participant brought up again how past context rather than knowledge of exact word meaning helped him identify word associations:

> I would try to remember times when I would hear the Chinese word that I didn't necessarily know the meaning of but sort of remember in context how this was said. So in this case maybe I would have heard someone said

> “珍贵的礼物,” but even if I didn’t know what it really meant, like I would have remembered the context … (Participant C)

Participant F talked about how she would try to recall in what context she had met a certain word once she was able to figure out its pronunciation:

> But I think most words I don’t know I try to at first I try to think of the audial, like the pronunciation, and then I try to think—because I stayed one year in Beijing for my … the language program… Yeah. It was like in year 2002, and then I think that’s—and then I saw a lot of TV shows, so sometimes when I don’t know the word, if I try—can guess the pronunciation, then I try to think of “Have I heard of this word?” “In which context?”—I think that’s my first approach. (Participant F)

4.3.4 Sound Mediation

Although this does not seem to be a strategy that directly facilitated identification of word associations, its help with word recognition seemed to bring an advantage to those who used it. The only participant (Participant F) who brought it up had some extended explanations and examples on how she utilized the strategy:

> … if there’s a word that I don’t know, I try to think of the pronunciation and try to … So there are lot of words that sometimes I do know the pronunciation, I can think of the exact letter, or vice versa. So if I don’t know the word, but if I can try to pronounce it, then I try to think of whether I have it in my mind, whether I’ve heard of the word long time ago, so like … it’s like an audial data I guess. (Participant F)
>
> … there are some words that I have not used very often, but I know if I hear it, or like in audial then I understand a lot faster than when I … when I see it on the text, so … so when I was working on the survey questions, I … I think there were some words like that—I saw the word and I feel like

> I know pronunciation and if I … if I don't know the word and if I don't know the pronunciation, if I don't know both, then it's harder for me to understand, like remember the meaning of it. But if I at least know how to pronounce it, then I think—that's when I'm able to like guess more if I should be… (Participant F)

As an instance, she talked about how she knew the character "当" in "恰当" (proper; suitable), but wasn't initially sure about "恰. " Once she recalled its pronunciation, however, she immediately identified the word—"then I remembered the pronunciation 'qià,' very luckily, so that's when I was like 'Oh, qiàdāng, I know this word…'" (Participant F)

She concluded that "as long as I know the pronunciation, then I … am better able to guess." When asked whether visual cues or audial cues helped her better, she responded that the latter was "a lot better."

On the other hand, failure to access pronunciation would often result in unsuccessful inference or identification of word meaning. This includes both the situation when the participants were unable to recall character pronunciation and the situation when what they recalled was incorrect. The description below about the same participant is a good case in point.

> Participant F: … sometimes I saw this word, I wasn't able to read this at first, so … and then I wasn't able to—I think I know this is something related to tax, but I wasn't able to … remember…
>
> Researcher: Yes, I see what you mean. The word "tax" is pronounced as "shuì," and this one is pronounced as "ruì."
>
> Participant F: Oh, so…
>
> Researcher: This means "ruìlì"—"sharp."
>
> Participant F: O…kay…

Researcher: And “shuì” has a different (Participant: to the right) component on the left.

Participant F: On the left (surprised)! Oh…So I guess—so this one I wasn’t able to come up with anything, so that’s why I just marked it … so if I remember this as “shuì,” if I believe that, I think I would have answered according to that pronunciation.

At first, of course, the participant’s error in word identification had more to do with the orthographic features of the word, i.e. she took the character 锐 (sharp) as another character 税 (tax) which has the same phonetic radical. Then her failure to recall how the character was pronounced brought further challenge to word recognition. If, on the other hand, she recalled an incorrect pronunciation, then her responses would have inevitably been wrong also.

However, the participant believed that while it was true for her, the sound mediation that was much needed for word recognition might not be necessary for other Koreans who had learned Chinese characters since elementary school. “So they know how—they know the meaning of the word, even though they can’t pronounce it, but for me, I think if I can’t pronounce it, I just can’t …” This, of course, is the participant’s own assumption, and it is questionable whether indeed learners highly proficient in Chinese characters do not need sound mediation to access meaning.

4.3.5 Strategy Use and Association Inference

CFL learners were found using four major types of strategies to varying extent in inferring associations among words: (1) orthographic knowledge, particularly knowledge of radicals or component characters; (2) their L1; (3) linguistic context either immediately available in the test or retrieved from earlier encounters of a word; (4) sound mediation in order to access word meaning.

Overall, despite the many situations when learners successfully inferred word association through the use of these strategies, their effectiveness seemed to be greatly restricted by learners' existing knowledge of word meaning or use. Appropriate use of strategies in itself did not always bring learners closer to the correct answer. As is indicated by one participant's response below, inference of word associations is often nothing more than educated guessing, which is different from the more automated association identification on the basis of sound word knowledge.

> Researcher: So if you do try hard and if you're given plenty of time to think about them, do you think you might be able to … like find some of the correct answers? Or is it if you don't know, you simply don't know?
> Participant D: I could sort of in… I could have looked at them and tried to guess more than I did. But other than that, no.

CHAPTER FIVE

DISCUSSION

Based on the results described in the previous chapter, this chapter will examine the development of CFL learners' knowledge of word association in the study, and how their association knowledge seems to be impacted by association type and word frequency level. The process and strategies involved in the inference of word association will also be discussed. Implications of these findings on CFL curriculum and pedagogical development, CFL proficiency standards, and vocabulary assessment design will then be further explored.

5.1 Overall State of CFL Learners' Vocabulary Breadth and Depth

In terms of vocabulary breadth, CFL learners in this study had a relatively high mastery of the target words, as they were able to recognize on average 28 words out of the 44 total, or approximately two thirds of the target words. Considering that one third of the target words were at a frequency level of below 1000, their vocabulary breadth appears to be fairly large. In comparison, their knowledge of word associations did not seem to be very well developed, as they were only able to average a 31% accuracy rate for the synonym association and a 37% rate for the collocational association. The unbalanced development of vocabulary breadth and depth indicates that while CFL learners with at least four years of study had acquired a large number of Chinese words, their "quality" of knowledge regarding many of these words still needed improvement. This finding was not surprising, however, considering earlier evidence that the vocabulary knowledge of advanced L2 learners with a large vocabulary size may still be "hazy" when it comes to understanding meaning ambiguities and "collocational restrictions" (Lennon, 1996, p. 35). Shen's study (2009) with advanced CFL learners also

indicated that the development of accurate word use or "vocabulary strength" tended to lag behind that of vocabulary size.

Moreover, CFL learners in the study may have had limited or non-existent exposure to some of the words tested in this study, especially words of lower frequency. As Participant E indicated, the "commonly-used words" in the test were the ones they were taught as Chinese learners, whereas he "had no clue" about other words that "aren't as commonly used." Compared with their performance for the whole test, their performance on just the high-frequency target words (with a frequency of 5000 and above) was much more satisfactory, with the participants being able to recognize on average nearly 90% of the target words in the high-frequency section, with average accuracy rates of 50% and 56% respectively for the synonym and collocational associations. These data indicate that their vocabulary breadth for the most frequently used Chinese words was very well-developed and beginning to resemble that of a Chinese native speaker. However, their associational knowledge lagged behind.

When identification of the two association types was compared, participants had an overall better mastery of collocational association than synonym association, with a significantly higher accuracy rate ($p = .02$) and rate of correctly-identified associates ($p = .001$) in collocational association than for synonym associations. However, there did not seem to be a significant difference in the rates of correctly-identified distracters for the two types of associations. This implies that while these CFL learners were better able to identify collocational associates than synonym associates, their competence in identifying distracters, or determining that a word was a non-associate, was comparable across the two association types.

The fact that collocational associates turned out to be easier for the CFL learners may be explained partly by the larger number of unknown word options in the synonym section. Participant E, for instance, indicated having "seen a lot more of the options" in the collocational section, whereas the synonym options "aren't as commonly used."

Another possible explanation is that the collocational options tended to provide mini language contexts where the target word could be used and might give additional clues that learners could employ in recalling or inferring word meaning. As Participant A indicated, “I was able to kind of recognize more words by context with other words rather than compared against other synonyms.” Even though the context available was rather limited compared with a sentence or longer unit of discourse, the participants could have found themselves recalling a situation when they read or heard about the collocation in a much more extensive context. Participant C explained how an earlier context in which a collocation was used might have helped him—“… maybe I would have heard someone said ‘珍贵的礼物’ (precious present), but even if I didn’t know what it really meant, like I would have remembered the context …” In daily language use, however, rarely would two synonyms be used together in the same language context, and seeing a target word and its synonym options does not provide any meaningful context whereby word meaning might be inferred or earlier encounters with the words could be recalled.

On the other hand, while it is true that the synonym options often share a common character with the target word, the shared character could sometimes become a hindrance rather than a facilitator of correct word identification. This is because (1) the other component character might be unknown to learners and (2) the meaning of the word is rarely the simple combined meaning of two Chinese characters. For CFL learners who had generally underdeveloped knowledge at the character level, their identification of words might as yet not benefit from the shared characters to a great extent. For this reason, their existing radical or character knowledge did not necessarily help them more in the synonym section than it did in the collocational section.

A more direct interpretation of the performance difference across association types is that CFL learners with comparable Chinese study background as those in this study had an overall unbalanced development of collocational and synonym association knowledge, with the former somewhat better developed than the latter. While

paradigmatic knowledge (or collocational knowledge in this study) "do(es) not in fact represent a higher degree of lexical development" (Wolter, 2001, p. 65), lack of competence in this area will undoubtedly deter CFL learners from greater accuracy, appropriateness, and sophistication in vocabulary use.

5.2 The Frequency Effect

Greidanus, Beks, and Wakely (2005) found that advanced L2 learners had deeper word knowledge regarding "frequent, earlier-acquired words" than "less frequent, more recently acquired words." Results of the study indicated a similar pattern. Overall, CFL learners' performance in the study decreased as word frequency decreased. That is, their accuracy rate dropped from 50% to 24% and to 21% for the synonym section and from 56% to 36% and to 21% for the collocational section with the decrease of word frequency. It also seemed that the more dramatic drop was from the high-frequency section to the mid-frequency section, implying the possibility that by the time CFL learners confronted mid-frequency level words, their word knowledge was already at a very low point, so that there was not much difference between their performance in mid- and low-frequency sections.

However, the effect of frequency on learner performance was not consistent between booklets, as the ANOVA results indicated a significant interaction between association type and word frequency level ($p = .004$). For collocational association, the differences in accuracy rate across any two word frequency levels were significant, whereas for synonym association, a significant difference was present between high- and mid-frequency and between high- and low-frequency levels, but not between mid- and low-frequency levels. As mentioned above, at the mid-frequency level of the synonym section, the less familiar target words when coupled with many unknown word options, made association identification very difficult, with participants performing almost equally poorly as the low-frequency section. On the other hand, with collocational association,

participants did not have as many unknown word options at the mid-frequency level as they did with synonym association, so that their collocational association performance was not impacted to the same extent. Another possibility was that the participants' synonym knowledge was only beginning to show prominence with the words they were most familiar with, whereas it was much less "visible" for mid- and low-frequency words. This provides evidence that instead of a syntagmatic-paradigmatic shift with the increase of CFL proficiency level, the roles of synonym and collocational association knowledge in the mental lexicon may be dependent upon the learners' familiarity with a given word, with the somewhat "delayed" emergence of synonym knowledge as learners gained familiarity with the word.

When comparing across association types with the frequency factor taken into account, it also seems that CFL learners' advantage in collocational association held true only at the mid-frequency level ($p < .001$) and barely at the high-frequency level ($p = .05$), given that there was no statistically significant difference between learners' performance in the two association types for low-frequency words. These results imply that with words of medium frequency, collocational association knowledge plays a clearly dominant role in CFL learners' vocabulary network. However, when it comes to words they are most familiar with, the gap between synonym and collocational association knowledge is not quite as discernable. Again, consistent with the explanation on the delayed emergence of synonym association with the growth of word familiarity, it is likely that CFL learners' synonym association knowledge begins to "catch up" as learners become more familiar with a given word.

As for the non-significant difference across association types at the low-frequency level, one possible explanation is that since they knew very little of these words, CFL learners' vocabulary network for these words had not yet been established, resulting in equally poor knowledge of both association types.

The significant interaction between association type and word frequency indicates that CFL learners' development of association knowledge is not purely dependent on proficiency level as some may believe. Rather, the profile of association knowledge for any given word in a CFL learner's mental lexicon is partly determined by how familiar they are with that word. For the participants in this study, the gap between synonym and collocational association knowledge was the largest at the mid-frequency level, with collocational association knowledge clearly taking the lead. This gap showed signs of closing with the high-frequency words, the words they were likely to be most familiar with. Compared with Lü's study (2010) which showed an evolution from a sound-oriented to meaning-oriented mental lexicon as beginning and intermediate CFL learners grew in word familiarity, this study provides some initial insights into how the meaning-related associations—synonym and collocational associations in this case—develop with word familiarity among CFL learners at higher proficiency levels.

5.3 Individual Vocabulary Knowledge Profile

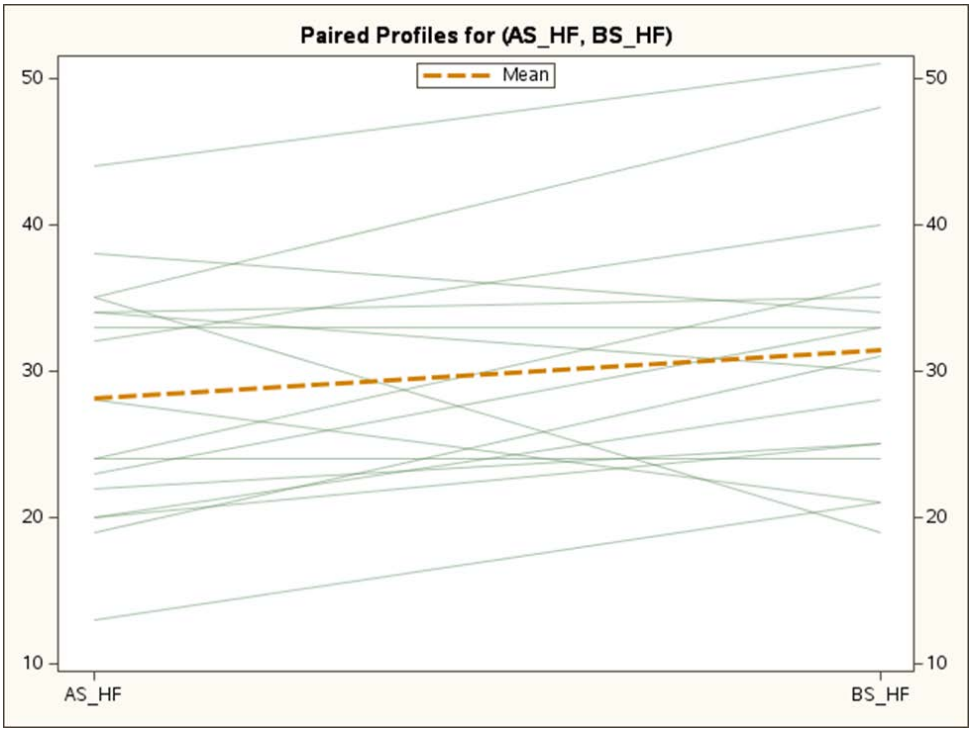
While the participants in this study had been learning Chinese for at least four years, there seemed to be very large differences among individual learners in how many of the tested words they knew and how well they knew them. The most and least competent learner in the study had a gap of over 20 words in the number of target words they were able to recognize. The overall accuracy rate for CFL learners ranged from 14% to 60% for synonym associations, and from 20% to 76% for collocational associations. While the gap in number of words known narrowed at the high frequency level, the variability in association knowledge remained almost the same, as indicated by the roughly equal standard deviations and similar gaps of 56% for synonym association and 57% for collocational association between the most and least competent learners.

Meanwhile, while overall CFL learners in the study knew collocational association better than synonym association, there were a few exceptions. The following

graph compares the scores individual participants received in the high-frequency section of Booklet 1 (AS_HF) as compared with those in the high-frequency section of Booklet 2 (BS_HF). This represents not only the variability among learners' performance in the high-frequency section of both booklets, but also the somewhat different directionalities in the synonym-collocational comparisons, with some learners scoring higher in the synonym section than in the collocational section and others scoring just the reverse. These idiosyncratic performance patterns pointed precisely to "the very individual nature" of vocabulary knowledge (Lessard-Clouston, 2006).

Figure 2

Individual Learner Profiles in High Frequency Words across Booklets



Note. AS_HF = Booklet 1 score for high-frequency words; BS_HF = Booklet 2 score for high-frequency words.

While heterogeneity can be present with any learner group, some of the larger gaps we observed are nonetheless worthy of concern. Factors that may help explain these variabilities include the actual contact hours learners have had with the Chinese language, the effectiveness of their Chinese curriculum, their study-abroad experiences, their personal strengths and learning styles, and the extent to which print exposure and reading have been an integral part of their overall Chinese language learning experience. Wolter (2001) raised a model of depth of individual word knowledge to account for the idiosyncratic nature of word knowledge on an individual level. In particular, depending on how well a word is known to a given learner, its association representation in his mental lexicon can be different, and a predominance of one type of association over another may result.

5.4 Process and Strategies in Word Association Inference

Previous literature (Greidanus & Nienhuis, 2001; Greidanus, Beks, & Wakely, 2005; Lessard-Clouston, 2006) on word association has focused more on whether association knowledge increases with L2 proficiency and to what extent L2 learners' association knowledge resembles that of native speakers. Rarely has any research probed the process that is involved in determining word association when association knowledge is not immediately available. Participant interviews in this study lend some rudimentary insights into this question.

The Chinese Word Associates Test in this study required learners to look at four options for an item and determine whether each option was an associate of the target word, either synonym or collocational. It is different from traditional free association tasks when learners need to freely produce words from their memory which they believe to be associates of a stimulus word. In the current test, there could be two major processes involved in identification (rather than production) of word associates: (1)

Learners looked at a target word and a word option and tried to decide whether the words were associated based on their existing association knowledge. For synonym association, this required the participants to know the meanings of both words as well as their meaning association. For collocational association, the participants needed to know whether the two words frequently appeared together from any previous exposures to the collocation. (2) Learners were either not sure about the meaning of at least one word in a pair (target word and word option) or not sure if the two words were related. They had to determine on the spot from their existing word knowledge and any available clues in the test whether the two words shared a synonym or collocational association. Even though there were clear directions in the test for them to mark a target word or word option as unknown when they simply did not know it, there were numerous instances whereby the participants tried to infer an answer when they did not know a word. The participants indicated that they would try to make an educated guess when they were "over 50% confident" (Participant C), when they "could maybe distinguish some similar radicals" (Participant D), or for "maybe a third of the time" (Participant C). Given the prevalence of guessing in the test, it appears that guessing or inference is a common practice in actual language use situations when automatic identification of word relationships is not attainable. In fact, learners may often be in an in-between state when they know a little or remember something vaguely about words or word relationships but are not completely certain, and therefore may need further verification from existing knowledge frameworks or external clues. The following processes or strategies were described by CFL learners in this study.

5.4.1 Word Recognition and Inferencing

Word recognition is typically the initial step before any word association can be identified. In order to determine whether two words are associated, learners usually have to understand initially what the two words mean. For this reason, strategies that are

conducive to word recognition are often key to word association identification. Interview data indicate that the participants used different strategies to either recall or infer word meaning.

Strategy One: When the participants knew the meaning of individual component characters, they would synthesize them in a way to infer the word meaning. Participant B inferred that since “远” was “far” and “久” was “a long time”, the word “久远” should be “dealing more with time than distance.”

Strategy Two: When the participants were not sure about individual character meaning, which was very common, they might rely initially on other words where these characters served as components to determine character meaning, and then use these inferences to further infer the word in question. One example is how Participant E inferred that the word “平常” probably meant “oftenly seen” or “really common” by knowing that “平时” (a word containing its first character) meant “often” or “common” and “常常” (a word containing its second character) meant “often.”

Strategy Three: When the participants had no knowledge about individual character meaning, they would sometimes rely on their radical components knowledge to infer character meaning, and subsequently the meaning of the whole word. A case in point would be the situation when Participant F recognized the sound component “皮” of the character “疲” as well as the outer part of the character (its meaning component), which means “sickness,” and was then able to infer that “oh, this must be like tire, fatigue.”

The three strategies above are telling evidence of the special nature of Chinese word recognition. The fact that compound words can be decomposed into individual (and frequently stand-alone) characters, which often further break down into radical components, requires CFL learners to build their word knowledge from knowledge of the very basic graphic components. Not having such fundamental radical knowledge will no doubt hinder learners from more effective learning of word meaning and consequently

relationships between words in either meaning or use. Multiple studies (Shen, 2000; Shen & Ke, 2007; Williams, 2013) have provided evidence of the benefits of radical knowledge in CFL word recognition and reading proficiency.

Other than the "graphic" clues of a Chinese character or word, some participants found that their word recognition was greatly facilitated by knowing its sound. Participant F in the study talked about how word pronunciation was like "audial data" that she tried to access every time she needed to identify a word's meaning. When she knew how to pronounce a word, she was able to "understand a lot faster" or "better able to guess" than if sound access was not available. On the other hand, if she was not able to correctly recall a word's pronunciation (as in the situation when she mistakenly pronounced the character 锐 (ruì) as shuì), she was not able to identify its meaning correctly. While none of the other participants highlighted the importance of sound access in their interviews, they were almost unanimously able to pronounce a word correctly when they could correctly identify its meaning. As some studies (Perfetti et al., 1992; Everson, 1998; Perfetti & Tan, 1998) indicated, sound mediation seems to be an essential step for CFL word recognition, with knowing the meaning of a word being closely associated with the ability to name it (Everson, 1998).

Another factor that facilitated the participants' recall or inference of word meaning is context of use. Participant A indicated that he was able to "recognize more words by context with other words" in Booklet Two, even though other participants believed that the context provided in the test was rather limited. More than one participant indicated that recalling an earlier context where a word was used helped them better identify word meaning. Questions such as those Participant F asked to herself when encountering a word—"Have I heard of this word?" "In what context?"—might be a common approach employed by many CFL learners in trying to recall word meaning. As Participant C also indicated, what they recalled might not be the exact word meaning, but rather their inference based on an earlier encounter. Therefore, language contexts, either

those readily available in the test or those to which learners had been previously exposed, could help with word identification. This is consistent with earlier findings (Bengeleil & Paribakht, 2004; Kondo-Brown, 2006; Qian, 2005) that contextual clues were an important knowledge source in word inferencing, except that in this study, contextual clues obtained in earlier language learning situations could turn out to benefit later word inferencing.

In addition, especially for the participants whose native language shared a similar orthographic system as Chinese, their L1 might facilitate word recognition when they had difficulty identifying word meaning in Chinese. An example is how a Korean participant identified the meaning of “惫” as in “疲惫” (tired; exhausted) through the same use of the character in the Korean language. However, these instances were limited to CFL learners whose first language shared considerable similarities with the Chinese writing system.

5.4.2 Inferencing of Word Association

A significant finding of this study is that correct identification or inferencing of word meaning (both target words and word options) does not necessarily lead to successful judgment on word associations. One reason for this is that learners’ knowledge of word meaning was often partial or imprecise, so that their decisions on word associations were also impacted. Another reason is that for especially collocational associates, if learners had never encountered a collocation in use, it was hard to determine whether the two words were associated by just knowing what they meant individually. As Participant E put it, “unless I’ve studied or heard of it before, I don’t know if a collocation exists or not in Chinese.” For these reasons, in addition to word identification, CFL learners employed other strategies in determining the relationship between two words.

For the synonym section, for example, CFL learners made frequent comparisons between the target word and a word option and across different word options in their component characters. A shared character was often taken as a clue of similarity in meaning. For example, Participant C explicitly indicated that once he was able to identify one character of a target word, he would "look for either a similar meaning in the four choices or even the exact same character." Since in most items more than one word option might have a shared character with the target word, the participants would often focus their attention on the non-shared characters. If they decided that the non-shared characters were also similar in meaning, they were more likely to determine that the two words were associated. Participant F, for instance, gave a detailed account on how she determined that "油滑" (slippery; cunning) instead of "光滑" (smooth) was a synonym associate of "圆滑" (slick and sly) by comparing the character "圆," "油," and "光"—the beginning characters of the three words. Such a judgment process actually bypassed the stage where they needed to identify full word meaning. In this case, the participant knew just the individual characters that composed the three words, but was not able to identify the full word meaning. Nonetheless, she still managed to make a judgment on word association by comparing the component characters of these three words.

For the collocational section, the participants might use an earlier context where they heard a collocation used to determine whether a particular association was legitimate. Participant C talked about how hearing someone say "珍贵的礼物" ("precious present") would have helped him make the judgment on their association—"even if I didn't know what it really meant, like I would have remembered the context."

When such contexts were not available, however, they would have to rely on their conceptual knowledge or word use in their first language to make a decision. Participant C talked about how he judged the legitimacy of the collocation "珍贵的资料" through determining whether it "made sense" in its L1 counterparts—"So like '资料' I think I wasn't sure if it was if '珍贵' was closer to 'precious' or 'valuable.' And in that case, if it

meant ‘precious’, like ‘precious information’ wouldn’t make sense. But if it was ‘valuable’, and like ‘valuable information’ would make a lot of sense.” While apparently he was examining the fit of two concepts (the concepts as represented by “珍贵” and “资料”), he was employing a strategy of resorting to his L1 as the criterion for whether these concepts worked together.

In conclusion, CFL learners used a wide variety of knowledge sources, such as radical and morphological knowledge, pronunciation, context, and L1 in the identification or inference of word associations. In some situations, they were able to combine the use of different sources in a fairly complicated way and seamlessly switch from one source to another as needed. This reveals that CFL learners at this stage of their literacy development were beginning to gain a high level of manipulation over these knowledge sources in understanding word associations. Proficiency in word knowledge, therefore, is represented not only by how much preexisting knowledge they have about words or word associations, but their ability to flexibly use such knowledge to make correct identification or inferences.

5.5 Implications for CFL Teaching and Learning

Researchers believe that one major characteristic of higher-level L2 proficiency is the “precision of lexicon” (Leaver & Shekhtman, 2002, p. 26) or knowledge of the “exact meaning and usage” of words, “both common and rare” (Kubler, 2002, p. 111). Results of the study indicated that while CFL learners with over four years of study had a relatively large vocabulary size especially at the high-frequency level, their vocabulary depth, or specifically association knowledge in this study, still lagged behind. Especially noticeable was that their synonym knowledge was particularly weak. One possible reason for their weak performance could be a lack of emphasis in classroom instruction on establishing and continuingly broadening the learners’ vocabulary networks. It is likely that words are still taught primarily as individual units, without giving due emphasis to

the relationship between one word and another. While collocational association between words may be something that CFL learners can "pick up" by themselves through repeated exposures to the same collocation, they may not be as sensitive to the similarities and subtle differences between the meanings of two words, unless they are overtly taught. One recommendation of this study, then, is to include systematic explicit instruction and practice on the connections among words as well as the building of learner sensitivities to these connections.

Another issue to consider is that vocabulary learning strategies that are helpful at lower proficiency levels need to be further strengthened at higher levels. For example, radical and character knowledge was one of the most frequently used strategies by the participants in this study. Interviews with the participants indicated that they were not only using these strategies voluntarily, but had developed a greater awareness of their effectiveness. In addition, they were able to use these strategies with caution, and recognize the fact that radical and character knowledge was often unreliable in word identification. Nonetheless, there were still situations when they tried to deduce word meaning through the simple combination of the meanings of its character components. There were also multiple situations when they were not able to make correct identification or inference because their knowledge of a certain character or word component was imprecise. In other words, while they were clearly at the "application" stage of radial or character knowledge (Shen & Ke, 2007), their knowledge base was still not solid enough to support more skilled or effective application. A second strategy that one participant (Participant F) indicated using was the reliance on sound recognition before accessing word meaning. She indicated that the "audial data" benefited her more than "visual" information of a word, and that when she failed to recall a word sound, she often had difficulties in identifying its meaning. This implies that correctly pronouncing a word is still an essential skill that learners of higher proficiency need to continue to build. For a writing system with a weak symbol-sound correspondence, developing such

competence for more effective word recognition is a laborious process and will continue to be a challenge for higher-level CFL learners.

For the above reasons, some of the more fundamental abilities comprising word knowledge in Nation's (2001) framework should continue to be stressed at the higher proficiency levels. This includes in particular what a word sounds like, what it looks like, and what parts are "recognizable" in a word (Nation, 2001, p.27). In the classroom, CFL learners may benefit from simple activities such as reading aloud in strengthening their sound-symbol connections, or by having reading and writing assignments that are accompanied by oral reports or discussions to ensure an integration of spoken and writing vocabulary. Emphasizing both sound and meaning input in vocabulary learning is also important. Meanwhile, CFL learners should also be encouraged to actively apply the orthographic knowledge they have accumulated in word recognition and retention, while being aware of its limitation and drawing upon other knowledge sources when necessary.

The participants' even lower performance in the low-frequency section reveals that they had not had adequate exposures to these low-frequency words either inside or outside the classroom. Even those who had extensive study-abroad experience failed to perform satisfactorily in this section. One explanation could be that while they were given plenty of opportunities to develop their Basic Interpersonal Communication Skills (BICS) (Cummins, 1999), their chances of developing Cognitive Academic Language Proficiency (CALP) were still rather limited. Many of the words at a lower frequency level are more often used in academic, literary, or professional situations rather than in every day communication. Due to their less frequent use, a more effective way to learn these vocabulary is through extensive reading of texts of different styles and genres.

In addition to offering repeated exposures to vocabulary, one other benefit of reading extensively is to help learners better acquire and retain word meanings in a varieties of language contexts. More than one of the CFL learners in the study indicated the importance of context in their recall or inference of word meaning. They either used

the context immediately available or turned to earlier contexts where they saw or heard a word or word association used. For learners of higher proficiency levels, varied language contexts help particularly with implicit learning of precise word meaning and use, complementing what explicit classroom instruction may fail to provide.

Due to this, extensive reading should be made an integral part of a CFL program. To initiate such a program, CFL learners at higher proficiency levels should read varied authentic materials at a level appropriate to their abilities, which may include literary works, newspapers and magazines, and academic or professional writings based on their individual interests and areas of expertise. Authentic materials would give them access to many of the low-frequency or less-commonly-used words that they do not typically learn in the classroom. In the meantime, well-written pedagogically altered materials may also be used to help CFL learners develop sight vocabulary and enhance their understanding of word meaning through repeated word exposures. CFL learners should be encouraged to infer word meaning out of contexts whenever possible rather than relying on the dictionary in the extensive reading process. Effective classroom reporting and evaluation activities should also be developed so that learners can monitor their progress and be guided in any future reading selections.

As an important part of CFL learners' earlier Chinese learning experience, the interview data collected in this study indicated that study-abroad also offered rich opportunities for vocabulary learning. However, more than one participant (Participant A and Participant E) indicated that study-abroad helped more with their spoken proficiency than their reading or writing proficiency. In her review of earlier literature, Llanes (2010) indicated that oral production was believed to improve the most in L2 learners' study-abroad experience, with Davidson (2010) also pointing out that development of spoken proficiency was often the primary motivation for students' study-abroad. While development of spoken proficiency may eventually benefit vocabulary learning, the emphasis is again on *koutouyu* (spoken language) rather than on *shumianyu* (written

language) so that vocabulary with a more academic, literary, or professional nature often does not receive adequate attention. It is essential, therefore, to carefully structure the study-abroad experience so that learners get more opportunities for vocabulary and reading development.

In conclusion, interview data from this study indicated that fundamental word recognition skills that prove to be worthwhile for beginning and intermediate CFL learners need to be further consolidated at higher proficiency levels. In the meantime, a better structured curriculum that integrates continuous vocabulary network building, extensive vocabulary learning activities, and rich reading opportunities both inside and outside the classroom is recommended.

5.6 Implications for CFL Proficiency Standards

While vocabulary knowledge is arguably an essential component of L2 proficiency, it is not given due emphasis in most proficiency standards. In the ACTFL Proficiency Guidelines (2012) for advanced-level reading, the only mention of vocabulary knowledge is in the following statement, with the key words underlined:

> "Readers are able to compensate for limitations in their lexical and structural knowledge by using contextual clues."

Similarly, in the Guidelines for superior-level reading, vocabulary knowledge was mentioned briefly in two statements:

> "Comprehension is no longer limited to the reader's familiarity with subject matter, but also comes from a command of the language that is supported by a broad vocabulary, an understanding of complex structures and knowledge of the target culture."
>
> "Superior-level readers understand texts that use precise, often specialized vocabulary and complex grammatical structures."

In the advanced-level statement, the term “lexical knowledge” was used without any elaboration of what it implied. The first superior-level statement had an emphasis on “a broad vocabulary” or vocabulary breadth. The only emphasis on vocabulary depth was through the word “precise” (vocabulary) in the second superior-level statement. Obviously, such statements are inadequate in describing the important role vocabulary depth needs to play in advanced- and superior-level reading. In order to establish concrete goals L2 learners need to achieve in vocabulary knowledge in order to attain the higher proficiency levels, we should have a better understanding of how vocabulary depth (including association knowledge) develops at various proficiency levels. We also need to know whether and how learners master synonym and collocational associations differently, and whether such mastery is subject to such variables as the influences of word frequency, word familiarity, part of speech, and stylistic classifications.

By understanding “precise” vocabulary use, for instance, we might indicate the ability to identify meaning similarities as well as subtle differences between two words and how the replacement of one word with another might change the text’s meaning. It may also refer to the ability to understand a word in context as it is used in combination with other words. In productive language use such as speaking and writing, precise vocabulary use in place of less appropriate words and in combination with the right words is also an important indicator of learners’ L2 proficiency.

In the assessment of L2 proficiency, it is also insufficient to test learners’ size of vocabulary alone. That is, assessment needs to take into consideration that a word may have multiple meanings and may take on different meanings in different contexts; moreover, it should be remembered that meaning boundaries between two words are not always clear-cut, and that there can be overlap as much as there are distinctions. Assessment also needs to highlight the fact that collocations and expressions that make sense in L1 may not always be acceptable in L2, and that highly-developed vocabulary proficiency should involve the ability to discern such inconsistency.

5.7 Limitations

5.7.1 Limitations in Test Design

As one of the earliest attempts to build a test of word association knowledge in Chinese, the test employed in the current study had some limitations. One limitation that might have impacted the validity of the test results is that learners' prior knowledge of the words tested had not been given more careful consideration. Some participants' relatively low performance on the test might have resulted partly from their lack of knowledge of the "unknown" words (both target words and word options), thus making it difficult to clearly differentiate vocabulary depth from vocabulary breadth. A second limitation is that the average word frequency level regarding word options was not parallel across booklets. The collocational section had clearly much higher average word frequency (for word options) than the synonym section. Therefore, performance differences as we observed between the two booklets might be partially explained by the unbalanced word frequency levels. Third, the option for the participants to fill out "unknown" for any given word complicated rather than facilitated data interpretation. The interview data showed that the participants had not used consistent criteria in determining when they should fill out "unknown" and when they were supposed to make a guess. Towards the end of each booklet where there were many unknown target words, many participants neglected to indicate whether a word option was unknown to them when they had marked a target word as "unknown." To put it more specifically, this test failed to draw a clear boundary between assessment for evaluation purposes and for research purposes. The participants also indicated confusion in this regard—while some tried to be "helpful" to the researcher by marking "unknown" conscientiously, others took it more as an actual test they needed to complete with the best of their efforts. This is probably due partly to the lack of clarity with the test directions, and partly to the fact

that guessing is an accepted practice with most selected-response questions learners work with.

On the other hand, some of the issues as mentioned above cannot be completely removed from such a test. For instance, even though a list of items can be preselected about which all participants have some knowledge, such a list is likely to be very narrow, and may contain only words of relatively higher frequency. In addition, since words that have multiple synonyms are very limited in number, creating an association test using the above list would be even more difficult. Likewise, creating a test with balanced word frequency between the two booklets is very challenging, if not impossible. This is because words in a synonym dictionary typically have a very small number of synonyms (rarely larger than four), whereas there can be a vast number of noun collocational associates for the adjectives tested. The only way to keep an appropriate balance between the two booklets is to select collocational associates (or distracters) of comparable frequencies with the synonym associates, which could be lower-frequency words. As a result, the test will end up having many low-frequency word options that the participants are not able to recognize.

In fact, traditional association tasks where learners are asked to freely provide associates on a prompt might be misleading for the same reason. Take, again, adjectives as an example. Learners in such tasks are very likely to produce more collocational associates than synonym associates simply because there is a larger base number of noun collocational associates and many noun collocational associates are of higher word frequency. Similarly, Read's WAT (Read, 1987, 1993, 1995, 2000) also failed to take into account the word frequency balance between the two types of associates or whether learners had some knowledge of all words tested in the first place.

While it is unlikely that the differences in overall word frequency between the two association sections can be eliminated completely, other issues in the current test may be better resolved. For instance, depending on the test's intended use and potential

test-taker population, a comprehensive examination of learners' vocabulary breadth can be conducted, and vocabulary items in the test could be selected out of those to which all or nearly all test takers had some previous exposures. In this case, the option of "unknown" could also be removed from both the target words and the word options, since it has been verified that these words were not completely unknown to most test takers. In addition, one possible way to work with the inconsistencies in word frequencies is through equating adjustment of the two section scores, so that learners' performance in the two association types can be reasonably compared despite the word frequency differences.

Developing a test that yields valid and reliable results on CFL learners' association knowledge may be a long way off. Nonetheless, while this test we used in the current study is imperfect, it still helps to provide an initial glimpse into the status of CFL learners' word association knowledge and point the way to possible curriculum and pedagogical approaches that might effectively develop their knowledge.

5.7.2 Other Limitations

This study has a very small sample size of seventeen CFL learners. These learners come from different CFL programs with diverse Chinese learning backgrounds. Therefore, any results in the study may not be representative of the CFL learners as a whole with the same length of study. Meanwhile, no common proficiency test scores were available for the participants, so there was no clear indication of their actual proficiency levels. In addition, only two-character adjectives were tested in this study, and the findings are not generalizable to words with other parts of speech or of other composition structures. The 44 target word total and 15 approximately for each proficiency level are also too small to be representative of the whole Chinese vocabulary of interest. Finally, while word frequency and word familiarity are two different

concepts, this study employed word frequency as a close-to-truth indicator of how familiar CFL learners are with a word for the sake of convenience.

5.8 Recommendations for Future Research

The significance of this study is closely tied to the need to better understand vocabulary development at higher levels of CFL proficiency, especially the development of synonym and collocational association knowledge. It also represents a small yet principled investigation to build an empirical base for understanding the vocabulary dimensions of "advanced" and even "superior" CFL proficiency. The design of a tentative measure of CFL association knowledge is another possible contribution to the CFL field where such assessment is lacking. However, as was mentioned in an earlier section, issues with the test instruments coupled with the small sample size hindered us from more meaningful interpretation of the study results. It is hoped that future research will continue to investigate assessment strategies that could represent CFL learners' vocabulary depth in a more valid and reliable way. Based on a more effective vocabulary depth measure, future research may explore topics such as CFL learners' association knowledge on Chinese words of other parts of speech, their knowledge on other dimensions of word association in addition to synonyms and collocations, a comparison of association knowledge across different levels of CFL proficiency, and how CFL learners' vocabulary breadth relates to their mastery of association knowledge. An answer to any of the questions above will potentially contribute to a more comprehensive and in-depth understanding of CFL learners' vocabulary development and lead to better-informed curriculum and pedagogical decisions on CFL vocabulary instruction.

APPENDIX A

CHINESE WORD ASSOCIATES TEST

Chinese Word Associates Test

General Directions:

There are two major types of relationships between words:

1. Collocational associates are words that appear frequently together in an expression.
2. Synonym or near-synonym associates are words that share the same or similar meaning with one another.

For example, the word 干净 (clean) has as one of its collocational associates the word 房间 (room) as these two words can occur together in the expression 干净的房间 (clean room), while 洁净 (clean, spotless) is its synonym associate as both mean "clean." This test assesses your ability to identify collocational and synonym associates in Chinese. You will be asked to work on two test booklets sequentially. Both assess your association knowledge of the same 45 words. Booklet I will focus on their synonym association, and Booklet II will focus on their collocational association.

A typical item will look like one of the following:

Decide if each of the four words in the bottom row forms a collocational

干净			
洁净	Y	N	U
清洁	Y	N	U
清静	Y	N	U
明净	Y	N	U

干净（的）			
意义	Y	N	U
食物	Y	N	U
成果	Y	N	U
房间	Y	N	U

association with the target word on top by filling in either the Y (Yes) or N (No) box next to it. If you simply don't know a certain word, fill in the box U (Unknown).

You will be given a maximum of 25 minutes to complete each booklet. After you have completed Booklet I, it will be collected and Booklet II will be distributed to you. Please ask for clarifications if you are unsure about the directions above.

Are you ready to start?

Booklet I

Directions:

In this section, you will be asked to decide if any of the four words printed beneath a target word is a **synonym associate** of the target word, i.e., if they share the same or similar meaning(s). Mark your answer by filling in either the **Y** (Yes) or **N** (No) box next to it. If you simply don't know a certain word, fill in the box **U** (Unknown).

In the example below, the target word means "clean," so choices 1 and 2 would be marked Y (Yes) because they mean "spotless" and "hygienic" which are synonyms to the target word. Choices 3 and 4 would be marked N (No) since they mean "quiet" and "clear and bright" which are not synonyms of the target word.

干净

洁净	Y	N	U
清洁	Y	N	U
清静	Y	N	U
明净	Y	N	U

Turn over the page to work on questions in Booklet I.

1. 稳定			U
确定	Y	N	U
安定	Y	N	U
特定	Y	N	U
坚定	Y	N	U

2. 广泛			U
广阔	Y	N	U
巨大	Y	N	U
众多	Y	N	U
普遍	Y	N	U

3. 安全			U
齐全	Y	N	U
周到	Y	N	U
平安	Y	N	U
安静	Y	N	U

4. 突出			U
突然	Y	N	U
出众	Y	N	U
非凡	Y	N	U
出名	Y	N	U

5. 特别			U
奇特	Y	N	U
单独	Y	N	U
唯一	Y	N	U
特殊	Y	N	U

6. 明显			U
显然	Y	N	U
明亮	Y	N	U
显要	Y	N	U
显著	Y	N	U

7. 高兴			U
激动	Y	N	U
幸福	Y	N	U
喜悦	Y	N	U
愉快	Y	N	U

8. 激烈			U
剧烈	Y	N	U
壮烈	Y	N	U
猛烈	Y	N	U
强烈	Y	N	U

9. 方便			U
大方	Y	N	U
随便	Y	N	U
轻便	Y	N	U
便利	Y	N	U

10. 紧急			U
急迫	Y	N	U
紧密	Y	N	U
着急	Y	N	U
迫切	Y	N	U

11. 普通			U
全面	Y	N	U
正式	Y	N	U
平常	Y	N	U
寻常	Y	N	U

12. 简单			U
单纯	Y	N	U
纯洁	Y	N	U
单调	Y	N	U
容易	Y	N	U

13. 清楚			U
明确	Y	N	U
清晰	Y	N	U
透明	Y	N	U
清醒	Y	N	U

14. 珍贵			U
宝贵	Y	N	U
昂贵	Y	N	U
高贵	Y	N	U
富贵	Y	N	U

15. 富裕			U
富足	Y	N	U
充裕	Y	N	U
宽绰	Y	N	U
宽容	Y	N	U

16. 成熟			U
老成	Y	N	U
老到	Y	N	U
老实	Y	N	U
老练	Y	N	U

17. 长远			U
长久	Y	N	U
边远	Y	N	U
久远	Y	N	U
长期	Y	N	U

18. 灵活			U
快活	Y	N	U
活泼	Y	N	U
灵巧	Y	N	U
矫捷	Y	N	U

19. 痛苦			U
刻苦	Y	N	U
难受	Y	N	U
辛苦	Y	N	U
痛楚	Y	N	U

20. 真诚			U
诚恳	Y	N	U
真挚	Y	N	U
真实	Y	N	U
忠诚	Y	N	U

21. 知名			U
名贵	Y	N	U
出名	Y	N	U
闻名	Y	N	U
著名	Y	N	U

22. 温暖			U
暖和	Y	N	U
温顺	Y	N	U
温柔	Y	N	U
和煦	Y	N	U

23. 出色			U
优秀	Y	N	U
神奇	Y	N	U
卓越	Y	N	U
杰出	Y	N	U

24. 细致			U
细腻	Y	N	U
仔细	Y	N	U
细小	Y	N	U
别致	Y	N	U

25. 自负			U
自傲	Y	N	U
自立	Y	N	U
自大	Y	N	U
自觉	Y	N	U

26. 锐利			U
锋利	Y	N	U
顺利	Y	N	U
精彩	Y	N	U
犀利	Y	N	U

27. 漂亮			U
标致	Y	N	U
响亮	Y	N	U
美丽	Y	N	U
可爱	Y	N	U

28. 恰当			U
得当	Y	N	U
合适	Y	N	U
正当	Y	N	U
适当	Y	N	U

29. 牢固			U
坚固	Y	N	U
结实	Y	N	U
顽固	Y	N	U
固定	Y	N	U

30. 热忱			U
热诚	Y	N	U
热门	Y	N	U
热闹	Y	N	U
热情	Y	N	U

31. 舒畅			U
舒坦	Y	N	U
舒服	Y	N	U
痛快	Y	N	U
畅快	Y	N	U

32. 详尽			U
安详	Y	N	U
实在	Y	N	U
详细	Y	N	U
详实	Y	N	U

33. 充沛			U
饱满	Y	N	U
充足	Y	N	U
满足	Y	N	U
旺盛	Y	N	U

34. 低廉			U
稀少	Y	N	U
便宜	Y	N	U
多余	Y	N	U
可怜	Y	N	U

35. 公道			U
公平	Y	N	U
公开	Y	N	U
公共	Y	N	U
公正	Y	N	U

36. 荒唐			U
荒芜	Y	N	U
荒诞	Y	N	U
荒凉	Y	N	U
荒谬	Y	N	U

37. 流畅			U
流利	Y	N	U
通畅	Y	N	U
通顺	Y	N	U
流行	Y	N	U

38. 尴尬			U
窘迫	Y	N	U
拮据	Y	N	U
狼狈	Y	N	U
肮脏	Y	N	U

39. 疲惫			U
勤劳	Y	N	U
疲乏	Y	N	U
疲倦	Y	N	U
疲劳	Y	N	U

40. 肤浅			U
粗糙	Y	N	U
简易	Y	N	U
轻浮	Y	N	U
浅薄	Y	N	U

41. 悲伤			U
哀伤	Y	N	U
悲切	Y	N	U
难过	Y	N	U
伤心	Y	N	U

42. 孤寂			U
孤独	Y	N	U
孤单	Y	N	U
寂寞	Y	N	U
孤立	Y	N	U

43. 尖刻			U
尖端	Y	N	U
深刻	Y	N	U
尖酸	Y	N	U
刻薄	Y	N	U

44. 凌乱			U
杂乱	Y	N	U
纷乱	Y	N	U
复杂	Y	N	U
狼藉	Y	N	U

45. 圆滑			U
世故	Y	N	U
圆满	Y	N	U
油滑	Y	N	U
光滑	Y	N	U

Booklet II

Directions:

In this section, you will be asked to decide if any given word is a **collocational associate** of a target word, i.e., if they appear frequently together in an expression. Mark your answer by filling in either the **Y** (Yes) or **N** (No) box next to it. If you simply don't know a certain word, fill in the box **U** (Unknown).

In the example below, the target word means "clean," so choices 2 and 4 would be marked Y (Yes) because they mean "food" and "room" which forms with the target word expressions such as "clean food" and "clean room." Choices 1 and 3 would be marked N (No) since they mean "meaning" and "achievement" which do not have collocational association with the target word.

干净（的）			
意义	Y	N	U
食物	Y	N	U
成果	Y	N	U
房间	Y	N	U

Turn over the page to work on questions in Booklet II.

1. **稳定**（的）			U
基础	Y	N	U
社会	Y	N	U
答复	Y	N	U
状态	Y	N	U

2. **广泛**（的）			U
应用	Y	N	U
关注	Y	N	U
兴趣	Y	N	U
道路	Y	N	U

3. **安全**（的）			U
场所	Y	N	U
思维	Y	N	U
成功	Y	N	U
环境	Y	N	U

4. **突出**（的）			U
成就	Y	N	U
地位	Y	N	U
贡献	Y	N	U
打击	Y	N	U

5. **特别**（的）			U
目的	Y	N	U
含义	Y	N	U
规定	Y	N	U
关系	Y	N	U

6. **明显**（的）			U
效果	Y	N	U
空间	Y	N	U
变化	Y	N	U
差别	Y	N	U

7. **高兴**（的）			U
神情	Y	N	U
印象	Y	N	U
样子	Y	N	U
思想	Y	N	U

8. **激烈**（的）			U
竞争	Y	N	U
愿望	Y	N	U
疼痛	Y	N	U
争吵	Y	N	U

9. **方便**（的）			U
成果	Y	N	U
条件	Y	N	U
负担	Y	N	U
功能	Y	N	U

10. **紧急**（的）			U
情绪	Y	N	U
神经	Y	N	U
关头	Y	N	U
任务	Y	N	U

11. **普通**（的）			U
生活	Y	N	U
朋友	Y	N	U
特点	Y	N	U
道理	Y	N	U

12. **简单**（的）			U
单纯	Y	N	U
纯洁	Y	N	U
单调	Y	N	U
容易	Y	N	U

13. **清楚**（的）			
认识	Y	N	U
立场	Y	N	U
眼睛	Y	N	U
目标	Y	N	U

14. **珍贵**（的）			
资料	Y	N	U
礼物	Y	N	U
物价	Y	N	U
友谊	Y	N	U

15. **富裕**（的）			
费用	Y	N	U
家境	Y	N	U
国家	Y	N	U
能力	Y	N	U

16. **成熟**（的）			
技术	Y	N	U
水果	Y	N	U
劳动	Y	N	U
经验	Y	N	U

17. **长远**（的）			
计划	Y	N	U
影响	Y	N	U
水平	Y	N	U
范围	Y	N	U

18. **灵活**（的）			
手指	Y	N	U
反应	Y	N	U
形式	Y	N	U
信息	Y	N	U

19. **痛苦**（的）			U
意见	Y	N	U
经历	Y	N	U
选择	Y	N	U
速度	Y	N	U

20. **真诚**（的）			U
利益	Y	N	U
祝福	Y	N	U
情形	Y	N	U
帮助	Y	N	U

21. **知名**（的）			U
人物	Y	N	U
作品	Y	N	U
表达	Y	N	U
结果	Y	N	U

22. **温暖**（的）			U
怀抱	Y	N	U
性格	Y	N	U
阳光	Y	N	U
态度	Y	N	U

23. **出色**（的）			U
表现	Y	N	U
人才	Y	N	U
要求	Y	N	U
作品	Y	N	U

24. **细致**（的）			U
经历	Y	N	U
服务	Y	N	U
毛病	Y	N	U
分析	Y	N	U

25. **自负**（的）			U
心情	Y	N	U
口吻	Y	N	U
神态	Y	N	U
道德	Y	N	U

26. **锐利**（的）			U
目光	Y	N	U
精神	Y	N	U
时机	Y	N	U
武器	Y	N	U

27. **漂亮**（的）			U
建筑	Y	N	U
竞争	Y	N	U
情感	Y	N	U
衣服	Y	N	U

28. **恰当**（的）			U
事实	Y	N	U
选择	Y	N	U
判断	Y	N	U
评价	Y	N	U

29. **牢固**（的）			U
计划	Y	N	U
身体	Y	N	U
联系	Y	N	U
观念	Y	N	U

30. **热忱**（的）			U
服务	Y	N	U
欢迎	Y	N	U
批评	Y	N	U
希望	Y	N	U

31. **舒畅**（的）			U
线条	Y	N	U
感觉	Y	N	U
气氛	Y	N	U
语言	Y	N	U

32. **详尽**（的）			U
资料	Y	N	U
概括	Y	N	U
介绍	Y	N	U
记录	Y	N	U

33. **充沛**（的）			U
精力	Y	N	U
进步	Y	N	U
资源	Y	N	U
赞扬	Y	N	U

34. **低廉**（的）			U
支持	Y	N	U
成本	Y	N	U
费用	Y	N	U
收入	Y	N	U

35. **公道**（的）			U
说明	Y	N	U
价格	Y	N	U
交易	Y	N	U
成绩	Y	N	U

36. **荒唐**（的）			U
观点	Y	N	U
行为	Y	N	U
成就	Y	N	U
解释	Y	N	U

37. **流畅**（的）			U
文字	Y	N	U
讨论	Y	N	U
动作	Y	N	U
声音	Y	N	U

38. **尴尬**（的）			U
局面	Y	N	U
道理	Y	N	U
势力	Y	N	U
滋味	Y	N	U

39. **疲惫**（的）			U
脚步	Y	N	U
身体	Y	N	U
品德	Y	N	U
道路	Y	N	U

40. **肤浅**（的）			U
眼光	Y	N	U
条件	Y	N	U
体会	Y	N	U
真理	Y	N	U

41. **悲伤**（的）			U
实际	Y	N	U
失败	Y	N	U
时刻	Y	N	U
情绪	Y	N	U

42. **孤寂**（的）			U
世界	Y	N	U
心灵	Y	N	U
事实	Y	N	U
概念	Y	N	U

43. **尖刻**（的）			U
批评	Y	N	U
思念	Y	N	U
语气	Y	N	U
记忆	Y	N	U

44. **凌乱**（的）			U
头发	Y	N	U
管理	Y	N	U
夜晚	Y	N	U
房间	Y	N	U

45. **圆滑**（的）			U
歌声	Y	N	U
色彩	Y	N	U
胜利	Y	N	U
手段	Y	N	U

APPENDIX B

DEMOGRAPHIC INFORMATION SHEET

Demographic Information Sheet

Directions: Please answer each of the questions below by either circling the corresponding letter or filling in the blanks. Feel free to give an approximation to the best of your knowledge when you are unsure of an answer. If a particular question makes you feel uncomfortable for any reason, you are not obliged to answer it. However, your response to each question is very important for our study and will be greatly appreciated.

1. Do any of your family members whom you grew up with speak Chinese as his or her native language? (**IF NO, SKIP QUESTION 2 AND GO TO QUESTION 3**.)
    a. Yes.
    b. No.
2. To what extent do you use Chinese to communicate with your family members?
    a. Never.
    b. Occasionally.
    c. Sometimes.
    d. Most of the time.
    e. Almost always.
3. For how many years have you studied Chinese formally as a foreign language in the classroom?
    a. Less than 2 years.
    b. 2-3 years.
    c. 4-5 years.
    d. More than 5 years.

4. While you study in the U.S., have you had any experience learning Chinese regularly outside class from a Chinese native speaker? If yes, please briefly describe your experience.
    a. Yes, and for ____years ____ months. I learned/used Chinese in the following activities/contexts:

       _____

       _____.
    b. No.

5. Have you ever lived/studied/worked abroad where Chinese is spoken as a native language? If yes, please briefly describe your in-country language-related experience.
    a. Yes, and for ____years ____ months. I learned/used Chinese in the following activities/contexts:

       _____

       _____.
    b. No.

6. Have you ever taken a standardized test in Chinese? If so, please provide your score and/or the highest level you have passed. If you cannot remember the exact score, give your best estimate.
    - AAPPL (The ACTFL Assessment of Performance Toward Proficiency in Languages) _____
    - BYU (Brigham Young University) Test _____
    - CATRC (Computerized Adaptive Test for Reading Chinese) _____
    - DLPT (Defense Language Proficiency Test) _____
    - HSK (Chinese Proficiency Test) _____
    - OPI (Oral Proficiency Interview) _____
    - TOP/TOCFL (Test of Chinese as a Foreign Language) _____

- Other

  _____

  _____

7. A follow-up Skype interview (no longer than 15 minutes) would be appreciated. If you are willing to do this, please PRINT your name and email below.

   Name: _____(First) _____ (Last)

   E-mail: _____

APPENDIX C

INTERVIEW QUESTIONS

Interview Questions (Semi-Structured)

Directions: You will be asked a few questions about the test you just took. If you are not comfortable with a particular question, let me know and you are not obliged to answer it. However, your response to each of these questions is very important to our study and will be greatly appreciated.

1. Did you have any difficulties understanding the test directions, i.e. how you were supposed to answer each question?

2. Did you feel that you had sufficient time completing the test? If you felt rushed, what do you think might have caused it (e.g., the length of the test, your test-taking strategies, etc.)?

3. Did you find yourself relying on guessing in answering some of these questions? If so, to what extent? Could you give me an example from each booklet?

4. probing questions on how they arrive at a response for a particular item

REFERENCES

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

American Council on the Teaching of Foreign Languages. (1982). ACTFL provisional proficiency guidelines. Yonkers, NY: ACTFL.

American Council on the Teaching of Foreign Languages. (1986). ACTFL proficiency guidelines. Yonkers, NY: ACTFL.

American Council on the Teaching of Foreign Languages. (2012). ACTFL proficiency guidelines. Yonkers, NY: ACTFL.

Bassetti, B. (2005). Effects of writing systems on second language awareness: Word awareness in English learners of Chinese as a Foreign Language. In V. J. Cook & B. Bassetti (Eds.), *Second language writing systems* (pp. 335-356). Clevedon, UK: Multilingual Matters.

Batty, A. (2012). Identifying dimensions of vocabulary knowledge in the Word Associates Test. *Vocabulary Learning and Instruction, 1*(1), 70-77.

Bengeleil, N. F., & Paribakht, T. S. (2004). L2 reading proficiency and lexical inferencing by university EFL learners. *The Canadian Modern Language Review, 61*(2), 225-249.

Bernhardt, E. B. (2011). *Understanding advanced second-language reading*. New York: Routledge.

Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen, (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). Cambridge: Cambridge University Press.

Cheng, R. (Ed.). (2010). *同义词大词典* [Synonym Dictionary]. Shanghai, China: Shanghai Lexicographic Publishing House.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Laurence Erlbaum Associates.

Cronbach, L. (1942). An analysis of techniques for diagnostic vocabulary testing. *The Journal of Educational Research, 36*(3), 206-217.

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly, 42*(3), 475-493.

Cummins, J. (1999). BICS and CALP: Clarifying the distinction. Retrieved from ERIC database. (ED438551)

Daller, H., Milton, J., & Treffers-Daller, J. (2007). Editors' introduction: Conventions, terminology and an overview of the book. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 1-32). Cambridge: Cambridge University Press.

Davidson, D. E. (2010). Study abroad: When, how long, and with what results? New data from Russian Front. *Foreign Language Annals, 43*(1), 6-26.

Everson, M. E. (1998). Word recognition among learners of Chinese as a foreign language: Investigating the relationship between naming and knowing. *The Modern Language Journal, 82*(2), 194-204.

Everson, M. E., & Ke, C. (1997). An inquiry into the reading strategies of intermediate and advanced learners of Chinese as a foreign language. *Journal of the Chinese Language Teachers Association, 32*(1), 1-20.

Fitzpatrick, T., & Izura, C. (2011). Word association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation. *Studies in Second Language Acquisition, 33*(3), 373-398.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographic depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 104-115.

Furman, N., Goldberg, D., & Lusin, N. (2010). *Enrollments in languages other than English in United States Institutions of higher education, Fall 2009*. Retrieved from http://www.mla.org/pdf/2009_enrollment_survey.pdf.

Greidanus, T., Beks, B., & Wakely, R. (2005). Testing the development of French word knowledge by advanced Dutch- and English-speaking learners and native speakers. *The Modern Language Journal, 89*(2), 221-233.

Greidanus, T., & Nienhuis, L. (2001). Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. *The Modern Language Journal, 85*(4), 567-577.

Haastrup, K., & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics, 10*(2), 221-240.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*(2), 303-317.

Ishii, T., & Schmitt, N. (2009). Developing an integrated diagnostic test of vocabulary size and depth. *RELC Journal, 40*(1), 5-22.

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 157-166.

Katz, L., & Frost, R. (1992). Reading in different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67-84). Amsterdam: Elsevier.

Ke, C. (1996). An empirical study on the relationship between Chinese character recognition and production. *The Modern Language Journal, 80*(3), 340-349.

Ke, C. (1998). Effects of strategies on the learning of Chinese characters among foreign language students. *Journal of the Chinese Language Teachers Association, 33*(2), 93-112.

Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing, 25*, 347-373.

Kondo-Brown, K. (2006). How do English L1 learners of advanced Japanese infer unknown Kanji words in authentic texts? *Language Learning, 56*(1), 109-153.

Kubler, C. C. (2002). Learning Chinese in China: Programs for developing superior- to distinguished-level Chinese language proficiency in China and Taiwan. In B. L. Leaver & B. Shekhtman (Eds.), *Developing professional-level language proficiency* (pp. 96-118). Cambridge: Cambridge University Press.

北京语言学院语言教学研究所 [Language Education and Research Center of Beijing Language Institute]. (1986). *现代汉语频率词典* [Modern Chinese frequency dictionary]. Beijing, China: 北京语言学院出版社 [Beijing Language Institute Press].

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing, 21*(2), 202-226.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15-30.

Lennon, P. (1996). Getting "easy" verbs wrong at the advanced level. *International Review of Applied Linguistics in Language Teaching, 34*(1), 23-36.

Leaver, B. L., & Shekhtman, B. (2002). Principles and practices in teaching Superior-level language skills: Not just more of the same. In B. L. Leaver & B. Shekhtman (Eds.), *Developing professional-level language proficiency* (pp. 3-33). Cambridge: Cambridge University Press.

Lee-Thompson, L. (2008). An investigation of reading strategies applied by American learners of Chinese as a foreign language. *Foreign Language Annals, 41*(4), 702-721.

Lessard-Clouston, M. (2006). Breadth and depth specialized vocabulary learning in theology among native and non-native English speakers. *The Canadian Modern Language Review, 63*(2), 175-198.

Liskin-Gasparro, J. E. (2003). The ACTFL Proficiency Guidelines and the Oral Proficiency Interview: A brief history and analysis of their survival. *Foreign Language Annals, 36*(4), 483-490.

Llanes, À. (2011). The many faces of study abroad: an update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism, 8*(3), 189-215.

Lü, C. (2010). The effects of word-knowledge depth, part of speech, and proficiency level on word association among learners of Chinese as a second language. In M. E. Everson & H. H. Shen (Eds.), *Research among learners of Chinese as a foreign language* (pp. 67-92). Honolulu, HI: National Foreign Language Resource Center.

Marzban, A., & Hadipour, R. (2012). Depth versus breadth of vocabulary knowledge: Assessing their roles in Iranian intermediate EFL students' lexical inferencing success through reading. *Procedia – Social and Behavioral Sciences, 46*, 5296-5300.

Maxim, H. (2009). An essay on the role of language in collegiate foreign language programmatic reform. *Teaching German, 42*(2), 123-129.

McBride-Chang, C., Shu, H., Zhou, A., Wat, C. P., & Wagner, R. K. (2003). Morphological awareness uniquely predicts young children's Chinese character recognition. *Journal of Educational Psychology, 95*(4), 743-751.

Meara, P. (1989). Matrix models of vocabulary acquisition. *AILA Review, 6*, 66-74.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35-53). Cambridge: Cambridge University Press.

Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*, 142-145.

Meara, P., & Milton, J. (2003). *X_Lex: The Swansea Vocabulary Levels Test*. Newbury: Express Publishing.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Clevedon, UK: Multilingual Matters.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindquist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 57-78). The European Second Language Association.

Mochizuki, M. (2012). Four empirical vocabulary test studies in the three dimension framework. *Vocabulary Learning and Instruction, 1*(1), 44-52.

Mori, Y. (2003). The roles of context and word morphology in learning new *kanji* words. *The Modern Language Journal, 87*(3), 404-420.

Namei, S. (2004). Bilingual lexical development: A Persian-Swedish word association study. *International Journal of Applied Linguistics, 14*(3), 363-388.

Nassaji, H. (2004). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *Canadian Modern Language Review, 61*, 107-134.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury house.

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

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin, 84*(1), 93-116.

Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics, 16*(3), 389-408.

Ordóñez, C. L., Carlo, M. S., Snow, C. E., & McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology, 94*(4), 719-728.

Packard, J. L. (2000). *Morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.

Paribakht, T. S., & Wesche, M. B. (1993a). The relationship between reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11*(1), 9-29.

Paribakht, T. S., & Wesche, M. B. (1993b). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady and T Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174-200). Cambridge: Cambridge University Press.

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(1), 101-118.

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a "universal" phonological principle. In R. Frost & L. Katz. (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 227-248). North-Holland: Elsevier.

Qian, D. D. (1998). *Depth of vocabulary knowledge: Assessing its role in adults' reading comprehension in English as a second language* (Unpublished doctoral dissertation). University of Toronto, Toronto, Ontario, Canada.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*(2), 282-308.

Qian, D. D. (2000). *Validating the role of depth of vocabulary knowledge in assessing reading for basic comprehension tasks in TOEFL 2000*. Princeton, NJ: Educational Testing Service.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536.

Qian, D. D. (2005). Demystifying lexical inferencing: The role of aspects of vocabulary knowledge. *TESL Canada Journal*, 22(2), 34-54.

Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21*(1), 28-52.

Read, J. (1987, August). *Towards a deeper assessment of vocabulary knowledge*. Paper presented at the Annual Meeting of the International Association of Applied Linguistic, Sydney, New South Wales, Australia. Retrieved from ERIC database. (ED301048)

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*(3), 355-371.

Read, J. (1995). Refining the word associates format as a measure of depth of vocabulary knowledge. *New Zealand Studies in Applied Linguistics, 1*, 1-17.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. J. Kunnan (Ed.), *Validation in language assessment: Selected papers from the17th Language* (pp. 41-60). Mahwah, NJ: Lawrence Erlbaum.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146-161.

Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing, 18*(1), 1-32.

Richards, J. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10*(1), 77-89.

Shen, H. H. (2000). Radical knowledge and character learning among learners of Chinese as a foreign language. *Linguistic Studies 语言研究, June*, 85-93.

Shen, H. H. (2005). An investigation of Chinese character learning strategies among non-native learners of Chinese. *System, 33*, 49-68.

Shen, H. H. (2008). An analysis of word decision strategies among learners of Chinese. *Foreign Language Annals, 41*(3), 501-524.

Shen, H. H. (2009). Size and strength: Written vocabulary acquisition among advanced learners. *Chinese Teaching in the World 世界汉语教学, 23*(1), 74-85.

Shen, H. H. (2010). Analysis of radical knowledge development among beginning CFL learners. In M. E. Everson & H. H. Shen (Eds.), *Research among learners of Chinese as a foreign language (Chinese Language Teachers Association Monograph Series, Vol. IV*) (pp. 45– 65). Honolulu, HI: University of Hawaii.

Shen, H. H., & Ke, C. (2007). Radical awareness and word acquisition among nonnative learners of Chinese. *The Modern Language Journal, 91*(1), 97-111.

Shu, H., & Anderson, R. C. (1999). Learning to read Chinese: The development of metalinguistic awareness. In J. Wang, A. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 1-18). Mahwah, NJ: Lawrence Erlbaum.

Spring, M. K. (2012). Languages for specific purposes curriculum in the context of Chinese-language Flagship programs. *The Modern Language Journal, 96,* Focus Issue, 140-157.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577-607.

Swender, E. (2003). Oral proficiency testing in the real world: Answers to frequently asked answers. *Foreign Language Annals, 36*(4), 520-526.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*, 217-234.

Walton, A. R. (1992). *Expanding the vision of foreign language education: Enter the less commonly taught languages*. Washington, DC: National Foreign Language Center. Retrieved from ERIC database. (ED345539)

Wang, M., & Yang, C. (2008). Learning to read Chinese: Cognitive consequences of cross-language and writing system differences. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages* (pp. 125-153). New York, NY: Routledge.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review, 53*(1), 13-40.

Wilkins, D. (1972). *Linguistics in language teaching*. London: Arnold.

Williams, C. (2013). Emerging development of semantic and phonological routes to character decoding in Chinese as a foreign language learners. *Reading and Writing, 26,* 293-315.

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition, 23*(1), 41-69.

Wolter, B. (2005). *V_Links: A new approach to assessing depth of word knowledge* (Unpublished doctoral dissertation). University of Wales Swansea.

Zhou, X., & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*(5), 1245-1265.

Zhu, Y. P. (1987). *Analysis of cueing functions of the phonetic in modern Chinese*. Unpublished manuscript.