Iowa Research Online
The University of Iowa's Institutional Repository

University of Iowa
Iowa Research Online

Theses and Dissertations

Spring 2010

Developing orthographic awareness among beginning Chinese language learners: investigating the influence of beginning level textbooks

Hui-Mei Fan
University of Iowa

Copyright 2010 Hui-Mei Fan

This dissertation is available at Iowa Research Online: http://ir.uiowa.edu/etd/496

Recommended Citation

Fan, Hui-Mei. "Developing orthographic awareness among beginning Chinese language learners: investigating the influence of beginning level textbooks." PhD (Doctor of Philosophy) thesis, University of Iowa, 2010.
http://ir.uiowa.edu/etd/496.

Follow this and additional works at: http://ir.uiowa.edu/etd

Part of the Teacher Education and Professional Development Commons

DEVELOPING ORTHOGRAPHIC AWARENESS AMONG BEGINNING CHINESE LANGUAGE LEARNERS: INVESTIGATING THE INFLUENCE OF BEGINNING LEVEL TEXTBOOKS

by

Hui-Mei Fan

An Abstract

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

May 2010

Thesis Supervisor: Associate Professor Michael E. Everson

ABSTRACT

The present study is based on the theoretical assumptions that frequency of characters and their structural components, as well as the frequency types of structural components, are important to enable learners of Chinese as a foreign language (CFL) to discover the underlying structure of Chinese characters. In the CFL context, since reliable target language input is limited largely to textbook materials and teacher instruction, it is important to more rigorously examine the inventory of Chinese characters that is typically presented in CFL textbooks.

The purpose of this study was to systematically describe and classify Chinese characters from ten CFL textbooks designed for college and adult beginning learners. The main focus was to compare the textbooks in the following areas: explicit orthographic decomposition instruction, character frequency selection, radical combination frequency, radical semantic transparency, radical positional regularity among different character graphic structures, phonetic element reliability, and phonetic component combination frequency. To accomplish the analysis required for this study, a special character database was created. Dictionaries were used to classify character characteristics, and documented frequency lists were used to classify the character usage frequency.

The findings revealed that most textbooks rarely include explicit orthographic decomposition instruction in the vocabulary lists or lessons, while over 40% of the characters in most of the textbooks did not combine with other characters to form words. In addition, analysis of frequency lists created over time revealed that the ten textbooks generally contain many high frequency characters. Furthermore, the results indicated that 60% of the characters in the database were classified by relatively few radicals (17%), with most radicals appearing on the left side of the characters. Relatively reliable phonetic analogy groups appeared frequently. About half of the characters are semantic-phonetic compound characters. Less than 10% of the characters in the textbooks are semantic-phonetic compound characters that contain semantically transparent radicals and reliable phonetic elements.

The results of the study suggest that textbook writers should consider integrating orthographic decomposition and component frequency materials into their textbooks, as systematic instruction in textbooks is generally lacking. Teachers should also be mindful of emphasizing the high frequency characters that are consistently featured in all textbooks, as well as the frequently appearing radicals and left-right internal structure of many of the characters. In this way, students will early on develop a firm foundation of the principles governing Chinese orthography.

Abstract Approved: ____________________________________
Thesis Supervisor

____
Title and Department

____
Date

DEVELOPING ORTHOGRAPHIC AWARENESS AMONG BEGINNING CHINESE LANGUAGE LEARNERS: INVESTIGATING THE INFLUENCE OF BEGINNING LEVEL TEXTBOOKS

by

Hui-Mei Fan

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

May 2010

Thesis Supervisor: Associate Professor Michael E. Everson

Copyright by

HUI-MEI FAN

2010

All Rights Reserved

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

Hui-Mei Fan

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy degree
in Teaching and Learning (Foreign Language and ESL Education)
at the May 2010 graduation.

Thesis committee: ______________________________
Michael E. Everson, Thesis Supervisor

____
Leslie L. Schrier

____
Helen H. Shen

____
Walter P. Vispoel

____
Lia Plakans

To my loved ones:
獻給我親愛的

Aunt	姑姑	范秀英 女士
Father	父親	范高橋 先生
Mother	母親	江春妹 女士

ACKNOWLEDGMENTS

So many people helped me write this dissertation that it is hard to know where to start my acknowledgments. Words cannot express how I am deeply grateful to my dissertation committee chair, Dr. Michael Everson for his guidance and encouragement during my study at Iowa and the dissertation process. He guided me to the Chinese as a Foreign Language (CFL) education world and helped me gain deep insights into Chinese orthography and CFL teaching and learning. He also showed me a great role-model of caring. I would not have been able to achieve the end of the Ph.D. process without him. I would also like to take this opportunity to express my sincere appreciation to my committee members, Dr. Leslie Schrier, Dr. Helen Shen, Dr. Walter Vispoel, and Dr. Lia Plakans. I thank them for providing me valuable feedback and encouragement. I would also like to remember Dr. Kathy Heilenman who was originally on my committee, and was so supportive of my original proposal. Her untimely passing left a void in all the lives of her students who were lucky enough to have had her as their professor.

My life in the United States would not have been possible without the support from my dearest aunt Hiuying, Dr. Peter Fries and Nan Fries. My dearest aunt saw my potential and encouraged me to start a new life in the United States. I am also indebted to Dr. Peter Fries and Nan Fries for the great love and guidance I receive from them. Their wisdom of life has had a great influence on my personal life.

Many thanks also to Dr. Leslie Grant, Dr. Ken Goodman, and Dr. Yeta Goodman for academic advice. I am thankful to the faculty members I worked for at the Education Technology Center, Dr. John Achrazoglou, Dr. Carolyn Wanat, and Dr. Larry Bartlett. I thank them for providing me opportunities and support to develop my professional skills working with professors and students.

In addition, special thanks to my friends and fellow graduate students, in particular: Chia-Hui Chiu, Hsinyen Yang, Fatima Baig, Chiemi Hanzawa, Yanrong Kang, Li Gu, David Byrd, Anita Jayachandran, Hsiaotung Lin, Lara Ross, and Dr. Roger Williamson. I thank them for their warm friendship and help.

Finally, I would like to thank my family for their unconditional support and love: to my parents, Gao-qiao Fan and Chun-mei Jiang; to my brothers, Senrong and Junyu; to my nieces and nephews, Kaiyun, Kaiwi, Kaijun, Yijun, and Wenlin; to my sister-in-law, Guichun and Tingxuan.

ABSTRACT

The present study is based on the theoretical assumptions that frequency of characters and their structural components, as well as the frequency types of structural components, are important to enable learners of Chinese as a foreign language (CFL) to discover the underlying structure of Chinese characters. In the CFL context, since reliable target language input is limited largely to textbook materials and teacher instruction, it is important to more rigorously examine the inventory of Chinese characters that is typically presented in CFL textbooks.

The purpose of this study was to systematically describe and classify Chinese characters from ten CFL textbooks designed for college and adult beginning learners. The main focus was to compare the textbooks in the following areas: explicit orthographic decomposition instruction, character frequency selection, radical combination frequency, radical semantic transparency, radical positional regularity among different character graphic structures, phonetic element reliability, and phonetic component combination frequency. To accomplish the analysis required for this study, a special character database was created. Dictionaries were used to classify character characteristics, and documented frequency lists were used to classify the character usage frequency.

The findings revealed that most textbooks rarely include explicit orthographic decomposition instruction in the vocabulary lists or lessons, while over 40% of the characters in most of the textbooks did not combine with other characters to form words. In addition, analysis of frequency lists created over time revealed that the ten textbooks generally contain many high frequency characters. Furthermore, the results indicated that 60% of the characters in the database were classified by relatively few radicals (17%), with most radicals appearing on the left side of the characters. Relatively reliable phonetic analogy groups appeared frequently. About half of the characters are semantic-phonetic compound characters. Less than 10% of the characters in the textbooks are semantic-phonetic compound characters that contain semantically transparent radicals and reliable phonetic elements.

The results of the study suggest that textbook writers should consider integrating orthographic decomposition and component frequency materials into their textbooks, as systematic instruction in textbooks is generally lacking. Teachers should also be mindful of emphasizing the high frequency characters that are consistently featured in all textbooks, as well as the frequently appearing radicals and left-right internal structure of many of the characters. In this way, students will early on develop a firm foundation of the principles governing Chinese orthography.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xii

CHAPTER 1 INTRODUCTION 1

1.1 Chinese Orthography 2
  1.1.1 Character Graphic Structure 5
  1.1.2 Radical Combination and Semantic Transparency 7
  1.1.3 Radical Positional Regularity among Character Graphic Structure 8
  1.1.4 Phonetic Element Reliability in Pronouncing Chinese Characters 9
  1.1.5 Phonetic Component Combination 11
1.2 A CFL Model of Orthographic Awareness 12
1.3 Chinese Character Selection in CFL Textbooks over Time 13
1.4 Explicit Orthographic Awareness Instruction 15
1.5 Purpose of the Study and Research Questions 17
1.6 Significance of the Study 18
1.7 Limitations of the Study 19
1.8 Definition of Terms 19

CHAPTER 2 LITERATURE REVIEW 21

2.1 Introduction 21
2.2 Word Recognition in Reading Comprehension 22
2.3 A Dual-Route Model of Reading Aloud in English 25
2.4 Beginning Orthographic Awareness Development in L1 Chinese 27
  2.4.1 Learning to Read in L1 Chinese 27
  2.4.2 The Interactive Constituency Model of Chinese Character Identification 31
2.5 Beginning Orthographic Awareness Development in CFL 33
  2.5.1 A CFL Model of Orthographic Awareness 35
  2.5.2 Studies on L2 Beginning Chinese Orthographic Awareness Development 37
2.6 Character Types for Orthographic Knowledge Research Purpose 40
  2.6.1 Chinese Character Decision Task 40
  2.6.2 Semantic Categorization Task 44
  2.6.3 Semantic Radical Judgment Picture-Matching Task 45
  2.6.4 Chinese Phonetic Reading Task 46
  2.6.5 Chinese Pseudo-Character Reading Task 46
  2.6.6 Chinese Pseudo-Character Construction Task 47
  2.6.7 Two-Character Word Reading Error Analysis 48
  2.6.8 Character Spelling Error Analysis 49
2.7 Chinese Character Selection in CFL Textbooks 49
2.8 Chinese Character Selection for Teaching and Learning Purposes 52
  2.8.1 Character Component Segment 52
  2.8.2 Character Component Analysis and Pedagogical Components 56

CHAPTER 3 RESEARCH METHODS 61

3.1 Textbook Selection 61
3.2 Dictionary and Frequency List Selection 65
3.3 Word Entry and Character Entry 68
3.4 Investigating Explicit Orthographic Decomposition Instruction Availability 71
3.5 Investigating Character Diversity and Repetition 72

3.6 Investigating Character Frequency Rank Selection 73
3.6.1 Investigating Character Frequency Rank Classified by Frequency Lists 75
3.6.2 Investigating Character Frequency Rank as Classified by the Yahoo and Google Search Counts 75
3.6.3 Investigating Character Frequency Rank as Classified by the HSK Proficiency Level List 76
3.6.4 Investigating Character Frequency Rank for Characters Appearing in All Ten Textbooks 76
3.7 Investigating Radical Diversity and Repetition 77
3.7.1 Investigating Radical Combination Frequency in Textbooks 79
3.7.2 Investigating Radical Regularity among Character Graphic Structure 81
3.8 Investigating Phonetic Component Diversity and Repetition 86
3.9 Investigating Ideal Semantic-Phonetic Compound Characters 88
3.9.1 Investigating the Classification of Semantic-Phonetic Compound Characters 89
3.9.2 Investigating the Classification of Reliable Phonetic Element Characters 89
3.9.3 Investigation of the Classification of Semantic Transparent Radical Characters 90
3.9.4 Investigation of the Classification of Ideal Semantic-Phonetic Compound Characters 93
3.10 Summary of Character Characteristics 93

CHAPTER 4 RESEARCH RESULTS 95

4.1 Results Investigating Explicit Orthographic Decomposition Instruction in the Textbooks 96
4.2 Results of Investigating Character Diversity and Repetition Comparison in the Textbooks 102
4.2.1 Descriptions of Characters and Words in the Ten Textbooks 103
4.2.2 Results Investigating Character Combination Frequency in Textbooks 105
4.3 Results of Investigating Textbook Character Frequency Rank Selection 111
4.3.1 Results of Investigating Character Frequency Rank as Classified by Yahoo and Google Search Counts 117
4.3.2 Results of Investigating Character Frequency Rank Classified by HSK Proficiency Level List 121
4.3.3 Results of Investigating Character Frequency Rank for Characters Appearing in All Ten Textbooks 122
4.3.4 Result Summary of Character Frequency Rank Selection in the Textbooks 125
4.4 Results of Character Radical Diversity and Repetition in the Textbooks 126
4.4.1 Investigating Radical Combination Frequency in the Textbooks 126
4.4.2 Results of Investigating the Most Commonly Appearing Character Graphic Structures in the Textbooks 131
4.4.3 Radical Regularity among Character Graphic Structures 135
4.4.4 Results Summary of Character Radical Diversity and Repetition in the Textbooks 138
4.5 Results of Phonetic Component Diversity and Repetition in the Textbooks 139
4.6 Results of Ideal Semantic-Phonetic Compound Characters in the Textbooks 143
4.6.1 Reliable Phonetic Element Characters in the Textbooks 145
4.6.2 Semantic Transparent Radical Characters in the Textbooks 147
4.6.3 Ideal Semantic-Phonetic Compound Characters in the Textbooks 149

CHAPTER 5 DISCUSSION 153

5.1 Overview of the Study 153
5.2 Summary of Primary Findings 154
5.3 Discussion 155
5.3.1 Integrating Orthographic Decomposition Materials into the Textbooks 155

        - 5.3.2 Becoming Familiar with the Processes of Chinese Character Combination 158
        - 5.3.3 Instructional Focus on High Frequency Characters and Low Frequency Characters 160
        - 5.3.4 High Combination Frequency Radicals Poorly Represented in Textbooks......... 162
        - 5.3.5 Identifying Radicals through Radical Position Regularity 163
        - 5.3.6 Using a Phonetic Component Analogy Strategy 165
        - 5.3.7 Ideal Semantically Transparent and Phonetically Reliable Characters in the Textbooks 166
    - 5.4 Limitations of the Study 169
    - 5.5 Recommendations for Future Research 170
        - 5.3.8 The Usefulness of the Chinese Orthography Introduction in the Textbooks....... 170
        - 5.3.9 Terminology of Components in Orthographic Decomposition Instruction......... 171
        - 5.3.10 Teachers' Knowledge About Language in Chinese Orthographic Awareness.. 173
    - 5.6 Recommendations for Textbook Writers 175
    - 5.7 Recommendations for Classroom Instructors 177
    - 5.8 Recommendations for Teacher Educators 179

APPENDIX A LIST OF CHARACTER FREQUENCY RANK APPEARED IN THE TEXTBOOKS IN WENLIN 180

APPENDIX B FREQUENCY DISTRIBUTION OF CHARACTER FREQUENCY RANK ACROSS TEN TEXTBOOKS 184

APPENDIX C CHARACTERS IN FIVE CHARACTER GRAPHIC STRUCTURES 185

APPENDIX D PHONETIC COMPONENT ANALOGY TYPES FOR EACH GROUP . 187

REFERENCES 196

LIST OF TABLES

Table 2.1 Component Lists in Fei (1996) and Xiao (1995) 55

Table 2.2 Pedagogical Phonetic and Semantic Components in Three Component Studie 59

Table 3.1 Screenshots of Character 你 in Four Vocabulary Lists 69

Table 3.2 Character and Word Entry in the Ten Textbooks 70

Table 3.3 Worksheet Example of Frequency Rank on the Chinese Character Frequency 74

Table 3.4 List of 214 Chinese Character Radical Meanings and Radical Categories .. 78

Table 3.5 Worksheet Examples of Character Radical Classification 80

Table 3.6 Non-character Component and Stroke List 83

Table 3.7 Character Component and Stroke List 83

Table 3.8 Classifications of Character Graphic Structure and Radical Position 85

Table 3.9 Classifications of Character Phonetic Component Reliability 87

Table 3.10 Classifications of Character Radical Semantic Transparency 92

Table 4.1 Screenshots of Character 你 in the Ten Vocabulary Lists 99

Table 4.2 Summary of Orthographic Decomposition Information in the Ten Textbooks 101

Table 4.3 Summary of Character and Word Distribution in the Vocabulary Lists 104

Table 4.4 Summary of Characters Combing to Form Different Words Across Textbook 106

Table 4.5 List of Characters that Appeared Only in One Particular Textbook 108

Table 4.6 Summary of First 20 High Character Combination Frequency Characters 110

Table 4.7 Descriptive Statistic Summary of Character Frequency Rank 114

Table 4.8 Descriptive Statistic of Character Frequency Rank in Yahoo and Google Search 118

Table 4.9 Descriptive Statistics of Characters Classified by HSK Level List 121

Table 4.10 Descriptive Statistics for Characters Classified by Six Frequency Lists 123

Table 4.11 Frequency Rank of 108 Core Characters Appeared in All Ten Textbooks 124

Table 4.12 List of High Combination Radicals in 中文字譜 and Ten Textbooks 130

Table 4.13 Frequency Distribution of Character Graphic Structures across Ten Textbooks 132

Table 4.14 Frequency Distribution of Characters in Five Character Graphic Structures 135

Table 4.15 Homophone Analogy Characters and Shared Components........................ 141

Table 4.16 List of Ideal Semantic-phonetic Compound Characters in the Textbooks . 151

LIST OF FIGURES

Figure 1.1 Stratified Analysis and Plane Analysis of Chinese Character Structure 6

Figure 1.2 Examples of Characters Graphic Structure.. 7

Figure 1.3 Examples of Characters Employing Different Radical Semantic Transparency.. 7

Figure 1.4 Examples of characters employing different radical positions...................... 9

Figure 1.5 Examples of Characters Employing the Same Phonetic Element "青 qīng".. 10

Figure 1.6 Examples of Words Which Mean "China" in Different Instruction Methods 15

Figure 2.1 Examples of Words Which Mean "China" in Different Instruction Methods 28

Figure 2.2 The Interactive Constituency Model of Chinese Character Identification .. 31

Figure 2.3 Ke's (1996) Stage Model for the Development of Orthographic Awareness among CFL Readers.. 36

Figure 3.1 An Example of Character Etymological Tree (Chart) and Character Entry in 中文字譜 Chinese Characters: A Genealogy and Dictionary 66

Figure 3.2 Character Graphic Structures and Radical Positions.................................. 84

Figure 4.1 A Screenshot of the Vocabulary List in BCR1977 96

Figure 4.2 Examples of Radical Activity in CICrw1994 98

Figure 4.3 An Example of Shared Phonetic Component Activity in CICrw1994......... 98

Figure 4.4 Screenshots of Chinese Character Decomposition in NPCR2002 100

Figure 4.5 Times Characters Combined to Form Different Words Across Textbooks 107

Figure 4.6 Trends of Most Frequently Occurring Characters in All Ten Textbooks... 112

Figure 4.7 Rank Distribution Differences for the Most Frequently Used Characters in Ten Selected Textbooks.. 114

Figure 4.8 Distribution of Most Frequently Used Characters in the Textbooks.......... 116

Figure 4.9 Trends of Cumulated Frequency Rank in Yahoo and Google Search Lists 117

Figure 4.10 Percentage Distribution of Yahoo and Google Search Frequency Across the Ten Selected Textbooks.. 120

Figure 4.11 Frequency of the Top 100 Frequency Rank Character in Yahoo and Google Search 120

Figure 4.12 Frequency Distribution of Characters Classified by HSK Levels 122

Figure 4.13 Distribution of Radical Combinations in the Database 127

Figure 4.14 Distribution of Radical Combination Frequency Across the Ten Textbooks 128

Figure 4.15 Frequency Distribution of Radical Combination 129

Figure 4.16 Percentage Distribution of Character Graphic Structure in the Ten Textbooks 132

Figure 4.17 Percentage Distribution of Left-right Graphic Structure in the Ten Textbooks 133

Figure 4.18 Percentage Distribution of Top-bottom Graphic Structure in the Ten Textbooks 134

Figure 4.19 Percentage Distribution of Half-enclosure Graphic Structure in the Ten Textbooks 134

Figure 4.20 Proportion Distribution of Radical Position in the Left-right Graphic Structure 136

Figure 4.21 Proportion Distribution of Radical Position in Top-bottom Graphic Structure 137

Figure 4.22 Proportion Distribution of Radical Position in the Half-enclosure Graphic Structure 137

Figure 4.23 Frequency Distribution of Phonetic Component Analogy Types for Each Group 140

Figure 4.24 Frequency Distribution of Phonetic Component Analogy Types for Each Group 142

Figure 4.25 Proportion Trend of Characters in The Six Book 144

Figure 4.26 Phonetic Component Reliability Comparison in the Database 145

Figure 4.27 Phonetic Component Reliability within Semantic-phonetic Compound Characters 146

Figure 4.28 Proportion Distribution of Radical Semantic Transparency across Ten Textbooks 148

Figure 4.29 Proportion Distribution of Number of Radical Meanings across Ten Textbooks 149

Figure 4.30 Proportion Distribution of Ideal Semantic-phonetic Compound Characters 150

Figure 5.1 Radical Regularity among Character Graphic Structures........................... 164

Figure 5.2 Defining Terminology of the Component within a Character.................... 172

CHAPTER 1

INTRODUCTION

Foreign languages that are linguistically unrelated to English are more challenging for native-English speakers to learn. According to the United States Foreign Service Institute (FSI) Scale, to reach basic professional reading competence in most common western European languages requires about 520 classroom hours. However, to reach the same level of reading competence in Chinese requires about 2,400 classroom hours (Kane, 2006). While there are many reasons for this difference, the non-alphabetic nature of the Chinese writing system is certainly one of the primary factors that makes learning Chinese such a time-intensive process. Because learning to read requires "becoming aware of the basic units of spoken language, the basic writing system, and the mapping between the two (Shu & Anderson, 1999, p. 1)", the experience of learning to read in languages employing alphabets is different from learning to read in non-alphabetic orthographies such as Chinese. Perhaps the most critical aspect of the non-alphabetic Chinese orthography for English-speaking foreign language learners is that it does not phonologically reflect the spoken language as it does in western alphabetic languages. In other words, learners cannot "sound out" the pronunciations of Chinese characters as characters are not composed of letters, and any phonetic components Chinese characters may contain to hint at a character's pronunciation are irregular and unsystematic.

Current research indicates that both students and teachers of Chinese as a foreign language (CFL) believe that character learning and writing are the most difficult tasks in learning Chinese at the college level (Everson, 1998; Ke, Wen, & Kotenbeutel, 2001). As previously stated, Chinese employs what is known as a logographic system of writing, where each character represents a word or morpheme. What Chinese characters lack are letter-sound correspondence which would help CFL learners to more easily access the sound and pronunciation of characters, thus enabling learners to make sense of unfamiliar or new words. To develop proficiency in reading Chinese, researchers have put forth the theory that native Chinese and CFL readers develop Chinese orthographic awareness to infer meaning and pronunciation of Chinese characters through repeated exposure to print and explicit orthographic instruction (Jackson, Everson, & Ke, 2003; Ho & Bryant, 1997; Ho, Yau, & Au, 2003; Li, Fu, & Lin, 2000; Shen, 2004, 2005; Shen & Ke, 2007). This awareness is reflected in a learner's ability to identify, analyze, and infer the sound and meaning of Chinese characters through analysis of their internal structural components.

To investigate the development of orthographic analysis, it is common to conduct a component analysis of Chinese characters which provides us with structural knowledge about Chinese characters. This in turn helps material developers, test developers, and teachers to select and adjust the characters that are to be learned, taught, and tested. A component analysis is widely used in modern sinographemics studies. The results of modern sinographemics provide basic statistics about use, compatibility, frequency, subgraphemes, construction, inventory and systematization of Chinese characters (Guder, 2007). Most of the component analysis studies have examined Chinese characters from dictionaries or corpus databases (Chen, 1997; Fu, 1989; Guder-Manitius, 1999; Zhu, 1993). Rarely, however, have studies investigated characters from beginning level textbooks (Everson & Fan, 2008). In the CFL context, since reliable target language input is limited largely to textbook materials and teacher instruction, there is a need to know which characters are typically presented in beginning level CFL textbooks so as to better understand whether or not CFL learners have opportunities to develop orthographic awareness. From such investigations, we will be able to build better models of how CFL learners develop orthographic awareness.

1.1 Chinese Orthography

The Chinese language is comprised of a wide range of dialects. The term 'Chinese' refers to a language containing a number of mutually unintelligible dialects, though a variation of the Beijing dialect, also known as Mandarin Chinese and Putonghua, is the language taught in Chinese schools and the preferred form of CFL learned in American schools. However, all the different varieties of the spoken Chinese dialects share a common writing system. Since the 1950s, the Mainland China government decided not to use traditional characters and came up with simplified characters (e.g., 国 *guó*, country), which were derived from traditional ones (e.g., 國 *guó*, country) and based on specific orthographic principles (Cheung & Ng, 2003). Many characters, however, were retained in their traditional forms.

In Chinese orthography, characters are not symbols that randomly combine different components. The Han Dynasty scholar 許慎 (*xǔ shèn*) wrote the etymological dictionary 說文解字 (*shuōwén jiězìto*) that explains the underlying logic of each character. His classification states that characters are differentiated into six types 六書 (*liùshū*, Six Book): pictographs 象形 (*xiàngxíng*), ideographs 指事 (*zhǐshì*), logical aggregates 會意 (*huìyì*), phonetic complex 形聲 (*xíngshēng*), associative transformations 轉注 (*zhuǎnzhù*), and borrowings 假借 (*jiǎjiè*) (Harbaugh, 1998). Pictographs 象形, for example, portraying objects, e.g., 火 (*huǒ*, fire) is a pictograph depicting rising flames, while ideographs 指事 suggest abstractions, e.g., a horizontal line 一 with a topped line which suggests 上 (*shàng*, top or up). These pictographs and ideographs combine to create logical aggregates 會意 and phonetic complexes 形聲. Logical aggregates combine the meaning of different characters to create a new meaning. Termed phonetic complexes by Rick Harbaugh, and semantic-phonetic compounds by John DeFrancis, these characters combine the meaning of one character 形旁 (*xíngpáng*, semantic component) with the sound of another 聲旁 (*shēngpáng*, phonetic element). For example, 清 (*qīng*, clear) is a semantic-phonetic compound character with the 氵 (*shuǐ*, water) semantic component, which provides a cue to the meaning of the character; and 青 (*qīng*) as the phonetic element, which provides a cue to the pronunciation of the character. Taylor and Taylor (1995) have estimated that about 80~90% of the characters are semantic-phonetic compound characters in modern Chinese. The final two types of characters represent transformations in the meanings of the first four types. Associative transformations 轉注 extend the meaning of a character to a related concept, as in the example of 父 (*fù*, father) and 爸 (*bà*, dad), a pair of characters that illustrate associative transformation. Borrowings 假借 give an unrelated meaning to a character, generally that of a spoken word which has the same pronunciation as the borrowed character but lacks its own character, e.g., the original meaning of 北 (*běi*) is "people turned back to back". 北 was borrowed to represent "north" because north is back to the sun. The etymological explanations of characters help learners to understand, appreciate and remember characters.

In most CFL textbooks, however, characters are introduced as integral or compound characters. An integral character is a single component 部件 (*bùjiàn*) which is a combination of strokes that cannot be decomposed, such as 中 (*zhōng*, middle). A compound character, on the other hand, can be further decomposed into different components; for example, 植 (*zhí*, plant) is a compound character which can be further decomposed into two components 木 and 直. A component analysis can be used to analyze the internal structure of modern Chinese characters although many scholars have debated the definition of "component"部件 (*bùjiàn*) (Fei, 1996; Fu, 1992, 1993; Guder-Manitius, 1999; Su, 1995; Xiao, 1993, 1994, 1995). According to Xiao (1994), four component structures are defined:

1. Strokes which go across each other: for example, the character 十 is composed of the stroke 一 and stroke 丨, that cross each other. Another example is the non-character 艹; composed of strokes 一, 丨, and 丨, that cross each other. Other examples include, 力, 丈, 女, and 扌. Components can be characters or non-characters.
2. Some strokes cross each other and some strokes are connected: for example, the character 千 is composed of strokes 丿, 一, and 丨; 丿 is connected with

十. 千 cannot be decomposed into 丿 and 十 because 丿 is a stroke which cannot stand alone as a component. The main rule is that a character can be decomposed into components, and components can be further decomposed into strokes.

3. All strokes are connected, and the components either have spaces within a character or appear in more than two characters: for example, character 竹 is easily decomposed into the two left and right components because there is space between them. Another example, 丆 is composed of two strokes 一 and 丿, and they are connected. In addition, 丆 can be found in more than two characters 而, 頁, and 百, so 丆 is a component.
4. All strokes are separated, but they always group together to form characters: for example, 氵, 刂, 巛, 小, and 心.

A component analysis of Chinese characters can provide us with structural knowledge about Chinese characters. This in turn can help material developers, test developers, and teachers to select and adjust the characters that are to be learned, taught, and tested. In the following section, five aspects of the structure of Chinese characters, which are the focus of the current study, are introduced.

1.1.1 Character Graphic Structure

As stated above, characters can be classified into integral and compound characters, and compound characters can be further classified into components 部件 (*bùjiàn*). Two methods are used to analyze character structure: one is stratified analysis 層次分析法 (*céngcì fēnxifǎ*); the other is plane analysis 平面分析法 (*píngmiàn fēnxifǎ*). In stratified analysis, a character is decomposed layer by layer until all components are minalcomponents 末級部件 (*mòjí bùjiàn*) which cannot be further decomposed. For example, the character 部 (*bù*) shown in Figure 1.1 is first decomposed into the left part 音 and the right part 阝, then 音 is further decomposed into 立 and 口. This analysis indicates that the character 部 is composed of three components 立, 口, and 阝, and that it is a left-right structure character. On the other hand, in the plane analysis, a character is decomposed in one step. For example, the character 部 is decomposed into three components 立, 口, and 阝 which is presented in the structure graphic in the Figure 1.1. Using plane analysis, more character structures (85) are defined, compared to 13 structures using stratified analysis (Fu, 1991).

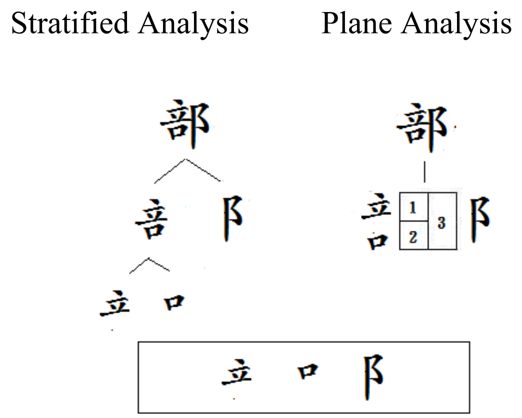


Figure 1.1 Stratified Analysis and Plane Analysis of Chinese Character Structure

In terms of component structure, Xiao (1993) defined three types of characters; a character composed of strokes only (i.e., integral characters), a character composed of components only (i.e., compound characters), and a character composed of both strokes and components (i.e., compound characters). Traditionally, in CFL instruction, compound characters have four graphic structures, which are based on rough stratified analysis. Therefore, five types of character graphic structure exist. Figure 1.2 presents these character structures with the illustration of structure graphics:

I = Integral characters, such as example 1 木

LR = Left-Right Structure, such as example 2 他 and example 3 吃

TB = Top-Bottom Structure, such as example 4 思 and example 5 望

HE = Half-Enclosure Structure, such as example 6 店 and example 7 區

E = Enclosure Structure, such as example 8 國

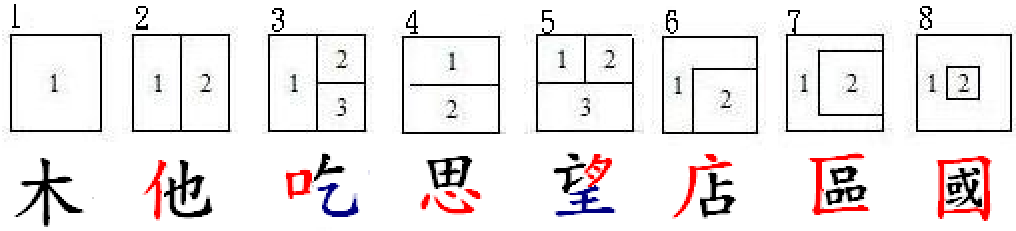


Figure 1.2 Examples of Characters Graphic Structure

Research results indicate that native Chinese learners' orthographic knowledge for characters with left-right structure is more highly developed than for characters with top-down or half-enclosure structures, and they take more time to recognize characters with the semantic radical in the right and bottom positions (Li & Chen, 1999; Li, Fu, & Lin (2000). In the current study, I investigated the transparency of the semantic radicals and the radical positions among the five character graphic structures.

1.1.2 Radical Combination and Semantic Transparency

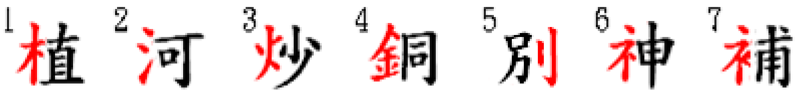


Figure 1.3 Examples of Characters Employing Different Radical Semantic Transparency

Traditionally, Chinese characters are grouped together according to their common components known as “radicals” 部首 (*bùshǒu*), with each character containing a radical.

In the example above, each character contains one radical, with some characters containing higher frequency radicals, such as 木 (*mù*, wood radical). Most of the radicals are helpful for understanding the meaning of the whole characters (Feldman & Siok, 1999), so they are semantically transparent radicals for those characters. For instance, the first four examples in Figure 1.3 contain the radicals 木 (*mù*, wood), 氵 (*shuǐ*, water), 火 (*huǒ*, fire), and 金 (*jīn*, metal) which are helpful to understand the meaning of the characters "plant", "river", "stir-fry", and "copper". Kang (1993) found that from 5,631 semantic-phonetic compound characters, there were 10 highly semantic transparent radicals including 氵 (water), 艹 (plant/grass) , 口 (mouth), 扌 (hand), 木 (wood), 金 (metal/gold), 亻 (human), 虫 (insect / reptile), 言 (words / to speak), and 土 (dirt / earth). However, some radicals are unrelated to the meaning of the whole characters, so they are considered to be semantically opaque radicals for those characters. In example 5 in Figure 1.3, for example, the radical 刂 (*dāo*, knife) in the character 別 (*bié*) is unrelated directly to the meaning "other, difference, differentiate, do not". In addition, there are some radicals that are visually similar, such as examples 6 and 7 in Figure 1.3 where the radicals 礻 (*shì* in 神) and 衤 (*yī* in 補) are differentiated by only one stroke. Some radicals have different shapes, such as with the water radical (水, 氵) when the altered radical 氵 is combined with other components to form a character that looks different from its free standing form 水.

1.1.3 Radical Positional Regularity among Character Graphic Structure

In principle, it is possible for a character component to appear in any position (left, right, top, down, or periphery) within a character. Among different character graphic structures, each character has an overriding structure where in the radicals can fit "legally" into a character. Knowing the legal positions that a radical may take can help to determine whether the combination of character components is the right combination of that character (Fu, 1992; Ho, Yau, & Au, 2003). For example, the 土 (*tǔ*, soil or land) radical usually occupies the left and bottom positions of a character, shown in the first four examples in Figure 1.4. If the 土 radical occupies the right position of the character 城 (*chéng*, city), we know from this rule that the new combination 成土 is not a character, and has no meaning or pronunciation. The legal positions of the 口(*kǒu*, mouth) radical can be found at the left, right, top, and down positions of a character, as shown in the last four examples in Figure 1.4.

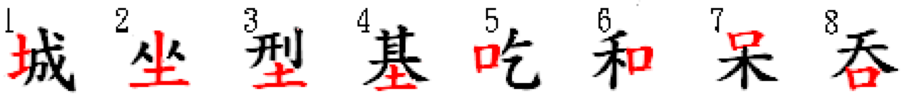


Figure 1.4 Examples of characters employing different radical positions

However, some radicals only occur in one position, such as the radicals 匚, 广, and 囗, shown in the last three examples in Figure 1.4: two half-enclosure graphic structure characters 區 with 匚 radical and 店 with the 广 radical and one enclosure graphic structure character 國 with the 囗 radical. These characters containing only one legal radical position are easier for readers to determine whether the combination of character components is the right combination for that character, as compared to characters containing radicals occurring in two, three, four or more positions within characters.

1.1.4 Phonetic Element Reliability in Pronouncing Chinese Characters

Not all compound characters include phonetic elements, but semantic-phonetic compound 形聲 characters must have phonetic elements. Semantic-phonetic compound characters combine the meaning of one character 形旁 (*xíngpáng*, semantic element) with the sound of another 聲旁 (*shēngpáng*, phonetic element). According to *Chinese Character: A Genealogy and Dictionary*, for instance, 規 is not a semantic-phonetic compound, so it does not have a phonetic element but it does have a radical 見 (*jiàn*, see) because all characters must have radicals. Another example, 植 (*zhí*, plant) is a semantic-phonetic compound character which can be further decomposed into its semantic element 木 (meaning “wood”) and its phonetic element 直 (pronounced “*zhí*”).

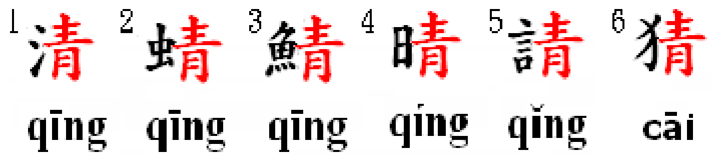


Figure 1.5 Examples of Characters Employing the Same Phonetic Element “青 qīng”

There are two ways to pronounce Chinese characters: using a derivation strategy and using an analogy strategy. The first way to pronounce Chinese characters via the phonetic components is by using a derivation strategy: that is, the pronunciation of the whole character is directly derived from the pronunciation of the phonetic element of the semantic-phonetic compound character. The six characters in Figure 1.5 above all share the same component 青 (*qīng*). The 清 (*qīng*), 蜻 (*qīng*), and 鯖 (*qīng*) characters are pronounced the same as its phonetic element 青 (*qīng*), so they are homophone characters. The phonetic elements of these characters are reliable. The characters 晴 (*qíng*) and 請 (*qǐng*) are partial homophone characters because they share the same phonetic element except for their tonal difference. There are also some characters that violate script-sound correspondence, so that the pronunciation of the whole is completely different from the pronunciation of its phonetic, such as the character 猜 (*cāi*) which is pronounced differently from its phonetic element 青 (*qīng*).

From an etymological viewpoint, Harbaugh (1998) stated that the phonetic element of some characters is not closely related to the character’s modern pronunciation, along with an occasional alteration in the shapes of the phonetic and semantic components. These differences reflect the evolution of the written and spoken languages over time.

1.1.5 Phonetic Component Combination

As stated, two ways can be used to figure out the pronunciation of Chinese characters. The second way to pronounce Chinese characters is by using an analogy strategy: that is, deducing the pronunciation of the whole character via analogy with other characters sharing the same component. The characters here refer to all characters and not just semantic-phonetic compound characters because not even native Chinese speakers can identify semantic-phonetic compound characters. Four types of analogy can be found (Lu, 2003):

1. Homophones: 璧 (*bì*) and 壁 (*bì*) have the same 辟 component and pronunciation.
2. Partial homophones: 清 (*qīng*) and 晴 (*qíng*) have the same 青 component, but with tonal difference.
3. Same rhymes: 板 (*bǎn*) and 返 (*fǎn*) have the same 反 component with the same rhyme in the final *ǎn*.
4. Same component but with completely different sounds: 煮 (*zhǔ*) and 奢 (*shē*) have completely different pronunciations although they have the same 者 zhě component.

In the current study, I investigated five aspects of Chinese orthography as presented in a number of CFL textbooks: radical combination frequency, radical semantic transparency, radical positional regularity among different character graphic structures, phonetic element reliability, and phonetic component combination frequency. There is growing evidence that the development of orthographic awareness among beginning CFL learners contributes to CFL reading ability (Jackson, Everson, & Ke, 2003; Ho & Bryant, 1997; Ho, Yau, & Au, 2003; Li, Fu, & Lin, 2000; Shen, 2004, 2005; Shen & Ke, 2007), so it is important to investigate and classify the characters beginning learners see in their textbooks so as to document this aspect of their orthographic awareness development. For this study, it is also important to describe how scholars have theorized and modeled this development.

1.2 A CFL Model of Orthographic Awareness

Ke (1996) proposed a stage-model of the development of CFL orthographic awareness, theorizing that during the first stage, CFL learners learn characters as wholes, and are not sensitive to the structures discussed in section 1.1. Consequently, they are unable to decompose characters because they have not yet accumulated enough characters in their mental lexicons to abstract the recurring components. Gradually, learners are able to make good guesses about the semantic components of novel characters in which the most perceptually and/or semantically salient and most frequently occurring radicals are embedded, such as the water radical (水, 氵), or the straw radical (艹), or the wood (木) radical. At the second stage, once they know substantial numbers of characters, CFL learners can guess the meaning and sound of most transparent novel semantic-phonetic compound characters quite accurately. They also can acquire more easily those characters with a high frequency of occurrence, with salient graphic features, or from neighborhoods that share few similar sounds and graphics. At the final stage, CFL learners' orthographic awareness is native-like. Their errors in character recognition and production tend to be primarily phonologically oriented.

Although there is no consensus among researchers on the best strategies for developing orthographic awareness, it is believed that L1/L2 Chinese readers develop orthographic awareness through frequent exposure to or encounters with non-character components (ex.氵 in 湖), character components (ex. 古 and 月 in 湖), character phonetic components (ex. 胡 *hú* in 湖 *hú*), and meanings of holistic Chinese characters (ex. "lake" in 湖). As stated, frequency of occurrence of printed characters also contributes to the degree of orthographic awareness development.

Researchers have found that after finishing one year of Chinese, some adult learners clearly could apply the principles while others performed at chance-level and apparently could not. Although learners, as a group, showed some knowledge of how to use both the semantic radical and phonetic components, they still had difficulty learning some distinctive perceptual features of Chinese orthography (Jackson, Everson, & Ke, 2003; Shen & Ke, 2007). Therefore, given the variability knowledge of beginning level CFL adult learners, I investigated beginning level CFL textbooks as important sources of input for beginning learners in developing orthographic awareness.

1.3 Chinese Character Selection in CFL Textbooks over Time

Looking at the history of language teaching methodologies in the past 100 years, language educators, applied linguists, and researchers have been in pursuit of the “best method”. Language learning methodologies or pedagogical principles influence textbook development and also influence the selection of words and characters to be learned in the textbooks over time (Nation, 2001; Richards, 2001). In the majority of textbooks, vocabulary has been selected from frequency lists. Researchers have found that words most commonly used are learned faster and remembered better (McCarthy, 1990; Nation, 2001; O’Dell, 1997; Sergent & Everson, 1992). The speed with which a reader can access a word’s meaning is related to how frequently that word has been encountered in the past. It seems clear that high frequency words are likely to predominate at the early stage of learning and teaching. In addition to being selected from frequency lists, vocabulary has been selected based on grammar points, communication skills, themes, and authentic materials, which means that relatively complicated characters are often among the first characters to be learned.

Most CFL textbooks provide options to CFL learners with traditional and/or simplified character editions being available. Some of the textbooks provide both traditional and simplified characters together in their vocabulary lists. Because characters are combined to form words, some textbooks introduce vocabulary as words (i.e., combination of characters); others introduce individual characters first, and then introduce words from the combination of the characters introduced. John DeFrancis (1977), a leading Chinese scholar and author of 12 series of popular materials for teaching spoken and written Chinese, stated that basic to developing reading skill is a familiarity with the processes whereby Chinese characters recombine with one another to form more complex vocabulary words. Such familiarity is best acquired by mastering several combinations of words formed from a limited number of characters rather than by learning one or two compounds comprised of many characters.

Guder-Manitius (1998) recommends that the selection of characters introduced should as far as possible consist of high frequency characters that also serve as components while containing as few strokes as possible. Using this approach, the students will later find it easier to remember more complex characters where these components are included. The students will then experience a high rate of recognition when they look through a standard text and be given higher motivation for learning.

As stated, native Chinese and CFL readers develop Chinese orthographic awareness through repeated exposure to print, and frequency of occurrence of printed characters contributes to the degree of orthographic awareness development. Therefore, it is important to compare the character selection in differing CFL textbooks. In the current study, I investigated the selections of Chinese characters in textbooks published between 1961 and 2008, as these textbooks were developed based upon different language teaching and learning methodologies. Each textbook will be further described in the textbooks section in Chapter 3 Methods.

1.4 Explicit Orthographic Awareness Instruction

In comparing Chinese reading development among four major societies that use Chinese (China, Hong Kong, Taiwan, and Singapore), Cheung and Ng (2003) reported rote memorization, phonetic transcription, and orthographic decomposition as the primary methods of Chinese reading instruction. Character examples in different instruction methods are shown in Figure 1.6.

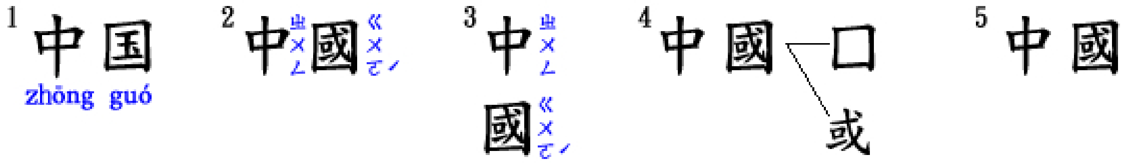


Figure 1.6 Examples of Words Which Mean “China” in Different Instruction Methods

In phonetic transcription, the teacher presents either a romanized transcription (in example 1 above, a system called “pinyin”) or a phonetic transcription (in example 2 above, termed “zhuyin) alongside the characters to be learned. Simplified Chinese characters are paired with Pinyin in example 1. Traditional Chinese characters are paired with Zhuyin written in horizontal and vertical directions in examples 2 and 3. Zhuyin is a system whereby the phonetic components composing a Chinese syllable are broken down into a phonetic rendering for each phonetic component. In the horizontal direction, sentences are written from left to right and top to down. In the vertical direction, sentences are written from top to down and right to left. To explain orthographic decomposition, the teacher points out that these characters are in many cases decomposable into orthographic components. In example 4, 中 is an integral character which cannot be decomposed, but 國 (*guó*, country) is a compound character which can be further decomposed into the semantic radical □ (*wéi*) and the phonetic component 或 (*huò*). In rote memorization, shown in example 5 in Figure 1.6, the teacher presents characters as holistic units and encourages students to memorize the pronunciations as unique character names. Children have to copy each new character at least ten times to make sure that they can reproduce them correctly (Chan & Wang, 2003).

One question that needs to be addressed is whether it is worthwhile to use precious class time to teach orthographic decomposition principles to CFL students. Jackson, Everson, and Ke (2003) found first year CFL college students were indeed able to take advantage of the systematic classroom instruction they had received in the use of semantic radicals to identify the meaning of some novel characters. Researchers also confirmed the benefits of explicit orthographic decomposition instruction (Jackson et al., 2003; Shen, 2004) leading researchers to believe that it is indeed worthwhile to implement orthographic decomposition instruction for CFL college students once learners have overcome the perceptual challenge of learning to recognize character components.

On the other hand, if explicit orthographic decomposition instruction in class is not possible, is the explicit orthographic decomposition information in the textbooks? If so, do students actually use and study the information? Or do teachers go over this information in class? Based upon a series of CFL classroom observations, Jackson et al., (2003) found that only three pages in the class text used during a particular lesson were devoted to discussion and examples of the roles of radicals as indicators of character meaning. Moreover, no information was identified that could be considered instruction in the use of phonetics. Therefore, the students had to rely primarily on classroom instruction for information about orthography principles and their application.

To summarize, more and more research points to the development of orthographic awareness as a key component in Chinese L1 and CFL reading development. Chinese readers can develop orthographic awareness through frequent exposure to or encounters with non-character components, character components, phonetic components, and learning the overall meaning of Chinese characters as holistic units. Both native Chinese and CFL readers can take advantage of orthographic decomposition instruction. Comparing L1 and CFL readers, CFL beginning college-level readers in the Untied States have less exposure to Chinese print outside commercial CFL textbooks, even through the internet has changed the potential for Chinese print exposure for CFL learners. In addition, non-heritage CFL beginning college-level learners have fewer opportunities to exposure to spoken and/or written Chinese outside the classroom as compared to heritage learners. If teachers give limited orthographic decomposition instructions based on personal preferences or program curriculum restriction, or if their classrooms have limited access to the world wide web, commercial CFL textbooks would be the primary resource where they can learn orthographic principles by themselves. However, whether CFL learners are provided with textual materials that facilitate their development of orthographic awareness is still unknown because only minimal research has addressed the language input provided by CFL textbooks (Everson & Fan, 2008; Jackson, Everson, & Ke, 2003).

1.5 Purpose of the Study and Research Questions

The purpose of this study is to systematically describe and classify Chinese characters in ten CFL textbooks for college and adult beginning learners. The main focus is to make a component inventory of characters and discuss textual materials availability in the following areas: (1) explicit orthographic decomposition instruction (research question 1); (2) character diversity and repetition (research question 2); (3) character frequency selection across textbooks (research questions 3); (4) radical component diversity and repetition (research questions 4, 5, and 7); (5) phonetic component diversity and repetition (research questions 6 and 7); and (6) ideal semantic transparent radicals and reliable phonetic elements (research question 7). More specifically, the present study was designed to answer the following research questions:

1. To what extent do textbooks provide explicit orthographic decomposition instruction to learners?
2. To what extent is a single character combined with other characters to form words in textbooks?
3. To what extent do textbooks contain high-frequency characters as documented by accepted Chinese character frequency lists over time?
4. To what extent is a radical combined with other components to form characters in textbooks?
5. For each character graphic structure, what is the most commonly appearing radical position?
6. To what extent is a phonetic component combined with other components to form characters in textbooks?
7. For semantic-phonetic compound characters, what percentages of the characters contain semantically transparent radicals and reliable phonetic components?

1.6 Significance of the Study

Further research on how Chinese orthographic awareness develops and specifically what factors facilitate this development is needed (Everson, 2007; Kupfer; 2007). A component analysis of Chinese characters provides us with structural knowledge about Chinese characters, and helps material developers, test developers, and teachers to select and adjust characters to be learned, taught, and tested. Most of the component analysis studies have examined Chinese characters from dictionaries or corpus database (Chen, 1997; Fu, 1989; Guder-Manitius, 1999; Zhu, 1993) and rarely have investigated characters from beginning level textbooks (Everson & Fan, 2008). In the CFL context, since reliable target language input is limited largely to textbook materials and teacher instruction, we have to know which characters are typically presented in beginning level CFL textbooks so we can discover whether CFL learners have opportunities to develop orthographic awareness. From such investigations, we will be able to build better models of developing CFL orthographic awareness.

A first step taken with this study was to investigate the types of Chinese characters CFL textbooks actually introduce to CFL students and describe the differences in Chinese character selection in textbooks published from 1961 to 2008. Various textbooks that feature all the modalities of listening, speaking, reading, and writing were examined. It is hoped that this study will facilitate curriculum and material development. As well, this study will provide pedagogical recommendations to Chinese teachers. In the future, this study can serve as a character and word inventory for generating test specification for Chinese orthographic knowledge assessment.

1.7 Limitations of the Study

The present study has the following limitations:

1. This study only examined widely used Chinese textbooks in the United States.
2. This study only focused on vocabulary (Chinese characters and words) in textbooks.
3. This study only focused on traditional Chinese characters in the CFL/CSL textbooks.

1.8 Definition of Terms

Orthography: The graphemic patterns of a written language and their mapping onto phonology, morphology, and meaning (Henderson, 1984).

Component 部件 (*bùjiàn*): a segment of a character.

Radical 部首 (*bùshǒu*): a common component that Chinese characters are grouped together in the dictionary.

Semantic element 形旁 (*xíngpáng*): for semantic-phonetic compound characters only, a semantic element provides the meaning of one character.

Semantically transparent radical: a radical helps to infer the meaning of the character. The relationships between the radical and the character could be category, directly related or indirectly related relationship.

Phonetic element 聲旁 (*shēngpáng*): for semantic-phonetic compound characters only, a phonetic element provides the sound of one character.

Reliable phonetic element: a phonetic element helps to infer the pronunciation of the character. The pronunciation of the phonetic element is the same as the pronunciation of the character.

Phonetic component: for all types of characters, a phonetic component provides pronunciation hint to characters shared the same component. Phonetic components include phonetic elements.

Ideal semantic-phonetic compound character: a semantic-phonetic compound character contains a semantically transparent radical and a reliable phonetic element.

Character graphic structure: a structural shape of character which is combined from the graphic shapes of their components, such as left-right, top-bottom.

Word recognition: the processes of obtaining a word's sound and meaning.

Decoding: the process of the extracting word's phonological information.

CHAPTER 2

LITERATURE REVIEW

The present study is based on the theoretical assumptions that frequency of characters and their structural components, as well as the frequency types of structural components are important to enable CFL learners to discover the underlying structures of characters. To explore this assumption, we have to know which characters are presented in CFL textbooks, so we can determine whether or not CFL learners have opportunities to develop orthographic awareness. In this study, I systematically described and classified the Chinese characters introduced in ten CFL textbooks for beginning college learners. The main focus was to make a component inventory of characters and discuss textual materials availability in the following areas: (1) explicit orthographic decomposition instruction; (2) character diversity and repetition; (3) character frequency selection across textbooks; (4) radical component diversity and repetition; (5) phonetic component diversity and repetition; and (6) characters with ideal semantically transparent radicals and reliable phonetic elements.

2.1 Introduction

Word recognition in reading comprehension is important, and reading involves processing specific orthographies. Since the focus in this study was on CFL learners who are native English speakers, it is important to learn how native English speakers read English orthography and how native Chinese children learn to read Chinese orthography. More important, when native English speakers learn Chinese as a foreign language, do they need to develop new strategies to read Chinese orthography? Therefore, the role of word recognition in reading comprehension is introduced and a dual-route model of reading aloud in English is explained to show how native English speakers read English words. This will be followed by a description of beginning orthographic development in L1 Chinese readers as well as CFL readers. Finally, for teaching and learning purposes, component analysis of characters is introduced, with these analyses serving as the comparison bases for the current study.

2.2 Word Recognition in Reading Comprehension

One important factor in reading comprehension is word recognition. Since it is not clear how readers obtain meaning from text, based on different L1 and L2 reading theories, scholars create reading models to visualize, represent, infer, or interpret available information about reading processes (Bernhardt, 1986). Word recognition or decoding in these models plays both an initial and essential role or a supportive role. For example, often the bottom-up models focus on the letter-sound correspondences or phonics rules that can be used to decode words. Gough (1972) proposed that reading progresses in a linear fashion through decoding letters and then moving to progressively larger units. No stage can be bypassed. LaBerge and Samuels (1974) believe that reading involves decoding and comprehension. Both of these require attention; however, decoding requires much of the available attention, at least for beginning readers. Goodman's (1968) model is the best known top-down reading model. In his view, readers rely on existing syntactic and semantic knowledge structures, so that reliance on the print and phonics rules can be minimized. In Bernhardt's L2 constructivist model (Bernhardt, 1986, 1991) reading includes the interaction between text-based and extratext-based factors. Text-based components include phonemic/graphemic decoding, word recognition, and syntactic feature recognition. Extratext-based components include intratextual perception, prior knowledge, and metacognition. In addition, Bernhardt added learners' L1 component in her constructivist model. These components are free to interact, but the contributions may differ by language, by orthography, by course level, or by a host of other variables (Everson, 1986).

Although the terms *word recognition* and *decoding* are often used interchangeably in research, the definitions of *word recognition* and *decoding* are quite different among these reading models described above. In the bottom-up model, *decoding* refers to letter-sound correspondence and translation of print to sound. In the top-down model, *decoding* refers to decode print to meaning. In the constructivist model, *word recognition* refers to the attachment of semantic value to words, and *phonemic/graphemic decoding* refers to the recognition of words based on sound or visual characteristics. In the current study, *word recognition* refers to the processes of obtaining a word's sound and meanings, and *decoding* deals specifically with the extraction of phonological information.

According to Koda (2005), word recognition includes three processes: orthographic processing, phonological processing, and contextual facilitation of semantic processing. In *orthographic processing*, learners must become aware that written symbols correspond to speech units, and then learn the specific ways in which symbols are combined to represent spoken words. Orthographic knowledge is acquired through cumulative exposure to visual word input and practice (Seidenberg & McClelland, 1989). *Phonological processing* is the ability to pronounce printed words, which is regarded as a powerful predictor of reading success among primary grade children (Koda, 2005). *Phonological processing* consists of three general areas: (1) phonological awareness, which emphasizes rhyming, phonemic segmentation, and phonemic blending; (2) verbal short-term memory, which focuses on maintaining phonemic information in working memory; and (3) rapid naming ability, which requires labeling common items. Phonological awareness has long been considered an essential component of beginning reading instruction. Children who have learned phonics get a better start in reading than children who have not learned phonics. Working-memory experiments show that phonological transformation is more efficacious than visual encoding in retaining visually presented information in working memory among native Chinese readers. Studies using tests that measure naming speed of familiar items have demonstrated that poor readers' rate of naming is significantly slower than the rate of good readers. In *semantic processing*, the empirical evidence demonstrated that all of a word's known meanings are activated by its orthographic input, even when the context imposes strong constraints. Contextual effects on word-meaning retrievals decrease as reading proficiency improves. Contextual reliance is a strategy that poor readers lean on to compensate for their underdeveloped visual-information sampling skills (Koda, 2005).

Current research on L2 word recognition has found three factors affecting L2 word recognition (Koda, 1997; Koda, 2005). The first factor is the impact of *L1 experience.* Cummins's (1981) Interdependence Hypothesis supports this assumption, which asserts that experience with either the L1 or the L2 can promote development of the capacities underlying both languages. In other words, literacy skills, such as the processing skills of word recognition, can transfer, given sufficient motivation and exposure to the L2. The second factor is the impact of *L2 experience*. Researchers found that the relationship between L1 and L2 reading may vary according to L2 proficiency level (Brisbois, 1995), which supports Cummins's (1981) Threshold Hypothesis. Cummins's Threshold Hypothesis asserts that language transfer in reading is possible only after a threshold level of L2 proficiency has been attained. Automaticity is a central concern in processing efficiency. The last factor is the impact of *L1 and L2 orthographic distance*. Cross-linguistic research on L2 learners with divergent L1 orthographic backgrounds repeatedly attests to the faster and more accurate recognition performance among those with related L1 orthographic backgrounds. Writing systems differ in terms of orthographic representation and depth. Orthographic representation refers to what each graphic symbol denotes. For example, in alphabetic writing systems, such as English and Spanish, each letter or combinations of letters represents a phoneme. In logographic writing systems, such as Chinese characters and Japanese Kanji, each symbol represents a word or morpheme; for example, the character 上 embodies the character sound "shàng" and means "top" or "up". Orthographic depth refers to the degree of regularity in symbol-sound correspondences. In shallow (i.e., transparent) orthographies, the symbol-sound relationships are highly regular, as in languages like Spanish and Czech. In deep orthographies, one-to-one sound-symbol correspondence is not consistent and reliable, as in languages like French and English. For example, the past tense morpheme *–ed* in English is pronounced in three different ways, as in *talked*, *visited*, and *called*.

In the previous discussion, most of the discussion of word recognition is not specific to particular languages. Since reading involves processing specific orthographies, and the focus of this study is on CFL learners who are native English speakers, it is important to investigate how native English speakers read English orthography and how native Chinese children learn to read in Chinese orthography. More importantly, when native English speakers learn Chinese as a foreign language, do they need to develop new strategies to read Chinese orthography?

2.3 A Dual-Route Model of Reading Aloud in English

Coltheart, Curtis, Atkins, and Haller (1993) discussed a dual-route of reading aloud to explain how native English speakers convert print to speech. In a dual-route model, readers can recognize words via one of two processing routes into their mental lexicons. The first route is the *orthographic route* (i.e., lexical route) that allows readers to retrieve a word's meaning directly from the print. The second route is the *phonological recoding rout* (i.e., nonlexical route) that allows readers to sound out words and then find a word matching the pronunciation of the words in the lexicon.

In addition, the pronunciation of words can be generated in two ways. One way is to apply the grapheme-phoneme (or letter-sound) correspondence rule to assemble phonological representations before accessing the word's meaning. The use of the letter-sound correspondence route is important for beginning readers and unskilled readers to sound out unfamiliar words. Although this pathway can be used to read nonwords and regular words which follow the regular letter-sound rules, words with unusual spelling-sound relations would be mispronounced. Therefore, another way of generating the pronunciation of words occurs after the readers have accessed the word's meaning through orthographic representations. That is, the word meaning is accessed directly, which in turn activates its pronunciation. Since this procedure is only used to retrieve the meanings of known words, this look up procedure cannot be used to read nonwords or pseudowords. However, researchers have argued that nonwords are read aloud by a process involving analogy (Glushko, 1979; Marcel, 1980). This involves a nonword activating the lexical entries for words that are orthographically similar to it, and so nonword reading is not nonlexical here and is not based on explicit rules. To overcome the shortcomings of the dual-route model, Seidenberg and McClelland (1989) proposed a connectionist model of reading aloud and designed a computer program to simulate reading aloud from print to speech using orthographic and phonological units. They emphasized that their model performed well on words that were exceptions, regular words, and nonwords, and they claimed that of the 2897 words in the model's training set, only 77 (2.7%) were wrongly read.

Knowledge of the word formation, word structure, and the sound system of English provide foundations for the development of orthographic awareness. By using English orthographic awareness, English learners can infer meanings of words by analyzing word structures and identifying morphemes within a word; they can also predict the pronunciation of a word given its spelling or come up with a possible spelling for a word given its pronunciation.

In addition, a dual-route model of reading aloud can be used to illustrate how English readers process word input, store word properties, and retrieve word properties when necessary. Through an orthographic route (i.e., lexical route), readers retrieve a word's meaning directly from the print; through a phonological recoding route, readers sound out words and then find a word matching the pronunciation of words in the lexicon. It is important to note that the notion of visual and phonological pathways to the lexicon also can be found in Chinese character studies (Ho & Bryant, 1997; Leck, Weekes, &

Chen, 1995). In the next section, how native Chinese children learn to read in Chinese orthography is discussed.

2.4 Beginning Orthographic Awareness Development in L1 Chinese

According to Taylor and Taylor (1995), roughly 6,000 characters are required for scholarly literacy, and 3,500 characters are designated as "modern Chinese characters for everyday use." Chinese children are required to learn 2,834 characters through formal instruction during their 6 years of primary education. In comparing Chinese reading development in four major Chinese societies, Cheung and Ng (2003) found 200 to 300 characters are learned in each term for First Grade in China, and children are expected to learn approximately 2,400 characters by the end of Fourth Grade. The students in Hong Kong are expected to know 460, 960, 1,490, and 2,080 characters upon finishing grades 1, 2, 3, and 4, respectively. The students in Taiwan are expected to know 1,600 characters at the end of Fourth Grade. The students in Singapore are expected to know 600 and 1,200 characters upon finishing grades 2 and 4.

2.4.1 Learning to Read in L1 Chinese

It goes without saying that to achieve literacy in Chinese, children are expected to know a large number of characters. How do native Chinese children learn to read in Chinese? As stated in section 1.3 , Cheung and Ng (2003) reported that rote memorization, phonetic transcription, and orthographic decomposition are the primary methods whereby children undergo Chinese reading instruction among the four major societies that use Chinese (China, Hong Kong, Taiwan, and Singapore). Character examples in different instruction methods are shown in Figure 2.1.

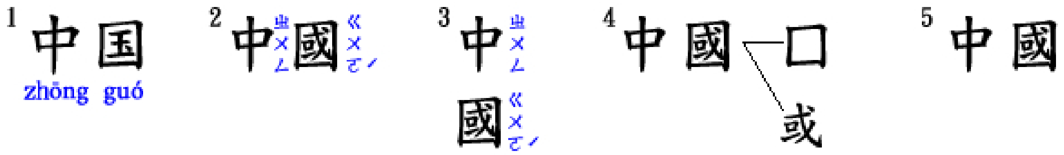


Figure 2.1 Examples of Words Which Mean “China” in Different Instruction Methods

In phonetic transcription instruction, the teacher presents the transcriptions (Pinyin or Zhuyin) alongside the to-be-learned characters, as shown in examples 1~3. Simplified Chinese characters, used in China, Singapore, and Hong Kong, are paired with the Pinyin pronunciation underneath in example 1. Traditional Chinese characters, used in Taiwan and Hong Kong, are paired with the Zhuyin written either in horizontal or vertical directions in examples 2 and 3. In the horizontal direction, sentences are written from left to right and top to down. In the vertical direction, sentences are written from top to down and right to left. With this reading instruction in print, children can access a word’s pronunciation directly from Pinyin or Zhuyin. The first term of 1st grade is spent learning and consolidating the Pinyin or Zhuyin symbols, and characters are not presented until the second term. Character pronunciations are always taught via Pinyin or Zhuyin in textbooks. By writing characters repeatedly with their Pinyin or Zhuyin alongside, children become familiar with sound to symbol associations. Therefore, they are expected to reproduce the character when presented with its Pinyin or Zhuyin, and they are also expected to reproduce the Pinyin or Zhuyin when they see the character.

In orthographic decomposition instruction, the teacher points out that these characters are in many cases decomposable into orthographic components. In example 4, 中 is an integral character which cannot be decomposed, but 國 is a compound character which can be further decomposed into the radical component □ and the phonetic component 或. With this instruction, native Chinese readers can use the orthographic knowledge of the radical to retrieve a word’s meaning. One way to pronounce Chinese characters via phonetic components is by using a derivation strategy: that is, the pronunciation of the whole character is directly derived from the pronunciation of its phonetic component. Another way to pronounce Chinese characters via the phonetic components is by using an analogy strategy: that is, deducing the pronunciation of the whole character via analogy with other characters sharing the same phonetic component. However, radical and phonetic component cues in characters are irregular and unsystematic.

In rote memorization instruction as used in Hong Kong depicted in example 5 in Figure 2.1, a teacher presents characters as holistic units and encourages students to memorize the pronunciations as unique character names. In most kindergartens, according to Chan and Nunes (1998), five-year old children are expected to learn one or two words every day, and rote memory is encouraged. Children have to copy each new character at least ten times to make sure that they can reproduce them correctly (Chan & Nunes, 1998; Chan & Wang, 2003).

As stated, children are expected to know large number of characters. Gough, Juel, & Griffith (1992) pointed out two problems faced by logographic readers at an initial learning stage. The first is a memory problem because it becomes difficult for children to differentiate and memorize great numbers of visually similar words. The second is the inability to read novel words. Fortunately, rote memorization is not the only learning strategy used to read novel words. A reader’s orthographic awareness is a powerful analytical tool to learn and read novel words. As readers become aware of the internal structures of the characters they are reading, they will develop new strategies for learning and reading novel words.

Koda (2005) illustrated two facilitative benefits of orthographic awareness. First, orthographic awareness helps literacy learning to occur. Readers must understand that written symbols correspond to speech units. They then must learn what each symbol represents, as well as how it can be combined with others to form a word. Second, an understanding of the segmental nature of language promotes analytical competence. With analytical competence, readers can extract partial information from a new string of symbols.

In general, as children learn more and more characters, the ability to recognize and capitalize on the orthographic structure of characters increases through the school years (Ho & Bryant, 1997; Shu & Anderson, 1999). Researchers have investigated, for example, how older learners start to use a phonological analogy strategy, but their findings are inconsistent. Ho, Wong, and Chan (1999) found that even first-graders make phonological analogies using the phonetic component when reading a novel character. However, Chan and Nunes (1998) have reported that a systematic use of the phonetic radical as a clue to pronunciation was not observed until nine years of age (grades 3 to 4). Ho, Yau, Au (2003) state that Chinese children develop some rudimentary orthographic knowledge quite early, but it takes a long time for them to develop semantic radical and phonetic component aspects of orthographic knowledge. In fact, they found that only 32% of the third-graders acquired some aspects of orthographic knowledge, including radical information knowledge, positional knowledge, and functional knowledge for both phonetic and semantic radicals. They concluded that it may take the whole period of primary education for children to develop a complete grasp of the different aspects of orthographic knowledge in Chinese.

Ho, Yau, and Au (2003) suspect the inconsistency of many research findings on when children start to use their orthographic knowledge after accumulating Chinese characters may be partly due to the different types of tasks employed in these studies. Regardless of the inconsistency of findings, the evidence indicates that children become aware of the internal structures of the characters as they gradually learn more characters.

2.4.2 The Interactive Constituency Model of Chinese Character Identification

Although there is no consensus among researchers on the best strategies for developing orthographic awareness, it is believed that L1/L2 Chinese readers develop orthographic awareness through frequent exposure to or encounters with non-character orthographic components (ex.氵 in 湖), character orthographic components (ex. 古 and 月 in 湖), character phonetic components (ex. 胡 *hú* in 湖 *hú*), and meanings of holistic Chinese characters (ex. "lake" in 湖). As stated, frequency of occurrence of printed characters contributes to the degree of orthographic awareness development.

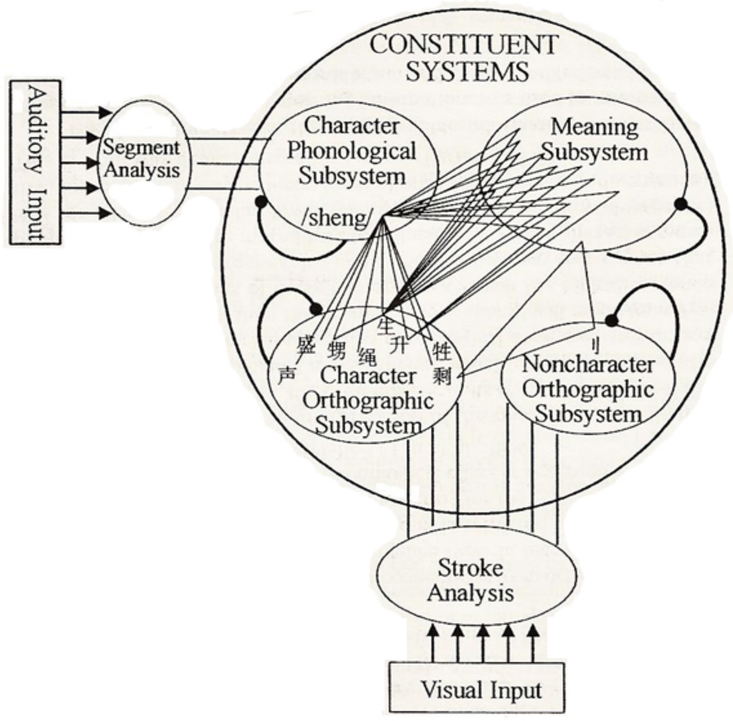

CONSTITUENT SYSTEMS
Auditory Input
Segment Analysis
Character Phonological Subsystem
/sheng/
Meaning Subsystem
盛
甥
绳
生
升
牲
声
剩
Character Orthographic Subsystem
刂
Noncharacter Orthographic Subsystem
Stroke Analysis
Visual Input

Figure 2.2 The Interactive Constituency Model of Chinese Character Identification (Perfetti & Tan, 1999, p127)

Perfetti and Tan (1999) proposed the Interactive Constituency Model of Chinese Character Identification presented as Figure 2.2 . This model can be used to explain how frequency of occurrence of printed characters contributes to the degree of orthographic knowledge development. The model includes four separate constituent representation subsystems that are interconnected: the character orthographic subsystem, the noncharacter orthographic subsystem including a few phonetics and many semantic radicals, the phonological subsystem, and the meaning subsystem. Each constituent subsystem consists of a set of representation units or nodes. A node whose activation value exceeds its threshold excites other nodes with which it is consistent and inhibits nodes with which it is not consistent. Character identification results from patterns of activation across the subsystems.

Perfetti and Tan (1999) believe the activation threshold of orthographic units (character and noncharacter) is determined by the frequency of occurrence of printed characters and components in daily usage. The threshold of phonological units is determined by the frequency of prior threshold level activations of the phonological form associated with the character, both through speech and print-speech experience. The model assumes that meaning is represented and organized in terms of semantic attributes. The activation threshold of a specific meaning node relies on the frequency of encounter with this meaning. The Interactive Constituency Model of Chinese Character Identification provides the theoretical explanation of frequency and how it contributes to the degree of orthographic awareness development. The model also illustrates characteristics of Chinese orthography.

The first stage of visual character recognition is stroke analysis. For example, each stroke of the compound character 甥 and its positional relationship with other strokes is detected. Once detected components (生, 田, 力, 男, 甥) exceed threshold (being recognized), detected components begin sending activation to the character orthographic (生, 田, 力, 男, 甥), phonological (*shēng*, *tián*, *lì*, *nán*, *shēng*), and meaning (to grow, field, power, male, nephew) subsystems but not to the noncharacter orthographic subsystem because all character components are free standing characters. In addition, with the phonological "*shēng*" unit being activated, consistent character orthographics (生, 甥, 牲, 声, 升) (i.e. homophones) and their meanings (to grow, nephew, domestic animal, sound, move upward) are excited. Tan and Perfetti (1997) believe that phonological nodes are more likely to influence access to meanings for characters having fewer homophones than for characters having many homophones. However, it is important to note that, as stated, the activation threshold of subsystem units is determined by the frequency of occurrence.

2.5 Beginning Orthographic Awareness Development in CFL

Languages that are linguistically unrelated to English are more challenging for native-English speakers to learn. The non-alphabetic nature of the Chinese writing system is certainly one of the primary factors that makes learning Chinese such a time-intensive process. Because learning to read requires "becoming aware of the basic units of spoken language, the basic writing system, and the mapping between the two (Shu & Anderson, 1999, p1)", the experience of learning to read in languages employing alphabets is different from learning to read in non-alphabetic orthographies such as Chinese. Perhaps the most critical aspect of the non-alphabetic Chinese orthography for English-speaking foreign language learners is that it reflects in only a limited way the phonology of the spoken language. In other words, learners cannot "sound out" the pronunciation of Chinese characters as they can in alphabetic scripts because phonetic elements available in Chinese characters are generally irregular and unsystematic.

DeFrancis (1977) explained that one of the beliefs underlying the *Beginning Chinese Reader* was that learning to read can be accomplished most efficiently by students who have some prior grounding in speech and who engage in simultaneous oral practice of what is read. Consequently, CFL learners would often go through the first few months of learning Chinese exclusively through Romanization, with the introduction of characters being delayed until the students achieved a firm grounding in the spoken language. Everson (1998) explained this pedagogical process when he pointed out that in the past, learning to read in the CFL setting involved a two-step process: one step involved the representation of Chinese for rapid acquisition of the spoken language through a system that represents Chinese via the Roman alphabet and tonal marks (e.g., Pinyin), and the other step involving the learning of actual Chinese characters. Following the DeFrancis philosophy stated above, some textbooks (e.g., *Beginning Chinese*, *Speak Mandarin*) presented sentences, dialogues, or paragraphs completely in Romanization without Chinese characters. Students, therefore, learned to "read" in Romanization first, and then slowly Chinese characters were substituted for the Romanized words as students began the learning of characters. However, do CFL learners actually rely upon their knowledge of the spoken language to help them remember the meaning of Chinese characters? Is it possible that CFL students develop a variety of strategies that may include learning the meaning of characters through largely visual means without learning their pronunciation? Everson (1998) conducted a study among beginning CFL college learners to investigate whether there is a relationship between correctly pronouncing and correctly identifying the meaning of words written in Chinese characters.. He found a very strong relationship between knowing a word's meaning and knowing its pronunciation. That is, when the participants knew the meaning of a two-character word, there was a mean probability of 91.4% that they also knew its pronunciation. On the other hand, when the participant knew the pronunciation of a two-character word, there was a mean probability of 90.7% that they also knew its meaning.

Current research indicates that both CFL students and teachers believe that character learning and writing are the most difficult tasks in learning Chinese at the college level (Everson, 1998; Ke, Wen, & Kotenbeutel, 2001). To develop proficiency in reading Chinese, researchers have put forth the theory that native Chinese and CFL readers develop Chinese orthographic awareness to infer the meaning and pronunciation of Chinese characters through repeated exposure to print and explicit orthographic instruction (Jackson, Everson, & Ke, 2003; Ho & Bryant, 1997; Ho, Yau, & Au, 2003; Li, Fu, & Lin, 2000; Shen, 2004, 2005; Shen & Ke, 2007, Shen, 2010).

Research on L2 learners of Chinese indicates that adult learners acquire orthographic awareness competences in a new language far more rapidly than beginning L1 readers, thus attesting to the developed cognitive maturity of adult CFL learners as well as there developed sense of language in general. L1 and L2 print-processing experiences especially contribute to L2 orthographic awareness development: “L1 experience establishes the scaffolding for foundation building, and L2 input instills a linguistic base necessary in fine-tuning (Koda, 2005, p.94)”. Jackson, Everson, and Ke (2003) described the CFL learners in their study as having two potential advantages over Chinese children:

> (a) they brought to the classroom the conceptual sophistication of adults already literate in another, albeit radically different writing system; (b) they were given beginning instruction in which the structure of semantic-phonetic compounds was made explicit. However, they were at great disadvantage, relative to Chinese children, in their knowledge of oral Chinese, their history of exposure to characters, and, perhaps, in having developed reading strategies that could interfere with learning a nonalphabetic language (Jackson, Everson, & Ke, 2003, p. 142).

2.5.1 A CFL Model of Orthographic Awareness

Ke (1996) proposed a model of the development of orthographic awareness which states that learners of Chinese acquire orthographic awareness in three successive stages (shown in Figure 2.3). Ke also believes certain recurring components and graphic features tend to be harder for the learners to acquire than others.

During the first stage, the Precomponent-processing Stage, CFL learners learn characters as wholes and are unable to decompose characters because they have not yet accumulated enough characters in their mental lexicons to abstract the recurring components. Gradually, learners may be able to make good guesses about the semantic components of novel characters in which the most perceptually and/or semantically transparent and most frequently occurring radicals are embedded, such as the water, straw, wood, or the animal radicals.

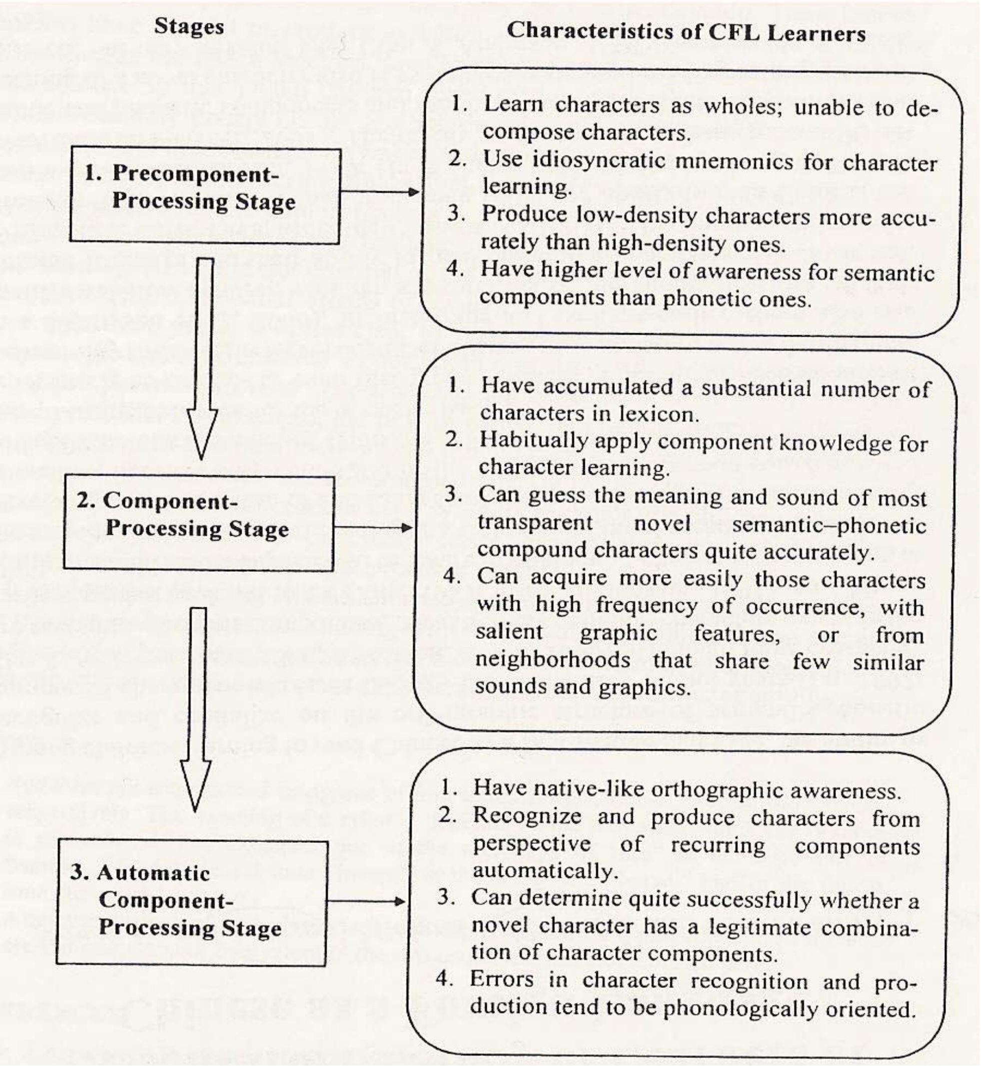

Stages
Characteristics of CFL Learners
1. Precomponent-Processing Stage
1. Learn characters as wholes; unable to decompose characters.
2. Use idiosyncratic mnemonics for character learning.
3. Produce low-density characters more accurately than high-density ones.
4. Have higher level of awareness for semantic components than phonetic ones.
2. Component-Processing Stage
1. Have accumulated a substantial number of characters in lexicon.
2. Habitually apply component knowledge for character learning.
3. Can guess the meaning and sound of most transparent novel semantic–phonetic compound characters quite accurately.
4. Can acquire more easily those characters with high frequency of occurrence, with salient graphic features, or from neighborhoods that share few similar sounds and graphics.
3. Automatic Component-Processing Stage
1. Have native-like orthographic awareness.
2. Recognize and produce characters from perspective of recurring components automatically.
3. Can determine quite successfully whether a novel character has a legitimate combination of character components.
4. Errors in character recognition and production tend to be phonologically oriented.

Figure 2.3 Ke's (1996) Stage Model for the Development of Orthographic Awareness among CFL Readers

At the second stage, the Component-processing Stage, CFL learners already know substantial numbers of characters and can guess the meaning and sound of most transparent novel semantic-phonetic compound characters quite accurately. They also can acquire more easily those characters that occur frequently and have salient graphic features, or characters that share few similar sounds and graphics with previous learned characters.

At the final stage, the Automatic Component-processing Stage, a CFL learner's orthographic awareness is native-like. Their errors in character recognition and production tend to be phonologically oriented.

2.5.2 Studies on L2 Beginning Chinese Orthographic Awareness Development

In a study investigating levels of cognitive processing encoding strategies affecting retention of Chinese characters (words), Shen (2004) found instructor-guided elaboration (deeper cognitive processing) resulted in a significantly better sound and meaning recall of words at a 20-minute interval than did student self-generated elaboration and rote memorization (shallow processing). In an instructor-guided elaboration condition, the instructor guided students to establish an explicit concept for a new word by using various means such as explaining the etymology if applicable, analyzing the radicals, creating anchors through which the new knowledge is connected with previously learned knowledge, and providing the example of word use in different contexts. However, the instructor-guided elaboration advantage seemed to disappear during a 48-hour interval. Shen (2004) cautioned that further review seemed to be necessary shortly after the learning. In addition, she found that self-generated elaboration from the CFL students, who completed their second-year of Chinese and had learned at least 100 radicals, were all semantically-based and relevant to the target word, though the degree of precision differed.

In a study by Shen and Ke (2007), a Chinese character decomposition task was used to measure how well CFL learners were able to decompose compound characters into radicals/components, and their ability to reproduce compound characters by using radicals/components when encountering an unknown character. They included four Chinese character graphic structures: right-left (e.g., 婉), top-bottom (e.g., 箕), half-enclosure (e.g., 匿), and enclosure (e.g., 国). First-through fourth-year CFL learners were asked to write out the number of radicals/components and each radical/component in order. Shen and Ke (2007) found that after a month of Chinese language study, the accuracy rate for visually decomposing compound characters into radicals reached 54%. After a full year's study, the mean accuracy rate reached 73%. Shen and Ke (2007) also used a semantic radical knowledge application task to assess CFL learners' knowledge of the meanings of semantic radicals. They asked CFL learners to circle the target character, based on the semantic radicals it contained, and which best fit the meaning provided in English. In addition, they asked CFL learners to restore a compound character by writing the missing semantic radical based on the meaning and sound given. The mean accuracy rate for one-year, two-year, and three-year learning levels were 53.85%, 56.62%, and 72.45%, respectively. They found significant progress in applying radical knowledge accurately on character learning beginning when students after completing 2- full-year study. They also concluded that at the end of 3 full years of study, most students have reached a high level of proficiency in using radical knowledge for learning new characters.

Jackson, Everson, and Ke (2003) investigated whether beginning CFL college learners can use the meanings of familiar semantic radicals to infer the meanings of novel compound characters which included the familiar semantic radicals. CFL learners were asked to choose the meanings of familiar semantic radicals and novel compound characters in multiple choice questions. The familiar semantic radicals were those present as free standing characters (e.g., 火 meaning “fire”), with the novel compound characters being those included the familiar semantic radicals (e.g., 燃 meaning “ to burn”). In addition, they used open-ended questions to assess the pronunciation of the phonetic components. To accomplish this, the CFL learners were asked to write down the pronunciations of familiar phonetics and novel characters in Pinyin. The familiar phonetic components were those present as free standing radical characters (e.g., 同 pronounced "*tóng*"), and the novel compound characters were those that included the familiar semantic radicals (e.g., 酮 pronounced "*tóng*"). Overall, after nearly one academic year study of Chinese, CFL learners' performances on both semantic radical and phonetic component tests were variable and rather poor. Some students clearly could apply their developing semantic and phonetic orthographic knowledge, while others could not.

Shen (2010) had investigated first-year college CFL students' learning behaviors in the study of Chinese semantic radicals. When asking the difficulties the beginning learners face in terms of mastering the sound, shape and meaning of a radial, although most of the CFL beginning learners in Shen's (2010) study believed that radical knowledge helped them to understand the meanings and sounds of compound characters, about 7% of the CFL learners had opposite responses. They complained that they could not relate a radical's meaning to a character containing the radical or that they never tried to use a radical's meaning to infer the character's meaning. The complaint reflects the frustration that beginning level CFL learners had when they encounter semantically opaque radicals of the characters when they are used to teach the application of radical knowledge.

In summation, reviewing the Interactive Constituency Model of Chinese Character Identification, Ke's stage model of orthographic awareness for CFL readers, and the evidence from L1 and CFL research, it seems that the development of orthographic awareness is a necessary process for learning to read in Chinese, yet is slow to develop in both L1 and CFL readers.

2.6 Character Types for Orthographic Knowledge Research Purpose

To assess native Chinese and CFL learners' orthographic knowledge, researchers have used different research tasks.. In these tasks, depending on the researchers' interests, the internal structures of the characters are manipulated and are used in creating research materials. Ho, Yau, and Au (2003) suspect that the inconsistency of many research findings may be partly due to the different types of tasks employed in these studies. It is important to find what tasks researchers used to investigate orthographic knowledge. Therefore, this discussion of Chinese character selection for orthography knowledge research is organized by task types.

Most of the studies focused on character decision latency and analyzed reaction-time, recall, and response errors (Feldman & Siok, 1997, 1999; Ho & Bryant, 1997; Ke & Wu, 2003; Shen & Bear, 2000; Shu & Anderson, 1999; Taft & Zhu, 1997). Target characters and distracters were created and matched based on several considerations, such as (1) character frequency: high- or low-frequency characters, (2) semantic radical combination frequency, or how often the semantic radicals enter into characters, (3) semantic radical position frequency, or how often the semantic radicals appear at which position in the character , (4) phonetic component combination frequency, or how often the phonetic components enter into characters. The most commonly used tasks and analyses to assess orthographic knowledge are introduced in the following section.

2.6.1 Chinese Character Decision Task

In a Chinese character decision task, learners make judgments about whether or not characters are legal Chinese characters, and both reaction-time and error rates are analyzed. Depending on the researchers' interests, the types of target and/or prime characters are varied. One type of study uses target characters which only include legal characters and pseudo-characters (Feldman & Soik, 1997, 1999; Li & Chen, 1999; Shu & Anderson, 1999; Taft & Zhu, 1997). Pseudo-characters are constructed either by taking legal characters and changing one or more strokes (i.e., ill-formed components), or by combining components in their wrong positions (i.e., ill-formed structure), or by combining components in their legal positions but using combinations that did not co-occur (i.e., well-formed structure) in any actual characters. In all cases, they look like real characters but have no meaning or pronunciation.

Character decision tasks can be used to investigate the age that L1 orthographic awareness develops; in other words, when do children become aware of the different aspects of the structure of characters? Shu and Anderson (1999) investigated 1st, 2nd, 4th, and 6th graders and found that the most interesting aspect of their results was the rate of false alarms that occurred on the pseudo-characters, i.e. when the respondents thought that a pseudo-character was a real character. Even for 1st and 2nd graders, the false alarm rate was very low on items with ill-formed structures (i.e., components are in their wrong positions). Shu and Anderson (1999) believe that this is evidence of children's' early developing insights into the basic structure of characters. In addition, there was a steady decline in the false alarm rate for items with ill-formed components (i.e., changing one or more strokes of a component) from 2nd to 4th to 6th graders. They consider this result to indicate the gradual development of insight into the detailed internal structure of characters.

The Chinese character decision task also can be used to investigate the relationship between position and function of a component. Taft and Zhu (1997) found character decision latencies were faster for characters containing higher-combination frequency components located on the right side of the character. They defined a component that appeared in more than 280 characters as a "high-combination component", and that appeared in fewer than 69 characters as a "low-combination component". Different from Taft and Zhu's (1997) findings, Feldman and Soik (1997) found a main effect for combinability for radicals on the left, but not for radicals on the right. The difference between these two studies is that Feldman and Soik (1997) separated character components into semantic radicals and phonetic components and conducted two separated analyses. On the other hand, Taft and Zhu (1997) did not distinguish a component by its function. In other words, they ignored the important difference in function of semantic radicals and phonetic components in Chinese orthography.

To investigate the radical semantic transparency among native college students, Li and Chen (1999) used the Chinese character decision task and found that characters with semantically transparent radicals were generally recognized more rapidly and more accurately than those with semantically opaque radicals. However, the effect of semantic radical transparency mainly affected low-frequency characters. The researchers suggested that perhaps high-frequency characters can be readily recognized on the basis of the character level activation, so the information carried by the radicals presumably has little chance to show its influence.

Another type of study that involves Chinese character decision tasks uses both target and prime characters to investigate the influence of prime characters on target characters. Usually a prime character is displayed first and followed by a target character with a delay time in msec. The delay time, also known as SOA (Stimulus Onset Asynchrony), is also one of the independent variables in the studies. For example, unlike Li and Chen (1999), who used only target characters in their investigation of radical semantic transparency, Feldman and Siok (1999) investigated native Chinese college students' judgment of whether a target character (e.g. 論 meaning "to say or to talk") contains a transparent semantic radical (e.g., 言 in 論) after seeing one of four prime types:

1. Target and prime characters share an identical semantic radical and are semantically related (e.g., 評 meaning "to comment");
2. Target and prime characters share an identical semantic radical but are semantically unrelated (e.g., 諸 meaning "some or various");

3. Target and prime character do not share an identical semantic radical but are semantically related (e.g., 述 meaning “to speak”);
4. Target and prime characters neither share an identical semantic radical nor are semantically related.

After controlling the character surface frequency and radical-combination frequency of all target characters, Feldman and Siok (1999) reported that primes which were visually similar to the target had a facilitative effect on target recognition latencies at 43 ms SOA and had no effect at 243 ms SOA. With the longer time interval between the onset of a first stimulus and the onset of a second stimulus, only semantically transparent radicals in primes facilitated target decision latencies. When the meaning of the radical was transparent in both the prime and the target, significant facilitation was observed. When the meaning of the radical was transparent in the target but opaque in the prime, inhibition was observed.

In summary, the character decision task has been used to obtain respondents’ judgments about whether or not characters are legal Chinese characters through using distracters as target characters or through using primes as possible influence effects. The research focuses are diverse. In Shu and Anderson’s (1999) study, they manipulated both component and character structures as target characters and found that learners display a gradual orthographic development for the internal structure of characters. Taft and Zhu (1997) and Feldman and Soik (1997) tested the position and function of components. While Taft and Zhu (1997) thought that right-position component knowledge developed faster, Feldman and Soik (1997) concluded that it is the function of radial component knowledge and not position. On the other hand, , Feldman and Soik (1999) carefully designing different types of primes to test whether they facilitated or inhibited judgments of target characters in radical semantic transparency. The researchers found that with the longer time interval between the prime and target character, only semantically transparent radicals in primes facilitated target decision latencies.

2.6.2 Semantic Categorization Task

The semantic categorization task investigates respondents' knowledge of radical meaning. In the task, a respondent would see a semantic category name first, followed by a stimulus, and the respondent had to decide whether or not the stimulus was a member of the semantic category that was previously presented. Three responses "yes," "no," and "do not know", as well as the respondent's reaction times were recorded.

In a study investigating the role of visual and phonological information in lexical access of Chinese compound characters among native Chinese students who were studying at Australian National University, Leck, Weekes, and Chen (1995) used ten semantic categories of compound characters including animals (with 犭 radical), buildings (with 宀 radical), body parts (with 月 radical), fruits (with 木 radical), plants (with 艹 radical), fuel (with 火 radial), cloth materials (with 糸 radical), emotions (with 忄 radical), areas with water (with 氵 radical), and actions (with 扌 radical). Besides the target character (e.g., 狐 hú, meaning "fox"), five distracters were designed in the following ways:

1. Visually similar, and phonologically identical to the target character, V+P+ (e.g., 弧 hú, meaning "arc" with the same phonetic component 瓜 *gū*),
2. Visually similar, but phonologically dissimilar to the target character, V+P− (e.g., 呱 gū, meaning "quack" with the same phonetic component 瓜 *gū*),
3. Visually similar and sharing the same radical component, but phonologically dissimilar to the target character, V_R+P− (e.g., 猜 cāi, meaning "to guess" with the same semantic radical 犭),
4. Visually dissimilar and phonologically identical to the target character, V−P+ (e.g., 湖 hú, meaning "lake" with the same pronunciation as 狐 *hú*),
5. Visually and phonologically dissimilar to the target character, V−P− (e.g., 愣 *lèng*) (Leck, Weekes, and Chen, 1995, p.470).

Leck et al., (1995) found respondents took significantly longer to reject a compound character distracter, but only when it was visually similar and phonologically identical (V+P+ & VR+P−) to its target exemplar. They also pointed out that the radical component is very important for retrieving the meaning of a compound character. However, this finding cannot be taken as evidence that compound characters are necessarily processed visually. If visual similarity between a distracter and a target exemplar (e.g., 狐 hú, meaning "fox") is mainly responsible for the delay in response latency, then both V+P− (e.g., 呱 *gū*) and VR+P− (e.g., 猜 *cāi*) distracters should cause similar delay in the rejection of the distracters. The fact is that only V_R+P− showed a longer delay in response latency. Leck, Weekes, and Chen (1995) also found that the recognition of a Chinese integral character depends primarily on visual information, whereas the recognition of a Chinese compound character relies on visual, phonological, and semantic information. They concluded that visual information plays a greater role in Chinese character recognition.

2.6.3 Semantic Radical Judgment Picture-Matching Task

The semantic radical judgment picture-matching task assesses a learners' knowledge of the meaning of semantic radicals by using pictures to represent the meanings of the radical. Ho, Yau, and Au (2003) used 24 test items, each consisting of a Chinese semantic radical and four picture-options. The target radical was placed in a box in its usual or legal character position next to four picture-options. The children were asked to choose a picture that was related to or best represented the meaning of the radical in each item. However, Ho et al., (2003) did not provide the types of semantic radical that were used in their study. Those radicals would have to, for the task to succeed, be semantically transparent and concrete radicals. The mean (standard deviation) results for kindergartners, 1st graders, and 3rd graders were 14(3.52), 17.8(3.24), 21.70(1.98). The results indicate that children make statistically significant gains in the knowledge of the meanings of individual semantic radicals as they advance to higher grades. This task is specifically designed to assess the children's knowledge of radical semantic transparency. As stated, one limitation on this study is the fact that the stimuli had to be restricted to radicals which expressed a concrete meaning and were semantically transparent.

2.6.4 Chinese Phonetic Reading Task

The Chinese phonetic reading task measures the learners' skill in reading phonetic components in isolation. In Ho, Yau, and Au's (2003) study, 21 phonetics that were within the children's reading vocabulary were selected. Some of the phonetic components were pronounceable in isolation while others were not. A phonetic component was considered to have been read correctly if it was read as a free standing radical character or had a pronunciation that was the same as another real character containing the phonetic. The mean (standard deviation) results for kindergartners, 1st graders, and 3rd graders were 7.65(3.80), 11.40(4.52), 18.40(1.27). The results again indicated that children made significant gains in knowledge of the pronunciation of individual phonetics as they advance to higher grades. However, Ho et al., (2003) did not provide the specific types of phonetic components that were used in their study. In addition, they did not separate the results according to whether the children read the phonetic component as a free stand character or read the phonetic component by providing another character that contained the phonetic component. Since not all characters have reliable phonetic components, rubrics in the task should be more specific to reflect the phonetic component reliability.

2.6.5 Chinese Pseudo-Character Reading Task

The Chinese pseudo-character reading task examines learners' composite knowledge of the position, function, and sound value of phonetic components (Ho & Bryant, 1997; Ho, Yau, & Au, 2003). In a study by Ho, Yau, & Au (2003) , 25 Chinese pseudo-characters, which had no real meaning, were created by combining a semantic and a phonetic radical in their legal positions. All the semantic radicals and phonetic components included were real characters in isolation and were within the learners' reading vocabulary. The phonetics were chosen from phonologically regular and consistent Chinese compound characters. A pseudo-character was considered to have been read correctly if it was pronounced by its phonetic radical or by the name of a character having the same phonetic radical as the pseudo-character. Although research results suggest that children may not use the phonetic radicals effectively to name novel characters until 3rd grade, tasks that simply ask respondents to read or invent pseudo-characters (a task to be introduced next) are not authentic. We therefore have to consider the possible misleading influence in character learning and teaching.

2.6.6 Chinese Pseudo-Character Construction Task

The Chinese pseudo-character construction task examines the learners' composite knowledge of the position, function, and sound value of character components by asking learners to invent new characters. Ho, Yau, and Au (2003) showed 12 pictures of strange objects to children. Twenty stroke-patterns, ten semantic components (each related to the meaning of one picture) and ten phonetic components (each related to the name of one picture), were printed on a sheet of paper to help the children invent new characters. In each trial, the experimenter pointed to a picture, provided the pronunciation of it, and asked the children to invent a new character for the object. Participants were asked to combine a semantic with a phonetic component in their legal positions to construct a compound character. Considerations were given to (1) the knowledge of character structure, (2) the knowledge of positional constraints, and (3) the knowledge of the relative functions of the semantic and phonetic components. The results showed that the kindergarteners did significantly less well than the 1st and 3rd graders, who did not differ from each other.

2.6.7 Two-Character Word Reading Error Analysis

Based on the assumption that response errors reflect the kinds of strategies that learners employ for reading Chinese characters, researchers have analyzed reading and spelling response errors to reveal character-recognition strategies (Ho & Bryant, 1997; Shen & Bear, 2000). Ho & Bryant (1997) asked 1st and 2nd graders in Hong Kong to read two-character words aloud one by one. The task was discontinued when the child failed to read 10 consecutive words. One point was given for each character correctly read in a word. They classified errors into six types:

1. Phonetic-derivation errors: the pronunciation of the phonetic component (e.g., 廣 *guǎng* in 擴) was used incorrectly as the pronunciation of the whole character (e.g., 擴 *kuò*);
2. Phonetic-analogy errors: the target character (e.g., 怕 *pà*) was read as another character (e.g., 伯 *bó*) having an identical phonetic component (e.g., 白 *bái*);
3. Radical-related errors: either pronunciation of the radical component (e.g., 火 *huǒ*) was used as the pronunciation of the whole character (e.g., 燈 *dēng*), or the target character (e.g., 燈 *dēng*) was read as another character having an identical radical component (e.g., 炷 *zhù*); and
4. Word-related errors: the target character (e.g., 歌 *gē*) was read as the other character (e.g., 唱 *chàng*) in a multicharacter word (e.g., 歌唱 *gē chàng*) that contained the target character .
5. Did not know: the child just said that s/he did not know how to read the character; and
6. Others: errors other than the above categories.

Ho & Bryant (1997) found that the most common type of error (34.3%) was phonetic-related errors, while radical-related errors accounted for only 3.8% of the total errors.

2.6.8 Character Spelling Error Analysis

Shen and Bear (2000) also used error analysis to examine Chinese 1st to 6th graders' writing samples. They classified three error categories: phonologically based, graphemic, and semantic spelling errors including 15 error types. They found that, similar to the previous findings, phonologically based errors predominated (79.6% of all errors). Phonologically based errors included pinyin substitution and homophonic character substitution. Pinyin substitution errors refer to children using pinyin (e.g., *děng*) to substitute for characters (e.g.,等). Homophonic character substitution errors refer to children substituting homophonic characters for the target characters. Shen and Bear (2000) reported that as grade level increases, the percentage of the use of phonological based errors decreased from 95.75% to 54.36%, while the use of graphemic and semantic strategies increased. In graphemic spellings, children tried to spell the target character with an invented or substitute character that was as visually similar as possible. In semantic spellings, children used meaning-similar characters or created morphologically related new characters to substitute for the target characters. They suggest that the reason for these results is that children in the lower grades know only a limited number of characters so when they needed to spell a new word that they did not know, they relied more on available phonological knowledge than on morphological knowledge. However, it is not clear why beginning spellers produced only a small percentage of graphemic errors (e.g., 3.75% for the first-graders) when these children had just started to learn the visually complicated Chinese characters.

2.7 Chinese Character Selection in CFL Textbooks

Looking at the history of language teaching methodologies over the past 100 years, language educators, applied linguists, and researchers have been in pursuit of the best method. Language learning methodologies or pedagogical principles influence textbook development and also influence the selection of words and characters to be included in textbooks (Nation, 2001; Richards, 2001). In addition to being selected from frequency lists, vocabulary was selected based on grammar points, communication skills, themes, and authentic materials that were introduced, which means that relatively complicated characters are among the first characters that learners encounter.

In the 1950s and 1960s, audiolingual textbooks used a set plan for selecting and limiting vocabulary. In addition to a frequency criterion, vocabulary was selected according to an expanding scope for the learners. The lessons began with vocabulary of the classroom, then school, home, community, and work. Later, the list was enlarged to include common vocabulary about the state, nation, and the world. In addition, the number of vocabulary items per lesson was kept to the minimum, so learners could concentrate on pronunciation and grammar (Hatch & Brown, 1995). For decades, the most widely used CFL textbooks in the United States were the conversation series and reading series by John DeFrancis. DeFrancis was a leading scholar and author of 12 series of textual materials for teaching spoken and written Chinese, and many of his textbooks are still used in language programs today. DeFrancis (1966) explained that one of the beliefs underlying the *Beginning Chinese Reader* was that learning to read can be accomplished most efficiently by students who have developed some level of oral proficiency and who engage in simultaneous oral practice of what is read. In the present study, I selected audiolingual textbooks *Character Text for Beginning Chinese* from the 2nd edition conversation series, *Beginning Chinese Reader* from the 2nd edition reading series, and *Read Chinese* series published in the 1960s and 1970s by Yale University Press.

In the late 1960s and early 1970s, language teaching turned to more communicative materials including three approaches. The first approach was survival-based, and vocabulary was selected for someone learning survival skills in the new country including units on obtaining housing, doing banking, shopping, getting a driver's license, and filling out job applications. The second approach was theme-based, with the selected vocabulary relating to social issues, global issues, world peace, the environment, health, and human rights etc. The third approach was notional-functional, and the vocabulary related to different notions and functions of the language, including exchanging factual and intellectual information, exchanging emotional and moral attitudes, etc. Pure communicative materials are rare, and most of the communicative materials were combined with audiolingual practice. In this study, I selected *Practical Chinese Reader: Componentary Course* (1995), *New Practical Chinese Reader* (2002), and *Far East Everyday Chinese* (2008).

Communicative approaches continued through the 1980s and 1990s with an increasing focus on authentic materials, and the vocabulary materials ideally reflect the needs and interests of the students and their teachers to include menus, labels, advertisements, various types of forms to be filled out etc. In this study, I selected *Communicating in Chinese: Reading and Writing* written by Cynthia Ning, published in 1994. I also selected *Integrated Chinese* 2nd Edition (2005) and *Integrated Chinese* 3rd Edition (2008) which were based on integrating teaching approaches and used simulated authentic materials.

By the late 1970s and early 80s, American government programs impacted the field of methodology as the oral proficiency interview of the Interagency Language Roundtable (ILR) was adapted and developed into the Oral Proficiency Interview (OPI) of the American Council on the Teaching of Foreign Languages (ACTFL) and the Educational Testing Service (ETS). This development opened up the area of oral proficiency testing in academia. Later, there were two important publications: *Standards for Foreign Language Learning: Preparing for the 21st Century* (ACTFL, 1996) and *Standards for Foreign Language Learning in the 21st Centur*y (ACTFL, 1999). These *Standards* specified the content that students should know and the things they should be able to do as a result of foreign language instruction. They did not identify or specify methods by which the students should come to such knowledge or capability. Hence, the use of a particular methodology has not been as significant since the beginning of the Proficiency Movement. In this study, I selected *Practical Audio-Visual Chinese* (2008) which was based on the considerations of the AP (Advance Placement) Language Curriculum and "5 C's" of the National Foreign Language Standards in the United States.

In summation, it seems clear that high frequency words are likely to predominate teaching materials at the early stage of learning and teaching. In addition to being selected from frequency lists, vocabulary was selected based on grammar points, communication skills, themes, and authentic materials that were introduced, which means that relatively complicated characters could be among the first characters to be learned.

2.8 Chinese Character Selection for Teaching and Learning Purposes

Research results indicate that both CFL students and teachers believe that character learning and writing are the most difficult tasks in learning Chinese (Everson, 1998; Ke, Wen, & Kotenbeutel, 2001). To ease the learning burden in the beginning, some educators suggest postponing the teaching of characters until learners have a base of spoken language upon which to build character learning; others are eager for a more analytical approach to teach characters. Researchers have conducted component analyses of Chinese characters and provided us with structural knowledge about Chinese characters (Chen, 1997; Fei, 1996; Fu, 1989, 1991, 1992, 1993; Guder-Manitius, 1999; Kang, 1993; Li & Kang, 1993; Xiao, 1995; Zhu, 1993), and structural knowledge about Chinese characters helps material developers and teachers to select and adjust the characters that students should learn. These analyses also provided test developers and researchers sets of characters that should be assessed.

2.8.1 Character Component Segment

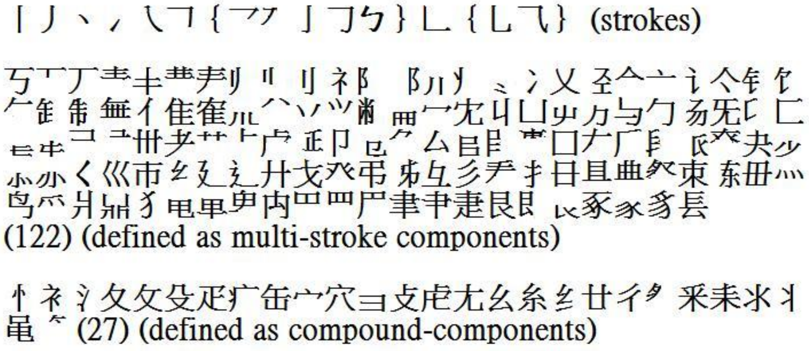
In most CFL textbooks, characters are introduced as integral or compound characters. In general, characters which can be decomposed into other components are defined as compound characters. However, what does it really mean when we say a "character can be decomposed" ? Researchers who study modern Chinese use the "character component segment" 部件 (*bùjiàn*) concept to distinguish integral and compound characters. Xiao (1994) stated that a character which can be decomposed into at least two independent component segments is a compound character. On the other hand, a character which is composed of only one component segment is seen as an integral character. In addition, if we decompose the component segment of an integral character, it would be reduced to strokes only. With this definition of compound characters, it is important to define "component" both in computer keyboard input and language education fields. Due to different usages of the component, there is still no consensus of the component definition (Cui; 1997; Fei, 1996; Fu, 1992; Guder-Manitius, 1999; Su, 1995; Xiao, 1993, 1994, 1995). For computer input usage, a character is decomposed into more component segments compared to that in language education usage, and many of them are non-character components. For language education usage, researchers are concerned about the "form, pronunciation, meaning 形音義" of a Chinese character, so a character needs to be decomposed into meaningful components as possible. Among researchers who discus the segment rules of Chinese characters, Xiao's (1993, 1994, 1995) component definition is systematic and easy to apply in decomposing characters. According to Xiao (1994), four component structures are defined:

1. Strokes go across each other: for example, the character 十 is composed of stroke 一 and stroke 丨, and they go across each other. Another example is the non-character 艹; composed of strokes 一, 丨, and 丨, and they go across each other. Other examples are 力, 丈, 女, and 扌. Components can be characters or non-characters.
2. Some strokes go across and some strokes are connected: for example, the character 千 is composed of strokes 丿, 一, and 丨; 丿 is connected with 十. 千 cannot be decomposed into 丿 and 十 because 丿 is a stroke which cannot stand alone as a component.

3. All strokes are connected, and the components either have spaces within a character or appear in more than two characters: for example, the character 竹 is easily decomposed into the left and right two components because there is space between them. Another example, 丆 is composed of two strokes 一 and 丿, and they are connected, in addition, 丆 can be found in more than two characters 而, 頁, and 百, so 丆 is a component.
4. All strokes are separated, but they always group together to form characters: for example, 氵， 刂, 巛, 小, and 心.

The main segment rule is that a character can be decomposed into components, and components can be further decomposed into strokes. As stated, the different component definitions contribute to different component lists. Table 2.1 presents component lists from Fei (1996) and Xiao (1995). Both studies investigated the same 3,500 most commonly used characters. Fei (1996) found 384 components including basic components and compound components (p.26), with the basic component including 162 character-components. Fei (1996) classified strokes as one-stroke components and mutli-stroke components. He also included multi-component segments as compound components, such as 曾比北熏曹惠鼠兼妻叟, while Xiao (1995) did not classify them as components. In addition, Fei (1996) separated some grouped-together non-character components into smaller non-character components. For instance, he classified the first top segment of character 虎 as a basic component and the first two top segments 虍 as compound components. The first segment classification is redundant because the first top is always grouped together with 匕 to form characters. Strangely, he classified the left part of 巸 as a character-component which is not a character itself. Xiao (1995) found 474 components but only listed 195 character-components in his paper (p.59). In the Xiao (1995) list, 水 is not listed in the 195 character-components, but 水 should be seen as a character-component based on his segment rule. Comparing both lists, 135 character-components were found in both studies, and many of the components from both studies are radicals. These characters should be introduced to learners at the initial stage of learning since they are also components that can be seen in many characters. In addition, the number of non-character components is about the same amount as character-components. Learners would expect to be frequently exposed to these non-character components. However, these studies investigated simplified characters, so it is important to know whether these components would be different compared to traditional characters.

Table 2.1 Component Lists in Fei (1996) and Xiao (1995)

Studies	Type	Component
Appeared in Fei (1995) and Xiao (1996)	Character-component	一二厂丰垂五儿八又及乃丈乙九丸几专丁于寸才乎事了子亥人入内两丹册甩甫舟门巾凸凹刀力刁勿匆匕七也屯毛氏民臣巴牙十井甘世卜四中串虫史吏日目且身曲更田由甲申电里重车皿母夷夫央女川州我书瓦心必手乐东柬为爪瓜片耳秉臼长西龙 (106) 户工土士王业广用山己巳禾止口白自血夕歹大犬小尸戈弓月木米火 (29) (defined as compound components in Fei (1996))
Appeared in Xiao (1996)	Character-component	万严玉三之义叉凡飞亡个丙亚半再刃办勺无巨匹千上正丐下不卫久尤丘乒乓太夹天夭失尺币平乡升开成互未末本术朱束来果丑灭牛生良韭 (60)
Appeared in Fei (1995)	Character-component	皮予鹿以年曾发冉丽农出司比北尹熏免争奂斥局其革堇弋戋弗韦曰曹曳水丞承永求乌黑惠见禹禺兼庸妻鼠叟酉雨面而豖象乍非熙? (56) (defined as character-components) 石麦青斗文风足走矛言立音辛方麻矢父谷羊羽己老舌卤齿角鱼欠斤示赤鸟马鬼页贝竹衣豆辰(40) (defined as compound components)
Appeared in Fei (1995)	Non-character component	 丨丿丶㇀㇏㇕{㇖㇇ 亅㇆㇉}㇗{㇄㇈} (strokes) (122) (defined as multi-stroke components) 忄衤氵夂攵殳𧾷疒缶宀穴彐支虍尢幺糸纟廿彳⺈ 釆耒氺丬龟⺈ (27) (defined as compound-components)

2.8.2 Character Component Analysis and Pedagogical Components

To ease the learning burden at the initial learning stage, some educators suggest postponing the teaching of characters until learners have a base of spoken language upon which to build character learning; others recommend using character components to teach characters in the CFL context (Cui, 1997; Guder-Manitius, 1999; 中華人民共和國教育部國家語言文字工作委員會, 2009). Cui (1997) pointed out four concepts to support component teaching:

1. In terms of memory chunking, fewer chunks can facilitate memorization. A character can be decomposed into strokes or components. The average number of strokes in a character is about 7 strokes. Cui (1996) investigated *A Level List of HSK Word and Character* 漢語水平詞彙與漢字等級大綱 *Hànyǔ shuǐpíng cíhuì yǔ Hànzì děngjí dàgāng* (1992) and found the average number of components in characters to be less than 3 components. Therefore, using components to memorize characters is better than using strokes.
2. In terms of phonological processing, the pronunciation of a component can facilitate character memorization. Therefore, those components that can be pronounced are easier to remember than those components that cannot be pronounced. Cui (1996) found 70% of character components can be pronounced.
3. In terms of semantic processing, the meaning of a component can facilitate character memorization. Cui (1996) found 68.5% of character components are meaningful.
4. Through error analysis, CFL learners' written errors are mostly related to components; i.e., they either use the wrong component or place the components in the wrong positions.

Conducting a component analysis is a time consuming task. Researchers have conducted component analysis not only to identify component segments, but also to investigate semantic and phonetic components. Kang (1993) and Li and Kang (1993) used 7,000 of the most commonly used characters from 現代漢語通用字表 (*Xiàndài Hànyǔ tōngyòng zìbiǎo*) to investigate the semantic and phonetic components in 5,631 semantic-phonetic compound characters, and they listed 246 semantic components (p.74) and 1,119 phonetic components (pp.87-91). Kang (1993) found 10 highly transparent semantic components including 氵 (water), 艹 (plant/grass) , 口 (mouth), 扌 (hand), 木 (wood), 金 (metal/gold), 亻 (human), 虫 insect / reptile, 言 (words / to speak), and 土 (dirt / earth). Li and Kang (1993) found 37.5% phonetic components pronounced the same as the semantic-phonetic compound characters and 18.2% phonetic components pronounced with different tones from the characters' pronunciation. Li and Kang (1993) concluded that most of the phonetic components are reliable (55.68%). While Li and Kang (1993) did not list the reliable phonetic components in their paper, Chen (1997) lists 151 reliable phonetic components, and he suggests this list can be used for character teaching and research reference (pp.32-33).

Fu (1989) edited the Attributive Dictionary 漢字屬性字典 *Hànzì shǔxìng zìdiǎn* to include 6,763 characters. These 6,763 characters were taken from GB2312-80 信息交換用漢字編碼字符集-基本集 (*Xìnxī jiāohuàn yòng Hànzì biānmǎzìfújí –jīběnjí*, Chinese character code symbol collection for information exchange – basic edition). The dictionary includes characters and radicals for computer information exchange usage. Each character was categorized with 24 attributives including pronunciation, stroke, stroke order, radical, component, structure, frequency of usage, national code, etc. In his Attributive Dictionary, Fu (1989) used stratified analysis 層次分析法 (*céngcì fēnxifǎ*) to analyze character structures. As stated in chapter 1, in the stratified analysis, a character is decomposed layer by layer until all components are minalcomponents which cannot be further decomposed. For example, the character 部 (*bù*), shown in Figure 1.1, is first decomposed into the left part 音 and right part 阝, then 音 is further decomposed into 立 and 口, so the character 部 is composed of three components 立, 口, and 阝, and is a left-right structure character. Fu (1991, 1993) defined 13 structures. In addition, Fu (1992, 1993) used 7,000 of the most commonly used characters from 現代漢語通用字表 (*Xiàndài Hànyǔ tōngyòng zìbiǎo*) to investigate the character component positions among character structures, and listed the high-combination components (Fu, 1993, pp. 123-152), but did not further explain any implications of his list. Both his Attributive Dictionary 漢字屬性字典 and the high-combination component positions among character structures list can be used for further quantitative and qualitative Chinese character research and character selection for teaching resources.

Guder-Manitius's (1999) component analysis of Chinese characters *Sinographemdidaktik* provides us with a complete structural knowledge about Chinese characters compared to the component analyses stated above. Guder-Manitius (1999) conducted a component analysis of 3,867 characters taken from the Attributive Dictionary 漢字屬性字典 *Hànzì shǔxìng zìdiǎn* (1989). He found that the 3,563 characters are compound characters which can be further divided into at least two components, with 1,403 components being identified. If a phonetic component appeared in at least 3 characters, according to Guder-Manitius's definition, the phonetic component was seen as being of pedagogical relevance, and he found 182 pedagogical relevant phonetic components. If a graphic component appeared in at least 3 characters and 5 out of 10 raters put them in similar semantic categories, the graphic component was seen as being of pedagogical relevance, and he found 122 pedagogical relevant semantic components. Guder-Manitius (1999) recommends that the selection of characters introduced should as far as possible consist of high frequency characters that also serve as components and which contain as few strokes as possible. Using this approach, the students will later find it easier to remember more complex characters where these components are included and will experience a high rate of recognition when they look through a standard text. Guder-Manitius also feels that this system will give students a higher motivation for learning.

However, among the 1,403 components identified by Guder-Manitius, there were 509 basic components which included only one component, and 304 out of these were characters. If we use the minimal component analysis approach and the component definition by Xiao (1994), many of the 509 basic components are composed of more than one component. Acceding to Xiao (1994), the main component identification rule is that a character can be decomposed into components, and components can be further decomposed into strokes. The minimal component structure is not the stroke. For example, Guder-Manitius includes 主 as one-component, but actually 主 is composed of two components 亠 and 土; other examples are 斗(冫 十), 可(丁口).

Table 2.2 Pedagogical Phonetic and Semantic Components in Three Component Studies

Studies	Component
Appeared in Guder-Manitius(1999) and Chen (1997)	Phonetic-component (76 listed in Guder-Manitius order) 包方[1]交干青中甫安半辟同及旁乔(喬[2])宛正帝胡巨仑(侖)曼朋宾(賓)冈(岡)朵勾留罗(羅)齐(齊)容太相星曾章直朱宗当(當)邦采成垂府旱皇焦考览(覽)末宁刃唐廷朝崔代段伐侯建竟卷康历(歷)蒙农(農)彭式署斯胃凶夜查丈
Appeared in Guder-Manitius (1999) and Kang (1993)	Semantic-component (93) 一人亻儿冫 刂刀力匚(匚匸[3])十厂口囗土女子宀尸山巾广弓彳忄心戈户(戶)手扌攵日月木歹气氵水火灬爪父犭王玉甘田疒皿目石礻禾穴米纟(糹)羽耒耳舟艹虫衤衣讠(言)贝(貝)走身车(車)辶阝[3](阝 邑阜)酉钅(釒)隹青革音页(頁)风(風)饣(飠)马(馬)骨髟鬼鱼(魚)鸟(鳥)黑齿(齒)光帛林男㔾刂

Note. [1]Shading component indicates the component is also a radical. [2] (齊) is the traditional form of the component. [3] Both studies treated 匚(basket) and 匸 (box) the same radical 匚. In addition, they treated 阝(邑) and 阝(阜) the same radical 阝.

Comparing Guder-Manitius's (1999) pedagogical phonetic and semantic components to Chen (1997) and Kang (1993), Table 2.2 presents the same components among these studies. For the phonetic component, 76 phonetic components were found in

Guder-Manitius (1999) and Chen (1997), and five of them are radicals. For the semantic component, 93 semantic components were found in Guder-Manitius (1999) and Kang (1993), and 87 of them are radicals. This indicates that many radicals are still severed as semantic radicals in modern Chinese characters. As mentioned, Cui (1997) stated that components which can be pronounced and have specific meanings facilitate character learning. Radicals are potential target components that can serve as pedagogical components because radicals have both pronunciation and meanings.

In summary, component analysis of Chinese characters provides us with structural knowledge about Chinese characters and with pedagogically relevant character components to be selected for teaching and learning Chinese characters. These analyses served as the comparison base for the current study. However, most of the component analysis studies have examined Chinese characters from dictionaries or corpus data, and they all targeted simplified characters. Rarely have studies investigated characters from beginning level textbooks. In the CFL context, since reliable target language input is limited largely to textbook materials and teacher instruction, a component analysis of the characters which are presented in beginning level CFL textbooks is needed, so we can discover whether CFL learners have opportunities to develop orthographic awareness. From such investigations, we will be able to build better models of developing CFL orthographic awareness and character learning pedagogy.

CHAPTER 3

RESEARCH METHODS

The purpose of this study is to systematically describe and classify Chinese characters and words in CFL textbooks for college and adult beginning learners, and draw conclusions as to their helpfulness in developing orthographic awareness among first-year CFL learners. The main focus was to make a component inventory of characters and to discuss textual materials availability in the following areas: (1) explicit orthographic decomposition instruction; (2) character diversity and repetition; (3) character frequency selection across textbooks; (4) radical component diversity and repetition; (5) phonetic component diversity and repetition; and (6) ideal semantic transparent radical and reliable phonetic element characters.

A component analysis was used to address these issues. This chapter describes the selection of the textbooks to be examined, the selection of documented frequency lists of character usage and dictionaries, the development of word entry and character database, the development of coding methods, analysis, and determining accuracy and reliability of coding.

3.1 Textbook Selection

Looking at the history of language teaching methodologies over the past 100 years, language educators, applied linguists, and researchers have been in pursuit of the best method. As stated in section 2.6, language learning methodologies or pedagogical principles influence textbook development and also influence the selection of words and characters to be learned in the textbooks over time. In addition to being selected from frequency lists, vocabulary has been selected based on grammar points, communication skills, themes, standards, and authentic materials, which means that relatively complicated characters are often among the first characters to be learned. In this study, the ten CFL textbooks that were selected for analysis are described below in order of their publication date:

1. *Read Chinese: A Beginning Text in the Chinese Character*, Book 1 華文讀本 (RC$_{1961}$): This textbook, published in 1961 by Yale University Press, is based on the assumption that before they start this textbook, beginners have covered the first 12 lessons of *Speak Chinese*, a text with the Chinese language presentation being done exclusively in Romanized Chinese. Therefore, students have a certain familiarity with sentence structure and a modest speaking vocabulary before they are introduced to the characters. Lessons begin with vocabulary and sentences, followed by story texts. Within sentences and texts, some characters are shown in Romanization only, and Romanization will not be treated as characters for analysis in the current character database. For example, kè in the word 上 kè (課) will not be included as a Chinese character for analysis.
2. *Character Text for Beginning Chinese* 初級漢語課本漢字本 (CTBC$_{1976}$): Twenty-four lessons are included in the *Character Text for Beginning Chinese*, and each lesson introduces between 19 and 33 characters. Each lesson begins by introducing vocabulary, followed by dialogue presentation, and sentence build-up and pattern drills. Such a pedagogical strategy is indicative of the Audiolingual methodology.
3. *Beginning Chinese Reader* (BCR$_{1977}$): DeFrancis (1966) explained the beliefs underlying the *Beginning Chinese Reader*, published in 1977 by Yale University Press, by saying that learning to read can be accomplished most efficiently by students who have some prior grounding in speech and who engage in simultaneous oral practice of what is read. Accordingly, *Beginning Chinese Reader* was matched with *Character Text for Beginning Chinese* and contains dialogues as well as narrative and expository material.

4. *Communicating in Chinese: Reading and Writing* 漢語交流-初級漢語閱讀書寫本 (CICrw$_{1994}$): This textbook, written by Cynthia Ning in 1994, aims first to teach learners to understand pieces of simple written Chinese texts, such as signs, schedules, advertisements, etc., then to teach learners to convey simple messages by writing Chinese. Students learn to form basic characters and string characters together in meaningful sequences. This is a significant book in that it is one of the first books based more on a communicative approach.
5. *Practical Chinese Reader: Componentary Course*, Book 1 實用漢語課本 (PCR$_{1995}$): This textbook, compiled by the Beijing Language Institute in 1995, aims to teach the learners speech forms, so most of the texts are written in a dialogue format, thus facilitating audiolingual practice. Lessons include texts, new words, notes, pronunciation and intonation, conversation, phonetics, grammar, reading and other exercises. In Book 1, the context is specifically designed for CFL learners to use Chinese in their own countries. This is not the first edition of PCR. When PCR was first published, it was actually the first Chinese language textbook to come out of the People's Republic of China, reflecting the thaw in relations between China and the United States.
6. *New Practical Chinese Reader 1* 新實用漢語課本 1 (NPCR$_{2002}$): Since the first edition was published in 1981 by Beijing Language and Culture University Press, *Practical Chinese Reader* has been used by Chinese language educators and learners worldwide. The new edition, published in 2002, is based on an integrated teaching approach to emphasize the communicative function of language and to ensure that learners obtain a firm grasp of language structure. Lessons include dialogues, vocabulary, notes, conversation practice and drill, reading comprehension and paraphrasing, phonetics and pronunciation drill, grammar, character, and cultural notes. The textbook was written in simplified characters with traditional characters in the

vocabulary index. In this study, traditional characters were used for characteristic classification.

7. *Integrated Chinese* 2nd Edition, Level 1 中文聽說讀寫 (IC2005): Since the first edition published in 1997, *Integrated Chinese* has been the best-selling Chinese introductory series in the United States. Based on integrating teaching approaches and using simulated authentic materials, this textbook is designed to develop four language skills and to use Chinese in real life situations. The integrative teaching approach mixes a communicative approach with grammar-translation and direct method. The 2nd edition, published in 2005, featured relatively minor changes and adjustments in the series.
8. *Integrated Chinese* 3rd Edition, Level 1 中文聽說讀寫 (IC2008): The 3rd edition of Integrated Chinese is the result of an extensive revision of the second edition (See item 7). For example, selected lessons were removed so teachers are able to finish all the lessons in Level 1 within one academic year, and students can have an appropriate amount of lessons and time to learn all the content. Some words and expressions that are used relatively less frequently have been deleted. In addition, this edition uses colors to highlight different components of each lesson and uses brand-new illustrations and photos to complement the content of the text.
9. *Practical Audio-Visual Chinese*, Book 1 新版實用視聽華語 (PAVC2008): Based on considerations of the AP (Advance Placement) Language Curriculum and “5 C’s” of the National Foreign Language Standards in the United States, this new edition textbook, published by National Taiwan Normal University in 2008, is designed to teach learners basic pronunciation, grammar, and vocabulary and practice, along with practical application

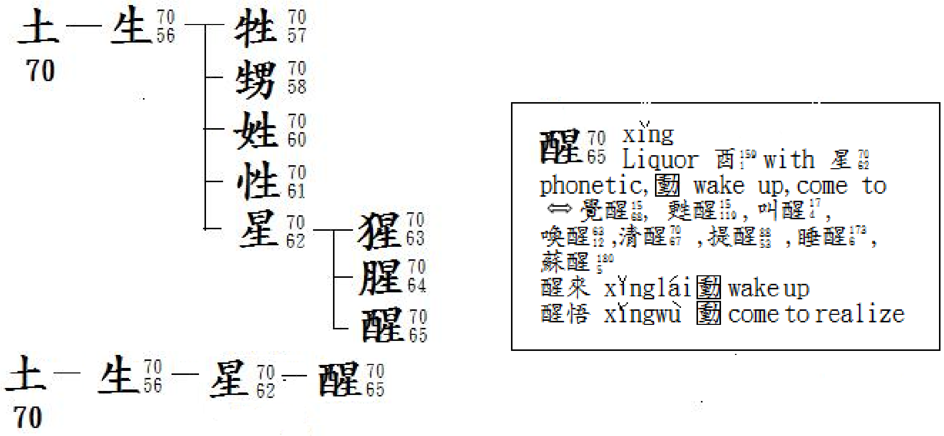
activities. Three phonetic symbol systems, including Zhuyin-Fuhao, Taiwan Tongyoung Pinyin, and Hanyu Pinyin, are presented in each lesson.

10. *Far East Everyday Chinese* 遠東生活華語 (FEEC2008): Based on a communicative task-based approach, this textbook, published in 2008, is designed to help learners develop the four language modalities-listening, speaking, reading, and writing. The lessons include dialogues, vocabulary, grammar, aural comprehension drills, games, role play, guessing and authentic materials.

3.2 Dictionary and Frequency List Selection

To categorize and investigate the Chinese characters selected in the textbooks described above, I used dictionaries to classify character characteristics and documented frequency lists to classify character usage frequency. The dictionaries and frequency lists were selected and described in the following:

1. *New Chinese Dictionary* 最新國語新辭典 (1987): This dictionary is a traditional radical-based dictionary using the Zhuyin system to obtain character pronunciation. This dictionary was used to classify the semantic radicals of characters, meanings of the semantic radicals, and Zhuyin of the characters.
2. *Chinese Characters: A Genealogy and Dictionary* 中文字譜 (1998): Most of the traditional radical-based or pinyin-based Chinese-English dictionaries do not provide the etymological explanations of the characters. In the current study, I used this etymological-based dictionary, compiled by Harbaugh (1998), to investigate phonetic component derivation and analogy. The genealogical tree (chart), as shown in Figure 3.1, highlights the connections between characters and allows a character to be found by analyzing any character component.



醒 70 65 xǐng
Liquor 酉 150 1 with 星 70 62
phonetic, 動 wake up, come to
⇔ 覺醒 15 68, 甦醒 15 119, 叫醒 17 4,
喚醒 63 12, 清醒 70 67, 提醒 99 53, 睡醒 173 6,
蘇醒 180 5
醒來 xǐnglái 動 wake up
醒悟 xǐngwù 動 come to realize

Figure 3.1 An Example of Character Etymological Tree (Chart) and Character Entry in 中文字譜 Chinese Characters: A Genealogy and Dictionary (Harbaugh, 1998)

For example, the etymological tree root (on the left side of Figure 3.1) for the character 醒 can be traced back from 星, 生, to 土. There are two numbers listed next to character 醒: 70 and 65. The top number 70 is the etymological tree root number which can be used to find characters that share the same components, such as 醒, 猩, and 腥 which share the 星 component. If we want to trace back further, 星, 性, 姓, 甥, and 牲 share the same 生 and 土 components. The bottom number 65 next to 醒 indicates this character is the 65th in that etymological tree family. This dictionary also was used to classify the three types of characters in 六書(The Six Books) and its phonetic component. Characters which belong to Pictograph, Ideograph, and Phonetic Complex (Semantic-Phonetic character 形聲字) categories were identified in this dictionary. From a character entry (on the right side of Figure 3.1 above), we can find the explanations of character's compositions. For example, 醒 is a combination of its semantic radical 酉 (meaning "liquor") and its phonetic component 星, so 醒 is a Phonetic Complex character which is indicated by its phonetic in the character entry. The usage of the character characteristics classified by this dictionary will be further explained in the data analyses sections.

3. Chinese Character Attributive Dictionary 漢字屬性字典 Hànzì shǔxìng zìdiǎn (1989): A code-oriented dictionary edited by 傅永和. This dictionary also contains frequency rank, and this frequency rank was used to classify frequency of usage.
4. Wenlin 文林 Software for Learning Chinese, version 3.4 (2007): In the current study, I used the "Characters by frequency" list from Wenlin 文林 Software for Learning Chinese to classify frequency of usage. Wenlin includes an expanded and improved version of the already huge ABC Chinese-English Dictionary edited by John DeFrancis, giving a total of over 10,000 characters and approximately 200,000 words and phrase.
5. A Frequency Dictionary of Mandarin Chinese (2009): A more recent frequency dictionary complied by Xiao, Rayson, & McEnery (2009) was used to classify frequency of usage as well. This dictionary is based on a 50-million-word corpus composed of spoken, fictional, non-fictional and news texts in current use. The dictionary also contains 30 thematically organized lists of frequently used words on a variety of topics such as food, weather, travel, and time expression.
6. A Level List of HSK Word and Character 漢語水平詞彙與漢字等級大綱 Hànyǔ shuǐpíng cíhuì yǔ Hànzì děngjí dàgāng (1992): HSK is a set of Chinese proficiency tests that aims to assess non-native speakers' capabilities of applying Chinese language in life, study, and work. In 1992, the department of HSK in China published this word and character level list as teaching and learning guidelines. There are 4 levels (甲 jiǎ, 乙 yǐ, 丙 bǐng, 丁 dīng) and 8,822 words and characters in this HSK list. There are 2,905 Chinese

characters in 4 levels. For the basic level, test-takers are expected to know 800 characters. In this study, I classified each character's level based on this HSK list to see whether all the characters from the ten beginning-level textbooks covered the basic level character expectation.

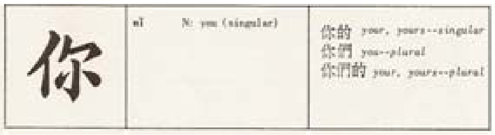
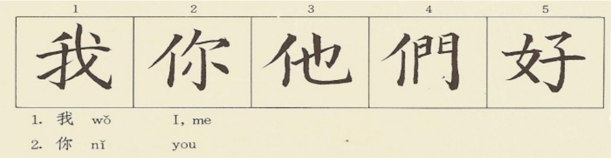
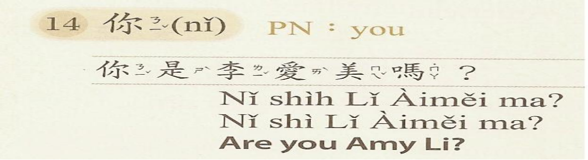
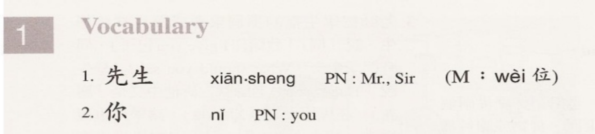
7. Chinese Character Usage Frequency Lists from Google and Yahoo Search Engines: 黄勇 Huang Yong (2009) used the most commonly used 2,500 characters, submitted each of them to the Google and Yahoo Search Engines, recorded the search counts, and sorted the counts to obtain these two character frequency lists. Since materials in the internet are available to CFL learners, these two character frequency lists can help us to discover the usefulness of the characters from the ten textbooks.

In summary, I used Zhuyin-based, etymological-based, and attributive-based dictionaries to serve as references in classifying character and component characteristics. In addition, to find the character usage frequency information, I used frequency lists constructed during different periods of time (1980s and 2000s) and for different purposes (proficiency test and internet search).

3.3 Word Entry and Character Entry

To accomplish the analysis required for this study, a special character database was created by using Microsoft Access and Excel. All of the vocabulary words in each lesson in the ten textbooks were first typed onto specifically designed worksheets in Excel. Later, the worksheets were converted into Access as a database for further analyses and inquiry. Classifications used in the character database were introduced in the following manner. First is the classification of word entry and character entry.

Table 3.1 Screenshots of Character 你 in Four Vocabulary Lists

Textbooks	Content	Screenshots of vocabulary list
RC 1961	Pronunciation Meaning Part of speech Word	 你 nǐ N: you (singular) 你的 your, yours--singular 你們 you--plural 你們的 your, yours--plural
BCR 1977	Pronunciation Meaning	 1 我 2 你 3 他 4 們 5 好 1. 我 wǒ I, me 2. 你 nǐ you
PAVC 2008	Pronunciation Meaning Part of speech Sentence	 14 你ㄋㄧˇ(nǐ) PN : you 你ㄋㄧˇ是ㄕˋ李ㄌㄧˇ愛ㄞˋ美ㄇㄟˇ嗎˙ㄇㄚ ？ Nǐ shìh Lǐ Àiměi ma? Nǐ shì Lǐ Àiměi ma? **Are you Amy Li?**
FEEC 2008	Pronunciation Meaning Part of speech Measure word	 1 Vocabulary 1. 先生 xiān·sheng PN : Mr., Sir (M : wèi 位) 2. 你 nǐ PN : you

The vocabulary lists in the ten textbooks were the essential units of analysis in this study. These vocabulary lists were organized in different formats. Table 3.1 displays screenshots for the character 你 (*nǐ*, you) in the four vocabulary lists to show the different types of information presented in different textbooks. For example, character pronunciation and meaning are essential pieces of information, with part of speech of the character appearing in textbooks RC 1961, PAVC 2008, and FEEC 2008. In RC1961, the character 你 was introduced first and later the character combination (i.e. word, such as 你的, meaning "yours") was provided. In PAVC2008, the character 你 was introduced first and an example sentence with the target character was provided later, such as "你是李愛美嗎?". In addition, character combinations (i.e. word) were introduced as main entries in textbooks, such as 先生 in FEEC2008; for this word, additional information, such as the measure word 位, for this word was also introduced.

As previously noted, DeFrancis (1977) stated that basic to developing reading skill in Chinese is a familiarity with the processes of character combination in Chinese. Such familiarity is best acquired by mastering several combinations for a limited number of characters rather than by learning one or two compounds for many characters. Since most of the words are combinations of characters, word entries are entered and are separated into character entries later, so each character still can be traced back to word entry if needed. Therefore, the character database included both the individual character (字 *zì*) in the character entry as well as its character combinations (詞 *cí*) in the word entry shown in Table 3.2.

Table 3.2 Character and Word Entry in the Ten Textbooks

Code	字 Character	詞 Word	Part of Speech	Meaning in English	中文聽說讀寫 LN in IC2008	新版實用視聽華語 LN in PAVC2008	新實用漢語課本 LN in NPCR2002
1	來	來			5	2	3
h	回	回來	vc	to come back	6	10	
h	來	回來	vc	to come back	6	10	
j	來	進來	vc	come in	5		4
l	來	來	v	to come	5	10	4, 7
x	來	想起來	vc	remember; recall	16		
y	來	越來越	adv	more and more	15		

The first column is *Code*. In the first row, Code 1 (one) indicates that it is the main character entry; the alphabetic code indicates the initial pinyin spelling of the word (character combination). The main character entry is filled in with color (in gray) to distinguish the main character entry from other character combinations. For example, in the second and third rows, word 回來 (*huílai*, to come back) in the third column includes two characters 回 (*huí*) and 來 (*lai*) in the second column, and the initial pinyin of 回 is "h", so the code is "h" in both the second and third rows. The fourth and fifth columns indicate the part of speech and meaning in English of the word. The following ten columns contain lesson numbers where the word appears in each textbook. For example, 回來 (*huílai*, to come back) appears in *Integrated Chinese* 中文聽說讀寫 (IC2008) in lesson 6 and appears in *Practical Audio-Visual Chinese* 新版實用視聽華語 (PAVC2008) in lesson 10. If the word or character is introduced in more than one lesson of a textbook, the lesson numbers are all put in the same cell. For instance, 來 (*lai*, to come) appears in lessons 4 and 7 in *New Practical Chinese Reader* 新實用漢語課本 (NPCR2002), so the cell is coded "4, 7". In addition, for each main character entry (in gray color in Table 3.2), the number in the column for each textbook is the number of word entry (character combinations). For example, 來 appears in 5 word entries in *Integrated Chinese* 中文聽說讀寫 (IC2008). After entering the character and word entry, the main character entry, the one in the gray color row and the code=1, was used for later character analyses.

3.4 Investigating Explicit Orthographic Decomposition Instruction Availability

The first focus of this study was the availability of the explicit orthographic decomposition instruction in the textbooks. As stated, research results have confirmed the benefits of explicit orthographic decomposition instruction (Jackson et al., 2003; Shen, 2004). If explicit orthographic decomposition instruction in class is not possible, is the explicit orthographic decomposition information formally included in the textbooks?

To answer research question 1, "to what extent do textbooks provide explicit orthographic decomposition instruction to learners?" each textbook was examined as to whether it contains explicit orthographic decomposition instruction. If so, the information the textbook supplied was also included. Information about orthographic decomposition could be found in five possible locations in the textbooks including the introduction lesson, interlude lessons, vocabulary section within the lesson, in the specially designed character learning section within the lesson, and/or in the textbook appendices. The amount of information could vary from being small to large, and could be a few sentences, paragraphs, or whole sections within the lesson. There could also be a whole lesson devoted to orthographic decomposition instruction.

Location, quantity, and quality of orthographic decomposition information in each textbook was therefore reported and compared through the use of location and quantity codes. Location codes include 1 = introduction; 2 = prelude / interlude lesson; 3 = vocabulary list in the lesson; 4 = character component section in the lesson; 5 = APPENDIX. Quantity codes include: S = sentences; P= paragraphs; C=column; SS= subpart of the section; WS= the whole section; FL= only in a few lessons.

3.5 Investigating Character Diversity and Repetition

The second focus of this study was to investigate the character diversity and repetition in the ten textbooks. DeFrancis (1977) stated that basic to developing reading skill in Chinese is becoming familiar with the processes of how characters combine in Chinese. Such familiarity is best acquired by mastering several combinations from a limited number of characters rather than by learning one or two compounds from many characters. Some textbooks introduce vocabulary in words (i.e., combination of characters) while others introduce individual characters first, and then introduce words from the combination of characters introduced. Therefore, it is important to know how frequently learners can find the same characters in the different words throughout the same textbook.

To answer research question 2, “to what extent is a single character combined with other characters to form words in textbooks?”, two investigation processes were employed. The first process was to answer “what is the distribution of the same character combining to form different words?” The frequency distribution of the number of word entries containing the same character was therefore examined. As stated in section 3.3 character and word entry, for each main character entry (in gray color in Table 3.2), the number in the column for each textbook is the number of word entries (character combinations). For example, for the character 來, the character combination frequency is 5 in IC2008 and is 2 in PAVC2008. For this research question, a finding that large numbers of character combinations were derived from a limited number of characters would indicate that learners can see specific characters in many words. From the character entry, the total number of different characters can be found. In addition, the total number of different words was also calculated, and each of them categorized into 1-character word, 2-character word, 3-character word, etc., further classifying the majority of the character combination types.

The other process was to find "which character have higher character combination frequency?" Characters appearing in all ten textbooks were listed, and are the core vocabulary in this database. Comparing their total number of the same character combining with other characters to form words, the characters having higher character combination frequency in the database and in each textbook were also identified.

3.6 Investigating Character Frequency Rank Selection

The third focus of this study was character selection in terms of frequency rank. In the majority of textbooks, vocabulary is selected from frequency lists. Researchers have found that words most commonly used are learned faster and remembered better (McCarthy, 1990; Nation, 2001; O'Dell, 1997; Sergent & Everson, 1992). It seems clear that high frequency words are likely to predominate at the early stage of learning and teaching.

To investigate research question 3, "to what extent do textbooks contain high-frequency characters as documented by accepted Chinese character frequency lists?", I used frequency lists constructed during different time periods (1980s and 2000s) and for different purposes (proficiency test and internet search), as stated in section3.2, dictionary and frequency list selection.

Table 3.3 Worksheet Example of Frequency Rank on the Chinese Character Frequency

			Frequency Rank in Frequency Lists					
Traditional Character	Simplified Character	Zhuyin /Pinyin	*Hànzì shǔxìng zìdiǎn* (1989)	Wenlin (2007)	Xiao et al (2009)	Yahoo Search Engine (2009)	Google Search Engine (2009)	HSK Level (1992)
不	不	ㄅㄨˋ/ bù	5	4	5	13	85	甲 1
不	不	ㄅㄨˊ/ bú	5	4	5	13	85	甲 1
家	家	ㄐㄧㄚ¯/ jiā	72	53	29	54	145	甲 1
傢	家	ㄐㄧㄚ¯/ jiā	72	53	29	54	145	甲 1
乾	干	ㄍㄢ¯/ gān	2251	2516	4001	674	676	甲 1
夥	夥	ㄏㄨㄛˇ/huǒ	?	3001	886	4001	4001	?

A character frequency rank worksheet was designed. Table 3.3 presents the frequency rank worksheet to classify frequency rank as documented by the Chinese character frequency lists. Several rules were applied to coding:

1. The same character with different pronunciations 多音字 (*duōyīnzì*), such as the characters 一, 不, 什, etc., had the same rank code because all lists treated them as the same. Therefore, the same characters with different pronunciations were combined to be one character entry.
2. Frequency lists contain both traditional and simplified characters, and they combine the two forms into one rank entry. Multiple traditional characters were combined into one simplified character. For example, characters 家 and 傢 under the 家 rank entry, and the Wenlin frequency rank is 53 for both of them. In this case, it is possible that the number of the top 100 characters can be over 100 if it contains different characters with the same frequency rank.
3. If a character could not be found in the frequency lists, the rank code was 4001, as with the character 乾 in Table 3.3. The Xiao et al. (2009) character frequency rank list contains 2,112 characters. For characters ranked in excess

of the top 2,112 rankings, the rank code was 4001 in order to compare it with other frequency lists.

4. The Wenlin character frequency rank list contains 3,881 characters, but it only provided specific frequency rank numbers up to 3,000. Therefore, for characters ranked in excess of the top 3,000 rankings, the rank code is 3001.

3.6.1 Investigating Character Frequency Rank Classified by Frequency Lists

To answer research question 3, the first investigation process is to classify the frequency rank of each character using three frequency lists which were constructed during the 1980s and 2000s. These lists were:

1. Chinese Character Attributive Dictionary 漢字屬性字典 *Hànzì shǔxìng zìdiǎn* (1989),
2. "Characters by frequency" list from Wenlin 文林 Software for Learning Chinese, version 3.4 (2007), and
3. A Frequency Dictionary of Mandarin Chinese (Xiao, Rayson, & McEnery, 2009).

Since the textbooks were published from 1961 to 2008, if we use the more recent frequency lists to classify the textbooks published before 2000, we need to understand whether or not those textbooks still contain high-frequency characters. If so, those textbooks still can be used now in terms of frequency usage. To accomplish this, the distributions of character frequency rank were examined by calculating the cumulative percentages of frequency rank for the top 100, 200, 300, 400, 500, 1000, 1500, and 2000 most frequently used characters across textbooks.

3.6.2 Investigating Character Frequency Rank as Classified by the Yahoo and Google Search Counts

To answer research question 3, the second investigation process was to classify each character using two frequency lists from the Google and Yahoo Search Engines.

Since materials on the internet are available to CFL learners, these two character frequency lists can help us discover the usefulness of the characters in the ten textbooks. The distributions of character frequency rank were examined by calculating the cumulated percentages of frequency rankings for the top 100, 200, 300, 400, 500, 1000, 1500, and 2000 most frequently used characters across textbooks.

3.6.3 Investigating Character Frequency Rank as Classified by the HSK Proficiency Level List

To answer research question 3, the third investigation process was to classify each character using *The HSK word and character level list* 漢語水平詞彙與漢字等級大綱 *Hànyǔ shuǐpíng cíhuì yǔ Hànzì děngjí dàgāng* (1992) which classifies 2,905 Chinese characters over 4 different levels. For the basic level (甲級, *jiǎjí*), test-takers are expected to know 800 characters. In this study, I classified each character's level based on this HSK list to see whether characters from the ten beginning-level textbooks all belong to the basic level. The frequency distribution of the HSK level characters was calculated individually to find out whether or not the distributions differed across textbooks.

3.6.4 Investigating Character Frequency Rank for Characters Appearing in All Ten Textbooks

To answer research question 3, the last process was to investigate the differences and similarities of frequency usage as documented by accepted Chinese character frequency lists. Characters appearing in all ten textbooks were identified and classified in terms of high frequency usage, as they represent the core vocabulary in the beginning level CFL textbooks. In the current study, I used six frequency lists: (1) Chinese Character Attributive Dictionary 漢字屬性字典 *Hànzì shǔxìng zìdiǎn* (1989), (2) "Characters by frequency" list from Wenlin 文林 Software for Learning Chinese, version 3.4 (2007), (3) A Frequency Dictionary of Mandarin Chinese (Xiao, Rayson, & McEnery, 2009), (4) A Chinese Character Usage Frequency List from Yahoo Search Engine

(Huang, 2009), (5) A Chinese Character Usage Frequency List from Google Search Engine (Huang, 2009), and (6) A level list of HSK word and character 漢語水平詞彙與漢字等級大綱 Hànyǔ shuǐpíng cíhuì yǔ Hànzì děngjí dàgāng (1992).

3.7 Investigating Radical Diversity and Repetition

Although there is no consensus among researchers on the best strategies for developing orthographic awareness, it is believed that L1/L2 Chinese readers develop orthographic awareness through frequent exposure to and encounters with components and meanings of characters. The fourth focus of the study was to investigate radical component diversity and repetition in character selection.

Traditionally, Chinese characters are grouped together according to their common components known as "radicals" 部首 (*bùshǒu*), and each character contains a radical that often helps classify its meaning. Table 3.4 lists the 214 radicals used to classify traditional form Chinese characters. Based on the radical meanings, I categorized each radical into its semantic category, and defined 21 categories. Some radicals relate to human surroundings such as 自然現象(nature), 五穀雜糧植物 (grain and vegetation),獸禽漁類 (animal and fish); some radicals relate to human and animal body and organ parts; some radicals relate to humans such as 人稱, 人生百態, 態度動作 ; some radicals relate to human activities and artifacts such as 食衣住行財 eat, cloth, live, move, and money; some radicals relate to culture and tools; others relate to numbers, measure words, and colors. However, a radical does not guarantee giving a clue to the meaning of the character. Only a semantic-phonetic compound character should have a semantic transparent radical that helps someone infer the meaning of the character, but it does not guarantee this because characters have evolved and been transformed over thousands of years.

Table 3.4 List of 214 Chinese Character Radical Meanings and Radical Categories

自然現象
日, sun
月, moon / month
气, air / breath / steam / vapor
雨, rain
風风, wind
夕, sunset / evening
音, sound
冫, frozen / ice

自然現象-地表
玉王, jade
石, stone
谷, valley
山, mountain
川巛, river

五行天干地支八卦
金钅, metal / gold
木, wood / tree
水氵氺, water
火灬, fire
土, dirt / earth
干, shield
乙乚乛, sprouting plant / second of the 10 stem
己, self /6th of the 10 stem
辛, bitter /8th of the 10 stem
辰, time (7-9am) /5th of 12 stem
酉, liquor (5-7pm) /10th of 12 stem
艮, obstinate
爻, intertwine
卜, to divine
玄, obscure, dark

五穀雜糧植物
禾, grain
米, rice
豆, bean
麥麦, wheat
黍, millet
屮, sprout
瓜, melon
竹, bamboo
艸艹, plant / grass
韭, leek
麻, hemp

獸禽漁類-六畜
牛牜, ox / bull / cow
犬犭,dog
羊, sheep
豕, boar / pig
馬马, horse
鳥鸟, bird

獸禽漁類
隹, bird
虫, insect / reptile
豸, clawed beast
虍, tiger
鹿, stag / deer
黽, toad / frog
鼠, mouse / rat
龍龙, dragon
魚鱼, fish
龜龟, tortoise

人稱
人亻, man
儿, person
女, woman
氏, family' clan
子孑, child
父, father
臣, prostrate / official
自, self
厶, selfish / private
士, person / knight / scholar
工, work
鬼, demon / spirit

人生百態
生, to live
老耂, old
疒, sickness / disease radical
尢, lame / thin
尸, body / corpse

數字
一, one
二, two
八丷, eight / separate
十, ten

器官-人動植物
身, body
手扌, hand
又, again /also / right hand
足, feet
爪爫, claw
皮, skin
肉月, flesh / meat
血, blood
骨, bones
心忄 (bottom of 忝), heart
首, head
頁页, head / page
面, face
目罒, eye
耳, ear
鼻, nose
口, mouth
舌, tongue
牙, tooth
齒齿, teeth / age
角, horn
彑彐, pig's snout
革, leather / raw leather
韋韦, pull / tanned leather
羽, feather / wing
毛, wool / hair
髟, long hair
彡, fine feathers / feather / hair
而, and /beard

量詞
大, big / great
小, small / little
高, high
長镸, long
寸, measured / inch
幺么, small
斗, measure / peck / bushel
斤, axe / pound
方, plow / squared / cardinal point
片, plank / strip / slip
里, village / mile

態度動作-口
言讠, words / to speak
凵, receptacle
曰, to say

態度動作-手
廾, hands folded
攴攵, strikes /to beat
鬥, to fight

態度動作-腳
入, to put / to enter
癶, feet / straddling feet
立, to stand
行, to walk / to go
走, to walk / to go away
辵辶, move / run fast and stop
夂, to march
夊, to walk slowly
廴, to march
彳, step with the left foot

態度動作
止, to stop
見见, to see
飛飞, to fly
力, strength / power
至, arrive
隶, reach / until
比, to compare
釆, to discern
勹, enveloped / wrapper
襾西, upper body / west
齊齐, even / complete

顏色
色, color
白, sunlight / white
赤, red
青, blue / green / black
黃, yellow
黑, black

食
食饣, food / to eat
甘, sweet
鹵卤, salt
香, fragrant / incense

衣物
衣衤, clothing / clothes
巾, cloth
疋, piece of cloth
糸糹纟, threads /silk
黹, to embroider

住-人造領域
囗, surroundings / enclosure
田, field
邑阝, city
阜阝, embankment / mound
冂, outside / border

住-建築傢俱
厂, cliff / slope
穴, hole / cave
广, shed / roof
亠, a roof
冖, covering / roof
宀, roof
几, table
戶, door /household
爿丬, planks / bed
門门, door

行
舟, ship
車车, cart / carriage
禸, track

錢財
貝贝, money / cowry / radical for financial terms
支, branch
用, to use

古今器具
匕, person / ladle
弋, stake / dart
弓, bow
戈, lance
殳, to beat / kill
矛, lance
矢, dart / arrow
鼓, drum
龠, (pan-) pipe
刀 刂, knife
匚, basket
匸, chest /box
皿, vessel / dish
网罒, web / net
耒, plough
臼, mortar
鬯, sacrificial wine
鬲, tripod, cauldron
鼎, tripod
缶, earthenware
瓦, pottery / tile

文明
文, lines /literature
聿肀, pen / writing brush
卩, part / seal
示礻, omen / to show / radical for religious terms

筆劃
丨, vertical downstroke
丶, flame / dot
丿, downstroke to the left
亅, dowstroke with a hook

負面意義
无, not have
欠, exhaled / to pant / to owe money / deficient
毋母, don't/not to be
非, wrong to be
歹, crushed bones / evil
舛, back to back / opposed

Etymological explanations of characters help students to understand, appreciate and remember characters. However, most of the traditional radical-based or pinyin-based Chinese-English dictionaries do not provide the etymological explanations of the characters. In the current study, I used the etymological-based dictionary 中文字譜 *Zhōngwén zìpǔ*, *Chinese Characters: A Genealogy and Dictionary*, compiled by Harbaugh (1998) and *New Chinese Dictionary* 最新國語新辭典 (1987) to code the semantic radicals and their meanings, and also the meanings of the characters containing these radicals.

Research questions 4, 5, and 7 were designed to answer questions dealing with the radical component, such as radical combination frequency, character graphic structure distribution, radical positional regularity within a character's graphic structure, and radical semantic transparency.

3.7.1 Investigating Radical Combination Frequency in Textbooks

To answer research question 4, "to what extent is a radical combined with other components to form characters in textbooks?", the frequency of each radical combination was calculated using three processes: the first process was to classify the radical components of the characters; the second process was to calculate the frequency of each radical component; and the final process was to find whether or not a radical is under-represented in the textbooks. Table 3.5 shows the classifications of the character radical components. The first column is a character in its traditional form, and the second column is its radical. This is followed by the radical's meaning, character's meaning in 中文字譜 *Zhōngwén zìpǔ*, and the number of the character's meaning which does not include surname, and surname.

Table 3.5 Worksheet Examples of Character Radical Classification

Traditional Character	Radical	Radical Meaning	Character Meaning in 中文字譜 *Zhōngwén zìpǔ*	Number of Character Meaning^1	Surname
木	木	wood	Pictograph of a tree. (n) tree (n) wood.	2	
泉	水	water	Modern form shows white 白 water 水. (n) spring.	1	
男	田	field	Field 田 strength 力. (n) male, man (adj) male.	2	
梁	木	wood	Cut 刅 (phonetic ㄔㄨㄤˊ) trees 木 over water 水. (n) bridge (n) beam (sur)a surname.	2	y

Note. [1]number of character meaning does not include surname. y= the character is surname.

The second process was to calculate the frequency of each radical component. For example, as shown in Table 3.5, the radical of the character 木 and 梁 is 木, so the frequency of the 木 combination is “2”. However, for the radicals 水 and 田, the frequency of each radical combination is “1” (the number is only based on the examples in Table 3.5). For each textbook, the frequency of the radical combination was calculated to find the most frequently used radicals.

The final process was to find whether or not a radical is under-represented in the textbooks. The radical combination frequency in 中文字譜 *Zhōngwén zìpǔ* served as a comparison base for the radical combination frequency in the ten textbooks. If a radical appeared in at least four characters in each textbook, it was judged to be a high combination radical. However, some of the radicals are combined with other components to form characters. For example, in the 中文字譜 *Zhōngwén zìpǔ* , only three characters 青, 靖, and 靜 contain the radical 青, so it is also important to know those radicals that have the same amount of radical combinations in textbooks and 中文字譜 *Zhōngwén zìpǔ*, which means the coverage is 100%. After these analyses, the 25 highest combination radicals in 中文字譜 *Zhōngwén zìpǔ* were compared with the radicals in each textbook to find similarities and differences.

3.7.2 Investigating Radical Regularity among Character Graphic Structure

Each radical has its legal position, and knowing the legal positions of the radical can help to determine whether a character is a real character (Ho, Yau, & Au, 2003). To answer research question 5, "for each character graphic structure, what is the most commonly appearing radical position?" three investigation processes were used: the first process was to decompose the character into components by using the component and stroke lists in Tables 3.6 and 3.7; the next process was to classify the character graphic structure by using the graphic structure and radical position reference in Figure 3.2; and the last process was to classify the position of the semantic radical in that character. As a result, for each character graphic structure, the most commonly appearing radical position could be found.

As stated above, the first process was to decompose the character into components by using the component and stroke list in Tables 3.6 and 3.7. and employing the two methods of component identification described in the first chapter of this study: stratified analysis 層次分析法 (*céngcì fēnxifǎ*) and plane analysis 平面分析法 (*Píngmiàn fēnxifǎ*). Using Xiao's (1994) definition of component structures as reference, the main rule is that a character can be decomposed into component parts, and the component parts can be further decomposed into strokes. After comparing the component lists from Chen (1997), Fei (1996), Guder-Manitius (1999), and Xiao (1995), as stated in section 2.7, a reference component and stroke list, as shown in Tables 3.6 and 3.7, was created to serve as reference to classify the traditional character graphic structures, with the simplified characters changed to traditional characters. While the four component lists did not take radicals into consideration, in this study, the components were classified as to whether or not they are radicals. The component list includes radicals and non-radical components categorized by non-character and character components. According to Xiao (1994), some characters are composed of components and strokes, so strokes such as 丶 乚 亅 丨 are also included, even though they are not components. Six types of components are defined and presented in Tables 3.6 and 3.7: (1) strokes that go across each other; (2) strokes that are connected; (3) strokes that go across with some being connected; (4) strokes that always group together; (5) compound components without spaces; and (6) compound components with spaces. However, when decomposing characters, the first four component types were given priority over the last two types. In addition, radical components were given priority over non-radical components.

After decomposing a character into components, the next process was to classify the character graphic structure by using the graphic structure reference in Figure 3.2, which is elaborated from the categories in Fu (1993) and Shen and Ke (2007). Five main types of character graphic structure were used:

I = Integral structure characters: characters that cannot be separated into components, such as 女, 子, 牛,車,千, 太, 而, 小, 心, 三, 丟, 非

LR = Left-right structure characters:

TB = Top-bottom structure characters:, 頁, 百, 全, 另, 主,

HE = Half-enclosure structure characters: 友在

E = Enclosure structure characters: 國

For example, the character 木 is also a character component which cannot be further decomposed into other components, so the character 木 belongs to the integral structure type; for this type, the coding is "I". In another example, character 安 can be further decomposed into 宀 and 女, and the two components are in the top and bottom positions in the character, so character 安 belongs to the top-bottom structure (TB) type 1, so the coding is "TB-1". Descriptive statistic information of character graphic structure was reported and compared so as to arrive at the most commonly appearing character graphic structures across textbooks.

Table 3.6 Non-character Component and Stroke List

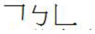
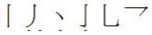
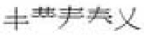
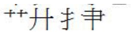
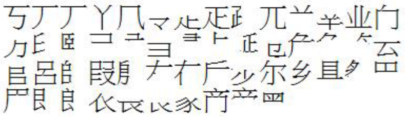
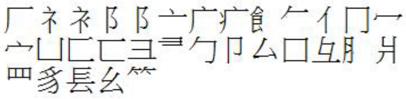
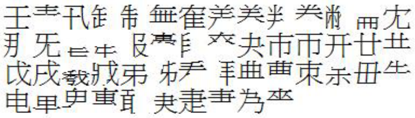
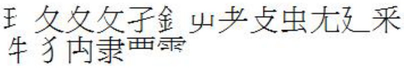
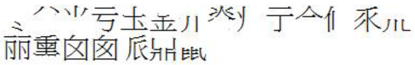
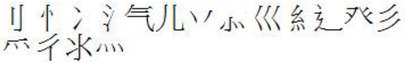
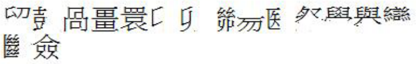
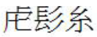
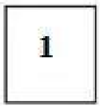
Types	Non-radical	Radical
Stroke:		
Strokes go crossed:		
Strokes are connected:		
Some strokes go crossed and some are connected:		
Strokes group together:		
Compound component:		

Table 3.7 Character Component and Stroke List

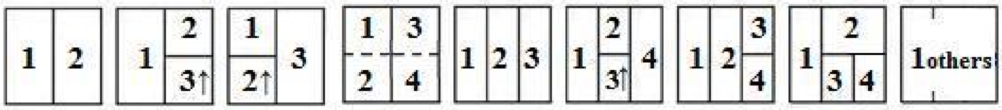
Types	Non-radical	Radical
Stroke:		一
Strokes go crossed:	丰丈九丸事冉七也屯尹夫井	又力十女聿
Strokes are connected:	之乏乃丁才了亡勿巨匹己已巴上正丐下不久四丘尺甚互且象乍凸凹予	乙卜工几立方人入山刀匕臣己止欠口斤夕歹尸弓日白自曰月片皿血臼豕韭疋金矛長
Some strokes go crossed and some are connected:	千太天夭戊戌成我丑良匆民南垂五及于內丙半平甩免斥失升由甲申重卑曲更禹禺兼秉再中串央史吏夷世冊出弗曳未末朱本東果柬	大士土王子羊用毛氏牙禾巾干甘手木米耒毋牛生田里目身耳西面而皮隹缶車父革艮
Strokes group together:	三元凡亦以刃刁勺北今令云尤乒乓少永求州氏丞叉乎兔兆必承丹甫卯	二玉斗寸言八谷舟么小川弋戈瓦心母水火爪瓜犬酉雨豆辰風鹵齒非食門鬥龍行龜飛麥鬲黽齊骨
Compound components without spaces:	主午年幸卒反	石文玄矢至足走辛首老支香舌色角見衣
Compound components with spaces:	共率亥曾周同向寅函可司奐局燕堇或典其系曹焉惠妻庸叟幾爭褒襄覽	示頁貝青竹比鹿黃爻殳高音麻龠肉穴羽韋黍舛鼠魚黑鳥馬鬼鼎鼓赤鼻

Integral Structure (I)



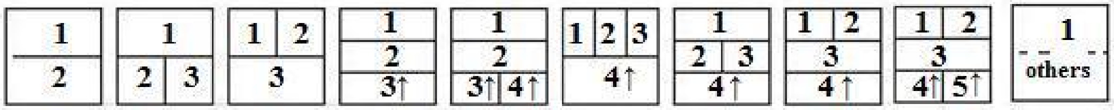
Character (C): 木州
Radical (R): 木川
Position number (P): 0 0 (cannot be decomposed)

Left-Right Structure (LR)


1 2
1 2 3↑
1 2↑ 3
1 3 2 4
1 2 3
2 1 4 3↑
3 1 2 4
2 1 3 4
1others

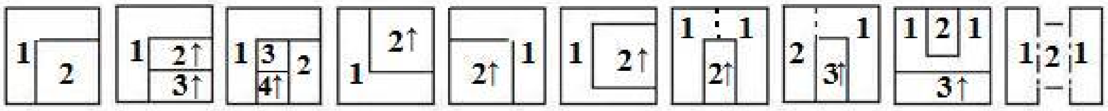
	LR-1	LR-2	LR-3	LR-4	LR-5	LR-6	LR-7	LR-8	LR-9
C:	他甜	吃號乾	聽能敏	能號	游鄉	鐵	搬	擺瑜	隨
R:	亻 甘	口虍乙	耳月攵	月虍	氵阝	金	扌	扌 王	阝
P:	1 2	1 2 3	1 2 3	2 3	1 3	1	1	1 1	1

Top-Bottom Structure (TB)


1 2
1 2 3
1 2 3
1 2 3↑
1 2 3↑ 4↑
1 2 3 4↑
1 2 3 4↑
1 2 3 4↑
1 2 3 4↑ 5↑
1 others

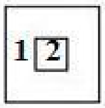
	TB-1	TB-2	TB-3	TB-4	TB-5	TB-6	TB-7	TB-8	TB-9	TB-10
C:	安思	花命前	準望照	菜愛當	蘿	變 響	籃 煎	幫	器	簡歲
R:	宀心	艹口 刂	氵月灬	艹心田	艹	言 音	竹 灬	巾	口	竹止
P:	1 2	1 2 3	1 2 3	1 3 3	1	2 4	1 4	4	1	1 1

Half-Enclosure Structure (HE)


1 2
1 2↑ 3↑
1 3 2 4↑
2↑ 1
2↑ 1
1 2↑
1 1 2↑
2 1 3↑
1 2 1 3↑
1 2 1

	HE-1	HE-2	HE-3	HE-4	HE-5	HE-6	HE-7	HE-8	HE-9	HE-10
C:	店右	疼慶腐	廚厭	起	氣可	區	閉向	戚威	興	街斑
R:	广口	疒心肉	广厂	走	气口	匚	門口	戈女	臼	行文
P:	1 2	1 3 3	1 1	1	1 2	1	1 2	1 2	1	1 2

Enclosure Structure (E)



C: 國
R: 囗
P: 1

Figure 3.2 Character Graphic Structures and Radical Positions

To answer research question 5, “for each character graphic structure, what is the most commonly appearing radical position?”, the last process was to classify the position of the semantic radical in that character by using the radical position reference in Figure 3.2. Examples of the character (C), the radical in that example character (R) and the position number of that character (P) are shown in Figure 3.2. Since an integral character cannot be decomposed into individual components, only one classification in this category and the radical position was coded “0”. For instance, 木 is an integral structure, and itself is the radical for a number of characters. Therefore, for the character 木 (see Table 3.8), the code is “I” in the character graphic structure, and “0” in the radical position. In another example, 他 is a left-right structure character type 1, so it is coded as LR-1. The radical is 亻, and its position is 1 in LR-1. Therefore, for the character 他 (see Table 3.8), the code is “LR-1” in the character graphic structure, and “1” in the radical position. Characters introduced in the textbooks were compiled as one database.

After the classification of all characters was completed, descriptive statistical information from the combined character database was calculated. In addition, characters in each graphic structure and the most commonly appearing semantic radical position for each character graphic structure were listed.

Table 3.8 Classifications of Character Graphic Structure and Radical Position

Traditional Character	Character Meaning	Radical	Radical Meaning	Number of Character Meaning	Radical Semantic Transparency	Character Graphic Structure	Radical Position
木	Wood	木	wood	2	1	I	0
州	land, state	川	river	1	3	I	0
他	he, other	亻	man	2	5	LR-1	1
看	See	目	eye	3	4	HE-1	2
男	Male	田	field	1	3	TB-1	1
國	Country	囗	enclose	1	4	E	1

In summary, the fourth focus of this study was radical component diversity and repetition in character selection. Research questions 4 and 5 were designed to answer radical component combination frequency, character graphic structure distribution, and radical positional regularity within character graphic structure so as to determine “the most frequently used radicals”, “the most commonly appearing character graphic structures”, and “the most commonly appearing radical position for each character graphic structure”.

3.8 Investigating Phonetic Component Diversity and Repetition

Although there is no consensus among researchers on the best strategies for developing orthographic awareness, it is believed that L1/L2 Chinese readers develop orthographic awareness through frequent exposure to or encounters with character phonetic components and other components. As stated, frequency of occurrence of printed characters also contributes to the degree of orthographic awareness development. One approach used to pronounce a whole character is directly derived from the pronunciation of its phonetic elements; the other approach used to pronounce a whole character is deduced via analogy with other characters sharing the same phonetic components (Lu, 2003; Chan & Wang, 2003). The fifth focus of the study was to investigate phonetic component diversity and repetition in character selection with research questions 6 and 7 designed to answer questions involving phonetic components.

To answer the research question 6, “to what extent is a phonetic component combined with other components to form characters in textbooks?”, three investigation processes were used: the first process was to classify basic characteristics of character phonetic components; the second process was to group the characters with the same phonetic components; and the last process was to compare the pronunciations of a character with its phonetic components.

The first process was to build the basic characteristics of character phonetic components. In the current study, the etymological-based dictionary 中文字譜, *Chinese Characters: A Genealogy and Dictionary*, compiled by Harbaugh (1998) was used to code etymological tree roots, tree root numbers, Six Books category, phonetic components, and the meaning of the characters (as explained in dictionary selection in Section 3.2). Table 3.9 below shows the characteristics of each character and classifications of its phonetic component reliability. The first column is the character followed by its pronunciation in Zhuyin and Pinyin. Based on Harbaugh's (1998) etymological dictionary 中文字譜, each character was traced back to its etymological tree root and category in The Six Books. If the character is a Phonetic Complex (or Semantic-Phonetic Compound 形聲字) character, the phonetic component was typed with Zhuyin and Pinyin. To this point, the basic characteristics of the character's phonetic component had been built.

Table 3.9 Classifications of Character Phonetic Component Reliability

1	2	3	4	5	6	7	8
Character	Zhuyin	Pinyin	Etymological Tree Root (Root Number)	Six Book	Phonetic Component	Phonetic Code	Phonetic Element Reliability
天	ㄊㄧㄢˉ	Tiān	大天(39)	Pictograph	-	p	
十	ㄕˊ	Shí	十(31)	Ideograph	-	i	
麻	ㄇㄚˊ	Má	木林麻(77)	Others	-	o	
池	ㄔˊ	Chí	也池(56)	Phonetic-C	也 ㄝˇ/ yě	100	
舞	ㄨˇ	Wǔ	十卌無舞(31)	Phonetic-C	無 ㄨˊ/ wú	110	2
姨	ㄧˊ	Yí	弓夷姨(65)	Phonetic-C	夷 ㄧˊ/ yí	111	1
煮	ㄓㄨˇ	Zhǔ	日白者煮(76)	Phonetic-C	者 ㄓㄜˇ/ zhě	2101	
媽	ㄇㄚˉ	Mā	馬媽(177)	Phonetic-C	馬 ㄇㄚˇ/ mǎ	2110	2
碼	ㄇㄚˇ	Mǎ	馬碼(177)	Phonetic-C	馬 ㄇㄚˇ/ mǎ	2111	1
校	ㄒㄧㄠˋ	Jiào	大交校(39)	Phonetic-C	交 ㄐㄧㄠ/ jiāo	30110	
快	ㄎㄨㄞˋ	Kuài	大夬快(39)	Phonetic-C	夬 ㄍㄨㄞˋ/ guài	30111	
貨	ㄏㄨㄛˋ	Huò	匕化貨(13)	Phonetic-C	化 ㄏㄨㄚˋ/ huà	31101	
腿	ㄊㄨㄟˇ	Tuǐ	匕艮退腿(12)	Phonetic-C	退 ㄊㄨㄟˋ/ tuì	31110	2
蕉	ㄐㄧㄠˉ	Jiāo	隹焦蕉(162)	Phonetic-C	焦 ㄐㄧㄠˉ/ jiāo	31111	1

One approach used to pronounce a whole character is deduced via analogy with other characters sharing the same phonetic components (Lu, 2003; Chan & Wang, 2003). To answer research question 6, the second process was to group the characters with the same phonetic components, while the last process compared the pronunciation of a character with its phonetic components. The fourth column in Table 3.9, *etymological tree root number* was used to identify the following four types of analogy:

1. Homophones: 璧 (*bì*) and 壁 (*bì*) have the same tree root 辟 (50) component and pronunciation.
2. Partial homophones: 清 (*qīng*) and 晴 (*qíng*) have the same root 青 (70) component, but with a tone difference.
3. Same rhymes: 板 (*bǎn*) and 返 (*fǎn*) have the same tree root 反 (21) component with the same rhyme *ǎn*. In other words, if the replaced initial syllables are b/p/f, d/t, j/q/x, g/k/h, ch/zh/sh/r, and z/c/s groups, the characters share the same rhymes (Guder-Manitius, 1999).
4. Same phonetic component but having completely different sounds: 煮 (*zhǔ*) and 奢 (*shē*) have completely different pronunciations although with the same tree root 者 (76) component.

Of the characters introduced in textbooks compiled as one database, characters that shared the same components were grouped into the four analogy types. After the classification of characters was completed, the frequency distributions of the four types were created.

3.9 Investigating Ideal Semantic-Phonetic Compound Characters

Semantic-phonetic compound characters 形聲(*xíngshēng*) combine the meaning of one character 形旁 (*xíngpáng*, semantic element) with the sound of another 聲旁 (*shēngpáng*, phonetic element). A ideal semantic-phonetic compound character is a semantic-phonetic compound character that contains a semantically transparent radical and a reliable phonetic element. A key concept of developing Chinese orthographic awareness is the frequency of exposure to the ideal semantic transparent and/or phonetic reliable characters. The last focus of the study is to investigate ideal semantic-phonetic compound characters in character selection.

To answer research question 7, "for semantic-phonetic compound characters, what percentages of the characters contain semantic transparent radical and reliable phonetic components?", four investigation processes were used: the first process was to identify semantic-phonetic compound characters; the second process was to classify characters with reliable phonetic elements; the third process was to classify characters with semantic transparent radical; the last process was to classify characters with both semantic transparent radical and reliable phonetic elements.

3.9.1 Investigating the Classification of Semantic-Phonetic Compound Characters

The first process was to identify the semantic-phonetic compound characters. The fifth column in Table 3.9, of The *Six Book* was used. The percentage of phonetic complex (Six Book Code = Phonetic-C) was calculated and compared across textbooks.

3.9.2 Investigating the Classification of Reliable Phonetic Element Characters

The second process was to classify characters with reliable phonetic elements. One approach used to pronounce a whole character is directly derived from the pronunciation of its phonetic element. Not all of the compound characters include phonetic elements-- only semantic-phonetic compound characters should have reliable phonetic elements, though this is not guaranteed because characters have been reformed over thousands of years. To investigate phonetic element reliability, the pronunciation of a semantic-phonetic compound character was compared with the pronunciation of its phonetic element (in the sixth column in Table 3.9). *Phonetic Code* in the seventh column in Table 3.9 was used to categorize the results of the comparison. In this study, Zhuyin was used as the comparison unit. The maximum number of syllables used in the Zhuyin system is three, with one tone: for example, one-syllable such as ㄕˊ, two-syllables such as ㄇㄚˊ, and three-syllables such as ㄊㄨㄟˇ. The first number of the *Phonetic Code* indicates the type of syllable structure. For instance, the comparison code of 舞ㄨˇ and its phonetic component 無 ㄨˊ is “110”. The first number “1”indicates that 舞 is one-syllable; the second number “1”indicates that it employs the same zhuyin character “ㄨ”; the last number “0”indicates that it does not have the same tone. Another example would be the comparison code of 蕉ㄐㄧㄠˉ and its phonetic component 焦 ㄐㄧㄠˉ is “31111”, indicating that the pronunciation of the phonetic component is the same as the whole character, so the phonetic component is very reliable. As a result, the number in the eighth column is “1”. If the comparison code is “31110”, it means that the character is three-syllable and only the tones are different, such as 腿 and its phonetic component 退. As a result, the number in the 8th column is “2”. To answer research question 7, the percentage of characters with reliable phonetic components (Code =1) was calculated and compared across textbooks. In addition, the tonal difference was also considered (Code =2), with the percentage of characters with reliable phonetic components were calculated .

3.9.3 Investigation of the Classification of Semantic Transparent Radical Characters

The third process was to classify characters with semantic transparent radicals. As stated in the introductory chapter, Chinese characters were grouped together according to their common components known as “radicals” 部首 (*bùshǒu*), with each character containing a radical. A radical does not guarantee that it will supply a clue to the meaning of the character. Only a semantic-phonetic compound character should have a semantically transparent radical to infer the meaning of the character, but this is not guaranteed because characters have changed over thousands of years. Yin (1994:25) and Guder-Manitius (1999:230) have discussed the four types of radical semantic transparency. In the current study, based on the assumption that a radical of a given character is known, the radical semantic transparency was classified, shown in the 6th column in Table 3.10. To classify this characteristic, five codes were used:

5 = semantically transparent and the radical is helpful for understanding all the meanings of the whole character because the radial defines the category of the character, such as 亻 (person) helps to understand character 他 (he; him) because 他 is a person. Other examples include: 銅 (copper) with 金 (metal) radical, 鯉 (carp) with 魚 (fish) radical, 柏 (cypress) with 木 (wood) radical, and 氧 (oxygen) with 气 (air) radical.

4 = semantically transparent and the radical is helpful for understanding all the meanings of the whole character because the radical has a direct relationship with the character, such as 木 (wood) helps to understand the character 梁 (bridge; beam) because 梁 is made of wood. Other examples include: 浸 (permeate) with 水氵 (water) radical, 扔 (throw; discard) with 手扌 (hand) radical, 扶 (support with hand; aid, help) with 手扌 (hand) radical, 杖 (staff, cane) with 木 (wood) radical, and 飯 (rice) with 食 (food) radical.

3 = semantically transparent and the radical is helpful for understanding all the meanings of the whole character because the radical has an indirect relationship with the character, such a 冫 (frozen) helps to understand the character 冷 (cold) because 冷 relates to freezing. Other examples include: 城 (city, town; city wall) with 土 (soil) radical, 助 (assist, help) with 力 (power; strength) radical, 滿 (full; satisfied; complete, fulfill) with 水氵 (water) radical, and 粒 (grain, pellet) with 米 (rice) radical.

2 = semantically opaque; the radical is unrelated to the meaning of the whole character, such as 水氵 (water) is not transparently related to 演 (evolve, practice, act, perform). Other examples include: 笑 (smile) with 竹 (bamboo) radical.

1 = the radical is a character itself, such as 木.

Table 3.10 Classifications of Character Radical Semantic Transparency

Traditional Character	Radical	Radical Meaning	Character Meaning in 中文字譜 *Zhōngwén zìpǔ*	Number of Character Meaning[1]	Radical Semantic Transparency
他	亻	man	Modern form shows person 人 with 也 (phonetic ㄝˇ, originally 它). (pron) he, (pron) other, another.	2	5
梁	木	wood	Cut 刅 (phonetic ㄔㄨㄤˊ) trees 木 over water 水. (n) bridge (n) beam (sur) a surname.	2	4
冷	冫	frozen	Cold 冫 with 令 (phonetic ㄌㄧㄥˋ). (adj) cold.	1	3
演	水氵	water	Water 水 with 寅 (phonetic ㄧㄣˊ). (v) evolve, (v) practice, (v) act, perform.	3	2
木	木	wood	Pictograph of a tree. (n) tree, (n) wood.	2	1

Note. [1] number of character meaning does not include surname.

When coding the radical semantic transparency of characters, two rules were established to deal with cases where the meaning of the whole character was not clear-cut. First, when there was uncertainty of the radical semantic transparency of a character due to it having multiple meanings, I consulted the etymological-based dictionary 中文字譜, *Chinese Characters: A Genealogy and Dictionary* (Harbaugh,1998) to classify the meaning and number of meanings for each character. If the radical of a character was classified as semantically transparent, all the multiple meanings of the characters had to directly or indirectly relate to the meaning of the radical. Second, it was not an easy task to classify the meaning of a single character because characters are combined to form words, and learners learn the meanings of words more often than they do the meanings of individual characters. Therefore, this study was based largely on the etymological meanings of characters in the dictionary to identify the relationship between a character's meaning and its radical. In addition, two native Chinese raters judged the semantic transparency of the radical with the percentages of semantically transparent characters (code = 5, 4, 3) being reported. The average number of semantic radical transparency codes was also compared across textbooks with a correlation analysis conducted to see whether the number of character meanings was related to semantic radical transparency.

3.9.4 Investigation of the Classification of Ideal Semantic-Phonetic Compound Characters

To answer research question 7, "for semantic-phonetic compound characters, what percentage of the characters contain semantic transparent radical and reliable phonetic components?", the last process was to classify characters with both semantic transparent radical and reliable phonetic elements. Based on the character characteristics from the previous three investigations, characters which are semantic-phonetic compound characters (Six Book code = Phonetic-C) that possess reliable phonetic elements (phonetic element reliability code =1) and semantically transparent radicals (radical semantic transparency code =5, 4,3) were classified as ideal semantic-phonetic compound characters. The percentage of ideal semantic-phonetic compound characters was compared across textbooks.

3.10 Summary of Character Characteristics

In summation, in the current study, I asked questions in the following six areas: (1) explicit orthographic decomposition instruction; (2) character diversity and repetition; (3) character frequency selection across textbooks; (4) radical component diversity and repletion; (5) phonetic component diversity and repetition; and (6) ideal semantic-phonetic compound character selection. For each character, the characteristics of the characters was classified to include:

1. Traditional characters
2. Pronunciation: Zhuyin and Pinyin
3. Character meaning

4. Words (character combination)
5. Frequency rank in Hanzi shuxing xidian, Wenlin (2007), Xiao, etal., (2009), Yahoo search counts list, Google search counts list, and HSK Level
6. Radical: radical code, radical meaning, radical semantic transparency code
7. Character graphic structure and radical position within the graphic structure
8. Six Book category
9. Etymological information: Tree root number 1 and 2
10. Phonetic element reliability: phonetic element, phonetic code, and reliability code
11. Phonetic analogy: shared phonetic component, analogy group code, homophone analogy, partial homophone analogy, same rhyme analogy, and different sound analogy

CHAPTER 4

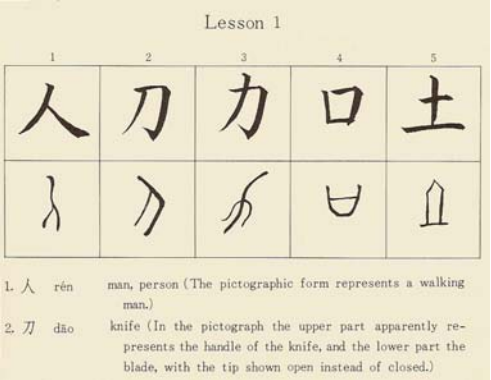
RESEARCH RESULTS

Based on the theory that native Chinese and Chinese as a foreign language (CFL) readers develop Chinese orthographic awareness to infer meaning and pronunciation of Chinese characters through repeated exposure to print and explicit orthographic instruction (Jackson, Everson, & Ke, 2003; Ho & Bryant, 1997; Ho, Yau, & Au, 2003; Shen, 2005; Shen & Ke, 2007), the purpose of this study was to systematically describe and classify Chinese characters in ten CFL textbooks for college and adult beginning learners. The main focus was to make an inventory of characters and discuss textual materials availability in the following areas: (1) explicit orthographic decomposition instruction (research question 1); (2) character diversity and repetition (research question 2); (3) character frequency selection across textbooks (research questions 3); (4) radical component diversity and repetition (research questions 4, 5, and 7); (5) phonetic component diversity and repetition (research questions 6 and 7); and (6) ideal semantic transparent radicals and reliable phonetic elements (research question 7). To perform the analysis required for this study, a special character database was created. All of the vocabulary characters in each lesson in the ten textbooks were first typed onto specifically designed worksheets. Each character was classified in terms of character frequency, radical combination frequency, radical semantic transparency, radical positional regularity among different character graphic structures, phonetic element reliability, and phonetic component combination frequency, with the different textbooks compared along these dimensions. The research results are presented in the following sections.

4.1 Results Investigating Explicit Orthographic Decomposition Instruction in the Textbooks

Research question 1: To what extent do textbooks provide explicit orthographic decomposition instruction to learners?

Each textbook was examined as to whether it contained explicit orthographic decomposition instruction. Information about orthographic decomposition was found in five possible locations in the textbooks including the introductory lesson, interlude lessons, vocabulary section within the lesson, designed character learning section within the lesson, and the APPENDIX in the textbooks.



Lesson 1

1	2	3	4	5
人	刀	力	口	土

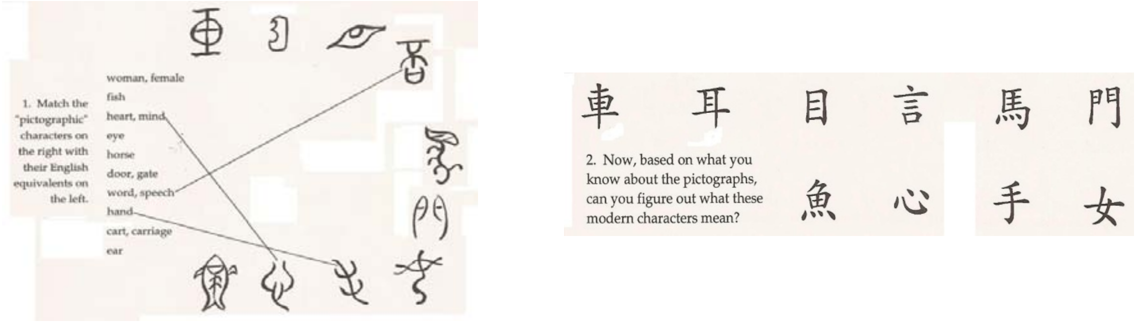
1.	人	rén	man, person (The pictographic form represents a walking man.)
2.	刀	dāo	knife (In the pictograph the upper part apparently represents the handle of the knife, and the lower part the blade, with the tip shown open instead of closed.)

Figure 4.1 A Screenshot of the Vocabulary List in BCR1977 (p. 3)

In the introductory part of the textbooks, the textbooks BCR1977 , IC2005, and IC2008 include materials about character formation, stroke, radical, and phonetic components. In BCR1977, DeFrancis (1977) suggests that learners spend a good deal of time in learning how to write characters, and “the component parts and similarities in structure of various characters should be pointed out as an aid toward memorization (p. xxxii)”. He also suggests that regardless of the approaches of learning to write characters (writing from memory or being aware of their components), it is advisable to pay special attention to the writing of the first 30 characters in the textbook because they are all radicals that can be stand alone as characters. In addition, these characters are presented with pictographic forms in the first three lessons; a screenshot of vocabulary list in BCR$_{1977}$ is shown in Figure 4.1. The first row is the modern form, and the second row is the pictographic form.

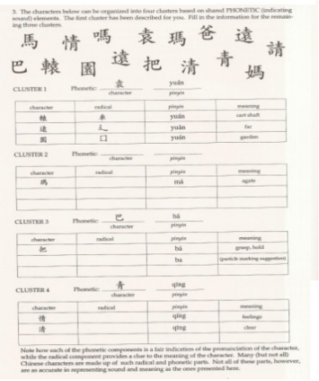
In one part of the introduction in IC$_{2008}$ and IC$_{2005}$, the Chinese writing system introduced The Six Books 六書 definitions, radical, and stroke subparts. The authors explained The Six Books and pointed out that the vast majority of Chinese characters are pictophonetic (semantic-phonetic compound) characters 形聲 consisting of a radical and a phonetic component. They state that "the radical often suggests the meaning of a character, and the phonetic component indicates its original pronunciation, which may or may not represent its modern pronunciation (p. 12)". However, these statements are only a general introduction and do not provide the students with many examples and explanations. Besides introducing The Six Books, they introduce 40 radicals. The authors believe that learners will find recognizing, remembering and reproducing characters much easier by knowing the radicals and other basic components well. In the last part of the Chinese writing system introduction, they state the basic strokes and rules of stroke order. They believe that following the rules of stroke order will make it easier for learners to accurately count the number of strokes in a character and knowing the exact number of strokes in a character will help them find the character in a radical-based dictionary.

Instead of providing brief statements in the introduction, textbook CIC$_{rw1994}$ includes one prelude and three interludes containing radical and phonetic activities. Some activities require students to recognize the radicals of the characters and connect them with the meanings of the characters. Two examples of radical activities in CIC$_{rw1994}$ are shown in Figure 4.2. For example, a "pictographic heart" is first matched with its English equivalent "heart, mind", and then connects them with the modern character "心". After these two basic activities, learners are asked to circle the radicals in the characters which are provided with English meanings. For example, 忘 means “to forget”, and learners are asked to circle “心” in “忘”. On the other hand, in Interlude 3, Ning (1994) provides a phonetic component activity to point out that some characters share the same phonetics, and that the phonetic components are a fair indication of the pronunciation of the character, as presented in Figure 4.3. For example, 轅, 遠, and 園 share the “袁” phonetic component, and the pronunciation of the three characters is the same except for one tonal difference 遠. To avoid providing overgeneralization to the learners, Ning (1994) cautions learners that not all of the radical and phonetic components are as accurate in representing sound and meaning as the ones in the activity she provided.


1. Match the "pictographic" characters on the right with their English equivalents on the left.

2. Now, based on what you know about the pictographs, can you figure out what these modern characters mean?

Figure 4.2 Examples of Radical Activity in CICrw1994 (p. 29)


3. The characters below can be organized into four clusters based on shared PHONETIC (indicating sound) elements. The first cluster has been described for you. Fill in the information for the remaining three clusters.

馬 情 嗎 袁 瑪 爸 遠 請
巴 轅 園 遠 把 清 青 媽

CLUSTER 1 Phonetic: 袁 (character) yuán (pinyin)

character	radical	pinyin	meaning
轅	車	yuán	cart shaft
遠	辶	yuǎn	far
園	囗	yuán	garden

CLUSTER 2 Phonetic: ____ (character) ____ (pinyin)

character	radical	pinyin	meaning
瑪		mǎ	agate

CLUSTER 3 Phonetic: 巴 (character) bā (pinyin)

character	radical	pinyin	meaning
把		bǎ	grasp, hold
		ba	(particle marking suggestion)

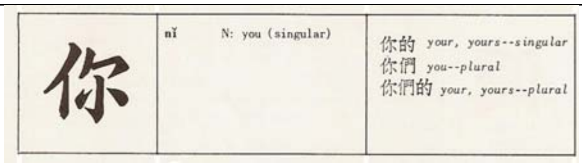
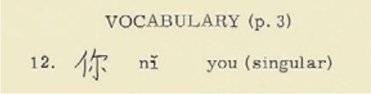
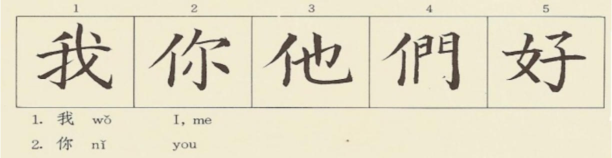
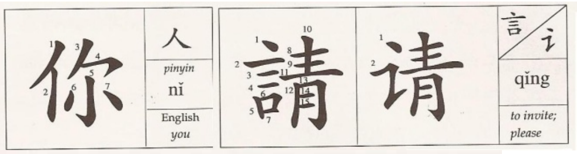
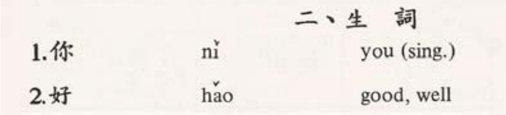
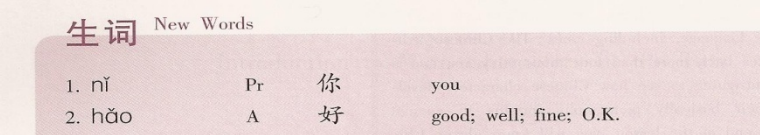
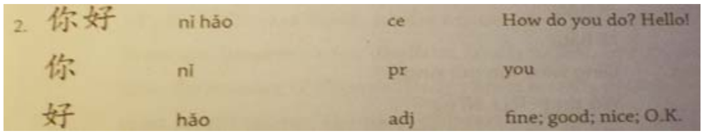
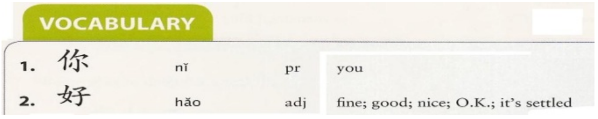
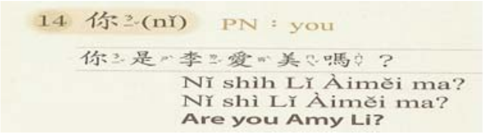
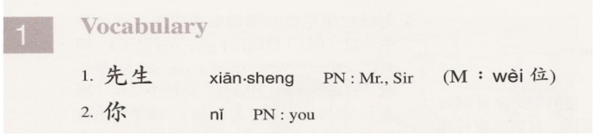
CLUSTER 4 Phonetic: 青 (character) qīng (pinyin)

character	radical	pinyin	meaning
情		qíng	feelings
清		qīng	clear

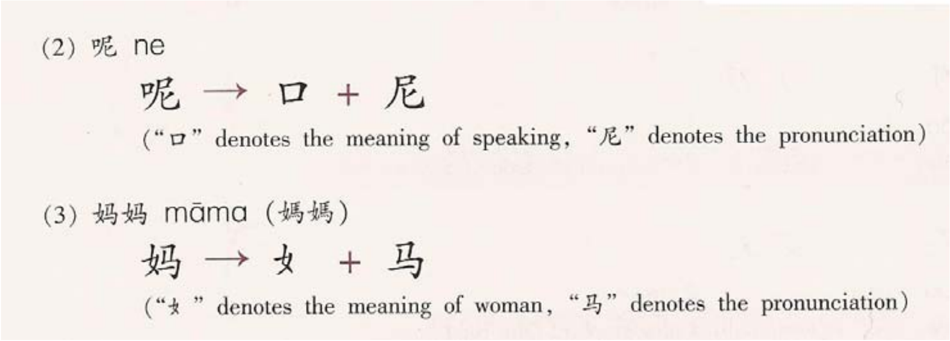
Note how each of the phonetic components is a fair indication of the pronunciation of the character, while the radical component provides a clue to the meaning of the character. Many (but not all) Chinese characters are made up of such radical and phonetic parts. Not all of these parts, however, are as accurate in representing sound and meaning as the ones presented here.

Figure 4.3 An Example of Shared Phonetic Component Activity in CICrw1994 (p.152)

Table 4.1 Screenshots of Character 你 in the Ten Vocabulary Lists

Textbooks	Content	Screenshots of Vocabulary List
RC 1961	Pronunciation Meaning Part of speech Word	 你 nǐ N: you (singular) 你的 your, yours--singular 你們 you--plural 你們的 your, yours--plural
CTBC1976	Pronunciation Meaning	 VOCABULARY (p. 3) 12. 你 nǐ you (singular)
BCR 1977	Pronunciation Meaning	 1 我 2 你 3 他 4 們 5 好 1. 我 wǒ I, me 2. 你 nǐ you
CICrw 1994	Pronunciation Meaning Radical Stroke-order	 你 人 pinyin nǐ English you 請 请 言 讠 qǐng to invite; please
PCR 1995	Pronunciation Meaning	 二、生 詞 1.你 nǐ you (sing.) 2.好 hǎo good, well
NPCR 2002	Pronunciation Meaning Part of speech	 生词 New Words 1. nǐ Pr 你 you 2. hǎo A 好 good; well; fine; O.K.
IC 2005	Pronunciation Meaning Part of speech	 2. 你好 nǐ hǎo ce How do you do? Hello! 你 nǐ pr you 好 hǎo adj fine; good; nice; O.K.
IC 2008	Pronunciation Meaning Part of speech	 VOCABULARY 1. 你 nǐ pr you 2. 好 hǎo adj fine; good; nice; O.K.; it's settled
PAVC 2008	Pronunciation Meaning Part of speech Sentence	 14 你(nǐ) PN : you 你是李愛美嗎 ? Nǐ shìh Lǐ Àiměi ma? Nǐ shì Lǐ Àiměi ma? Are you Amy Li?
FEEC 2008	Pronunciation Meaning Part of speech	 1 Vocabulary 1. 先生 xiān·sheng PN : Mr., Sir (M ： wèi 位) 2. 你 nǐ PN : you

An important place to find explicit orthographic decomposition instruction is the vocabulary list in each lesson. Table 4.1 below presents screenshots of the character 你 in the vocabulary lists in the ten textbooks. All of them mainly focus on the pronunciation and meaning of the character 你. They do not introduce the internal structure of the character except for CICrw$_{1994}$ that provides columns for the radical (i.e. 人) and the stroke-order of the character 你. No phonetic components are introduced in the vocabulary lists in the ten textbooks.



(2) 呢 ne

呢 → 口 + 尼

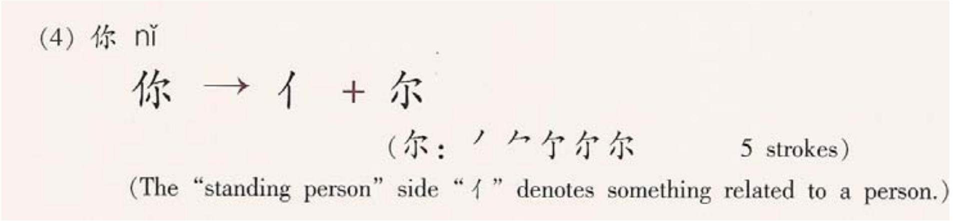
(“口” denotes the meaning of speaking, “尼” denotes the pronunciation)

(3) 妈妈 māma (媽媽)

妈 → 女 + 马

(“女” denotes the meaning of woman, “马” denotes the pronunciation)

(p.22)



(4) 你 nǐ

你 → 亻 + 尔

(尔: 丿 ⺈ 亇 尓 尔 5 strokes)

(The “standing person” side “亻” denotes something related to a person.)

(p.34)

Figure 4.4 Screenshots of Chinese Character Decomposition in NPCR$_{2002}$ (p.22 and p.34)

Only one textbook NPCR$_{2002}$ devoted a fixed section in each lesson to preset orthographic awareness development material. This text contains three subparts; the first subpart presents fundamental rules and structures of Chinese orthography including stroke order, combination of strokes, graphic structure of the character, internal structure of character, and radical and phonetic components; the remaining two subparts in the section are designed to help students “learn and write basic Chinese characters”, and "learn and write the Chinese characters appearing in the texts". Figure 4.4 displays examples of character decomposition in NPCR2002. For example, the character 媽 is decomposed into 女 and 馬, and 女 denotes the meaning of woman and 馬 denotes the pronunciation. Character 呢 is decomposed into 口 and 尼, with 口 denoting the meaning and 尼 denoting the pronunciation.

The last location to find orthographic decomposition materials is in the APPENDIX. RC1961, CTBC1976 and BCR1977 provide the stroke-order of each character in the APPENDIX. No orthographic decomposition materials were found.

Table 4.2 Summary of Orthographic Decomposition Information in the Ten Textbooks

Location in Textbook	RC 1961	CTBC 1976	BCR 1977	CICrw 1994	PCR 1995	NPCR 2002	IC 2005	IC 2008	PAVC 2008	FEEC 2008
1.Introduction										
Six Book							SS	SS		
Radical			P				SS	SS		
Stroke							SS	SS		
2.Prelude/Interlude				WS						
3.Lesson_Section										
Vocab List										
Radical			FL*	C						
Stroke				C						
Phonetic										
4.Lesson_Section										
Character Formation						WS				
Stroke-order						SS				
5.APPENDIX										
Stroke-order	SS	SS								

Note. S = sentences; P= paragraphs; C=column; SS= subpart of the section; WS= the whole section; FL*= only in the first three lessons.

Table 4.2 presents a summary of the orthographic decomposition information in the five locations in the ten textbooks. Three textbooks, PCR1995, PAVC2008, and FEEC2008 do not provide explicit orthographic decomposition information. While

BCR1977 contains a few paragraphs mentioning the importance of radicals in the introduction, IC2005 and IC2008 contain three subparts introducing The Six Books, radical, basic strokes, and stroke order in the introduction. RC1961 and CTBC1976 only provide stroke-order of characters introduced in the texts in the APPENDIX. Compared with the other eight textbooks, CICrw1994 and NPCR2002 systematically provide explicit orthographic decomposition materials throughout the textbooks by using columns and subsections in each lesson. Unlike CICrw1994 using prelude, interlude, and columns in the vocabulary list to present materials, NPCR2002 provides comprehensive materials within each lesson, with the orthographic decomposition materials being more relevant to the characters introduced in the text. These materials show learners that characters can be further decomposed into different functional components. Therefore, learners can see some explicit examples that can help them develop orthographic awareness while learning the characters in each lesson.

In conclusion, the results from examining whether or not textbooks contain explicit orthographic decomposition instruction revealed that most textbooks rarely include explicit orthographic decomposition instruction in the vocabulary lists and/or in each lesson. It should be noted, however, that two of the more recently published texts, CICrw1994 and NPCR2002, contained systematic and explicit instruction into character decomposition, indicating a recent awareness among textbook designers as to the importance of providing this instruction to students.

4.2 Results of Investigating Character Diversity and Repetition Comparison in the Textbooks

Research question 2: To what extent is a single character combined with other characters to form words in textbooks?

The second focus of this study was to investigate character diversity and repetition in the ten textbooks. DeFrancis (1977) stated that basic to developing reading skill in Chinese is for learners to become familiar with the processes of character combination in Chinese. Such familiarity is best acquired by mastering several combinations for a limited number of characters rather than by learning one or two compounds for many characters.

Vocabulary lists in the ten textbooks were the essential data source in the current study. Since most words in modern Chinese are composed of combinations of characters, word entries were entered in the database and were separated into character entries later, so each character still can be traced back to word entry if needed. Therefore, the character database includes both character (字 *zì*) in the character entry and its character combinations (詞 *cí*) in the word entry. In this database, 6,648 character entries were created including its character combinations in one of the columns.

4.2.1 Descriptions of Characters and Words in the Ten Textbooks

Table 4.3 presents a summary of character and word distribution in the ten textbooks. Vocabulary lists were not included in all lessons. The last lesson in RC$_{1961}$, for example, only contained reading texts without a vocabulary list. In CTBC$_{1976}$ and BCR$_{1977}$, after every five lessons, one lesson is devoted to exercises and drills for the previous five lessons, so vocabulary lists were not included in all lessons as shown in result type 1 in Table 4.3.

In CTBC$_{1976}$ and IC$_{2005}$, there are supplementary characters in the supplementary lessons at the end of the book. RC$_{1961}$ is the only textbook that combined romanization and characters in combination, as with “一路平 ān”, “ān” stands for the character “安”. Some other examples are listed in the following with characters in parentheses representing the missing character:

> tán(談)一 tán (談), lóu (樓)上, 上 haǐ (海), 上 kè (課), shìyè(世
> 界 jiè)上, syiāng (鄉 xiāng)下, 下 wǔ (午), 不 gǎn 當, jì (記)不住,
> dyoū1(丟 diū)臉, 九個 bàn (半)月, gānjìng (乾淨), 說 gù (故)事,
> lǐ (李)二, 北 jīng(京)人, 衣 shang(裳), ràng 讓 , etc.

These characters were not counted as characters found in this database, but these words did count as words.

Table 4.3 Summary of Character and Word Distribution in the Vocabulary Lists

Result types	RC	CTBC	BCR	CIC$_{\text{rw}}$	PCR	NPCR	IC	IC	PAVC	FEEC
	1961	1976	1977	1994	1995	2002	2005	2008	2008	2008
1.Number of Lesson including vocabulary lists	19/20	20/24	28/33	14/14	30/30	14/14	23/23	20/20	12/12	12/12
2.Total number of different word	797	791	947	814	547	370	1,180	847	451	530
1-character word	40.2%	64.0%	30.6%	26.8%	39.1%	48.1%	33.4%	31.8%	49.0%	40.2%
2-character word	45.7%	32.1%	53.3%	48.5%	53.6%	46.2%	52.5%	55.1%	44.6%	48.5%
3-character word	12.5%	3.8%	13.3%	15.7%	6.8%	4.6%	11.4%	9.0%	5.5%	7.7%
4-character word	1.6%	0.1%	2.5%	6.6%	0.4%	1.1%	2.7%	3.8%	0.9%	3.0%
5-character word			0.2%	1.6%	0.2%			0.1%		0.2%
6-character word				0.6%				0.1%		0.4%
7-character word				0.1%				0.0%		
8-character word								0.1%		
Above 1-character	59.8%	36.0%	69.4%	72.5%	60.9%	51.9%	66.6%	68.0%	51.0%	59.4%
3.Total number of character repeated	1,300	1,176	1,826	1,703	954	678	2,306	1,609	733	956
4.Total number of different character	308	509	281	695	533	354	811	759	314	449
5.Average number of character repeated	4.22	2.31	6.50	2.45	1.79	1.92	2.84	2.12	2.33	2.13
6.Standard deviation of character repeated	3.98	1.58	5.53	2.36	1.50	1.38	3.05	2.07	2.00	1.81
7.Total number of character found including different pronunciations of the same character	321	523	295	708	543	360	831	774	323	454

Looking at result type 2 in Table 4.3, CTBC$_{1976}$ contains the highest proportion of 1-character words (64%) as compared to other textbooks (26.8% ~ 49%). In other words, single character introduction is the focus of CTBC$_{1976}$. Seven out of ten textbooks contained more two-character words than others, with the proportions ranging from 45.7% to 53.6%. After separating words from characters, character repetition numbers are varied and range from 682 to 2,306, as is shown in result type 3. Eventually,the total number of different characters was found, shown in result type 4, with half of the ten textbooks containing less than 500 characters. The average numbers and standard deviations of character repetition were shown in result type 5 and 6. BCR$_{1977}$ and RC$_{1961}$ had the higher average number of characters repeated, with the repeated numbers being more varied than other textbooks. If the same character with different pronunciations is taken into account, the total number of different characters is shown in result type 7.

4.2.2 Results Investigating Character Combination Frequency in Textbooks

Research question 2: Research question 2: To what extent is a single character combined with other characters to form words in textbooks?

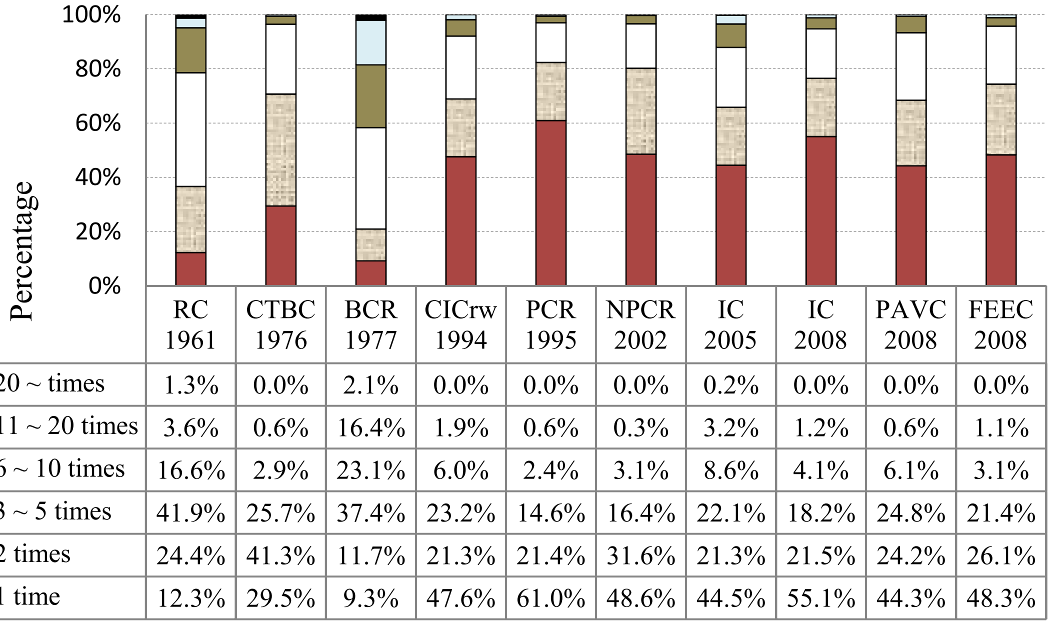
To answer this question, the times characters combined to form different words were calculated, as shown in Table 4.4 and Figure 4.5. Comparing the total number of times the same character combined with other characters to form words identified the characters having the highest character combination frequency in the database.

Table 4.4 Summary of Characters Combing to Form Different Words Across Textbook

Comb[1] Times	RC	CTBC	BCR	CIC$_{rw}$	PCR	NPCR	IC	IC	PAVC	FEEC
	1961	1976	1977	1994	1995	2002	2005	2008	2008	2008
36			1							
34			1							
33										
32	1									
29	1									
25			2							
24	1									
23	1									
22			2							
21							2			
20	2		2				1			
19			4	1			3	1		
18			2				2			
17	1		5	1			3		1	
16	2		5					2		
15			4				4	2		1
14	2		5	1	2		1			1
13	2	2	3	6			2			1
12	1	1	6	3			4	2		1
11	1		10	1	1	1	6	2	1	1
10	1	1	8	7		1	4	4	2	
9	6	1	7	8		1	11	4	4	1
8	11	6	16	2	3		13	4	3	2
7	8	4	18	9	5	2	16	9	5	7
6	25	3	16	16	5	7	26	10	5	4
5	29	18	32	28	7	6	36	25	9	15
4	35	26	30	45	25	10	59	43	17	24
3	65	87	43	88	46	42	84	70	52	57
2	75	210	33	148	114	112	173	163	76	117
1	38	150	26	331	325	172	361	418	139	217
Total Char.	308	509	281	695	533	354	811	759	314	449

Note. [1] Combination times: Number of times the character combined to form different words.

[2] The proportion of one time character combination in RC1961.



	RC 1961	CTBC 1976	BCR 1977	CICrw 1994	PCR 1995	NPCR 2002	IC 2005	IC 2008	PAVC 2008	FEEC 2008
■ 20 ~ times	1.3%	0.0%	2.1%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%
□ 11 ~ 20 times	3.6%	0.6%	16.4%	1.9%	0.6%	0.3%	3.2%	1.2%	0.6%	1.1%
■ 6 ~ 10 times	16.6%	2.9%	23.1%	6.0%	2.4%	3.1%	8.6%	4.1%	6.1%	3.1%
□ 3 ~ 5 times	41.9%	25.7%	37.4%	23.2%	14.6%	16.4%	22.1%	18.2%	24.8%	21.4%
▨ 2 times	24.4%	41.3%	11.7%	21.3%	21.4%	31.6%	21.3%	21.5%	24.2%	26.1%
🟥 1 time	12.3%	29.5%	9.3%	47.6%	61.0%	48.6%	44.5%	55.1%	44.3%	48.3%

Times of character combined to form different words

Figure 4.5 Times Characters Combined to Form Different Words Across Textbooks

As illustrated in Figure 4.5, BCR$_{1977}$, RC$_{1961}$ and IC$_{2005}$ have higher numbers of character combinations that appear more than 20 times. Both BCR$_{1977}$ and RC$_{1961}$ introduced individual characters first, such as the character 你, and later the character combinations (i.e. words, such as 你的) were provided. Therefore, learners were guided to notice the character combinations involving each character. However, across textbooks, most of the characters appeared one or two times to form different words. The proportions of one time ranged from 9.3% in BCR$_{1977}$ to 61.0% in PCR$_{1995}$, and the proportions of two times ranged from 11.7% in BCR$_{1977}$ to 41.3% in CTBC$_{1976}$. Combining proportions of one and two times, the proportions in PCR$_{1995}$ and NPCR$_{2002}$ were above 80%; the proportions in IC$_{2008}$, FEEC$_{2008}$, and CTBC$_{1976}$ above 70%; and the proportions in CIC$_{rw1994}$, PAVC$_{2008}$, and IC$_{2005}$ above 60%.

In terms of character diversity, the characters that appear in all ten textbooks were also determined , as they are the core vocabulary in this database. One hundred and eight characters appeared in all ten textbooks, and they are listed below from higher character combination frequency to lower frequency:

兒一子不上天學國人文　生下中大來麼飯小好有　年的家日是書意法了
會西開以話多點水說可看　要明這本去時為東那得　對過名在問們女晚字
坐事回見氣請個常你友教　吃朋沒思太幾快還今所　到再張我鐘想他姓道
都先現買貴寫少四念經二　歡知叫喜易容每誰

Table 4.5 List of Characters that Appeared Only in One Particular Textbook

Times	BCR 1977	RC 1961	PAVC 2008	NPCR 2002	FEEC 2008	CTBC 1976	PCR 1995	CICrw 1994	IC 2008	IC 2005
1	版 省 朝 縣	鋪		術楊 聖聚 梁煉 貝宋 誕壽 炎嗓 牀纔	傳領 決章 臨弄 玉貿 稍爬 帳填 裹鹹 叉椒 姊欸	於府 興划 蘇故 繼永 困遇 獎晨 尤仁 拼虎 仰曉 駕獅 賬唉 杭笨 瑰玫 杏夥	相使總 解團布 廠誌官 村待聞 答竟武 判詩雜 互蘭翻 招齊魯 贊努譯 迅裁閱 郭朗誼 覽茅輔 沫帕釣 檳佔酪 鍊	此由入什臺性至 任形即品基取未 類選需參派落般 例推巴陽修細止 察狀討龍警肯執 彩弱麥耳絲普禁 購暴甲訊操籍兄 袋側咱伍齡炸鍋 粗黎坡森池澳淺 朵仔桂盒乙灘斜 錦盲厭堡丙煮敦 蒸巾勿垃浴圾橡 蔥薯肆煎茄仟壹 捌柒玖蕃薑貳飩 餛	管准響 器勝顧 靠草免 亂胡托 稅野博 橋恩返 珠挺植 叔摩鹽 棋餅塵 桃酷懶 晴斑匹 兜蘿葱 寵毯枕 梨艙碟 涕瑜珈 莓萄蹟	各及颱指 變講滿調 稱苦存鄉 樹傷葉庭 減彈順鮮 淚雄探醒 擺慶鎮吹 慣閉辛琴 肥賓冠丟 泰扣飾罰 漲賺颳菲 橄欖
Sum	4	1	0	14	18	28	43	99	48	46

However, 297 characters appeared only in a particular textbook, as presented in Table 4.5. Compared with the other textbooks with similar total amounts of characters, CICrw$_{1994}$ contained more characters (99) that appeared only in the textbook compared to 48 characters in IC$_{2008}$ and 46 characters in IC$_{2005}$. For textbooks containing around 500 characters, more characters (43) appeared only in PCR$_{1995}$ than in CTBC$_{1976}$ (28).

Comparing the total number of characters repeated, the characters having higher character combination frequency in each textbook were identified. Table 4.6 presents the first 20 characters having higher character combination frequency (at least 4) in the database and in each textbook. The number next to a character is the character combination frequency. In addition, if a character has the same character combination frequency as the 20th character, it was listed in Table 4.6 as well. For example, the 20th character is 年 in the database, but character 有 has the same character combination frequency, so 有 was listed as well. Among these characters, character 兒 *ér* has the highest character combination frequency because it is a noun suffix in the words 有空兒, 餐館兒, 事兒, 孫女兒, 有點兒, and 一下兒, and because adding it to the end of many words is a unique characteristic of the Beijing dialect. The character 子 *zǐ* is also a common noun suffix, as in the words 位子, 刀子, 包子, 叉子, 單子, and 女孩子. The character 學 *xué* has a higher combination frequency which may due to its tendency to combine into words that depict school settings, such as 大學生 (college student), 大學 (college or university), 學校 (school), 學期 (semester), 同學 (student or classmate), and 學習 (study).

The results from examining character diversity and repetition indicate that the design of character instruction and content settings may influence character selection, which further affects character combination frequency.

Table 4.6 Summary of First 20 High Character Combination Frequency Characters

Comb. Rank	Database[1]	RC 1961	CTBC 1976	BCR 1977	CIC$_{\mathrm{rw}}$ 1994	PCR 1995	NPCR 2002	IC 2005	IC 2008	PAVC 2008	FEEC 2008
1	兒 168[2]	一 32	子 13	兒 36	一 19	兒 14	學 11	國 21	子 19	兒 17	子 14
2	一 142	兒 29	學 13	學 34	天 17	子 14	生 10	不 21	兒 16	一 11	上 13
3	子 138	不 24	兒 12	不 25	人 14	人 8	兒 9	兒 20	文 16	子 10	一 12
4	不 121	上 23	文 10	人 25	兒 13	年 8	天 7	天 19	天 15	上 10	下 8
5	上 119	下 20	不 9	文 22	生 13	學 8	麼 6	一 19	上 15	天 9	好 8
6	天 117	的 20	一 8	一 20	子 13	一 7	上 6	子 19	學 12	國 9	不 7
7	學 117	子 17	麼 8	家 20	日 13	天 7	國 6	中 18	國 11	下 9	麼 7
8	國 107	天 16	小 8	上 19	國 13	生 7	大 6	文 18	大 11	文 8	大 7
9	人 96	國 16	中 8	生 19	大 12	上 7	們 6	好 17	一 10	飯 8	開 6
10	文 95	來 14	天 7	會 19	你 12	國 6	中 5	來 17	中 10	書 8	天 5
11	生 93	人 13	國 7	中 18	中 11	大 6	不 5	生 15	飯 10	不 7	國 5
12	下 88	有 13	得 7	說 18	不 10	有 6	問 5	上 15	不 9	好 7	學 5
13	中 88	了 12	東 7	小 17	來 10	書 5	可 5	人 15	生 9	麼 7	家 5
14	大 84	麼 10	以 6	的 17	飯 10	文 5	為 4	下 14	下 9	學 6	生 5
15	來 76	大 9	上 5	法 17	年 10	下 5	一 4	學 13	好 8	人 6	去 5
16	飯 74	生 8	的 5	意 17	點 10	們 5	有 4	西 13	人 8	那 6	明 5
17	麼 73	飯 8	人 5	子 16	西 10	會 5	小 4	小 12	麼 8	有 6	飯 4
18	小 71	是 8	了 5	來 16	學 9	日 4	多 4	麼 11	西 7	這 6	有 4
19	好 70	去 8	飯 5	大 16	話 9	中 4	他 4	大 11	可 7	中 5	這 4
20	年 69	坐 8	年 5	日 16	書 9	不 4		年 11	水 7	年 5	中 4
21	有 69	回 8	法 5		東 9	飯 4		家 11	本 7	看 5	年 4
22			明 5			小 4		到 11		女 5	請 4
23			那 5			是 4					們 4
24			時 5			家 4					小 4
25			見 5			開 4					多 4
26						以 4					對 4
27						教 4					以 4
28						問 4					話 4
29						友 4					名 4
30											水 4
31											東 4

Note. [1] Combining character combination data from all ten textbooks as one database.

[2] Character combination frequency.

4.3 Results of Investigating Textbook Character Frequency Rank Selection

Research question 3: To what extent do textbooks contain high-frequency characters as documented by accepted Chinese character frequency lists over time?

The third focus of this study was character selection in term of frequency rank. In the majority of the textbooks, vocabulary was selected from frequency lists. Research question 3 was designed to answer whether textbooks contain high-frequency characters. To find this information, I used frequency lists constructed during different periods of time (1980s and 2000s) and for different purposes (proficiency test and internet search), as stated in section3.2 . A character frequency rank worksheet was designed, as shown in Table 3.3. with the results presented in the following sections.

To answer research question 3, the first investigation process was to classify the frequency rank of each character using three frequency lists that were constructed in the 1980s and 2000s: Chinese Character Attributive Dictionary 漢字屬性字典 *Hànzì shǔxìng zìdiǎn* (1989), "Characters by frequency" list from Wenlin 文林 Software for Learning Chinese, version 3.4 (2007), and A Frequency Dictionary of Mandarin Chinese (Xiao, Rayson, & McEnery, 2009). Since the textbooks were published from 1961 to 2008, if we used the more recently compiled frequency lists to classify the textbooks published before 2000, there is a possibility that the textbooks might not contain the same high-frequency characters.

Of the characters introduced in the textbooks compiled as one database, the total number of characters is 1,166, including 27 characters having different pronunciations. Figure 4.5 gives the trends of the amount of characters across high to low frequency of usage in the three frequency lists. High frequency characters were found more than low frequency characters. The top 100 most frequently occurring characters were all included.

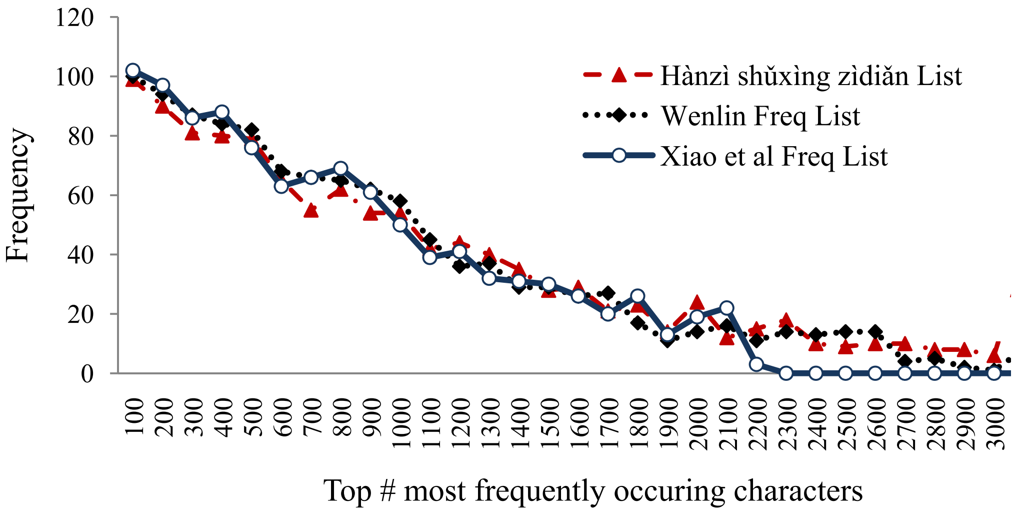


Figure 4.6 Trends of Most Frequently Occurring Characters in All Ten Textbooks

Based on the 漢字屬性字典 *Hànzì shǔxìng zìdiǎn* (1989) frequency list, a declining amount of characters was found from higher frequency to lower frequency characters, with the frequency rank ranging from the top 1 (的) to top 7464 (柒). Three characters (欸, 檳, 磯) could not be found in the dictionary (code 4001 in Figure 4.5), and they only appeared in particular textbooks; 欸 *èi* in FEEC$_{2008}$, 檳 (香檳酒) in PCR$_{1995}$, and 磯 (洛杉磯) in CICrw$_{1994}$ and IC$_{2008}$. The frequency ranks of 38 characters were above the top 3000 (code 3001 in Figure 4.5): 襪姊餃乒乓啤筷逛涕磅碟嗽澡廁蕃恤圾垃酪橄欖 ranked between 3001 and 4000, 娜瑜橘單餛 ranked between 4001 and 5000, 玖壹貳噓伯莓 ranked between 5001 and 6000, 廠佰仟珈 ranked between 6001 and 7000, and 捌柒 ranked above 7000.

Based on the Wenlin (2007) frequency list, a declining amount of characters was found from the top 500 to 600 characters (80s to 60s). Eight characters 餃橘恤嗽藩夥檳瑜 were classified above the top 3000 in the Wenlin frequency list, and 27 characters could not be found in the Wenlin frequency list: 佰仟壹貳柒捌玖莓萄蕃薑橄欖餛飩酪裏淨週遊傢噓欸珈蹟鍊佔. A complete list of the characters appearing in the ten textbooks by Wenlin exposure frequency is shown in APPENDIX.A, with the shaded characters being radical characters.

Based on the Xiao et al. (2009) frequency list, a declining amount of characters was found from the top 400 to 600 characters (80s to 60s). Because there are fewer characters in the Xiao et al., (2009) frequency list compared with the Wenlin and 漢字屬性字典 lists, more characters (106) could not be found in Xiao et al., (2009): 5 characters 布划游洲歐 were in the top 1000 rank in Wenlin; 19 characters 魯趙俄紐韓黎貝郭洛澳泰悶桂姆辣颳堡碧丙 were between the top 1001 and 2000 rank; 53 characters 糟娜葡裹敦萄菲茅蝦臘鹹芝匆唉沫兜蘿橡葱蔥杭汁寵薯拌癢夷毯淹肆筷帕逛瑰乾釣煎玫枕梨椒艙杉杏碟磅噢涕茄蕉襪磯姊 were between the top 2001 and 3000 rank; 7 characters 嗽恤橘檳瑜藩餃 were above the 3000 rank, and 22 characters 佰仟壹貳柒捌玖莓蕃薑酪橄欖餛飩欸佔傢噓蹟鍊珈 could not be found in either Wenlin or Xiao et al.(2009).

Combining characters in terms of the most frequently used 500, 1000, 1500, 2000, 2500, 3000 and above characters, except for the characters that could not be found in the frequency lists, the proportion distributions were similar, based on the three lists. This is shown in Figure 4.6, where about 87.4%, 88.9%, and 88.8% of the characters belong to the top 2000 characters in 漢字屬性字典, Wenlin, and Xiao et al.,(2009) lists. About 97% of the characters in the database belong to the top 3000 most frequently used characters in the 漢字屬性字典 and Wenlin frequency lists while 91% of them are found in the Xiao et al., (2009) frequency list.

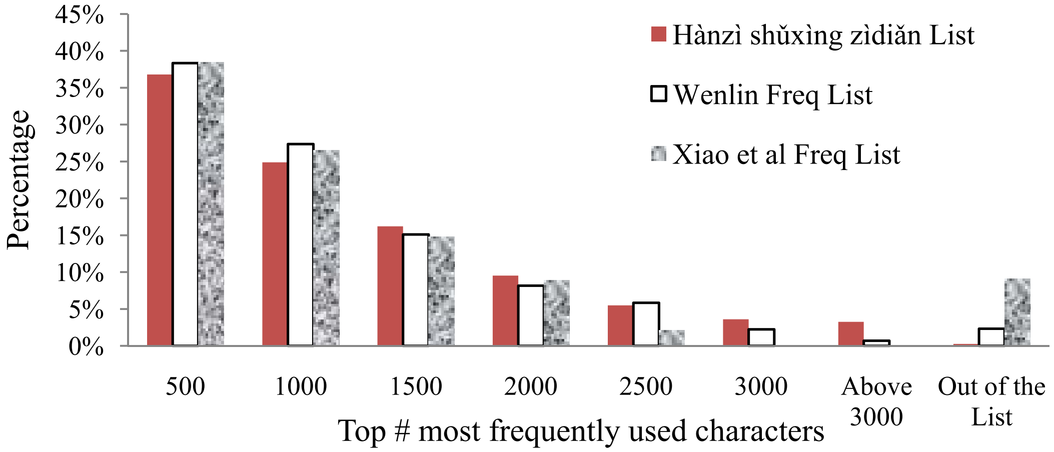


Figure 4.7 Rank Distribution Differences for the Most Frequently Used Characters in 10 Selected Textbooks

Table 4.7 Descriptive Statistic Summary of Character Frequency Rank

Textbooks		RC 1961	CTBC 1976	BCR 1977	CICrw 1994	PCR 1995	NPCR 2002	IC 2005	IC 2008	PAVC 2008	FEEC 2008
Character No.		308	509	281	695	533	354	811	759	314	449
漢字屬性字典	Mean	426.4	635.7	363.6	846.5	720.9	674.1	830.2	871.1	530.5	675.0
	SD	399.5	651.1	406.0	941.1	744.0	728.3	799.0	867.0	546.7	677.2
	Max	2433	4298	4298	7464	5949	4113	6148	6817	5476	4298
	Min	1	1	1	1	1	1	1	1	1	1
	Miss				1	1		1			1
Wenlin	Mean	337.0	548.8	294.0	691.9	605.6	546.7	707.1	740.1	406.6	569.3
	SD	303.6	571.3	274.7	637.9	586.8	558.9	620.6	670.5	373.4	572.8
	Max	1798	3001	1715	3001	3001	2784	3001	3001	2701	2979
	Min	1	1	1	1	1	1	1	1	1	1
	Miss	1	1	1	14	5	2	9	11		2
Xiao et al.	Mean	329.3	487.7	299.6	623.2	541.9	508.8	643.0	646.2	398.8	515.0
	SD	301.8	464.5	275.3	533.0	489.6	509.1	524.0	543.1	375.4	480.7
	Max	1892	2112	1440	2107	2106	2089	2112	2112	1994	2107
	Min	1	1	1	1	1	1	1	1	1	1
	Miss		13		42	20	9	36	44	4	11

Note. SD = standard deviation of frequency rank. Max = largest number means the lowest frequency rank. Min = smallest number means the highest frequency rank. Miss = missing data means that frequency rank cannot be found.

Comparing characters across the ten textbooks, as shown in Table 4.7, IC$_{2005}$, CTBC$_{1976}$, and IC$_{2008}$ contained more characters than the other textbooks, as shown in Figure 4.10. The most characters in BCR$_{1977}$ were high frequency characters because their mean of frequency rank was the lowest mean (i.e. 363.6, 294, and 299.6) among textbooks in all three frequency lists. On the other hand, IC$_{2008}$ had more low frequency characters because its mean of frequency rank was the highest mean (i.e. 871.1, 740.1, and 646.2). Based on the 漢字屬性字典, character frequency ranks in CIC$_{rw1994}$ were more varied (i.e. SD = 941.1) than the other textbooks.

In addition, comparing the top 1000 frequently used characters across textbooks given in Figure 4.7, BCR$_{1977}$ contains the highest percentage (i.e., 94.7%, 97.9%, and 98.6%) of the top 1000 frequently used characters, with less than 75% of the characters in CIC$_{rw1994}$, IC$_{2005}$, and IC$_{2008}$ being the top 1000 most frequently used characters. APPENDIX B presents the frequency distribution of character frequency rank across the ten textbooks. Looking at the top 100 most frequently used characters across ten textbooks, except for that in PAVC$_{2008}$ classified by 漢字屬性字典, most of the textbooks contained at least 70 of the top 100 most frequently used characters. From the top 100 to top 200 characters, a decline was detected of at least 20 characters in all ten textbooks. Looking at the characters ranked above the top 3000, CIC$_{rw1994}$ contains the highest number of characters (i.e., 21) ranking above the top 3000characters ranked in the 漢字屬性字典.

The results indicated that, regardless of when the frequency lists were created, the ten textbooks contained many high frequency characters. If we use the more recently compiled frequency lists to classify the textbooks published before 2000, these textbooks still contain high frequency characters, which implies that these textbooks can still be used today in terms of character frequency usage. In particularly, BCR$_{1977}$ contains the highest percentage (i.e., 94.7%, 97.9%, and 98.6%) of the top 1000 most frequently used characters.

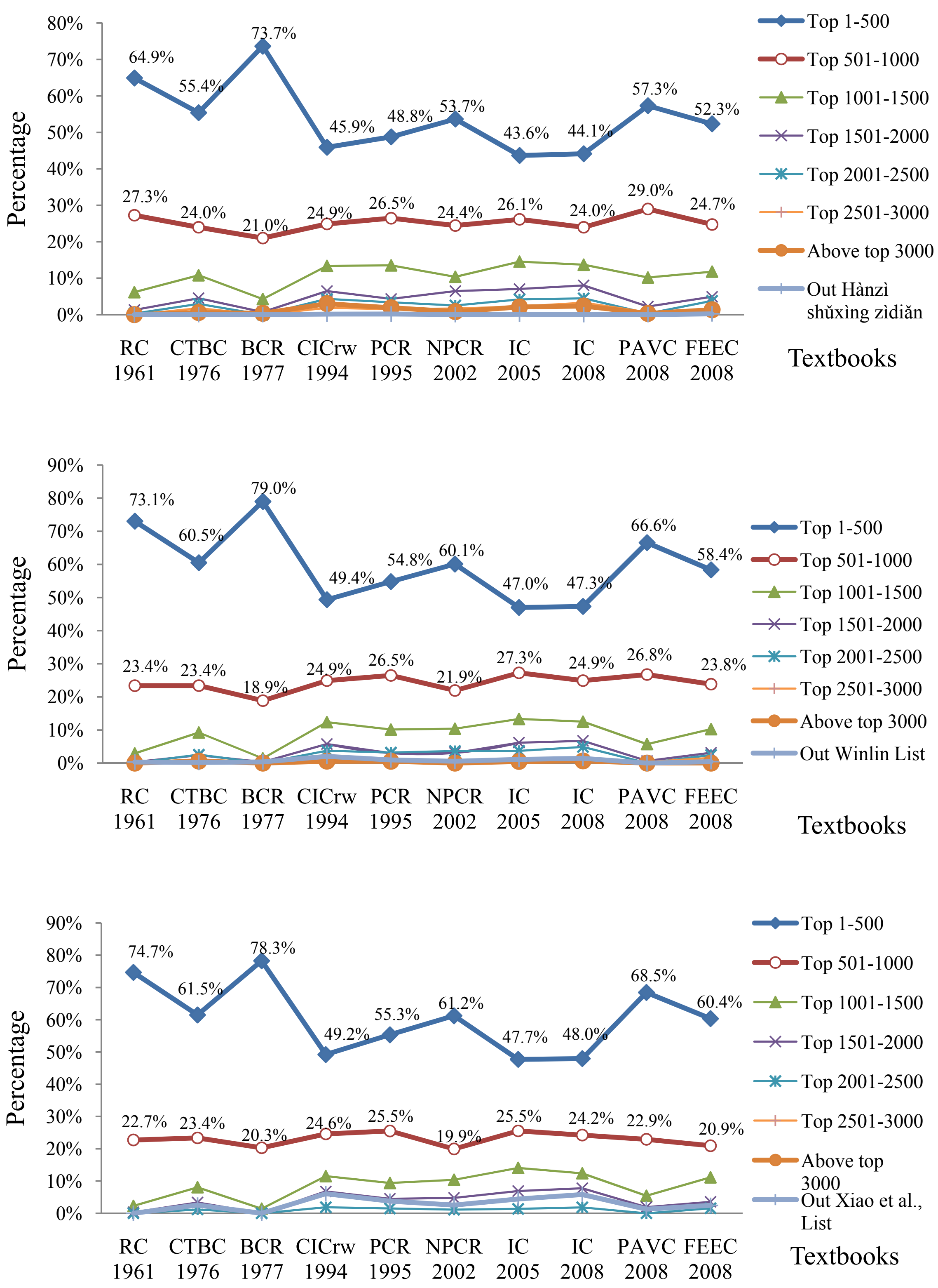


Figure 4.8 Distribution of Most Frequently Used Characters in the Textbooks

4.3.1 Results of Investigating Character Frequency Rank as Classified by Yahoo and Google Search Counts

To answer research question 3, the second investigation process was to classify each character using two frequency lists: the Chinese character usage frequency lists from Google and Yahoo Search Engines. Huang Yong (2009) took the most commonly used 2,500 characters, submitted each of them to Google and Yahoo Search Engines, recorded the search counts, and sorted the counts to obtain these two character frequency lists. Since materials on the internet are available to CFL learners, these two character frequency lists can help us to discover the usefulness of the characters in the ten textbooks.

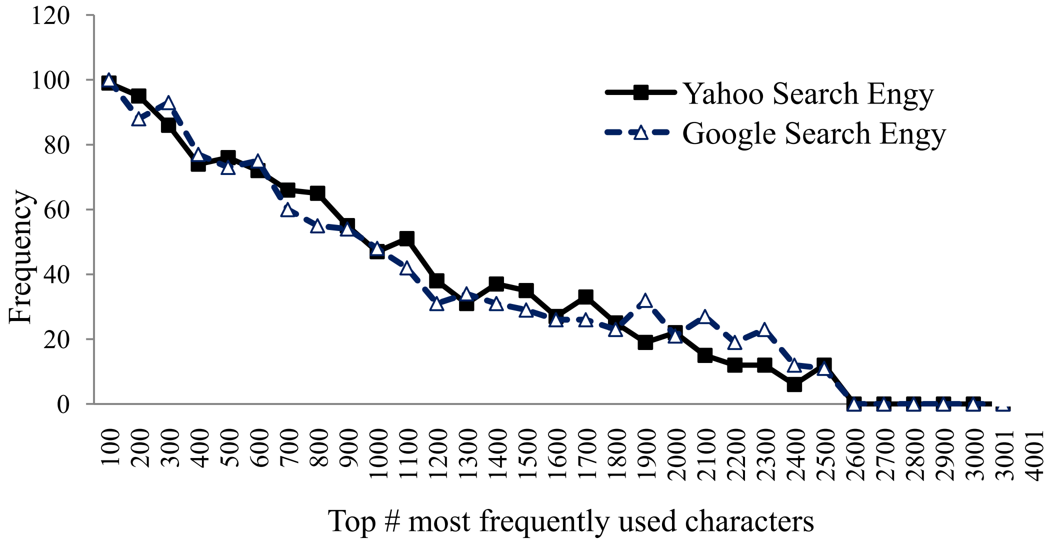


Figure 4.9 Trends of Cumulated Frequency Rank in Yahoo and Google Search Lists

Of the characters introduced in the textbooks compiled as one database, the total number of characters was 1,166 including 27 characters having different pronunciations. Figure 4.8 presents the trends of characters across high to low frequency of usage in the two frequency count lists. High frequency characters were found more than the low frequency characters. Combining amounts of characters in terms of the top 500, 1000, 1500, 2000, and 2500 most frequently found characters, except for the characters that could not be found in the lists (code 4001 in Figure 4.8), the proportion distributions were similar in the Yahoo and Google search lists; that is, about 63.0% and 62.0% of the characters belong to the top 1000 most frequently used characters. About 95.2% of the characters in the database belong to the top 2500 most frequently used characters in both the Yahoo and Google search lists. The results therefore indicate that the database contained many high frequency characters.

Table 4.8 Descriptive Statistic of Character Frequency Rank in Yahoo and Google Search

Textbooks		RC 1961	CTBC 1976	BCR 1977	CIC$_{rw}$ 1994	PCR 1995	NPCR 2002	IC 2005	IC 2008	PAVC 2008	FEEC 2008
Character	No.	308	509	281	695	533	354	811	759	314	449
Yahoo	Mean	395.0	572.0	358.3	678.6	639.8	625.4	710.5	719.1	430.4	563.9
	SD	387.3	520.1	365.3	572.2	572.3	597.9	571.4	592.6	410.6	520.6
	Max	2103	2449	1929	2484	2482	2477	2482	2482	2103	2408
	Min	1	1	1	1	1	1	1	1	1	1
	Miss		8		25	8	3	20	21	1	8
Google	Mean	456.6	625.1	369.8	715.0	664.8	647.8	759.6	768.4	485.2	609.6
	SD	468.8	604.4	399.9	633.2	615.1	646.1	634.4	650.7	493.5	587.2
	Max	2189	2387	2362	2494	2492	2491	2492	2492	2283	2405
	Min	1	1	1	1	1	1	1	1	1	1
	Miss		8		25	8	3	20	21	1	8

Note. SD = standard deviation of frequency rank. Max = largest number means the lowest frequency rank. Min = smallest number means the highest frequency rank. Miss = missing data means that frequency rank cannot be found.

Comparing characters across the ten textbooks, as summarized in Table 4.8, most characters in BCR$_{1977}$ are high frequency characters because its mean of frequency rank is the lowest (i.e. 358.3 and 369.8) among textbooks in both frequency lists. On the other hand, IC$_{2008}$ has more low frequency characters because its mean of frequency rank is the highest (i.e. 719.1 and 768.4). Character frequency ranks in NPCR $_{2002}$ and IC$_{2008}$ are more varied (i.e. SD = 597.9 and 650.7) than other textbooks. CIC$_{rw1994}$ contains the highest number of characters (i.e., 25) that could not be found in rankings above the top 2500 rank in the Yahoo and Google search count lists.

Comparing characters across ten textbooks, as shown in Figure 4.9, for the Yahoo search engine, BCR$_{1977}$, RC$_{1961}$, and PAVC$_{2008}$ contained higher percentages of the top 1000 most frequently used characters (above or around 90%), and CIC$_{rw1994}$, IC$_{2005}$, and IC$_{2008}$ contained less than 73% of the top 1000 most frequently used characters. For the Google search engine, BCR$_{1977}$, contained the highest percentage of the top 1000 most frequently used characters (about 93.6%), and IC$_{2008}$ contained the lowest percentage of the top 1000 most frequently used characters (about 70.4%). CIC$_{rw1994}$ contained the highest percentage (i.e., 3.6% in Yahoo and 3.6% in Google) of characters out of these two frequency lists. However, looking at the top 100 most frequently found characters across ten textbooks, IC$_{2005}$, CTBC$_{1976}$, and IC$_{2008}$ contained more characters than the other textbooks, as depicted in Figure 4.10. NPCR$_{2002}$ contained the fewest characters both in the Yahoo and Google search engines.

The results indicated that when using character usage frequency lists from Google and Yahoo search counts to classify the character frequency rank, the ten textbooks contained many high frequency characters, which implies that these textbooks still can be used today in term of character frequency usage on the Yahoo and Google search engines.

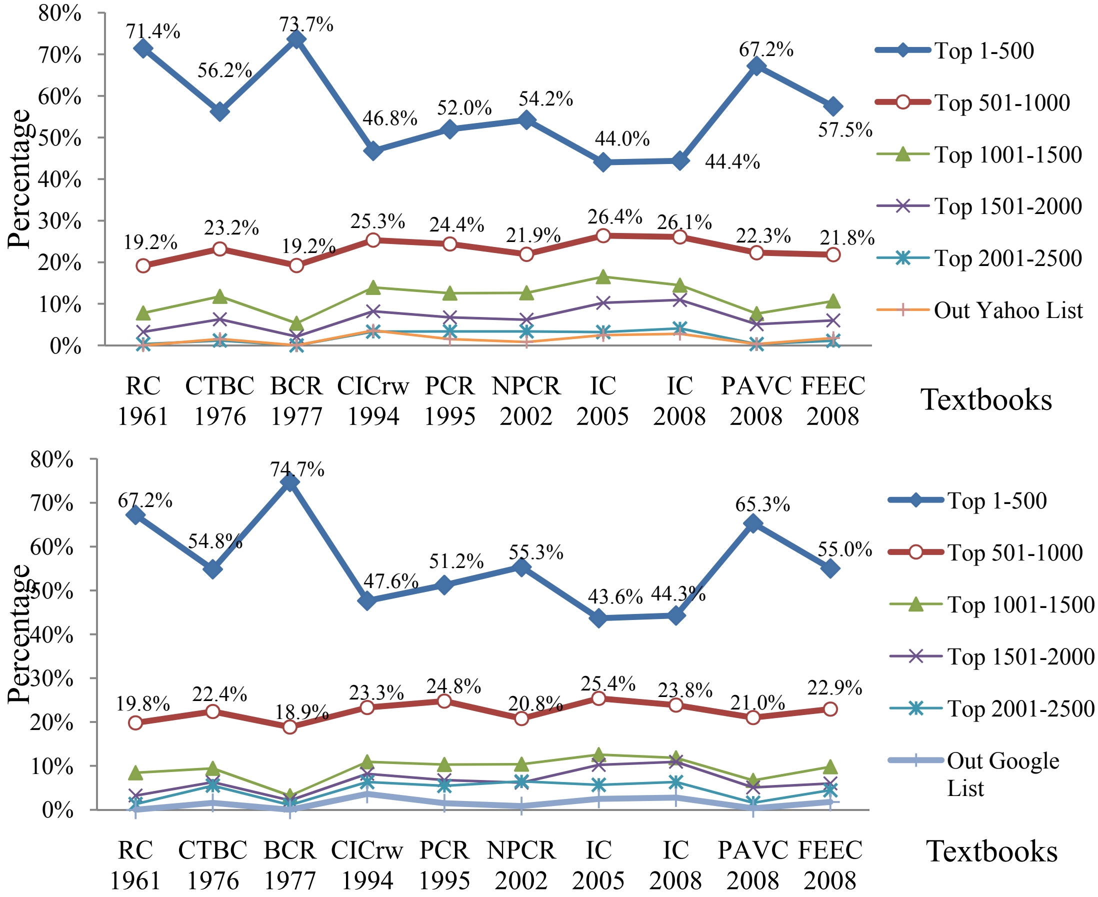


Figure 4.10 Percentage Distribution of Yahoo and Google Search Frequency Across the 10 Selected Textbooks

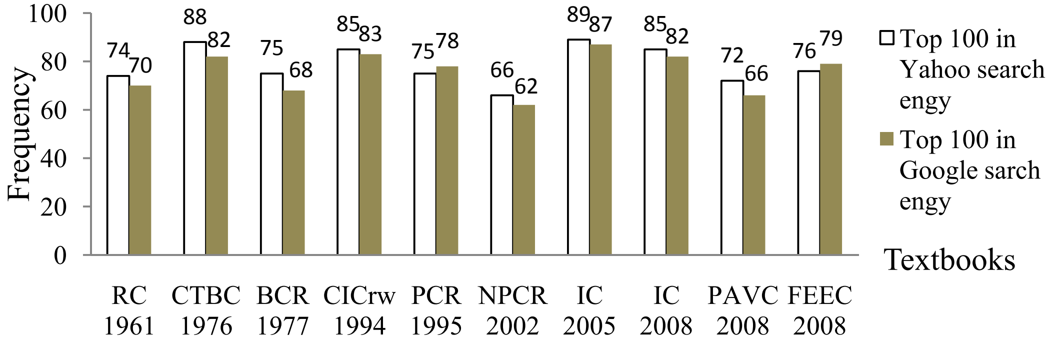


Figure 4.11 Frequency of the Top 100 Frequency Rank Character in Yahoo and Google Search

4.3.2 Results of Investigating Character Frequency Rank Classified by HSK Proficiency Level List

To answer research question 3, the third investigation process was to classify each character using *A level list of HSK word and character* 漢語水平詞彙與漢字等級大綱 *Hànyǔ shuǐpíng cíhuì yǔ Hànzì děngjí dàgāng* (1992). The HSK list contains a total of 2,905 Chinese characters in 4 levels (甲乙丙丁級). For basic level (Level 1 甲級), test-takers are expected to know 800 characters and words. In this study, I classified each character's level based on this HSK list to see whether characters from the ten beginning-level textbooks all belong to the basic level.

Table 4.9 Descriptive Statistics of Characters Classified by HSK Level List

Textbooks		RC 1961	CTBC 1976	BCR 1977	CIC$_{rw}$ 1994	PCR 1995	NPCR 2002	IC 2005	IC 2008	PAVC 2008	FEEC 2008
Character	No.	308	509	281	695	533	354	811	759	314	449
HSK	Mean	1.3	1.6	1.4	1.8	1.6	1.7	1.7	1.8	1.4	1.6
	SD	0.8	1.0	0.8	1.0	1.0	1.0	1.0	1.0	0.8	0.9
	Max	4	4	4	4	4	4	4	4	4	4
	Min	1	1	1	1	1	1	1	1	1	1
	Miss		3		16	6	7	13	13	2	3

Note. SD = standard deviation of frequency rank. Max = largest number means the lowest frequency rank. Min = smallest number means the highest frequency rank. Miss = missing data means that frequency rank cannot be found.

Comparing characters across the ten textbooks as displayed in Table 4.9, all characters in RC$_{1961}$ and BCR$_{1977}$ can be found in the four HSK levels, with CIC$_{rw1994}$ containing the highest number of characters (i.e., 16) that could not be found in the HSK levels. Figure 4.12 presents a frequency distribution of the characters classified by HSK levels. IC$_{2005}$ and IC$_{2008}$ contained more HSK basic level characters (i.e., 433 and 396) while RC$_{1961}$, PAVC$_{2008}$, NPCR$_{2002}$, and BCR$_{1977}$ contained less than 300 HSK basic level characters). Since test-takers are expected to know 800 characters in the HSK basic level

(Level 1 甲級), the results indicate that these beginning level CFL textbooks obviously do not cover all the HSK basic level characters.

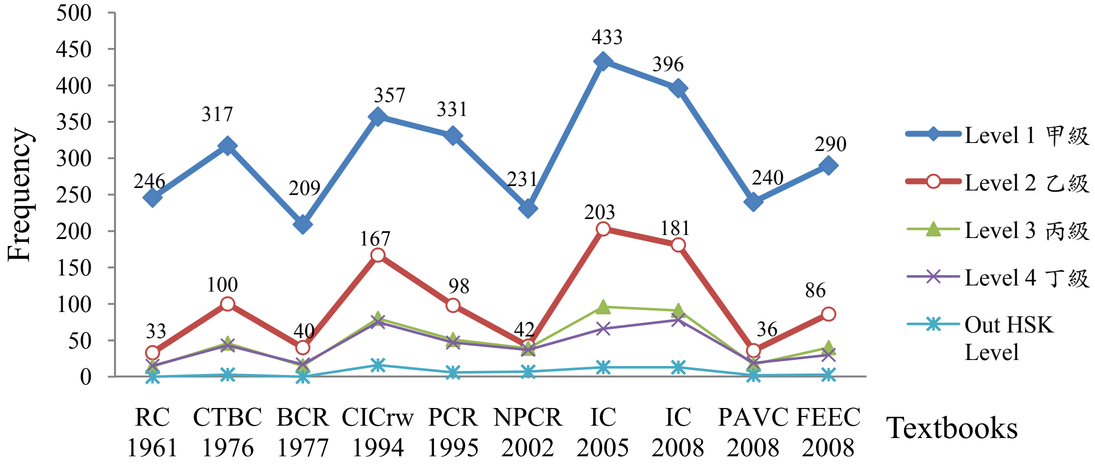


Figure 4.12 Frequency Distribution of Characters Classified by HSK Levels

4.3.3 Results of Investigating Character Frequency Rank for Characters Appearing in All Ten Textbooks

To answer research question 3, the last investigation process was to find the differences and similarities of frequency usage as documented by accepted Chinese character frequency lists. Characters appearing in all ten textbooks were identified and classified in terms of frequency usage, as they were seen as the core vocabulary in the beginning level CFL textbooks. In the current study, I used all six frequency lists used to answer research question 3.

Overlapping degree of frequency rank across textbooks was calculated to identify core vocabulary, and 108 core characters were classified in terms of their frequency rank in the six frequency lists. These frequency ranks can be compared within a frequency list, but they cannot be compared across different frequency lists. However, the descriptive statistic information indicates that, as shown in Table 4.10, these 108 characters rank within the top 300 most frequently used characters, with most of them being within the HSK levels.

Table 4.10 Descriptive Statistics for Characters Classified by Six Frequency Lists

Type	Hanzi shuxing zidian (1989)	Wenlin (2007)	Xiao etal (2009)	Yahoo (2009)	Google (2009)	HSK Level (1992)
Mean	250.1	175.3	165.2	277.9	278.8	1.3
SD	298.2	205.4	193.3	361.1	375.6	0.8
Max	1424	1025	956	1732	2033	4
Min	1	1	1	1	1	1
Miss	0	0	0	0	0	0

Note. SD = standard deviation of frequency rank. Max = largest number means the lowest frequency rank. Min = smallest number means the highest frequency rank. Miss = missing data means that frequency rank cannot be found.

Table 4.11 presents 108 core characters and their frequency ranks in the six frequency lists. Comparing core characters in the Wenlin (2007) frequency list, 57 characters belong to the top 100 (的一是不了人在我有中這大國上個來他為到時們年生會子要以說學得對下那可過多小你天家去都好日還經沒麼本所事現想開二道看), 41 characters belong to the top 101-500 rank (法文意點四問名知回明氣話幾女水見常西少東太再書先每教請今張思吃寫兒叫友字容快晚易買), 9 characters belong to the top 501-1000 (歡喜飯坐念朋誰鐘貴), and only 1 character is above the top 1000 (姓).

Table 4.11 Frequency Rank of 108 Core Characters Appeared in All Ten Textbooks

CH	HZ	WL	XF	YA	GO	HSK	CH	HZ	WL	XF	YA	GO	HSK	CH	HZ	WL	XF	YA	GO	HSK
的	1	1	1	1	12	1	得	62	35	34	90	205	1	先	254	235	160	72	75	1
一	2	2	2	6	17	1	經	63	69	62	389	391	2	回	255	144	144	79	33	1
是	3	3	3	12	73	1	家	72	53	29	54	145	1	東	280	202	270	748	208	1
在	4	7	6	11	47	1	水	77	167	149	68	44	1	思	292	306	223	647	782	4
不	5	4	5	13	85	1	二	81	93	193	62	1	1	再	309	211	182	73	143	1
了	6	5	4	25	88	1	小	84	49	67	38	15	1	教	318	237	205	316	301	1
有	7	9	8	15	72	1	現	86	82	57	455	288	3	每	331	236	271	156	279	1
人	9	6	7	7	28	1	都	90	63	50	29	38	1	名	347	137	161	32	43	2
這	10	11	10	100	150	1	點	97	128	82	95	89	1	張	351	301	354	321	255	1
中	11	10	33	5	21	1	本	100	76	114	17	7	1	話	355	151	116	344	268	1
大	12	12	13	9	39	1	好	104	66	41	41	140	1	今	361	267	202	108	154	1
為	13	18	21	33	68	1	開	106	87	65	264	303	1	叫	379	363	306	363	739	1
上	14	14	14	8	23	1	還	109	68	49	97	195	1	書	392	230	191	504	120	1
個	15	15	11	60	24	1	天	117	52	54	67	232	1	容	395	409	355	1021	639	3
國	16	13	63	167	102	1	四	119	129	253	87	3	1	兒	396	359	107	1015	296	1
我	17	8	9	19	22	1	日	120	67	230	3	8	1	太	440	203	261	159	52	1
以	18	32	25	66	112	2	那	121	38	38	176	404	1	寫	451	343	368	301	423	1
要	19	30	16	34	113	1	事	124	81	51	80	189	1	快	456	416	260	210	322	1
他	20	17	12	45	69	1	你	143	51	31	46	64	1	易	468	453	451	543	452	3
時	21	22	28	116	11	4	明	144	145	113	507	430	1	女	515	159	119	127	127	1
來	22	16	15	92	133	1	看	145	98	64	65	222	1	字	551	396	327	179	141	1
們	24	23	17	255	376	1	麼	148	74	43	872	1063	1	吃	627	342	304	235	647	1
生	25	25	32	184	251	2	氣	154	150	199	896	148	2	請	681	261	394	64	29	1
到	26	19	19	35	94	1	道	157	95	76	183	269	1	坐	701	590	589	489	202	1
對	33	36	27	70	108	1	沒	163	70	61	243	45	1	念	718	593	441	731	1269	1
會	35	26	22	44	35	1	問	166	134	79	403	157	1	友	723	371	308	390	561	1
可	36	39	44	56	16	2	意	167	119	99	639	591	1	歡	827	518	445	1438	1821	1
年	39	24	46	2	13	1	去	170	54	52	118	310	1	晚	838	440	468	544	650	1
下	45	37	45	21	10	1	想	180	85	85	98	258	1	貴	848	832	869	785	816	1
過	46	40	39	152	27	1	常	198	189	166	293	327	1	鐘	849	828	748	1732	1030	1
子	47	29	36	203	142	3	文	199	115	128	225	104	4	誰	871	627	465	242	274	1
說	48	33	18	136	230	1	少	220	196	169	291	406	1	喜	877	579	491	644	991	4
多	55	48	35	22	57	1	知	225	141	129	368	411	3	買	886	484	512	259	56	1
學	58	34	48	244	153	1	見	230	188	159	478	247	1	飯	996	583	601	1276	1345	1
法	59	102	115	139	119	4	幾	242	153	163	207	63	1	朋	1339	605	626	1710	2033	1
所	60	79	81	55	9	2	西	248	191	227	298	266	1	姓	1424	1025	956	1350	1510	1

Note. CH = character; HZ = Hànzì shǔxìng zìdiǎn frequency rank; WL = Wenlin frequency rank; XF = Xiao et. al. (2009), frequency list; YA = Yahoo search engine count rank; GO = Google search engine count rank; HSK = HSK level.

4.3.4 Result Summary of Character Frequency Rank Selection in the Textbooks

The third focus of this study was the character selection in term of frequency rank. In the majority of the textbooks, vocabulary was selected from frequency lists. Research question 3 was designed to answer whether textbooks contained high-frequency characters as compared to established Chinese character frequency lists. Four investigation processes were used.

The results for process 1 indicated that, regardless of using frequency lists from the 1980s or 2000s to classify character frequency rank, the ten textbooks contain a high degree of high frequency characters. If we use the more recently compiled frequency lists to classify the textbooks published before 2000, these textbooks still contain a high percentage of high frequency characters, which implies that these textbooks still can be used now in terms of character frequency usage.

The results for process 2 indicated that when using character usage frequency lists from Google and Yahoo search counts to classify the character frequency rank, the ten textbooks contain many high frequency characters, which implies that these textbooks still can be used now in term of character frequency usage on Yahoo and Google search engines.

The results for process 3 indicated that all ten textbooks contain HSK basic level characters. However, since test-takers are expected to know 800 characters in the HSK basic level (Level 1 甲級), the results indicate that these beginning level CFL textbooks do not cover all the HSK basic level characters.

The results for process 4 indicated that 108 core characters which appeared in all ten textbooks rank within the top 300 most frequently used characters and most of them are within the basic HSK level (Level 1 甲級).

4.4 Results of Character Radical Diversity and Repetition in the Textbooks

The fourth focus of the study was to investigate radical component diversity and repetition in character selection. Research questions 4, 5, and 7 were designed for this purpose, and include radical combination frequency, character graphic structure distribution, radical positional regularity among character graphic structure, and radical semantic transparency. The etymological-based dictionary 中文字譜 *Zhōngwén zìpǔ*, *Chinese Characters: A Genealogy and Dictionary* (Harbaugh, 1998) and *New Chinese Dictionary* 最新國語新辭典 (1987) was used to code semantic radicals, meanings of the semantic radicals, and meanings of the characters.

4.4.1 Investigating Radical Combination Frequency in the Textbooks

Research question 4: To what extent is a radical combined with other components to form characters in textbooks?

To answer this question, the frequency of each radical combination was calculated. Three investigation processes were used: the first process was to classify radical components of characters; the second process was to calculate the frequency of each radical component; and the final process was to determine whether or not a radical was under-representative in the textbooks.

Of the characters introduced in the textbooks compiled as one database, the total number of characters was 1,166, and the total number of radical types was 179. The frequency of radical combination distribution is shown in Figure 4.13.

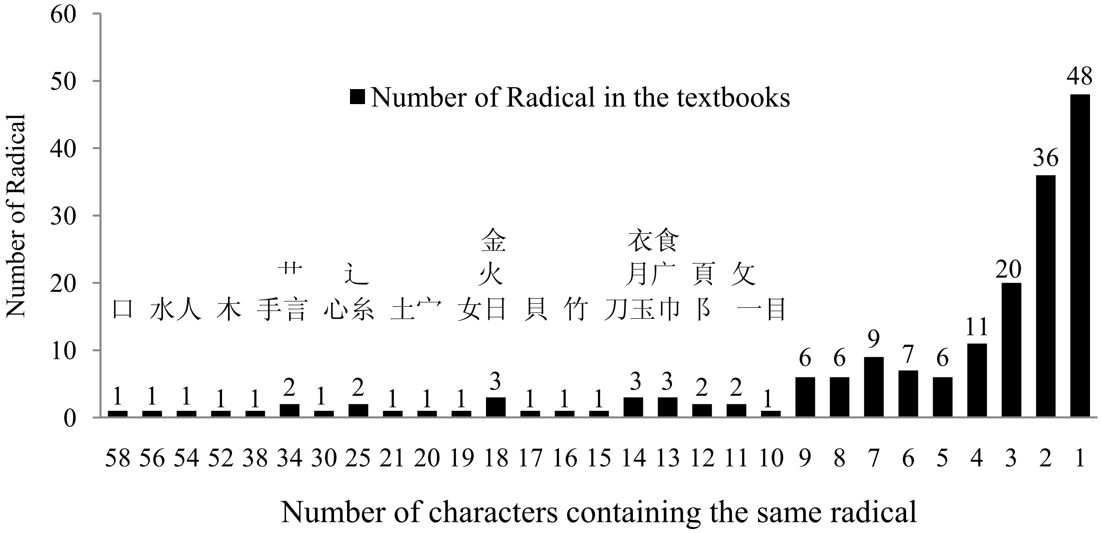


Figure 4.13 Distribution of Radical Combinations in the Database

The X-axis presents the number of characters containing the same radical, and the Y-axis presents the number of radicals. The radical having the highest radical combination number in the database is the 口 (mouth) radical, with 58 characters in the database containing the 口 radical. In contrast, 48 characters do not share their radical with other characters. Thirty out of the total 179 radicals, listed in Figure 4.13, combined with other components to form more than 10 characters in the database, and 60% (705 out of 1166) of the characters in the database included these 30 radicals. In other words, 60% of the characters in the database combined with relatively few radicals (17%), while 83% of the radicals were seen in only a limited amount of characters.

Comparing characters across the ten textbooks, as depicted in Figure 4.14, the most commonly appearing radical combination frequency is 1 character containing the same radical, which means that around 40 ~60 radicals across the ten textbook contain only one character example.

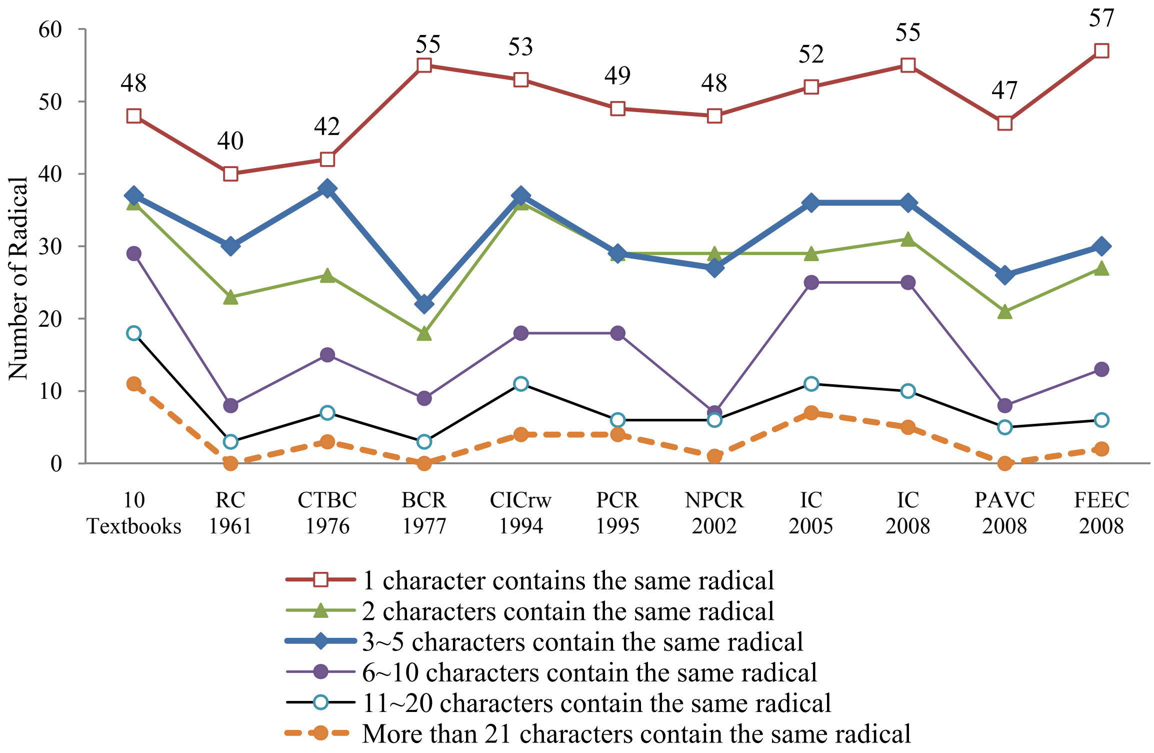


Figure 4.14 Distribution of Radical Combination Frequency Across the Ten Textbooks

Comparing the radical combination in this database with that in 中文字譜, *Chinese Characters: A Genealogy and Dictionary* (1998), some radicals that do not frequently appear in the textbooks were found. For example, 58 characters containing the 囗 or “enclosure” radical were found in the database while 174 characters containing this 囗 radical were found in 中文字譜, so the percentage of radical combination comparison was 33.3% (58/174=33.3%). If characters containing these 83% of sparsely seen radicals in the database were not frequently included in the textbooks, the average percentage of the radical combination comparison should be small. However, the average radical combination comparison percentages for the radicals forming less than 10 characters (40.6%) is higher than the average radical combination comparison percentages of the radicals forming more than 10 characters (30.5%). This is due to the fact that some radicals just do not have many radical combinations.

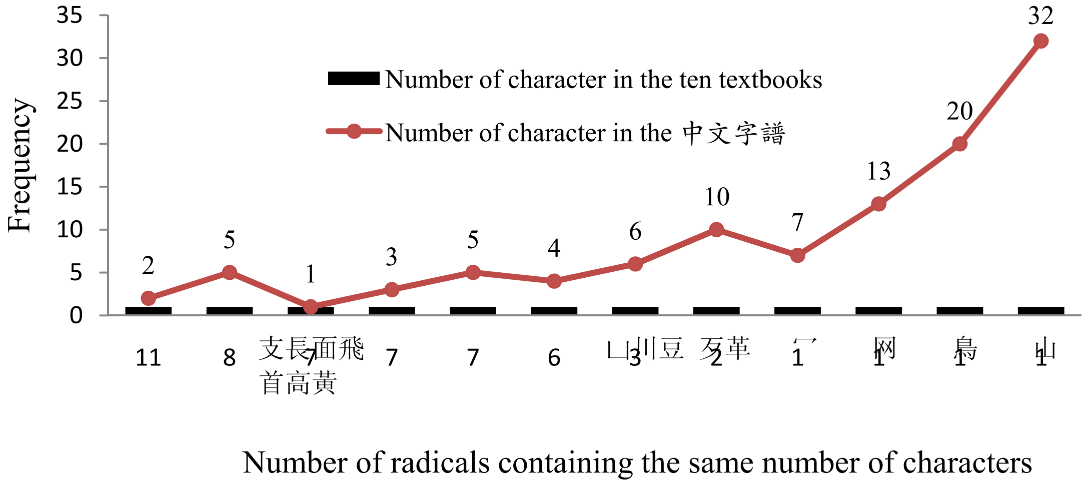


Figure 4.15 Frequency Distribution of Radical Combination

As shown in Figure 4.15 above, for example, 7 radicals 支長面飛首高黃 are free standing radical characters, and they are the only character in that radical category in 中文字譜. They were included in the database and in 中文字譜, so the percentage of the radical combination comparison for each of them is 100%. In addition, 11radicals 弋爿牙瓜皮自色香鼻齊龍 have 50% radical combination comparison because 1 out of 2 radical combination in the database. The radicals with higher radical combination comparison percentages are not the radicals that we should be aware of. It is the radical with lower radical combination comparison percentage that we should pay attention to. In the database, the percentages of radical combination comparisons less than 10% are the radicals 歹 (10%), 革 (10%), 网 (8%), 鳥 (5%), 虫 (4%), and 山 (3%).

Table 4.12 List of High Combination Radicals in 中文字譜 and Ten Textbooks

Radical & meaning	中文字譜	RC 1961	CTBC 1976	BCR 1977	CICrw 1994	PCR 1995	NPCR 2002	IC 2005	IC 2008	PAVC 2008	FEEC 2008
水氵氺 water	216[1]	5	14	7	32	22	11	41	33	6	12
手扌 hand	209	7	14	4	15	7	9	28	25	1[2]	13
口 mouth	174	17	26	14	36	31	22	45	44	20	28
人亻 man	166	16	23	16	34	27	20	40	35	20	22
木 wood/tree	160	7	15	8	28	18	11	30	31	11	14
艸艹 plant/grass	142	4	8	3	19	12	7	15	15	6	4
心忄 heart[3]	137	17	21	10	16	18	14	22	20	14	15
言讠 words/speak	108	9	20	11	18	21	16	25	19	8	13
糸糹 threads/silk	100	4	14	4	15	14	6	19	16	3	6
土 dirt/earth	83	7	8	7	16	8	5	9	10	5	7
肉月 flesh/meat	79	2	3	1	6	3	1	11	10	1	4
辵辶 move	79	9	12	9	18	11	7	18	20	7	13
女 woman	78	3	9	6	11	13	12	15	14	7	10
火灬 fire	71	3	6	3	15	5	7	11	11	4	7
金 metal/gold	69	4	7	4	8	7	3	12	9	4	5
日 sun	63	9	12	7	13	9	8	13	13	11	8
竹 bamboo	61	5	9	2	5	8	2	11	12	3	6
貝 money	56	4	6	3	7	6	4	10	6	4	7
虫 insect/reptile	55		1		2	1	1		2		1
宀 roof	50	6	7	7	12	10	6	15	16	6	10
刀刂 knife	49	5	9	5	9	8	5	11	10	6	7
目罒 eye	49	4	4	5	5	4	3	6	7	3	3
衣衤 clothing	49	3	2	2	9	8	2	10	10	2	6
阝阜 embankment	46		4		6	2	3	9	7	1	2
玉王 jade	45	3	5	1	7	6	5	7	9	4	6
…											
Total Radical	214	104	131	107	159	135	118	160	162	107	135
Total Character	4,166	308	509	281	695	533	354	811	759	314	449

Note. [1] 216 characters contain the 水(water) radical in the 中文字譜. Double underline 216 indicates that 26 is the highest combination number.

[2] A combination number less than 3 is underlined.

[3] The heart radical has another different shape which is the bottom part of the character 忝.

Table 4.12 lists the 25 highest combination radicals classified by 中文字譜, listing the combination numbers of each radical in each textbook. Eleven high combination radicals were included in the 中文字譜 and the textbooks: 水口人木心言土辶日宀刀. The 口 (mouth) radical is the highest combination radical in 9 textbooks, and the 2nd highest in BCR1977. In contrast with the other textbooks, PAVC2008 contains only one character 找 in the hand 手(扌) radical category, which indicates that learners using PAVC2008 would not be exposed to characters with the hand radical 手(扌) frequently. Similar cases were found with radicals, such as the 肉 (月) radical in BCR1977, NPCR2002, and PAVC2008. The jade 玉 (王) radical in BCR1977. In addition, the 阝 (阜) and insect 虫 radicals are under-represented in many textbooks.

To summarize, the results for research question 4 indicate that, of the characters introduced in the textbooks compiled as one database, 30 out of the total 179 radicals combined with other components to form more than 10 characters in the database, with 60% of the characters in the database including these 30 radicals. In other words, 60% of the characters in the database combined from relatively few radicals (17%), while 83% of the radicals appeared in only a few characters. Among the 25 highest combination radicals in 中文字譜, 11 radicals 水口人木心言土辶日宀刀 were included in all textbooks. However, 5 radicals 手(扌), 肉(月), 玉(王), 阝 (阜), and 虫 were under-represented in some textbooks, particularly in PAVC2008.

4.4.2 Results of Investigating the Most Commonly Appearing Character Graphic Structures in the Textbooks

Research question 5: For each character graphic structure, what is the most commonly appearing radical position?

Based on the assumption that a radical of a given character is known, three investigation processes were used: the first process was to decompose the character into components by using the component and stroke lists in Tables 3.6 and 3.7; the next process was to classify the character graphic structure by using the graphic structure and radical position reference in Figure 3.2; the last process was to classify the position of the semantic radical in that character. As a result, for each character graphic structure, the most commonly appearing radical positions were identified.

Table 4.13 Frequency Distribution of Character Graphic Structures across Ten Textbooks

Textbooks	RC 1961	CTBC 1976	BCR 1977	CICrw 1994	PCR 1995	NPCR 2002	IC 2005	IC 2008	PAVC 2008	FEEC 2008
Integral	56	74	65	99	61	43	96	85	48	68
Left-Right	123	235	104	326	257	161	406	375	134	207
Top-Bottom	91	142	77	193	153	106	218	210	92	120
Half-Enclosure	34	51	30	70	56	41	84	82	35	48
Enclosure	4	7	5	7	6	3	7	7	5	6
Sum	308	509	281	695	533	354	811	759	314	449

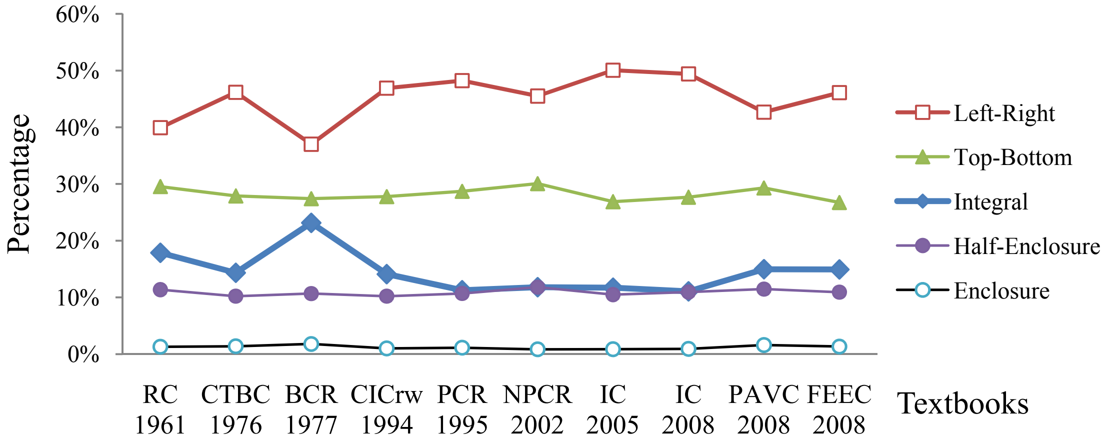


Figure 4.16 Percentage Distribution of Character Graphic Structure in the Ten Textbooks

Comparing characters across the ten textbooks, as depicted in Table 4.13 and Figure 4.16, the most commonly appearing character graphic structure is the left-right structure. The percentages comparing the top-bottom (around 30%), half-enclosure (around 10%), and enclosure graphic (around 1%) structures in all ten textbooks are similar. The percentage distribution of the five graphic structures in BCR1977, however, is different from the other nine textbooks, with a higher percentage of integral structure and a lower percentage of left-right structure appearing in BCR1977.

Among the left-right graphic structure, presented in Figure 4.17, types LR-2 and LR-1 are the most commonly appearing character graphic structures across the ten textbooks. In type LR-2, one component appeared on the left-side of the character, with the right-side containing at least two components in the top-bottom position, as in characters such as 旅孩. Type LR-3 contained the opposite structure to LR-2, with one component appearing on the right-side of the character, and the left-side containing at least two components in top-bottom position, such as in the characters 部朝. In type LR-1, one component appeared on the left-side of the character, while the right-side contained only one component, such as in the characters 好他.

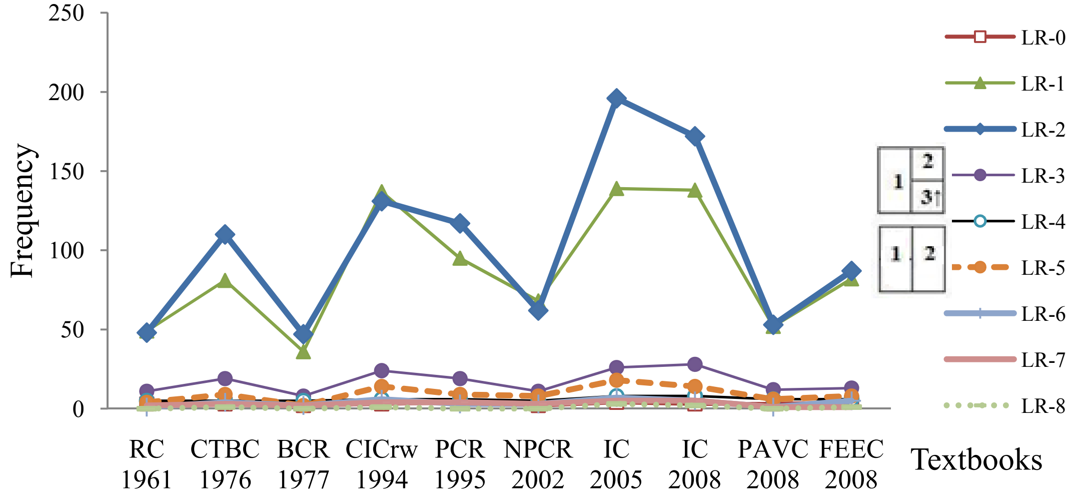


Figure 4.17 Percentage Distribution of Left-right Graphic Structure in the Ten Textbooks

Among top-bottom graphic structures, as shown in Figure 4.18, types TB-1 and TB-4 were the most commonly appearing character graphic structures across the ten textbooks. The amount of type TB-1 structures was more than the amount of type TB-4 structures across all textbooks. In type TB-1, only two components composed the character, and they appeared on the top-bottom positions, such as in the characters 思泉.

The difference between types TB-1 and TB-4 is that type TB-4 contains more than two components, such as in the characters 蒽薯.

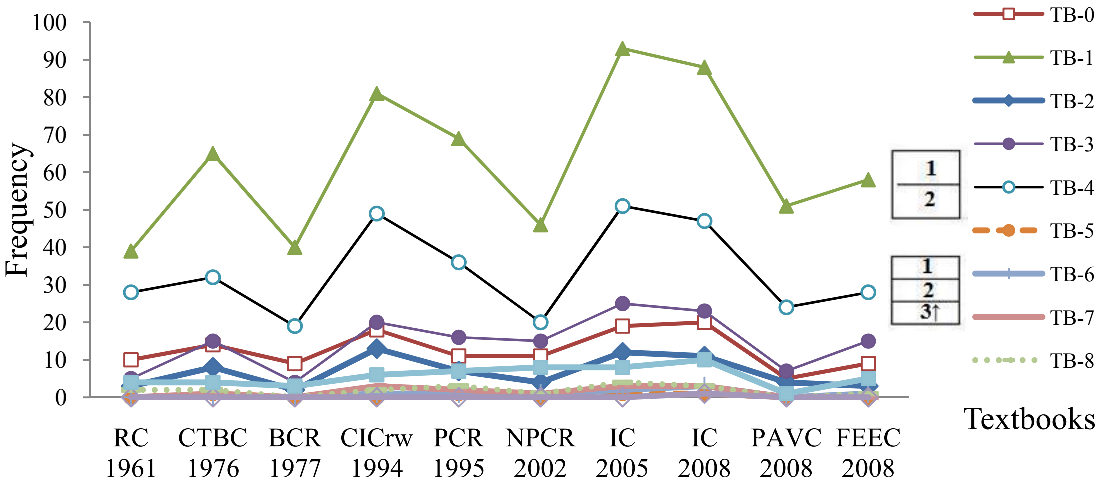


Figure 4.18 Percentage Distribution of Top-bottom Graphic Structure in the Ten Textbooks

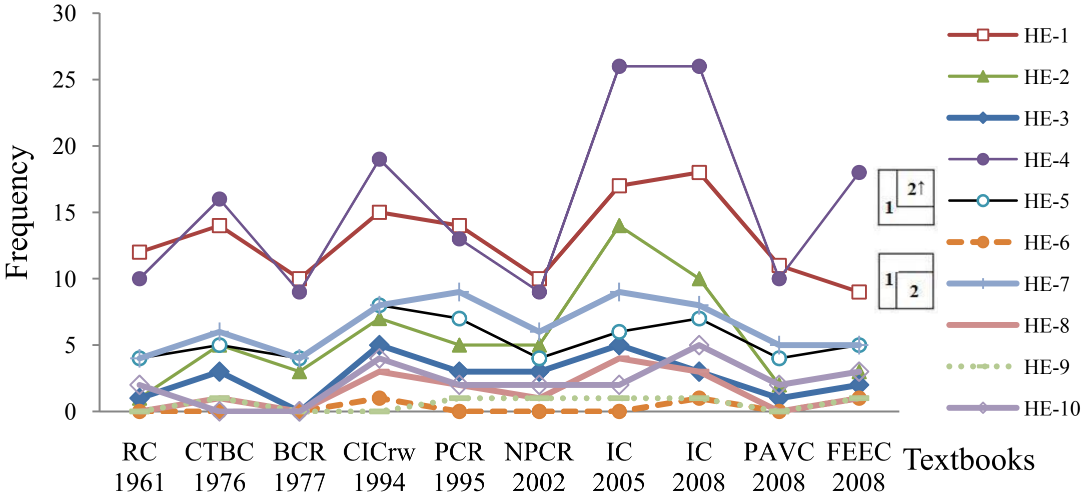


Figure 4.19 Percentage Distribution of Half-enclosure Graphic Structure in the Ten Textbooks

Among the half-enclosure graphic structures displayed in Figure 4.19, types HE-4 and HE-1 were the most commonly appearing structures across the ten textbooks. In BCR$_{1977}$, PCR$_{1995}$, NPCR$_{2002}$, and PAVC$_{2008}$, the amount of HE-4 and HE-1 was about the same. In type HE-4, one component appeared on the left-bottom of the character, such as in components 辶 and 走 in characters 送起. The difference between types HE-4 and HE-1 is that one component appeared on the left-top of the character in type HE-1, such as in the characters 屋左.

4.4.3Radical Regularity among Character Graphic Structures

Each radical has its legal position and knowing legal positions where radicals may occur can help to determine whether a character is real (Ho, Yau, & Au, 2003). Once the graphic structure is identified, the next classification is the position of the semantic radical in the character. Figure 3.2 presents the semantic radical position examples in each graphic structure. In addition, examples of characters, the radical in that example character and position number of that character are shown.

Table 4.14 Frequency Distribution of Characters in Five Character Graphic Structures

Types	Radical	0.5	1	2	3	4	5	6	7	8	9	10	Sum
Integral	72	48											120
Left-Right	5		212	268	47	10	28	10	8	4			592
Top-Bottom	19	7	120	22	44	81	2	4	5	5	1	14	324
Half-Enclosure			25	16	8	36	10	2	12	4	1	6	120
Enclosure			10										10

As previously mentioned, a total number of 1,166 characters were introduced in the textbooks and compiled as one database. The frequency of characters in five character graphic structures is shown in Table 4.14., revealing that about half of the characters (592/1166 ≈ 50.7%) employ a left-right structure, with the most common types within the left-right structure being LR-1 and LR-2. About 27.8% of the characters employ a top-down structure, with the most common type within the top-down structure being TB-1. Enclosure characters are rare (0.9%), and half-enclosure characters (10.3%) total the same amount as integral characters (10.3%). Except for the integral characters not being able to clearly separate the positions of the characters, the radical positions of the other four types were discussed in the following. For enclosure graphic structure characters, except for the free standing radical characters, all radical positions were located at the outside of the enclosure characters.

For the 592 left-right graphic structure characters, the proportion distribution is shown in Figure 4.20. The most common radical positions are LR-21 (44.6%), LR11 (28.0%), LR-33 (7.6%), and LR-12(7.4%), which indicates that the most common radical position for LR is on the left side of the character.

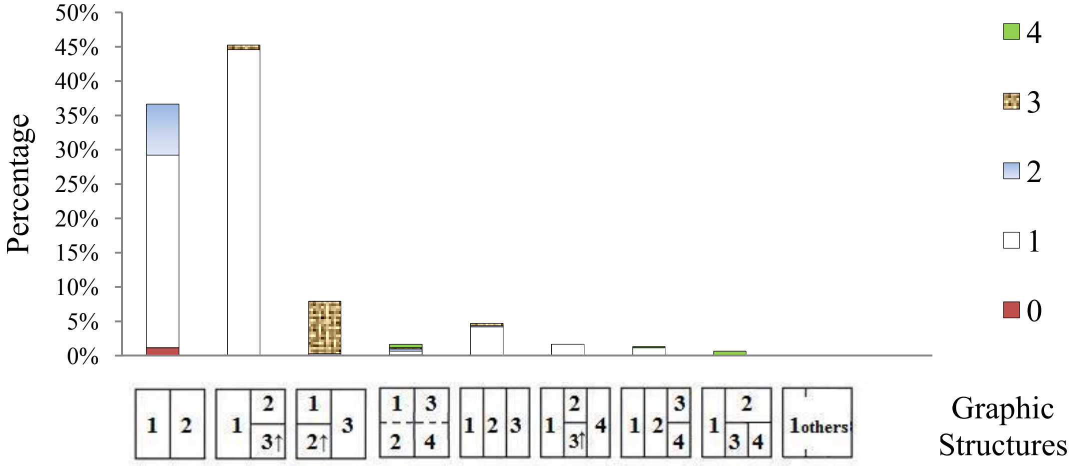


Figure 4.20 Proportion Distribution of Radical Position in the Left-right Graphic Structure

For the 324 top-bottom graphic structure characters, the proportion distribution is depicted in Figure 4.21. For top-bottom character type TB-1, the radical positions could be on the bottom (20.1%) or top (16.4%). Other common radical positions are TB-41 (14.5%), TB-33 (11.5%), and TB-43 (8.3%). Combining these types, about 30.9% of the radicals are on the top position (TB-11 andTB-41), and about 39.9% of the radicals on the bottom position (TB-12, TB-33, and TB-43). The results indicate that for top-bottom graphic structures, no particular position regularly appears more than others.

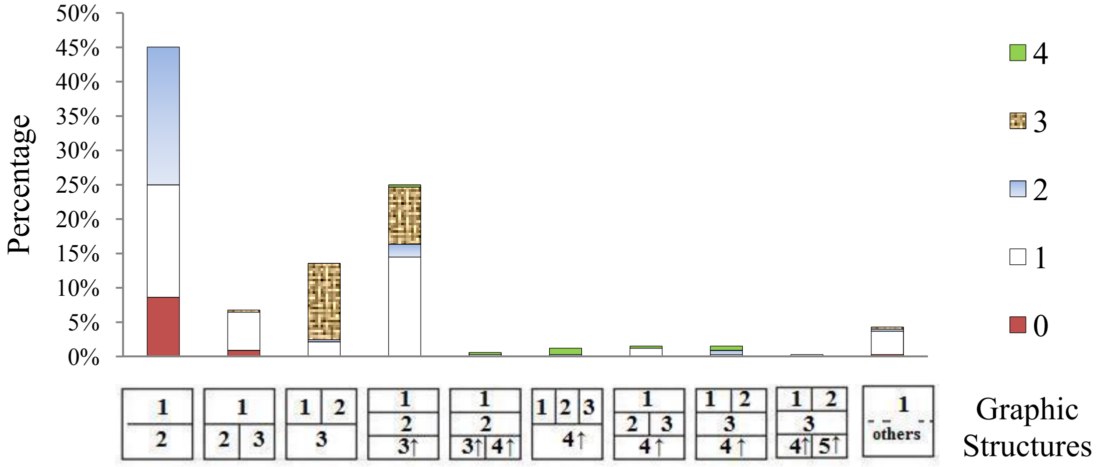


Figure 4.21 Proportion Distribution of Radical Position in Top-bottom Graphic Structure

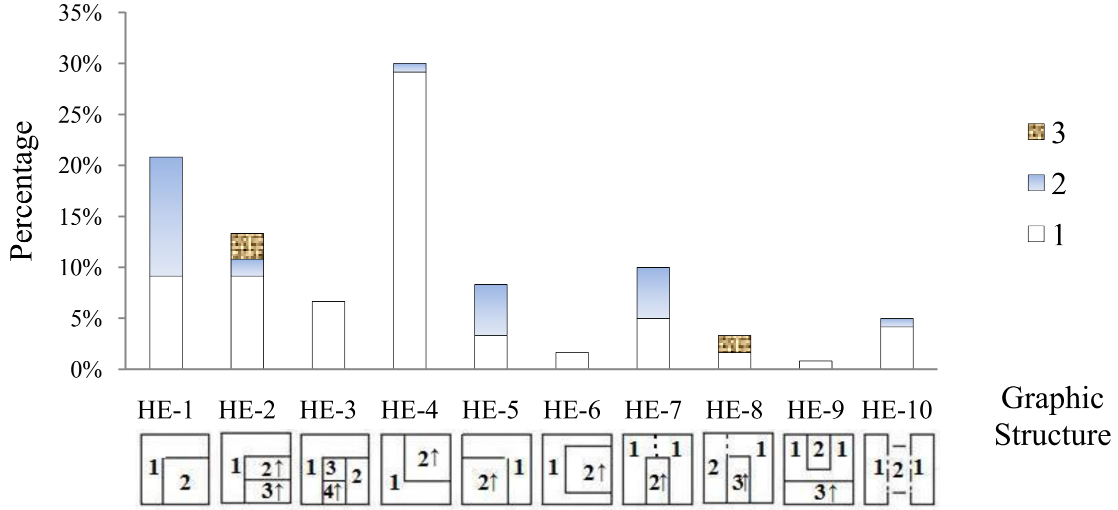


Figure 4.22 Proportion Distribution of Radical Position in the Half-enclosure Graphic Structure

For the 120 half-enclosure graphic structure characters, the proportion distribution is shown in Figure 4.22. The most common radical position is HE-41 (29.2%),which is on the left-bottom position. Other common radical positions are HE-11 (9.2%), HE-12 (11.7%), HE-21 (9.2%), and HE-31 (6.7%). Combining the HE-1, HE-2, and HE-3 structures, about 25% of the radicals are on the left-top position. However, except for types HE-1, HE-2, and HE-3, other graphic structures are quite different from each other. For half-enclosure graphic structure characters, depending on the graphic structure, the radical positions are quite different.

In summation, for enclosure graphic structure characters, except for the free standing radical characters, all the radical positions were located on the outside of the enclosure characters. Left-right graphic structure characters are the most commonly appearing and the radical positions for left-right characters are more on the left side than on the right side of the character. Other graphic structures do not have particular positions. APPENDIX C lists the character graphic structures and their radical positions for all 1,166 characters.

4.4.4 Results Summary of Character Radical Diversity and Repetition in the Textbooks

The fourth focus of the study was radical component diversity and repetition in character selection. Research questions 4 and 5 were designed to answer radical component questions including radical combination, character graphic structure distribution, and radical positional regularity among character graphic structure. In term of radical diversity and repetition, the results indicate that in the beginning level CFL textbooks, 60% of the characters in the database combined from few radicals (17%), and 83% of the radicals do not frequently being exposed to be noticed the functions of the radical. Some high combination frequency radicals were under-represented in the textbooks. The most commonly appeared character graphic structure is left-right structures, and the most common radical position is on the left side than on the right side of the left-right structure character. For top-bottom graphic structure characters, no particular position appeared regularly more than others. The most common radical position for half-enclosure structure character is HE-41 (29.2%), such as radical 辶 in character 近, which is on the left-bottom position. However, for other half-enclosure graphic structure characters, depend on the graphic structure, the radical positions are quite different. For enclosure graphic structure characters, except for the free stand radical characters, all the radical positions were located at the outside of the enclosure characters.

4.5 Results of Phonetic Component Diversity and Repetition in the Textbooks

Research question 6: To what extent is a phonetic component combined with other components to form characters in textbooks?

One approach used to pronounce a whole character is deduced via analogy with other characters sharing the same phonetic components. The fifth focus of the study was phonetic component diversity and repetition in character selection. To answer the research question 6, "to what extent is a phonetic component combined with other components to form characters in textbooks?", three investigation processes were used: the first process was to build basic characteristics of character phonetic components; the second process was to group the characters with the same phonetic components; and the last process was to compare the pronunciations of characters that shared the same phonetic components.

The characters here refer to all characters not semantic-phonetic compound characters because not every one can identify semantic-phonetic compound characters, even as native Chinese speakers. To answer research question 6, *etymological tree root number* was used to identify the following four types of analogy:

1. Homophones: "璧 bì" and "壁 bì" have the same tree root 辟 (50) component and pronunciation.

2. Partial homophones: “清 qīng” and “晴 qíng” have the same root 青 (70) component, but with a tone difference.
3. Same rhymes: “板 bǎn” and “返 fǎn” have the same tree root 反 (21) component with the same rhyme ǎn. In other words, if the replaced initial syllables are b/p/f, d/t, j/q/x, g/k/h, ch/zh/sh/r, and z/c/s groups, the characters share the same rhymes (Guder-Manitius, 1999).
4. Same component but have completely different sounds: “煮 zhǔ” and “奢 shē” have completely different pronunciations although with the same tree root 者 (76) zhě component.

Of the characters introduced in the textbooks compiled as one database, the total number of characters is 1,196 including 27 characters with different pronunciations. Because the research question is related to the pronunciation of the character, the same characters with different pronunciation are viewed as different entry. After grouping characters that have the same components, subgroups were identified within each group in terms of the four types.

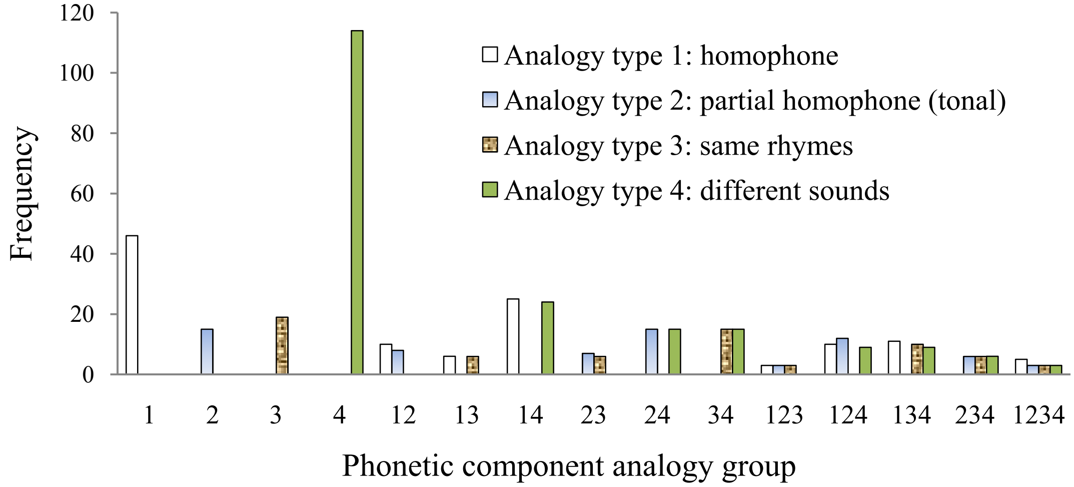


Figure 4.23 Frequency Distribution of Phonetic Component Analogy Types for Each Group

For each group, a frequency distribution of phonetic component analogy types was presented in Figure 4.23, and 298 groups across 15 types were identified. The largest group is 114 groups in type 4, where characters have the same component but have different pronunciations. Only three groups has all four phonetic component analogy types, denoted type 1234 in Figure 4.22, including (古) 胡湖/古姑故/苦/做個, (巴)巴吧/把爸/爬/肥, and (青) 青清/晴情/精睛/請/猜. Except for type 4, the combined group number of the single phonetic analogy types (1, 2, 3) is 80, and the amount is fewer than the groups of combination phonetic component analogy types (104), such as type 12, 13, 123 etc..

Table 4.15 Homophone Analogy Characters and Shared Components

Type	Characters and shared components *(Shared component, Zhuyin) Character*
Non-reliable phonetic component, but characters included this component pronounced the same	(卂ㄒㄧㄣˋ)訊迅,(斿ㄧㄡˊ)遊游,(气ㄑㄧˋ)汽氣,(矣ㄧˇ)唉欸,(亢ㄎㄤˋ)航杭,(爭ㄓㄥˉ)淨靜,(玨ㄐㄩㄝˊ)班斑,(監ㄐㄧㄢˉ)藍籃,(則ㄗㄜˊ)廁側
Reliable phonetic component and characters included this component pronounced the same	(岡ㄍㄤˉ)剛鋼,(介ㄐㄧㄝˋ)介界,(番ㄈㄢˉ)蕃翻藩,(象ㄒㄧㄤˋ)象像橡,(七ㄑㄧˉ)七柒,(州ㄓㄡˉ)州洲,(千ㄑㄧㄢˉ)千仟,(久ㄐㄧㄡˇ)久玖,(考ㄎㄠˇ)考烤,(畫ㄏㄨㄚˋ)畫劃,(票ㄆㄧㄠˋ)票漂,(冒ㄇㄠˋ)冒帽, (志ㄓˋ)志誌,(直ㄓˊ)直植,(式ㄕˋ)式試,(代ㄉㄞˋ)代袋,(母ㄇㄨˇ) 母姆,(夷ㄧˊ)夷姨,(史ㄕˇ)史使,(坐ㄗㄨㄛˋ)坐座,(支ㄓˉ)支枝, (賓ㄅㄧㄣˉ)賓檳,(末ㄇㄛˋ)末沫,(保ㄅㄠˇ)保堡,(勿ㄨˋ)勿物,(受ㄕㄡˋ)受授,(貴ㄍㄨㄟˋ)貴櫃,(由ㄧㄡˊ)由油,(永ㄩㄥˇ)永泳,(養ㄧㄤˇ)養癢,(表ㄅㄧㄠˇ)表錶,(敢ㄍㄢˇ)敢橄,(面ㄇㄧㄢˋ)面麵,(員ㄩㄢˊ)員圓,(具ㄐㄩˋ)具俱,(家ㄐㄧㄚˉ)家傢,(師ㄕˉ)師獅

For groups which have only homophone analogy type 1 characters, 46 components were identified including three non-character components (卂, 气, and 斿)

and 43 character components. In some analogy groups, the pronunciations of the shared components are not the same as the pronunciation of the characters, but characters included the shared components pronounced the same, as shown in Table 4.15, such as (亢) 航杭, and (爭)淨靜. On the other hand, for reliable phonetic components, the pronunciations of the shared components are the same as the pronunciation of the characters, such as (岡) 剛鋼, and (番)蕃翻藩. The reliable phonetic component type should be seen as pedagogically useful components.

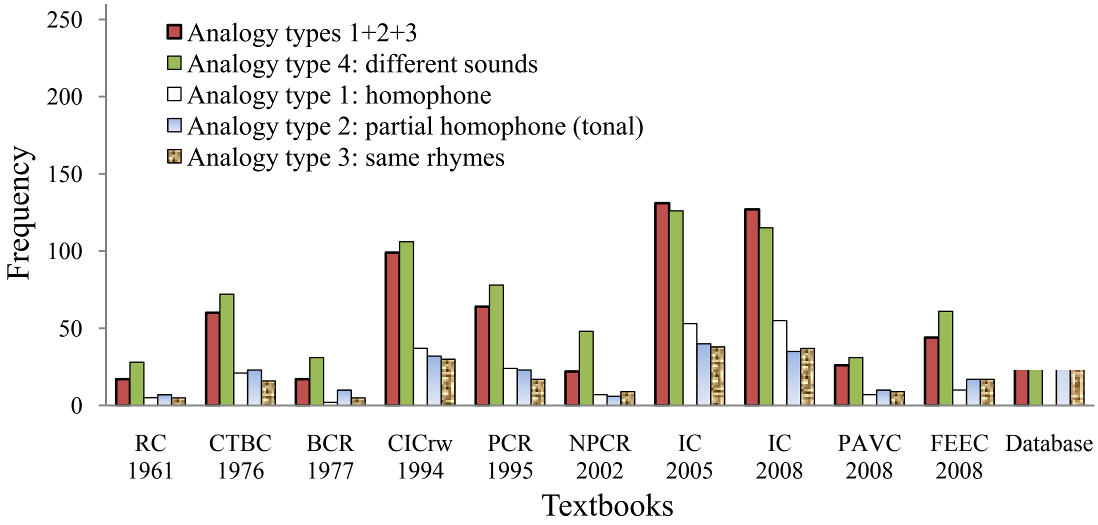


Figure 4.24 Frequency Distribution of Phonetic Component Analogy Types for Each Group

Separating the whole component analogy groups into individual four-type analogy subgroups, about 116 subgroups are homophone analogy, 69 subgroups are partial homophone analogy, 68 subgroups have the same rhymes, and 195 subgroups have completely different pronunciations. The list of the phonetic component group and subgroups is shown in APPENDIX D. The results indicated that it is possible to use a phonetic component analogy approach for character pronunciation because the number of relative reliable analogy groups (116+69+68=253) are more than the unrelated groups (195).

Comparing phonetic component analogy across the ten textbooks, the different sound analogy trend is the same as the characters combined as one database, shown in Figure 4.24. The amounts of reliable analogy groups (type 1, 2, 3) are about the same within the majority of textbooks. However, the amounts of homophone type are fewer than the other groups in BCR_{1977}, RC_{1961}, PAVC_{2008}, and FEEC_{2008}.

In summary, for characters contain the same components, 298 groups across 15 types were identified. The results indicated that it is possible to use phonetic component analogy approach for character pronunciation because the number of relative reliable analogy groups (113+67+70=250) are more than unrelated groups (195).

4.6 Results of Ideal Semantic-Phonetic Compound Characters in the Textbooks

Research question 7: For semantic-phonetic compound characters, what percentages of the characters contain semantically transparent radicals and reliable phonetic components?

The last focus of the study was to investigate the presence of ideal semantic-phonetic compound characters across the ten textbooks. Semantic-phonetic compound characters 形聲 (*xíngshēng*) combine the meaning of one character 形旁 (*xíngpáng*, semantic element) with the sound of another 聲旁 (*shēngpáng*, phonetic element). A key concept of developing Chinese orthographic awareness is that learners can be exposed to ideal semantically transparent and/or phonetically reliable characters. To answer research question 7, “for semantic-phonetic compound characters, what percentages of the characters contain semantically transparent radical and reliable phonetic components?”, four investigation processes were used: the first process was to identify semantic-phonetic compound characters; the second process was to classify characters with reliable phonetic elements; the third process was to classify characters with semantically transparent radicals; the last process was to classify characters with both semantically transparent radicals and reliable phonetic elements. The inter-rater reliability after the initial coding of radical semantic transparency was .88. To resolve disagreement, the raters discussed until 100% interrater agreement.

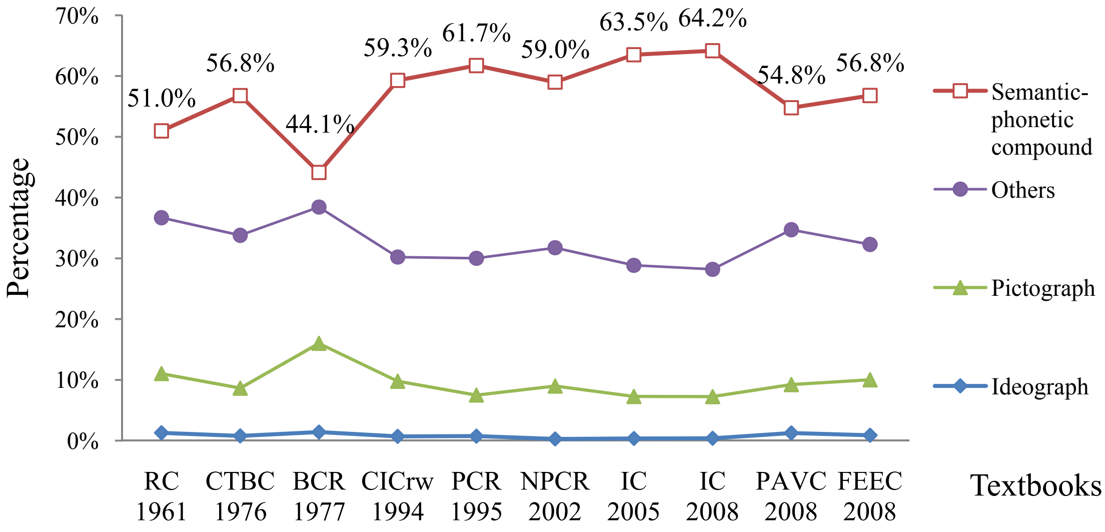


Figure 4.25 Proportion Trend of Characters in The Six Book

Of the characters introduced in the textbooks compiled as one database, the total number of characters is 1,166 regardless of the same characters with different pronunciation. About 65.8% characters are semantic-phonetic compound characters, 7% are pictograph characters, 0.4% are ideographs, and 26.8% are others which the Six Books did not classify in the etymological-based dictionary. Comparing characters across the ten textbooks, as shown in Figure 4.25, most of them were semantic-phonetic compound characters containing from 44.1% to 64.2% in each textbook. BCR$_{1977}$ and RC$_{1961}$ contained lower percentages of semantic phonetic compound characters. The percentages of semantic-phonetic compound characters in the beginning level textbooks were lower than the research has reported. Taylor and Taylor (1995), for instance, estimated that between 80~90% of the characters are semantic-phonetic compound characters in modern Chinese.

4.6.1 Reliable Phonetic Element Characters in the Textbooks

One approach used to pronounce a whole character is directly derived from the pronunciation of its phonetic component. To investigate phonetic component reliability, the pronunciation of a character was compared with the pronunciation of its phonetic component. *Phonetic code* and *Phonetic component reliability code* were used to categorize the results of the comparison. In this study, Zhuyin was used as the comparison unit.

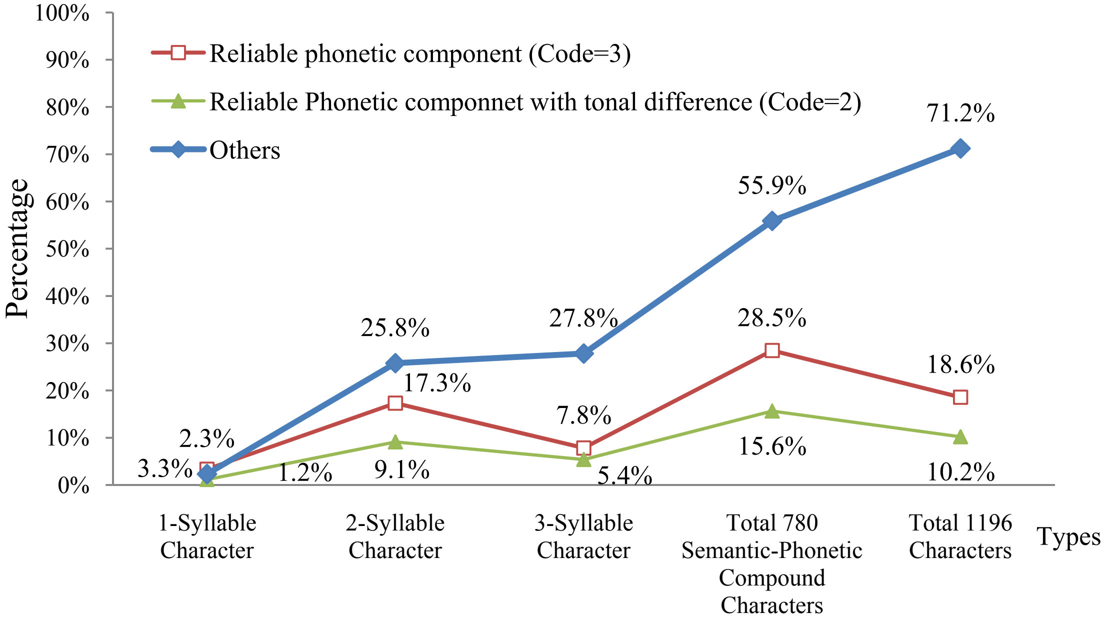


Figure 4.26 Phonetic Component Reliability Comparison in the Database

Of the characters introduced in the textbooks compiled as one database, the total number of characters was 1,196 including 27 characters with different pronunciations. Among the 780 semantic-phonetic compound characters, as presented in Figure 4.26, 28.5% semantic-phonetic compound characters have reliable phonetic components, including 3.3 % in 1-syllable, 17.3% in 2-syllable, and 7.8% in 3-syllable. In addition, 15.6 % semantic-phonetic compound characters have reliable phonetic components with tonal difference. On the other hand, among all 1,196 characters in the database, a low percentage of 18.6% characters have reliable phonetic components.

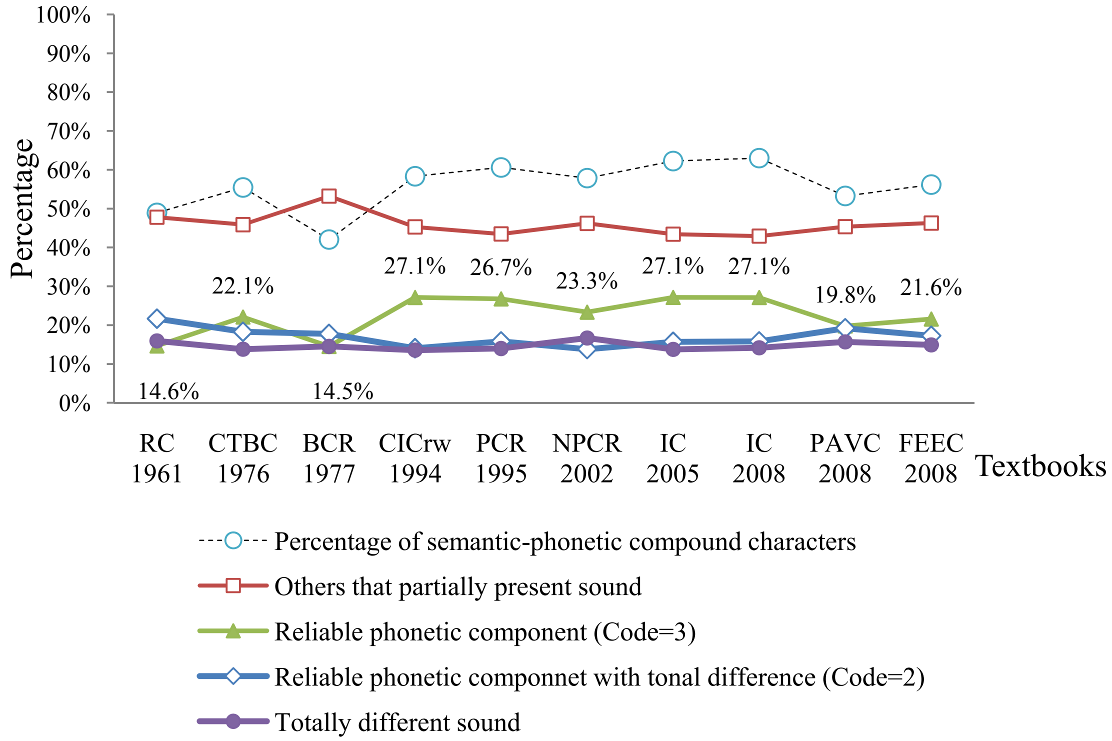


Figure 4.27 Phonetic Component Reliability within Semantic-phonetic Compound Characters

Comparing characters across the ten textbooks, the low reliable phonetic components trend is the same as the characters combined as one database, shown in Figure 4.27. Within the semantic-phonetic compound characters, the percentage of characters containing reliable phonetic components range from 14.5% to 27.1% across textbooks;. Several textbooks contained similar amounts of Code 3 and Code 2 semantic-phonetic compound characters, such as CTBC$_{1976}$, BCR$_{1977}$, PAVC$_{2008}$, and FEEC$_{2008}$. If textbooks contained few semantic-phonetic compound characters, do they contain more reliable phonetic components to emphasize the function of phonetic components? The results indicate that textbooks which contained lower semantic-phonetic compound characters still contained lower reliable phonetic components. BCR$_{1977}$ and RC$_{1961}$ contained both fewer semantic-phonetic compound characters and reliable phonetic components. Learners would have fewer opportunities to notice the function of semantic-phonetic compound characters.

4.6.2 Semantic Transparent Radical Characters in the Textbooks

As stated in the introduction, Chinese characters are grouped together according to their common components known as “radicals” 部首 (*bùshǒu*), and each character contains a radical. A radical does not guarantee giving a clue to the meaning of the character. Only semantic-phonetic compound characters should have semantically transparent radicals that can give a hint as to the meaning of the character, but this is not guaranteed because characters have evolved over thousand of years. Based on the assumption that a radical of a given character is known, the radical semantic transparency of every character was classified. Five codes were used, as shown in section 3.9.3 (p. 105). Two native Chinese raters judged the semantic usefulness of the radical.

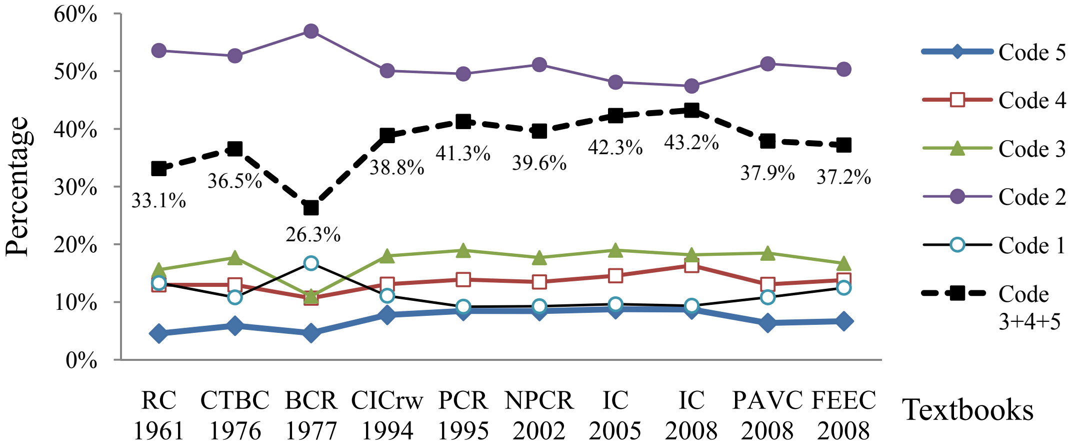


Figure 4.28 Proportion Distribution of Radical Semantic Transparency across Ten Textbooks

Of the characters introduced in the textbooks compiled as one database, about 43.1% of the characters in the database contain useful semantic radicals to infer the meaning of the character. Figure 4.28 illustrates the trends of radical semantic transparency across textbooks, with all ten textbooks containing more semantically opaque characters (i.e., code 2) than semantically transparent (i.e., code 5+4+3) characters. Across all textbooks, less than 44% of the characters contain useful radicals to infer the meanings of the characters (i.e., code 3+4+5). In addition, across all textbooks, the amount of semantically transparent radicals which define the category of the character (i.e., code 5) is less than the amount of semantically transparent radicals which have a direct (i.e., code 4) and indirect (i.e., code 3) relationship with the characters. BCR$_{1977}$ contains the smallest proportion of semantically transparent characters (26.3%) and the highest proportion of free standing radical characters (16.7%).

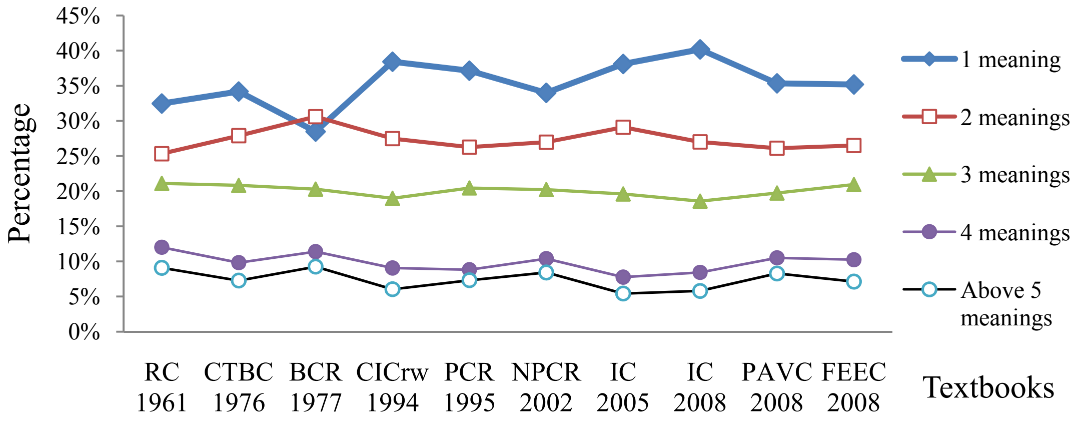


Figure 4.29 Proportion Distribution of Number of Radical Meanings across Ten Textbooks

The means and standard deviations of the radical semantic transparency across ten textbook are similar and range from 2.30 (1.02) to 2.68 (1.12). In addition, the means and standard deviations of the number of radical meanings are similar as well, and range from 2.19 (1.39) to 2.54 (1.59). Figure 4.29 illustrates the proportion distribution of the number of radical meanings. To determine whether the numbers of character meanings related to semantic radical transparency, a correlation analysis was conducted. The correlation coefficient was -0.18 which indicates no significant relationship between the numbers of character meaning and their semantic radical transparency. The results indicate that less than 44% of the characters featured in the ten textbooks investigated in this study contain useful radicals with which to infer the meanings of these characters.

4.6.3 Ideal Semantic-Phonetic Compound Characters in the Textbooks

To answer research question 7, “for semantic-phonetic compound characters, what percentages of the characters contain semantic transparent radical and reliable phonetic components?”, four investigation processes were used. After the first process, semantic-phonetic compound characters were identified. Further, characters with reliable phonetic elements were indentified. Then, characters with semantic transparent radical were identified as well. The last process here had classified characters with both semantic transparent radical and reliable phonetic elements, shown in Figure 4.30.

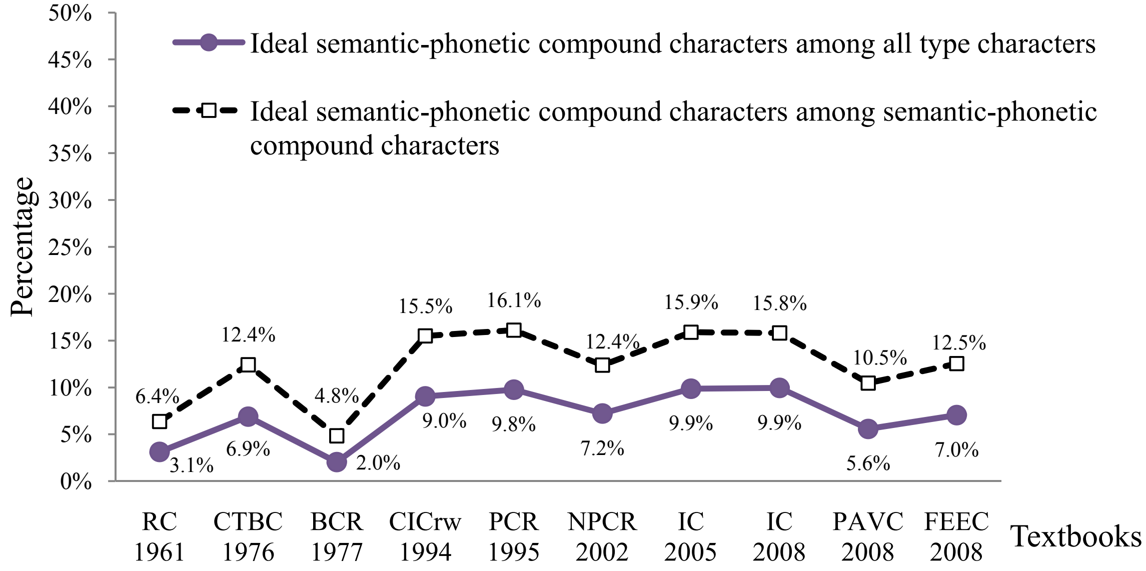


Figure 4.30 Proportion Distribution of Ideal Semantic-phonetic Compound Characters

Across all textbooks, less than 17% of the semantic-phonetic compound characters are ideal semantic-phonetic compound characters with useful radicals to infer the meanings of the characters and with reliable phonetic elements to infer the pronunciations of the characters. In other words, less than 10% of all type characters are ideal semantic-phonetic compound characters. Couple textbooks contained less than 5% of the ideal semantic-phonetic compound characters (RC_{1961} and BCR_{1977}). Among the 134 ideal semantic-phonetic compound characters, 35 characters have category radials, and they are 伯傅 (radical 亻, human), 嘴 (radical 口, mouth), 奶姆姊姨 (radical 女, female),指 (radical 扌, hand), 杉植楊橄橡檳欖 (radical 木, wood/tree), 氣 (radical 气, air/breath/steam/ apour), 洋湖 (radical 氵, water), 獅 (radical 犭, beast), 珈珠瑜 (radical

斜玉, jade), 盒 (radical 皿, vessel/dish), 腦 (radical 肉月, flesh/meat), 茅草葉葱蔥蘭 (radical 艹, plant/grass), 衫褲 (radical 衤, clothing/clothes), 帽 (radical 巾, cloth), 鋼 (radical 金, metal/gold), 麵 (radical 麥, wheat).

Table 4.16 List of Ideal Semantic-phonetic Compound Characters in the Textbooks

Times	BCR 1977	RC 1961	PAVC 2008	NPCR 2002	FEEC 2008	CTBC 1976	PCR 1995	CICrw 1994	IC 2008	IC 2005
Ideal	6	10	18	26	32	36	53	64	77	82
	氣湖們城得懂	氣站們吧城底懂界關得	伯氣廳枝站視錶們停剛吧啊底急懂汽洲灣	傅奶楊氣麵廳掛烤糕萄葡蕉蘋雲們剛吧啊喂得復泳洲漂鍛關	姊氣洋腦衫包寓廳掛換搬碗站筷視們剛吧味啊啦喂嗯址懂漂灣燈糖腐裹關	氣湖獅鋼麵園廳掛晨枝烤碗站筷糕露們剛劃功吧啊喂城婚得懂汽油漂糖聽腐越遇關	奶帽檳氣茅蘭衫褲麵包園廠廳枝站筷箱網萄葡蕉蘋視觀錶雲倆們停健冰吧啊喂址城婚座復懂汽沫泳深清漂碼糖裁誌譯鍛關	伯奶帽杉橡氣湖盒蔥衫褲麵包園圓廳換炸烤站箱糕蕃薑薯蘋袋視購雲飩們停冰剛功吧啊啦嗯址城堡婚律恤汽油泳深清溫澳灣燈營牆界碼糖腐遊關騎	傅嘴奶姆姨帽植氣洋珈珠瑜腦草葱衫褲鋼麵包噴園圓寓廳拌換搬枝橋烤癢碗站箱糕網艙萄葡蕉蘋蘿視觀倆們健冰剛劃功吧味啊喂城律復恤懂汽泳洲淹清漂潮燈碼糖聽腐越遊關騎	伯傅嘴奶姆姨帽指杉橄欖氣腦葉衫褲鋼噴園圓寓庭廳拌掛換搬烤癢碗站箱糕網萄葡視錶雲露颱倆們停健冰剛劃功吧味啊啦喂城婚座律得復急恤懂汽泳淹清溫漂潮灣燈營牆界碼糖聽腐越遊關

While 2 ideal semantic-phonetic compound characters 氣們 appeared in all ten textbooks, 41 ideal semantic-phonetic compound characters appeared in particular textbook only. Ideal semantic-phonetic compound characters appeared in textbooks are shown in Table 4.17. The results showed that textbooks included more characters contained more ideal semantic-phonetic compound characters.

In summation, the last focus of the study is ideal semantic-phonetic compound characters in character selection. Of the characters introduced in the textbooks compiled as one database, about 65.8% of the 1,166 characters are semantic-phonetic compound characters. Comparing characters across ten textbook, they contained about half of the characters are semantic-phonetic compound characters, which is lower than research reported. In addition, they contained 14.5% to 27.1% reliable phonetic components. Finally, less than 17% of the semantic-phonetic compound characters are ideal semantic-phonetic compound characters. The results indicate that beginning level CFL textbooks contained low percentage of ideal semantic-phonetic compound characters with useful radicals to infer the meanings of the characters and with reliable phonetic elements to infer the pronunciations of the characters.

CHAPTER 5

DISCUSSION

In this final chapter, an overview of the study is presented. Next, the implications of the research findings for Chinese orthographic awareness development are discussed, followed by the limitations of the study. Finally, I will recommend some areas for future research in Chinese orthographic awareness development.

5.1 Overview of the Study

To develop proficiency in reading Chinese, researchers have put forth the theory that native Chinese and CFL readers develop orthographic awareness to infer meaning and pronunciation of Chinese characters through repeated exposure to print and explicit orthographic instruction (Ho & Bryant, 1997; Ho, Yau, & Au, 2003; Jackson, Everson, & Ke, 2003; Shen, 2005; Shen & Ke, 2007). Orthographic awareness is the ability to identify, analyze, and infer the structure of words and their internal components.

A component analysis of Chinese characters provides us with structural knowledge about Chinese characters, and helps material designers, test developers, and teachers to select and evaluate the characters to be learned, taught, and tested. Most of the component analysis studies have examined Chinese characters from dictionaries or corpus databases (Chen, 1997; Fu, 1989; Guder-Manitius, 1999) and rarely have investigated characters from beginning level textbooks (Everson & Fan, 2008). In the CFL context, since reliable target language input is limited largely to textbook materials and teacher instruction, it is important to more rigorously examine the inventory of Chinese characters that is typically presented in CFL textbooks. From such investigations, we will be able to build better models of how CFL orthographic awareness develops.

The purpose of this study was to systematically describe and classify Chinese characters from ten CFL textbooks, published from 1961 to 2008, for college and adult beginning learners. The main focus was to compare the different textbooks in the following areas: explicit orthographic decomposition instruction, character frequency selection, radical combination frequency, radical semantic transparency, radical positional regularity among different character graphic structures, phonetic element reliability, and phonetic component combination frequency. To accomplish the analysis required for this study, a special character database was created by using Microsoft Access and Excel. Dictionaries were used to classify character characteristics, and documented frequency lists were used to classify the character usage frequency.

5.2 Summary of Primary Findings

The findings of the present study revealed that:

1. Most textbooks rarely include explicit orthographic decomposition instruction in the vocabulary lists and/or in each lesson;
2. In terms of character diversity and repetition, seven out of the ten textbooks contained more two-character words than others, with the proportions ranging from 45.7% to 53.6%. When character combinations for each character were tallied, the results showed that over 40% of the characters in the textbooks published after 1990 did not combine with other characters to form words.
3. In terms of character inclusion and frequency, 108 characters appeared in all ten textbooks, while 297 characters appeared only in a particular textbook. Whether frequency lists from the 1980s to 2000s were used to classify the character frequency rankings, the ten textbooks still contained a sizeable amount of high frequency characters;
4. In terms of the number of semantic radicals that were used to classify characters in the textbooks, the results indicated that 60% of the characters in the database were classified by relatively few radicals (17%). Moreover, across all textbooks, less than 44% of the characters contained useful radicals to infer the meanings of the characters. In terms of where radicals appeared in

the characters, the left-right graphic structure was the most commonly appearing graphic structure, with the radical positions for left-right characters appearing more on the left side than on the right side of the character;

5. About half of the characters featured in the 10 textbooks were semantic-phonetic compound characters, a figure lower than what is reported in the research literature when taking into account all characters in the modern Chinese lexicon. Within the semantic-phonetic compound characters, the percentage of characters containing reliable phonetic components ranges from 14.5% to 27.1% across all textbooks. Although relatively reliable analogy groups (homophone, partial homophone, and same rhymes) appeared frequently, less than 17% of the semantic-phonetic compound characters in the textbooks were ideal semantic-phonetic compound characters.

5.3 Discussion

5.3.1 Integrating Orthographic Decomposition Materials into the Textbooks

The results of this study revealed that most textbooks rarely include explicit orthographic decomposition instruction in the vocabulary lists and/or in each lesson, an unfortunate finding given the research results which support the benefits of explicit orthographic decomposition instruction for CFL learners (Jackson et al., 2003; Shen, 2004). This result indicates a general neglect of supplying students through their textbooks with explicit information dealing with the principles of orthographic decomposition. The result also imply that to develop Chinese orthographic awareness, learners may have to rely primarily on classroom instruction for information about orthography principles and their applications, a situation that is problematic since in first year instruction, class time should be used for activities that stress target language usage by and for the learners. If time is taken away from these activities so as to provide lengthy instruction in English as to how characters are constructed, instruction that can be better provided through textbook content, language development can suffer.

Should learners be introduced to Chinese orthography in the textbooks at the beginning of their entry level courses? The Chinese writing system is certainly one of the primary factors that makes learning Chinese such a time-intensive process. The experience of learning to read in languages employing alphabets is different from learning to read in non-alphabetic orthographies such as Chinese. In the current study, only three out of ten beginning level textbooks (BCR$_{1977}$, IC$_{2005}$, and IC$_{2008}$) introduced character formation, strokes, radicals, and phonetics in the introduction of the textbooks. It is to be applauded that these textbooks set the stage for Western students to gain insights into the non-alphabetic Chinese writing system at the beginning of the courses. The introduction of the Chinese writing system is not only " nice to know" knowledge about the historical and cultural aspects of the Chinese writing system, but also provides foundations for learners to know that there is actually a system to the Chinese writing system, one that will facilitate their learning if they are able to master its basic principles. Future research should look at how learners value the introduction of Chinese orthography principles as featured in their beginning-level textbooks.

In addition, Shen (2004) has maintained that deeper processing including self-elaboration and guided-elaboration encoding strategies in learning Chinese characters results in better retention and recall than rote memorization. The advantages of deeper processing is that it provides additional cues for recall and making information more meaningful by "means of visual imagery, by relating new material to known information, and by arranging information into a meaningful structure" (Shen, 2004, p. 169). The results of the current study indicate a general neglect for supplying students through their textbooks with explicit information dealing with the principles of orthographic decomposition, leaving beginning level CFL students little choice but to employ a primary strategy of rote memorization to learn characters. As a result, learning Chinese characters becomes a labor-intensive process and requires huge demands on their memories, time, and study capabilities, with the expected result that students will often drop the course out of frustration and concern about the demands on their time. Related to this is the important concept of stroke-order, a topic that varied considerably in terms of where and how it was introduced in the ten textbooks in this study. Among the three textbooks published before the1990s, RC1961 and CTBC1976 presented the stroke-order of characters introduced in each lesson in the APPENDIX, while BCR1977 did not present character stroke-order at all. Interesting enough, both CTBC1976 and BCR1977 were written by DeFrancis, so it was unclear why he decided to include stroke-order material in one textbook but not in the other. On the other hand, more recently published textbooks did not include the stroke-order of all the characters introduced in the vocabulary lists. The reason of this decision is probably because additional character workbooks were designed to accompany the textbooks. Some workbooks, such as IC2005, IC2008, and FEEC2008, for example, included the correct stroke order and components for the characters. More recently published language materials tend to divide the materials into different volumes, such as textbooks, workbooks, and character workbooks as authors feel that textbooks alone do not provide sufficient orthographic principles for students to learn.

Based on the findings in this study, to promote orthographic awareness development among beginning CFL learners, we need to integrate into our textbooks materials that deal with orthographic decomposition. Among the ten textbooks in this study, NPCR2002 is the only textbook that devoted a fixed section in each lesson to preset orthographic awareness development material, with the orthographic decomposition materials more relevant to the characters introduced in the text. The good news is that textbook writers seem to be beginning to see the need to integrate the character decomposition information into their textbooks. However, simply providing character decomposition information without further explanation is questionable in contributing to orthographic awareness development. Further explanations are needed to distinguish the different types of characters. For example, on page 22 in NPCR2002, instead of introducing that 女 denotes the meaning of woman and 馬 denotes the pronunciation, the character 媽 should be introduced as a semantic-phonetic compound character with a semantically transparent radical 女 and a reliable phonetic component 馬 with tonal difference. For the character 呢, 尼 is not a reliable phonetic component that denotes the same pronunciation of 呢. Without such explanation, students will not learn that there is a system, albeit imperfect, for inferring the meaning and sound of Chinese characters. The effort of simply providing character decomposition information without further explanation is questionable, as it does not contribute to orthographic awareness development.

5.3.2 Becoming Familiar with the Processes of Chinese Character Combination

DeFrancis (1977) stated that basic to developing reading skill in Chinese is a familiarity with the processes of character combination in Chinese. Such familiarity is best acquired by mastering several combinations for a limited number of characters rather than by learning one or two compounds for many characters. Ke (1998) investigated what strategies seem to predict success in CFL character learning, and found that an overwhelming majority of beginning CFL learners felt that practicing characters in the context of character combination was more effective than practicing characters individually. In addition, those participants performed better on character recognition tasks. Ke and Everson (1999) explained that practicing characters in the context of character combination provides larger and more meaningful units for character learning, and might indicate that words are more salient than individual characters for CFL learners until they develop more orthographic awareness.

A second focus of this study was to investigate character diversity and repetition in the ten textbooks. Looking at the results of the total number of different words in the textbooks, the majority words are 1-character and 2-character words with the proportions ranging from 75.3% to 96.1%. On the other hand, excluding 1-character words, CFL learners have more than 50% opportunities (51.0% ~ 72.5%) to see how characters are combined to form words in these textbooks, except for CTBC$_{1976}$ which contains the highest proportion of 1-character words (64%) compared to the other textbooks (26.8% ~ 49%).

However, when character combinations for each character were investigated, the results showed that over 40% of the characters in the textbooks published after 1990 did not combine with other characters to form words. One implication is that CFL learners have limited opportunities to see these characters in other character combinations, although the other implication is that beginning CFL learners would have a clearer meaning of the single character than the meaning of the character within a word. However, according to Ke (1998) and Ke and Everson (1999), these single-characters may not be as meaningful for word learning, though this might depend on the proficiency level attained by the learner.

The results of this study also showed that over 40% of the characters in the majority of the textbooks can be seen about 2 to 5 times in character combinations. In some textbooks (BCR$_{1977}$, RC$_{1961}$ and IC$_{2005}$), CFL learners can see some characters more than 20 times in character combinations. Both BCR$_{1977}$ and RC$_{1961}$ introduced individual character first, such as the character 你, and later the character combinations (i.e. word, such as 你的) were provided. Some textbooks introduced words first and later introduced each character which were combined to form words. In this way, learners were guided to notice the character combination of each character. Therefore, CFL beginning learners would learn the processes of Chinese character combination.

Moreover, in terms of character diversity, 108 characters appeared in all ten textbooks, but 297 characters appeared in only one particular textbook. CICrw$_{1994}$ contained more characters (99) that appeared in only one textbook , with 55 of these being low frequency characters. For CFL learners, if and when they learn these 297 unique characters, it is possible they would forget them easily since they would not see them often. On the other hand, the characters 兒一不上天學國人文生 had the highest character combination frequency among textbooks. Among these characters, the character 兒 *ér* had highest character combination frequency, perhaps due to its unique use in representing an important phonological aspect of the Beijing dialect. The character 學 *xué* had the highest combination frequency which may due to this character's inclusion in words dealing with school settings. The results from examining the character diversity and repetition imply that the contexts of character instruction and content settings may influence character selection, which further affects character combination frequency.

5.3.3 Instructional Focus on High Frequency Characters and Low Frequency Characters

Researchers have found that high frequency words are learned faster and remembered better (McCarthy, 1990; Nation, 2001; O'Dell, 1997; Sergent & Everson, 1992). It seems clear that high frequency words are likely to predominate at the early stage of learning and teaching. From a pedagogical standpoint, Guder-Manitius (1999) suggests that, at the beginning level, the selection of characters introduced should as far as possible consist of high frequency characters that also serve as components and contain as few strokes as possible.

In this study, the results confirmed that whether one uses frequency lists constructed over different periods of time or for different purposes, all ten textbooks contain many high frequency characters, confirming that frequency rank plays an important role in selecting characters and words in the textbooks. Specifically, if we use the more recently compiled frequency lists to classify the textbooks published before 2000, these textbooks still contain high frequency characters, which implies that these textbooks still can be used now if one only considers character frequency usage. The finding of the existence of high frequency characters in these textbooks would follow Ke's (1996) CFL model of orthographic awareness which states that during the component-processing stage, learners can acquire more easily those characters with high frequency of occurrence because most of the CFL textbooks in the current study provide sufficient high-frequency character examples to learners.

It should be noted, however, that the results indicate that these beginning level CFL textbooks do not cover all the HSK basic level (甲級) characters, implying that CFL learners would not be able to meet the HSK basic level requirement after studying any one of the beginning level textbooks sampled in the current study. Characters in RC$_{1961}$ and BCR$_{1977}$ can be found within the HSK four levels. As mentioned, CICrw$_{1994}$ contained many characters which particularly appeared in that textbook, and the results of the current study also revealed that CICrw$_{1994}$ contained the highest number of characters that could not be found in the HSK four levels. In addition, 108 core characters which appeared in all ten textbooks ranked within the top 300 most frequently used characters with most of them falling within the basic HSK level. These 108 characters can, therefore, be seen as pedagogically useful characters for learning.

In general, however, the results of this study, demonstrated that all ten textbooks contained many high frequency characters and some low frequency characters. Many of these low frequency characters are commonly used in spoken language or are used in special contexts. Learning to read in the CFL setting involves both the learning of the spoken language and the learning of Chinese characters. The low frequency characters may represent words in the spoken language that are not necessarily printed all that often in terms of frequency, and may therefore represent words in "proficiency" contexts (foods, travel sights, etc.) that employ low frequency characters. In classroom instruction, how should the varying degrees of frequency vocabulary be resolved? Nation (2008) maintains that high frequency words need to be learned before low frequency words. High frequency vocabulary deserves the teacher's attention and deserves direct teaching because each high frequency word occurs very often, so the effort of learning it will be repaid by opportunities to meet and use it. When learners are at the stage of working on low frequency vocabulary, the teacher should attend to strategies that are needed to deal with low frequency vocabulary, and make sure students are given opportunities to practice and review these characters, as they will not be sufficiently reoccurring in other print situations. The strategy of using word parts supports the development of orthographic awareness. Through frequently being exposed to and taught about components in high frequency characters, readers develop abilities to look for the internal structure of the characters. Therefore, when they encounter low frequency characters or novice characters, they would be able to use their orthographic knowledge to infer the meanings and pronunciations of the characters.

5.3.4 High Combination Frequency Radicals Poorly Represented in Textbooks

Chinese characters are grouped together according to their common components known as "radicals" 部首 (*bùshǒu*), and each character contains a radical. The frequency of a radical could affect the speed and accuracy of character recognition. The predominant research results indicate that native Chinese and CFL students recognize characters containing higher frequency radicals faster than characters containing lower frequency radicals. In addition, beginning CFL learners consider radical knowledge helpful to learn Chinese characters (Ke, 1996, 1997; Li & Chen, 1999; Shen, 2010).

The results of this study indicate that, comparing the radical combination in the textbooks with that in 中文字譜, some radicals that do not occur frequently in the textbooks were found, such as the radicals 歹, 革, 网, 鳥, 虫, and 山. In addition, among the 25 highest combination frequency radicals in 中文字譜, 11 radicals 水 (water), 口人木心言土辶日宀刀 were included in all textbooks, while 5 radicals 手(扌), 肉(月), 玉(王), 阝(阜), and 虫 tended to be under-represented in some of the textbooks. The characters in this study which contain the radical 虫 (insect) are 蛋 (egg) and 蝦 (shrimp), but their meanings have nothing to do with insects. Although those textbooks contain the characters 蛋 and 蝦, they are not as useful for learning these characters as they are for learning the characters 蚊 (mosquito) and 蟻 (ant) which are more easily associated with the 虫 (insect) radical category. Another example is that six out of ten textbooks contain only the character 山 in the 山 (mountain) radical category, which indicates that CFL learners would not be able to be exposed to other characters with the 山 radical, such as commonly used characters 峰 and 岩 on maps. Unfortunately, textbooks which do not contain the character 山 at all are recently published textbooks.

We should be aware of these under-represented high combination frequency radicals because beginning CFL learners would not be exposed to them in the beginning level textbooks. Shen (2010) found that beginning CFL learners reported that they did not have enough opportunities to encounter the radicals after initially learning them. The results of this study provide evidence to support the lack of opportunities for students to gain frequent exposure to radical combinations in the beginning level textbooks.

Not only do beginning level CFL teachers have to be aware of the under-representation of high combination frequency radicals, but researchers must be as well when they construct their instruments to investigate character perception and production. When researchers select high combination frequency radicals from corpus or dictionaries, those high combination frequency radicals may not be truly high frequency radicals to CFL learners because they do not encounter them frequently in textbooks.

5.3.5 Identifying Radicals through Radical Position Regularity

Research results indicate that native Chinese learners' orthographic knowledge for characters with a left-right structure is more highly developed than for characters with top-down or half-enclosure structures. They take more time to recognize characters with the radical in the right and bottom positions (Li & Chen, 1999; Li, Fu, & Lin (2000) , which has led researchers to suspect that readers experience left-right graphic structure characters and radicals in the left position more than other graphic structure characters. The results of this study confirmed that the most commonly appearing character graphic structure present in the ten textbooks is the left-right structure, followed by top-bottom, integral, half-enclosure, and enclosure structures.

Researchers have found that beginning learners considered radical knowledge helpful to learning Chinese characters (Ke, 1996, 1997; Li & Chen, 1999; Shen, 2010). The prerequisite, however, is that readers have to recognize the radicals of the characters which can be a problem for learners who do not understand Chinese character structural principles. In this study, after combining the most frequently appearing radical positions with the most commonly appearing graphic structures in the textbooks, a possible regularity was identified, shown in Figure 5.1.

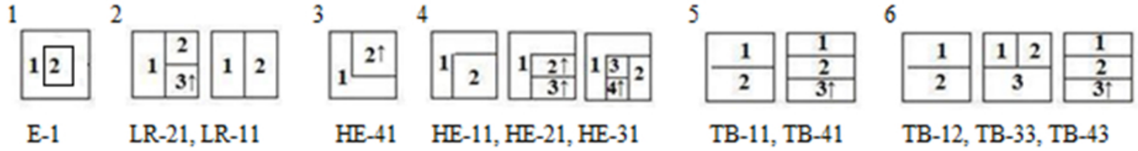


Figure 5.1 Radical Regularity among Character Graphic Structures

For the enclosure structure, radicals were mostly located outside of the character structure, such as E-1 in example 1. For left-right characters, the radicals were mostly located on the left side, such as in LR-21 and LR-11 in example 2. For the half-enclosure structure, the radical positions were quite different. For the most frequently appearing type HE-4, the radicals were mostly located on the left-bottom position, such as HE-41 in example 3. In addition, combining types HE-1, HE-2, and HE-3, the radicals were mostly located on the left-top position in example 4. However, for top-bottom characters, no particular position could be identified. The radicals could be located on the top, such as

TB-11 and TB-41 in example 5. Radicals also could be located on the bottom, such as TB-12, TB-33, and TB-43 in example 6.

5.3.6 Using a Phonetic Component Analogy Strategy

One approach used to pronounce a whole character is deduced via analogy with other characters sharing the same phonetic components. The characters here refer to all characters and not just semantic-phonetic compound characters because not everyone can identify semantic-phonetic compound characters, even native Chinese speakers. Chan and Wang (2003) found native Chinese children preferred analogy to derivation strategies. However, researchers also found the most common errors are phonetic-related errors (Ho & Bryant, 1997). For example, the target character 怕 (*pà*) was read as another character 伯 (*bó*) because they have an identical component 白 (*bái*).

The results of the current study provide evidence to explain this phonetic-analogy error. In this study, characters were grouped by the same phonetic components, and pronunciations of the grouped characters were compared. Of the characters introduced in the textbooks compiled as one database, 298 groups across 15 types were identified. About 40% of the groups were characters that have the same component but have different pronunciations. In addition, the amount of pure homophones, tonally-different-homophones, and same-rhyme analogy types is fewer than the groups of combination phonetic component analogy types, where characters in the same group could have the same pronunciation, tonal difference, same rhyme or totally different pronunciations. Three groups have all four phonetic component analogy types, denoted as type 1234 in Figure 4.22, including (古) 胡湖/古姑故/苦/做個, (巴)巴吧/把爸/爬/肥, and (青) 青清/晴情/精睛/請/猜. Based on the results of this study, it is no surprise that readers would misprounce a character with other characters sharing the same components.

If we are afraid of making phonetic-analogy errors, should we encourage learners not to use phonetic analogy strategies? Based on the data of this study, the answer is not a simple yes or no answer, because 46 homophone groups and 15 tonally-different-homophone groups were identified as well. Among these 61 groups, 49 shared components are character components, and they are reliable phonetic components, such as: 七久介代保具冒勿千受史員坐夷家州師式志支敢末母永由畫直票表象貴賓面養岡番丙九分原將星本牙那鄉采袁前. These character components should be viewed as pedagogically useful components and be introduced to beginning CFL learners to encourage them to use phonetic-analogy strategy when they encounter these components. Using an analogy strategy, we have no need to determine first whether the target character is a semantic-phonetic compound character or whether it is an ideally semantic and transparent and phonetically reliable character. The pedagogical implication for this investigation is that material developers and Chinese teachers can provide learners with analogy examples when they encounter a character which belongs to a reliable analogy group. Therefore, they can guide learners to look inside of the character not just look at the character.

5.3.7 Ideal Semantically Transparent and Phonetically Reliable Characters in the Textbooks

A key concept for developing Chinese orthographic awareness is the frequency of exposure to the ideal semantically transparent and/or phonetically reliable characters. Although each character contains a radical, a radical does not guarantee giving a clue to the meaning of the character. The findings of this study support this fact. Across the ten textbooks, less than 44% of the characters contained useful radicals to infer the meanings of the characters. The findings of this study are similar to the findings of Kang (1993) (43.79%) who used the 7,000 most common characters from 現代漢語通用字表 (*Xiàndài Hànyǔ tōngyòng zìbiǎo*) to investigate semantic and phonetic components in the 5,631 semantic-phonetic compound characters.

Taylor and Taylor (1995) have estimated that about 80~90% of the characters are semantic-phonetic compound characters in modern Chinese. However, in this study, not only were ideal semantically transparent and/or phonetically reliable characters somewhat rare in the ten beginning level CFL textbooks, but sizeable percentages of characters, approximately 35% to 56%, which the beginning level CFL learners initially encounter are not semantic-phonetic compound characters.

Within the semantic-phonetic compound characters, the percentages of characters containing reliable phonetic components range from 14.5% to 27.1% across textbooks. The findings of this study are much lower than the findings of Li and Kang (1993) and Yin (1991). Li and Kang (1993) found that 37.51% of phonetic components were pronounced the same as the semantic-phonetic compound characters, and Yin (1991, as cited in Yin & Butterworth, 1992) found that 36% of phonetic elements completely represent the characters' sound.

Further, only characters with both reliable phonetic elements and transparent semantic radicals were counted, with the percentages ranging from 2.0% to 9.9% across textbooks. In other words, less than 10% of the characters in the textbooks contained ideal semantic-phonetic compound characters with useful radicals to infer the meanings of the characters and with reliable phonetic elements to infer the pronunciations of the characters. These results are similar to Guder-Manitius's (1999) findings that ideal semantically transparent and phonetically reliable characters are rare. For example, Guder-Manitius (1999) found that only 11.6% of the 3,867 characters taken from the Attributive Dictionary 漢字屬性字典 *Hànzì shǔxìng zìdiǎn* (1989) are ideal semantic-phonetic compound characters, and also found that only 8% of the 1,600 basic level HSK characters (漢語水平詞彙與漢字等級大綱, *Hànyǔ shuǐpíng cíhuì yǔ hànzì děngjí dàgāng*) are ideal semantically transparent and phonetically reliable characters (p. 314).

According to Ke (1996), CFL learners have a higher level of awareness for semantic components than phonetic ones. In addition, Shen and Ke (2007) found that the development of skills in decomposing compound characters into radical units and reproducing compound characters by radical units emerged at the very beginning stage of learning. However, they also found that students' radical knowledge had a significant increase at each learning level, but that the learners' ability to apply radical knowledge to the learning of new characters seemed to plateau. They further explained that the plateau period could be longer or shortened for each individual learner depending on the amount, quality, and frequency of practice in knowledge application, as well as the strength and frequency of linguistic cues. Both radicals and phonetic elements are linguistic cues to the target Chinese characters. The results of the current study reveal that the semantically transparent radical characters having reliable phonetic elements were rare across the ten textbooks, which implies that the quality and quantity of the linguistic cues are not particularly helpful in developing knowledge application skills through these beginning level textbooks.

Shen and Ke (2007) recommend that classroom instruction should not focus merely on introducing and reviewing radical knowledge, but creating opportunities for students to practice and use their knowledge purposefully in their everyday learning. Since ideal semantically transparent and phonetically reliable characters are rare in beginning CFL textbooks, students have limited opportunities to notice the functions of semantic and phonetic components. When learners are trying to build their own understanding of the Chinese character system, what CFL textbooks actually provided them with are counter-examples or irregular-examples, which is further confusing them when they are developing their own orthographic awareness. No wonder that CFL learners believe that character learning and writing are the most difficult tasks in learning Chinese at the college level (Everson, 1998; Ke, Wen, & Kotenbeutel, 2001). It may be due to the nature of the Chinese character but also due to the unsystematic and limited character examples introduced in the CFL beginning level textbooks.

5.4 Limitations of the Study

The purpose of this character frequency and component study was to systematically describe and classify Chinese characters in ten beginning level CFL textbooks, and some limitations of the study exist. The first limitation of the study is the selection of CFL textbooks. In this study, I only examined ten widely used Chinese textbooks in the United States. I might also neglected some textbooks which provide pedagogically sound materials to learners. In addition, other CFL materials from other English-speaking countries should be examined in the future. Further, CFL materials which are written in other languages should be put into consideration in the future as well.

The second limitation of this study is the exclusion of the character workbooks. This study only investigated the materials in the textbooks but not in the character workbooks. Textbooks published before the 1990s did not have separated character workbooks, so I decided to examine textbooks only. Further, whether character workbooks are required in the language course is another consideration for textbook selection. Future research should investigate what students and teachers actually do with the character workbooks in the basic curriculum.

Another limitation of the study is from the method restrictions of documented Chinese character frequency lists. The same characters with different pronunciations were combined to be one character entry, so the same character with different pronunciations 多音字 (*duōyīnzì*), such as characters 樂, 行, 爲, etc., had the same rank code because all lists treated them as the same. In addition, frequency lists contain both traditional and simplified characters, and they combined the two forms into one rank entry. Therefore, multiple traditional characters are combined into one simplified character and their frequency rank are the same. As well, some characters were not included in these frequency lists. Lastly, actual character frequency in the text was not calculated in this study. Instead, character combination frequency in the vocabulary lists was examined to investigate the process of character combination familiarity.

It should be remembered that this study focused its analysis on the characters in the textbooks, and not on how the textbooks have been used by students and teachers in everyday use. Clearly, it would be impossible to do this, but the fact remains that students, especially beginning ones, are often very reliant on their textbooks to see their way through their course. As well, foreign language teachers have often been criticized for the "textbook-as-curriculum" syndrome whereby a well thought out and planful process of curriculum design is thought to be unnecessary if a popular textbook series is available for immediate use.

Lastly, the amounts of characters in the vocabulary lists in the ten CFL textbooks were different; that is, they were less than 300, less than 500, and more than 700 characters. Whether these textbooks are used for a semester or a academic year was not specified in the textbooks. The focus of this study was not to compare component content among them, but the component content within each textbook. Future research, therefore, can investigate the significance of classifying each textbook into the number of hours it generally takes to cover the content.

5.5 Recommendations for Future Research

A component analysis of Chinese characters provides us with structural knowledge about Chinese characters, and it helps material designers, test developers, and teachers to select and evaluate the characters that to be learned, taught, and tested. The purpose of this study was to systematically describe and classify Chinese characters in ten CFL textbooks for college and adult beginning learners. The results of this study revealed some areas that are needed for further investigation in the future. Some of them were stated above, but they are listed together in the following.

5.3.8 The Usefulness of the Chinese Orthography Introduction in the Textbooks

Should learners be introduced to Chinese orthography explanations in the textbooks at the beginning of the courses? The introduction of the Chinese writing system is not only “nice to know knowledge” about the historical and cultural aspects of the Chinese writing system, but also provides foundations for learners to know that there is actually a systematic property of the Chinese writing system. Future research should look at how learners value the introduction of Chinese orthography information in the textbooks at the beginning of the courses.

5.3.9 Terminology of Components in Orthographic Decomposition Instruction

During the component analysis of Chinese characters, I found that giving names to some of the components was a challenging task because there is no consensus of the terminology of components and the structural graphic description of components. If a component is a character-component, I can use the pronunciation of the character as the name of the component. However, if a component is a non-character component and is neither a semantic nor a phonetic element, I would not be able to pronounce the component. In this case, we can imagine how hard it is for Chinese teachers to systematically explain these components to students. Fu (1993), Guder-Manitius (1999), and Kupfer (2007) have pointed out the need for a more scientific and appropriate terminology. To promote explicit orthographic decomposition instruction, I think the first step is to determine the terminology of components.

To fulfill the needs for a more scientific and appropriate terminology and to promote explicit orthographic decomposition instruction, I propose a strategy, shown in Figure 5.2, which contains the steps to name components within a character in orthographic decomposition instruction. The rules are set to put the radical, semantic element, phonetic element, altered component, and others into consideration.

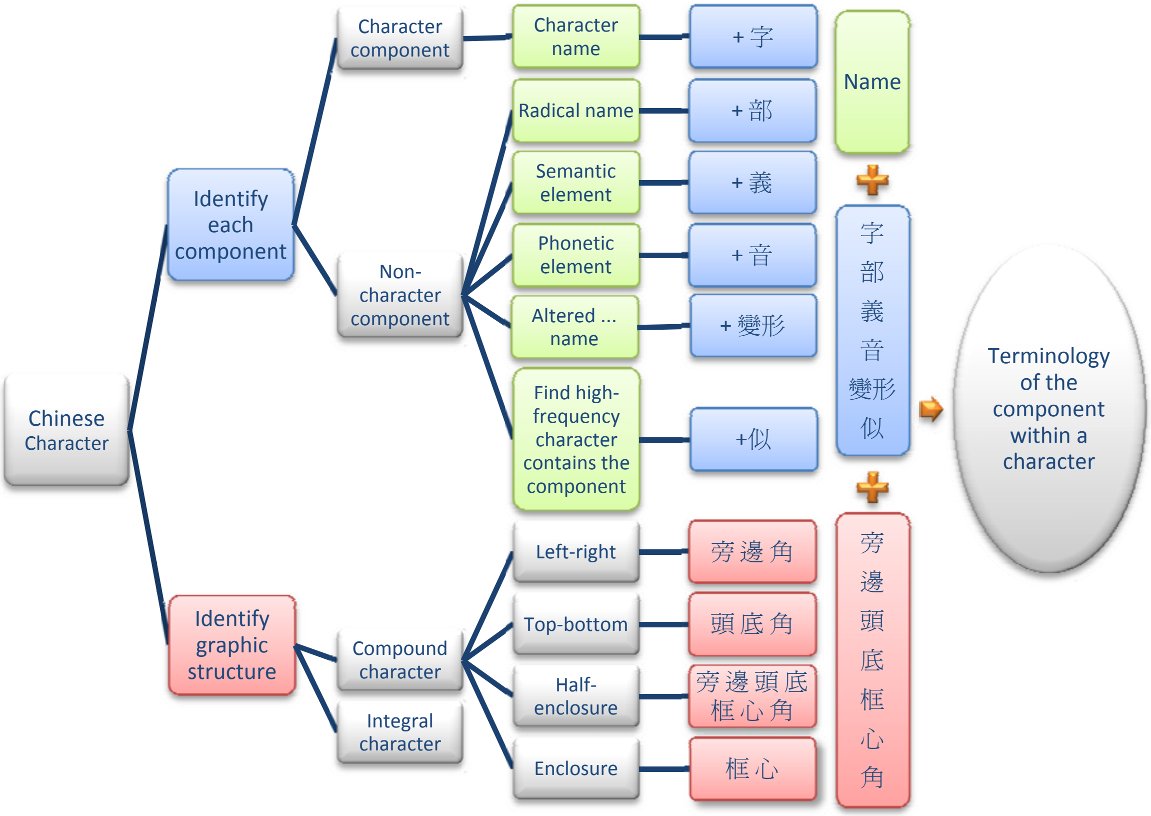

Chinese Character
Identify each component
Character component
Non-character component
Character name
Radical name
Semantic element
Phonetic element
Altered ... name
Find high-frequency character contains the component
+字
+部
+義
+音
+變形
+似
Name
字 部 義 音 變形 似
Identify graphic structure
Compound character
Integral character
Left-right
Top-bottom
Half-enclosure
Enclosure
旁 邊 角
頭 底 角
旁 邊 頭 底 框 心 角
框 心
旁 邊 頭 底 框 心 角
Terminology of the component within a character

Figure 5.2 Defining Terminology of the Component within a Character

Therefore, we can use 字, 部, 形, 聲, 變形, and 似 to infer the characteristics of the component. In addition, we can use 旁, 邊, 頭, 底, 框, 心, and 角 to infer the position of the component within a character. We also can list the terminology of component within a character next to every character introduced in the vocabulary lists in the textbooks, so we can set up standards, and teachers do not need extra class preparation time to find out the decomposition materials. However, this proposal needs to be further investigated in the future research. In addition, I would like to compare the component naming of this proposal to the naming rules of components published by the mainland Chinese government in June, 2009 現代常用字部件及部件名稱規範 (Specification of common modern Chinese character components and component names, GF 0014-2009).

The document includes the four naming rules of components: using pronunciation, stroke, convention name (俗稱), and position to name components. The 現代常用字部件及部件名稱規範 naming rules did not consider the difference within character structures.

5.3.10 Teachers' Knowledge About Language in Chinese Orthographic Awareness

In the CFL context, the three main sources of input for learners are materials, other learners, and the teachers themselves. Since all of the CFL beginning level textbooks in this study did not include the information about semantic transparency of characters, where should teachers look it up and structure input for learners? In this case, teachers' knowledge about the language (KAL) plays an important role. For future research, we should first investigate teachers' knowledge about language (KAL) in Chinese orthographic awareness.

Research had confirmed that the knowledge which teachers have of the underlying systems of the language can impact upon their pedagogical practice (Andrews, 1999; Andrews & McNeill, 2005, Xiao, 2009). Andrews (1999) believes that the teacher's explicit knowledge, her confidence in her own knowledge, and her awareness in making use of her knowledge can affect structuring input for learning, both negatively and positively. As stated, the three main sources of input for learners are materials, other learners, and the teachers themselves. Andrews (1999) found that teachers' reactions to textbooks or the school designed standardized exercises varied from the unaware, uncritical, diffident acceptance of all that the materials say to the rather more aware and self-confident modification of perceived textbook inadequacies. Andrews and McNeill (2005) further investigated whether 'Good Language Teachers' possess highly developed levels of declarative knowledge of the language systems. Declarative knowledge is the knowledge that one can talk about (declare) and describe (Nation, 2001, p361). The 'Good Language Teachers' in their study were teachers whose classroom L2 teaching had been rated as exceptional on the basis of at least two observed lessons. To investigate the

‘Good Language Teachers’ knowledge about language, they were asked to take both grammar and vocabulary components of the language awareness test. In the vocabulary component, they asked L2 teachers of English to divide words into morphemes, count the number of morphemes in each word, describe the lexical relations between words, correct vocabulary errors, and explain the errors. Andrews and McNeill (2005) found that these three teachers performed at a very similar level on the grammar component, but the two Hong Kong teachers performed far worse on the vocabulary component. Andrews and McNeill (2005) suspected that the low performance on the vocabulary component is associated with the emphasis traditionally placed on grammatical competence in L2 teaching and learning in Hong Kong, and the relative lack of attention paid to vocabulary. One teacher could correct all but one of the 15 sentences containing a vocabulary error (93.3%), but was able to score only 23.3% for her explanations of those same errors; the other teacher could correct only 66.6% of the sentences, but performed marginally better (26.6%) in her explanation of those corrections. Further, Andrews and McNeill (2005) investigated these teachers’ application of their knowledge of language (KAL) in their pedagogical practice. They found that teacher’s limitations (low performance) in their subject-matter knowledge became apparent in the ways they made language input available to the students although they were rated as ‘Good Language Teachers’.

Different from investigating rated good language teachers, Xiao (2005) had investigated novice CFL teacher’s explicit knowledge of Chinese orthography and their use of such knowledge in instructional decision making. Xiao (2005) found the six novice CFL teachers were scored on the average 92.67% , indicating that these novice CFL teachers largely possessed the needed analytical orthographic skills to deal with learners’ orthographic errors including graphemic, phonological, semantic, combined phonological/grahemic, and combined phonological/semantic errors. In terms of corrective strategy articulation, Xiao (2005) found that these novice CFL teachers preferred to use intracharacter component analysis and stroke analysis and to ask students to repeat writing the characters. Further, in response to a question asking how to prevent errors in students' future learning, over 50% of the responses were in favor of raising the learners' orthographic awareness and using explicit orthographic knowledge for explanation.

Xiao's findings, however, cannot be over interpreted because these novice CFL teachers were investigated after taking Xiao's CFL pedagogy course to develop their knowledge in Chinese orthography including character structure and configuration, character density (number of strokes) effect, word superiority (orthographic unit recognition) effect, graphic/semantic/phonetic similarity effect, homophone interference, character encoding processes, etc. For those CFL teachers who do not receive explicit Chinese orthography pedagogy, should we assume that they posses the declarative knowledge of Chinese orthodoxy to support their CFL teaching? Therefore, for future research, we should first investigate teachers' knowledge, beliefs and awareness about Chinese orthography and their impact upon teachers' pedagogical practice.

5.6 Recommendations for Textbook Writers

When learners are trying to build their own understanding of the Chinese character system, the CFL textbooks examined in the current study often provided them with counter-examples or irregular-examples. Textbook writers, therefore, should consider integrating orthographic decomposition and component frequency materials into the textbooks, so CFL learners have more opportunities to develop orthographic awareness through characters in the textbooks. To write textbooks that give opportunities for students to develop orthographic awareness, beginning level textbook writers should consider the following issues:

1. Chinese writing system: In the introduction of the textbooks, important features of the Chinese writing system such as etymological character formation (The Six Books), radicals, phonetic elements, strokes, and

components should be introduced to CFL learners. Research indicates that information about the Chinese writing system is not only “nice to know” knowledge about the historical and cultural aspects of the Chinese writing system, but also provides declarative knowledge that helps develop word recognition proficiency.

2. Explicit orthographic decomposition instruction: Each character introduced in the vocabulary list should be decomposed into different orthographic components along with its graphic structure. Radical and phonetic components could also be color coded. In this way, CFL learners would be given meaningful clues from the visual presentations of characters. In addition, textbook writers should provide character decomposition information with further explanation. For example, character 媽 should be introduced as a semantic-phonetic compound character in left-right graphic structure with a semantically transparent radical 女 and a reliable phonetic component 馬 with tonal difference. Without such explanation, students will not learn that there is a system, albeit imperfect, for inferring the meaning and sound of Chinese characters.
3. Character diversity, repetition, and frequency of usage: To develop Chinese reading proficiency, DeFrancis (1977) recommended that learners should become familiar with the character combination process by mastering several character combinations for a limited number of characters as opposed to learning one or two character combination for many characters. Therefore, textbook writers should provide more high-frequency character combination (high-frequency word) examples to learners in the vocabulary list or in the APPENDIX. In this way, learners are exposed to both high-frequency words and high-frequency characters.

4. Radical component diversity and repetition: Textbook writers should introduce characters with high-frequency combination and semantically transparent radicals in the first few lessons in the textbooks with the radicals color-coded, so learners can be directed to notice the function of the radical. In addition, radicals that are characters themselves should be emphasized, for once learners master these radical characters, learners can use them to infer unfamiliar or novel characters. The radical locations of characters should be emphasized as well, such as whether they occur on the left side of left-right structured characters.
5. Phonetic component diversity and repetition: Textbook writers should introduce derivation and analogy strategies to learners, so they can use these two strategies to infer the pronunciation of characters through shared phonetic components. While phonetically reliable elements of characters are introduced, other characters in the homophone and partial homophone analogy groups should be mentioned as well.

To promote reading in Chinese, orthographic awareness-based textbooks should be developed because the main goal is to help CFL learners become independent readers. If textbooks are designed as communicative or theme based, textbook writers still can put explicit orthographic awareness materials in the APPENDIX, so both learners and teachers can learn from them.

5.7 Recommendations for Classroom Instructors

The results of this study suggest that to develop Chinese orthographic awareness, learners may have to rely primarily on classroom instruction for information about orthographic principles and their applications, since overt orthographic awareness instruction in the textbooks investigated in this study was generally lacking. To help learners to develop orthographic awareness, classroom instructors should consider the following issues:

1. High frequency characters: Classroom instructors should pay attention to and direct their students' attention to high frequency characters. Through frequently being exposed to and taught about the components in high frequency characters, readers would develop the ability to look for the internal structure of the characters. Therefore, when they encounter low frequency characters or novel characters, they would be able to use their orthographic knowledge to infer the meaning and pronunciation of the characters. Orthographic awareness development at the beginning stage is not so much a function of the emphasis we should provide learners with, but the thinking processes or strategies we should teach students so that they can become independent readers.
2. Building the orthographic awareness thinking process: With the findings that semantically transparent and phonetically reliable characters are rare in beginning CFL textbooks, can CFL instructors still teach the semantic and phonetic functions? Guder (2007) cautioned that the teaching of components and their semantic and phonetic functions is an important step for building a learner's graphemic competence, but it should not be overemphasized in the first months of study. Through the coding process used to investigate the semantic and phonological components of characters for this study, we must consider that the orthographic awareness development at the beginning learning stage is not so much a function of the emphasis we should provide learners with, but the thinking processes or strategies we should teach students so that they become independent readers.

Equipped with the proper strategies, CFL learners can leave the classroom and find that they are able to make reasonable and principled inferences as to the pronunciation and meanings of many unknown characters. When students start to use orthographic awareness outside the classroom, they open their world to self-learning and continue to develop their own Chinese orthographic awareness.

5.8 Recommendations for Teacher Educators

The results of this study imply that to develop Chinese orthographic awareness, learners may have to rely primarily on classroom instruction for information about orthography principles and their applications. Therefore, teachers' knowledge about the language plays an important role, but we should not assume that teachers possess a working knowledge of Chinese orthography. Teacher educators, therefore, should provide pedagogy courses that educate pre-service teachers about Chinese orthodoxy that can support their CFL teaching. Course content should include the following areas: explicit orthographic decomposition instruction, character frequency selection, radical combination frequency, radical semantic transparency, radical positional regularity among different character graphic structures, phonetic element reliability, phonetic component combination frequency, and terminology used to describe the component within a character. To ensure that they successfully structure input for learning, pre-service teachers should possess explicit knowledge about the language, have confidence in their own knowledge, and be aware of how to make use of their knowledge to support their CFL teaching in the future (Andrew, 1999).

APPENDIX A LIST OF CHARACTER FREQUENCY RANK APPEARED IN THE TEXTBOOKS IN WENLIN

	Exposure Frequency in the Ten Textbooks										
Top Rank	10	9	8	7	6	5	4	3	2	1	Sum
100	的一(ㄧˋ)一(ㄧˊ)是不了人在我有中這大(ㄉㄚˋ)國上個來他為到時們年生會子要以說學對下那可過多小你天家去都好日還經沒麼本所事現想開二道看 (57)	一(ㄧˉ)不不地出就也得後能工十用作分方起行(ㄒㄧㄥˊ)前美外高 (22)	電當從 (3)	自她心然三面 (6)	和得發加業 (5)	得成民同如麵種 (7)	了為行(ㄏㄤˊ) (3)	和那裡里而還動 (7)	的這大(ㄉㄞˋ)地主 (5)	於 (1)	116
200	法文意點四問名知回明氣話幾女水見常西少 (19)	老月兩進很機等定間九車給口 (13)	些公把頭已因平位關打 (10)	者只長(ㄔㄤˊ)最又新己表五比次 (11)	正手第或員身 (6)	但理長(ㄓㄤˇ)力重華體特 (8)	實情化社活 (5)	政全錶將代內度 (7)	其部它相軍向被象 (8)	此由相使總入什各及 (9)	96
300	東太再書先每請今 (8)	應真別八百件路教邊北走難愛聽 (14)	報果門做才共白期 (8)	萬信題教放六便七認 (9)	海台場山數更親接直帶 (10)	市係安結元便記金色收 (10)	利物計 (3)	合提原界許處隊系目流風交運 (13)	聲世受光區眼 (6)	臺颱性至任解指變 (8)	89
400	張思吃寫兒叫友字 (8)	遠候告住英半王錢 (8)	覺覺南辦師馬該近千影服片 (12)	條紅識語司 (5)	務京言完往讓院算研視單照 (12)	必程感節空步商林連 (9)	花思非廣清轉級談 (8)	治保科導且越歷式 (8)	濟觀深 (3)	形傳術領決即品基取講管滿府 (13)	86
500	容快晚易買 (5)	早號怎房找男 (6)	離拿孩毛校火圖星 (8)	音李病 (3)	熱約息始黑母究興 (8)	夜笑樂樂輕參局 (7)	德極證包準克夫復 (8)	青線史士較整精拉死專備亞價黃 (14)	際裝農料志育首支隨 (9)	末團調准類布選稱需廠響參誌派官落興 (17)	85
600	歡喜飯坐念 (5)	呢客嗎吧 (4)	歲送站船 (4)	城飛幫剛左 (5)	費像球願父 (5)	香緊希游習河久 (7)	足灣功注雖醫春留 (8)	驗須除古律劃演屬排護初 (11)	另查食具油底溫份土續 (10)	苦存般版划蘇例推故 (9)	68

	Exposure Frequency in the Ten Textbooks										
Top Rank	10	9	8	7	6	5	4	3	2	1	Sum
700	朋誰 (2)	跟樣哪訴 (4)	媽畫票衣賣刻哥 (7)	夠右甚塊 (4)	考漢班怕您 (5)	雙簡室 (3)	助雲停米 (4)	職速角印港腦波超奇陳 (10)	句勞壓層紀靜洋爺若 (9)	村巴待鄉省陽器章樹修聞傷細止答繼勝竟 (18)	66
800			錯試酒啊 (4)	差(ㄔㄚˋ)差(ㄔㄚˉ)筆店樓 (5)	汽亮紙雨冷 (5)	夏歌園唱跑貨短舊錄卡 (10)	州臉換假銀街 (6)	皮吸訂適味沙鐵洲 (8)	阿板營養景座田衛木急檢素陸散腳 (15)	察顧武靠判永詩葉雜草朝免庭 (13)	66
900	鐘貴 (2)	課弟姐謝玩慢 (6)	旅穿菜睡午忙 (6)	忘介 (2)		禮肉練舞 (4)	封劇著燈壞 (5)	預著堂雪威藥畢寄煙戶概付伯 (13)	頓擔露登娘訪牌遍敢險架 (11)	狀亂減互彈蘭翻討龍困招順警縣 (14)	63
1000		館 (1)	魚 (1)	牛爸零紹 (4)	休 (1)	綠迎秋借旁跳 (6)	簽盤哭婚航窗 (6)	湖套怪季賽末途束 (8)	趣輸退戲屋嘴頂億拍趕危歐篇恐晴秘 (16)	肯遇執鮮臨淚胡托楊彩弄雄稅弱玉 (15)	58
1100	姓 (1)	喝 (1)	懂茶 (2)	床洗 (2)	雞燒附丁 (4)	詞狗冬 (3)	碼麻冰抱掛 (5)	圓呀牙折康楚鋼 (7)	孫牆盛潮 (4)	探麥野醒齊獎擺聖慶博耳鎮絲橋晨普 (16)	45
1200			妹 (1)	餐 (1)	桌典什 (3)	累豆宜奶 (4)	授煩(2)	通健寒 (3)	姑刀針陰尺 (5)	禁購暴尤魯贊聚吹努恩返甲梁譯迅慣貿仁 (18)	37
1300			廳 (1)	杯 (1)		舍藍腐 (3)	售婆蛋箱拜 (5)	網輛粉礦幣腿滑冒啦俄倆灰舒 (13)	趙貼戴 (3)	訊珠操籍兄閉袋鋪側咱伍 (11)	37

	Exposure Frequency in the Ten Textbooks										
Top Rank	10	9	8	7	6	5	4	3	2	1	Sum
1400					郵糖 (2)	顏瓶租鞋 (4)	豬碗紐枝斤 (5)	倫拖涼址掃 (5)	剩 (1)	辛琴齡炸鍋稍肥裁挺賓粗拼 (12)	29
1500		昨 (1)			湯祝漂 (3)	宿 (1)	酸 (1)	韓暖搬飽 (4)	濕鼻宮肚洛 (5)	植煉叔黎虎仰貝閱曉坡冠丟森(14)	29
1600						望 (1)	褲 (1)	聰夾(2)	寸飲噴騎瓜敏 (6)	池澳宋摩泰誕淺朗扣飾罰漲爬帳朵鹽 (16)	26
1700					喂 (1)		甜帽餓啤 (4)	泉刷贏羊墨炒吵 (7)	悶躺髒瘦 (4)	棋仔壽桂駕盒乙填餅塵灘 (11)	27
1800				廚 (1)		鴨衫(2)		紫疼(2)	姆册聊賀拾貓嘗 (7)	誼桃斜炎酷 (5)	17
1900						椅 (1)		櫃傅 (2)	猜俱 (2)	覽錦賺懶盲厭 (6)	11
2000							臥辣 (2)	踢渴拳 (3)	戚碧姨棒 (4)	颳堡丙晴煮 (5)	14
2100							葡萄 (2)	糟押籃 (3)	娜胖醬 (3)	斑裹匹敦嗓蒸獅菲 (8)	16
2200						泳襯 (2)	蝦 (1)		臘芝 (2)	巾茅賬鹹勿垃 (6)	11
2300						糕烤 (2)		乒拐 (2)	壺帥 (2)	唉浴輔圾沫兜叉蘿 (8)	14
2400							蘋裙 (2)	乓汁 (2)	咳歉拌 (3)	橡葱蔥杭寵薯 (6)	13

	Exposure Frequency in the Ten Textbooks										
Top Rank	10	9	8	7	6	5	4	3	2	1	Sum
2500						咖啡 (2)	廁 (1)	夷暑寓筷哎(5)	癢鍛淹 (2)	毯肆笨 (1)	14
2600						醋 (1)	乾 (1)	鉛 (1)	逛杉(2)	帕瑰釣煎玫枕梨椒艙 (9)	14
2700									磅嗯(2)	杏碟 (2)	4
2800								蕉襪(2)	噢 (1)	涕茄 (2)	5
2900						澡 (1)			磯 (1)		2
3000										姊 (1)	1
Above 3000						餃 (1)	橘 (1)	恤 (1)	嗽藩(2)	夥檳瑜 (3)	8
Out of Wenlin List			裹 (1)				佰淨 (2)	週遊 (2)	傢噓 (2)	仟壹貳柒捌玖欸珈莓萄蕃薑蹟酪鍊佔橄欖餛飩 (20)	27
Sum	107	80	68	59	70	99	91	165	156	299	1194

Note. Shading characters are radical characters.

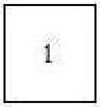
APPENDIX B FREQUENCY DISTRIBUTION OF CHARACTER FREQUENCY RANK ACROSS TEN TEXTBOOKS

Top # Rank	Textbooks																													
	RC 1961			CTBC 1976			BCR 1977			CIC$_{\text{rw}}$ 1994			PCR 1995			NPCR 2002			IC 2005			IC 2008			PAVC 2008			FEEC 2008		
Lists	HS	WL	XF	HS	WL	XF	HS	WL	XF	HS	WL	XF	HS	WL	XF	HS	WL	XF	HS	WL	XF	HS	WL	XF	HS	WL	XF	HS	WL	XF
100	70	81	86	85	93	95	73	82	82	83	92	93	87	89	92	72	77	77	93	95	98	92	95	97	66	77	78	75	84	84
200	49	52	54	63	67	76	50	55	56	67	68	74	60	62	68	37	42	52	73	84	87	70	77	80	39	44	45	49	55	61
300	32	42	39	47	60	52	32	36	32	57	68	62	37	47	50	25	32	33	61	74	71	54	60	63	28	34	36	46	50	46
400	30	26	30	48	52	53	33	30	32	63	63	64	41	46	45	29	39	28	68	68	73	63	67	71	26	30	30	38	43	46
500	19	24	21	39	36	37	19	19	18	49	52	49	35	48	40	26	22	26	59	60	58	56	60	53	21	24	26	27	30	34
600	17	22	19	30	38	33	16	17	16	41	41	34	35	32	34	17	20	19	52	54	45	38	46	40	18	19	13	24	23	20
700	19	18	12	31	26	26	16	13	14	31	40	39	30	30	24	13	20	13	41	40	48	37	35	40	14	18	19	19	27	22
800	20	12	21	25	18	24	12	9	11	47	35	45	26	24	33	24	10	13	41	47	41	41	42	40	24	19	18	25	17	22
900	21	14	13	24	25	23	10	7	9	30	34	34	27	31	29	20	20	18	40	42	41	37	36	37	22	17	14	23	23	19
1000	7	6	5	12	12	13	5	7	7	24	23	19	23	24	16	13	8	8	38	38	32	29	30	27	13	11	8	20	17	11
1100	4	5	2	10	13	11	4	2	2	18	21	23	16	15	18	7	12	11	24	34	30	22	28	27	5	8	9	11	15	19
1200	4	1	3	14	9	4	2	2		24	15	18	20	13	11	10	7	6	30	16	26	30	17	17	11	4	5	14	8	12
1300	1	2	2	7	7	9	2		1	20	22	16	13	12	9	5	7	8	24	24	23	15	20	18	6	4	2	9	8	6
1400	4			13	8	7	2			17	17	13	14	7	6	6	5	6	22	19	20	21	15	18	5	1		12	9	7
1500	6	1		11	10	10	2		1	14	11	10	9	7	6	9	6	6	18	15	15	16	15	14	5	1	1	7	6	6
1600	1			7	2	3	1			14	10	6	5	2	6	6	2	3	14	13	12	15	11	13	2		1	8	3	6
1700	1			4	6	4				8	12	13	6	5	3	6	3	1	12	13	11	9	16	12	1		1	2	7	5
1800	1	1		5	2	4		1		13	7	10	4	5	5	3	5	4	11	8	14	17	10	17	3	1		6	2	
1900	1		1	3	1	3	1			5	4	8	3	2	6	4	1	4	6	6	8	6	5	6	1	1	3	2	1	2
2000				4	1	3				5	7	10	5	2	4	4		5	14	10	11	14	9	11			1	4	1	3
2100				2	3	4				6	3	11	5	4	7	2	4	4	6	7	8	6	8	12				2	2	5
2200				2	2	2				7	7	2	4	3	1		1		10	4		10	4					6	3	2
2300				6	5					10	4		4	3		1	2		7	5		5	7		1			7	2	
2400				3	1					4	6		1	3		3	2		5	5		8	8						1	
2500	1			2	2					3	6		4	4		3	4		6	9		5	10					2	3	
2600				3	4					4	3		1	3		1			6	4		7	7		1			1	4	
2700					1					5	2		4			3			6	1		8	1					1	1	
2800				3	1					1	2		1	1			1		2	1		3	3			1		1		
2900				1						2	2		1	1		1			2	2		4	1		1			1	1	
3000				2						1			1			1			2			3							1	
Above 3000				3	3		1			21	4		10	3		3			17	4		18	5		1			6		
Out of the list		1			1	13		1		1	14	42	1	5	20		2	9	1	9	36		11	44			4	1	2	11
Total num	308			509			281			695			533			354			811			759			314			449		

Note. HZ =漢字屬性字典 *Hànzì shǔxìng zìdiǎn* frequency rank; WL = Wenlin frequency rank; XF = Xiao et. al., frequency list.

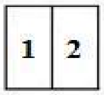
APPENDIX C CHARACTERS IN FIVE CHARACTER GRAPHIC STRUCTURES

I

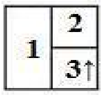


Radical character : 一乙二人入八刀力十又口土士大女子寸小山工己巾心戶手斤方日月木止毛火片牙牛玉王瓜生用田白皮目米羊老而耳肉自色衣西角走足身車辛里金長雨面風飛首馬麻龍 (72/120)
丁七上下不且世丙中久九也了事互五亞內兩册再出勿半及史夫夷州已巴平幾必我才更未末本東柬果母民由甲重(48/108)

LR

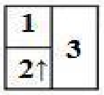


LR-1
八朋林比門非
1⃞打訂杯紐姐租粗
叫奶玖托作昨炸池
他地她伍忙臥瑜臉
檢險驗桃跳倆輛賺
枕加物此死什汁計
針訊仟拌胖啤牌外
吸圾級假蝦如扣知
使肚社睡任快決姨
姊樓仔好付討你飩
洲紅紀記吧把肥機
代衫秘找城俄餓拜
玫所折訴訪便喝渴
明休味妹沫珠辣陳
煉練鍊吹飲址姆紙
冰球漢秋呀件姓性
輔鋪細油押伯帕怕
拍波坡被植野位垃
拉洋鮮律稍咱很眼
跟銀恤現像橡理裡
種帳張賬們准推誰
啡排煩碼塊瑰麵
(166)
2⃞頂助收預刻化領
歉次北判印即郵數
頓剛順功帥形須划
敢放相夥政酒難狀
鴨視利和料取到初
頭那朗動肆 (44)



LR-2
1⃞隊住往注吃拖臨飾旅
仁語極流杭航校較餃咳
孩該涼停從於冷齡除陰
倫輸艙船鉛玩洗稅說晚
棋俱稱帽講購解份粉釣
的約抱跑飽德許估站貼
派板版飯授祝拐姑個狗
招紹河阿椅騎始治詞洛
路酪拾給結裙保酷調減
鹹鍋操澡糖橋彈溫牆法
怪桂鞋陸煙曉燒程酸復
淹棒換澳噢接院館誼傢
演村待時特詩傅博傳轉
吵沙炒躺呢鋼腦經輕臘
港掃錦拼餅譯後磯褲礦
滿研試涕灣綠錄杉認誌
嗯聰總億淺錢淚枝科繼
磅得但指唱婚踢借醋錯
傷場楊湯陽醒提糟慢服
課碟嗓沒般海汽泳線談
淨靜暖備喂噓燈證佰填
鎮橘唉漂際鐘絲係孫網
探深糕樣烤豬活話娘哎
貓髒襪錶壞讓欖嘴信擔
浴短禮體慣蹟櫃續噴價
檳懂護情晴清猜睛精請
嗎媽滑點濟 (264)

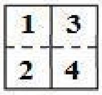
3⃞務勝將師 (4)



LR-3
1⃞執

2⃞裁

3⃞對剩
雜郭敦
新親斜
散期歐
朝韓願
雄別故
胡歌部
戴封雞
教叔刷
瓶顏影
敏亂翻
劃欸離
類都甜
戲雖劇
彩顧歡
觀



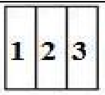
LR-4
行弱

1⃞夠報
聽

2⃞能舒

3⃞號

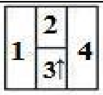
4⃞乾就
縣



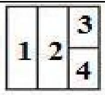
LR-5
水
1⃞附例
側珈咖
仰聊啊
謝樹獅
誕挺健
漲啦橄
概來嗽
腳腿哪
娜灘

2⃞夾

3⃞以鄉

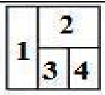


LR-6
1⃞捌椒
做湖鐵
掛識潮
職襯



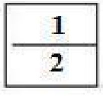
LR-7
1⃞修候
假搬游
隨鍛

4⃞條

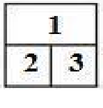


LR-8
①擺濕碗
餛

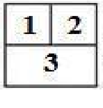
TB



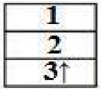
TB-1
支父至言豆貝
青音食香高魚
麥黃黑鼻齊文
見多炎象年弟
冒永甚直



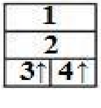
TB-2
①冠宿
寵最爺
發筷箱



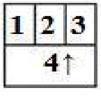
TB-3
①哭單
楚準琴
聖聚雙



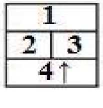
TB-4
①交京今
員套完容



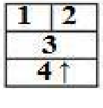
TB-5
①蘿簽



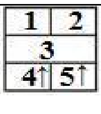
TB-6
②變



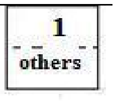
TB-7
①察簡
籃藍藥
蘭



TB-8
①營

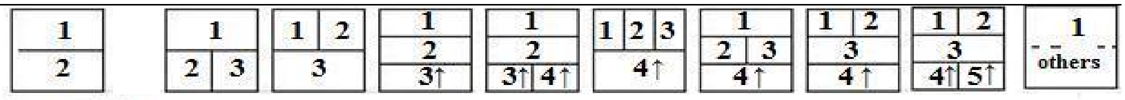


TB-9
①器



TB-10
黎
1⃞萄葡
蔔南參
壽歲蘭
裏裹簡

TB


1
2
1
2 3
1 2
3
1
2
3↑
1
2
3↑ 4↑
1 2 3
4↑
1
2 3
4↑
1 2
3
4↑
1 2
3
4↑ 5↑
1
others

TB-1
①華宜主少
芝介簽全穿
公男易它早
杏另只宮美
笑英安要李
官爸第雪魯
宋笨步星電
畢定是習茅
寓萬究窗空
需筆室登家
員晨套菲

2⃞百卡千系
怎午每元六
市忘盲合分
共其具典軍
朵古台兄先
告恩去志冬
各名天學覺
字尤費光專
盛拳旁春泰
農書肯正舞
弄思累蛋泉
季買養息兒
舍票貴售素
表

TB-2
品森夜
1⃞冠茄
花節最
落露發
筷籍藩
宿箱蘇
爺罰蘋
寵

3⃞前

TB-3
1⃞琴聚
單準楚
聖雙

2⃞望

3⃞坐袋
您堡贊
梨架賀
駕貨熱
裝努獎
醬恐幣
警照然
禁些紫
餐盤聲
醫柒梁
婆碧留
貿想緊
覽

TB-4
1⃞三乓乒
今雲交亮
京景茶零
完容界寒
葱蒽丟苦
奇寄若客
答壺管等
篇暴章草
竟葉查莓
蒸菜蕃薑
算暑薯著
寫壹實蕉

2⃞夏真喜
臺嘗畫

3⃞業普育
會念克免
基舊桌急
拿盒常靠
賣賓希帶
受愛堂當
意煮導墨

4⃞賽

TB-5
1⃞蘿
4⃞羸

TB-6
2⃞變
4⃞樂響兜

TB-7
1⃞藥籃
藍察

4⃞煎

TB-8
1⃞營

2⃞哭整

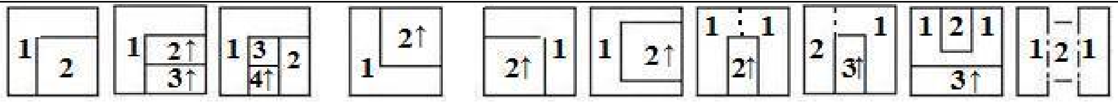
4⃞勞幫

TB-9
1⃞器

TB-10
2⃞哥

3⃞鹽

HE


1
2
1
2↑
3↑
1
3
4↑
2
2↑
1
2↑
1
1
2↑
1
1
2↑
2
1
3↑
1
2
1
3↑
1
2
1

HE-1
1⃞考者床
房病康虎
麼屋底尺

2⃞友右左
在存差布
有灰省看
塵歷摩

HE-2
1⃞原局層
屬度座廣
疼瘦癢店

2⃞危慶

3⃞壓應腐

HE-3
1⃞厭府
庭廁廚
廠廳處

HE-4
1⃞毯爬起超
越趕趙趣迅
迎近返退送
途這通逛速
連週進遇遊
運遍過道遠
適選還邊颱
颳
2⃞題

HE-5
1⃞包式
或氣

2⃞句可
司武
為貳

HE-6
1⃞匹區

HE-7
1⃞太閉
開間閱
關

2⃞同向
問悶聞
商

HE-8
1⃞成戚

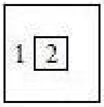
3⃞威感

HE-9
1⃞興

HE-10
1⃞術街
衛班辦

2⃞斑

E



叉四回因困國園圓圖團 (10)

APPENDIX D PHONETIC COMPONENT ANALOGY TYPES FOR EACH GROUP

Type	The whole group with the same component	Analogy type 1=homophone	Analogy type 2=partial homophone (tonal)	Analogy type 3=same rhymes	Analogy type 4=different sounds
1	(七ㄑㄧ¯)七柒	七柒			
1	(久ㄐㄧㄡˇ)久玖	久玖			
1	(介ㄐㄧㄝˋ)介界	介界			
1	(代ㄉㄞˋ)代袋	代袋			
1	(保ㄅㄠˇ)保堡	保堡			
1	(具ㄐㄩˋ)具俱	具俱			
1	(冒ㄇㄠˋ)冒帽	冒帽			
1	(勿ㄨˋ)勿物	勿物			
1	(千ㄑㄧㄢ¯)千仟	千仟			
1	(受ㄕㄡˋ)受授	受授			
1	(史ㄕˇ)史使	史使			
1	(員ㄩㄢˊ)員圓	員圓			
1	(坐ㄗㄨㄛˋ)坐座	坐座			
1	(夷ㄧˊ)夷姨	夷姨			
1	(家ㄐㄧㄚ¯)家傢	家傢			
1	(州ㄓㄡ¯)州洲	州洲			
1	(師ㄕ¯)師獅	師獅			
1	(式ㄕˋ)式試	式試			
1	(志ㄓˋ)志誌	志誌			
1	(支ㄓ¯)支枝	支枝			
1	(敢ㄍㄢˇ)敢橄	敢橄			
1	(末ㄇㄛˋ)末沫	末沫			
1	(母ㄇㄨˇ)母姆	母姆			
1	(永ㄩㄥˇ)永泳	永泳			
1	(由ㄧㄡˊ)由油	由油			
1	(畫ㄏㄨㄚˋ)畫劃	畫劃			
1	(直ㄓˊ)直植	直植			
1	(票ㄆㄧㄠˋ)票漂	票漂			
1	(考ㄎㄠˇ)考烤	考烤			
1	(表ㄅㄧㄠˇ)表錶	表錶			
1	(象ㄒㄧㄤˋ)象像橡	象像橡			
1	(貴ㄍㄨㄟˋ)貴櫃	貴櫃			
1	(賓ㄅㄧㄣ¯)賓檳	賓檳			
1	(面ㄇㄧㄢˋ)面麵	面麵			
1	(養ㄧㄤˇ)養癢	養癢			
1	(岡ㄍㄤ¯)剛鋼	剛鋼			
1	(矣ㄧˇ)唉欸	唉欸			
1	(亢ㄎㄤˋ)航杭	航杭			
1	(卂ㄒㄧㄣˋ)訊迅	訊迅			
1	(气ㄑㄧˋ)汽氣	汽氣			
1	(爭ㄓㄥ¯)淨靜	淨靜			
1	(玨ㄐㄩㄝˊ)班斑	班斑			
1	(斿ㄧㄡˊ)遊游	遊游			

Type	The whole group with the same component	Analogy type 1=homophone	Analogy type 2=partial homophone (tonal)	Analogy type 3=same rhymes	Analogy type 4=different sounds
1	(監ㄐㄧㄢˉ)藍籃	藍籃			
1	(則ㄗㄜˊ)廁側	廁側			
1	(番ㄈㄢˉ)蕃翻藩	蕃翻藩			
2	(丙ㄅㄧㄥˇ)丙病		丙病		
2	(九ㄐㄧㄡˇ)九究		九究		
2	(分ㄈㄣˉˋ)分份粉		分份粉		
2	(原ㄩㄢˊ)原願		原願		
2	(將ㄐㄧㄤˉ)將獎醬		將獎醬		
2	(星ㄒㄧㄥˉ)星醒		星醒		
2	(本ㄅㄣˇ)本笨		本笨		
2	(牙ㄧㄚˊ)牙呀		牙呀		
2	(那ㄋㄚˋ)那娜哪		那娜哪		
2	(鄉ㄒㄧㄤˉ)鄉響		鄉響		
2	(戔ㄐㄧㄢˋ)錢淺		錢淺		
2	(廷ㄊㄧㄥˊ)庭挺		庭挺		
2	(采ㄘㄞˇ)彩菜		彩菜		
2	(袁ㄩㄢˊ)園遠		園遠		
2	(咼ㄎㄨㄚˉ)鍋過		鍋過		
3	(前ㄑㄧㄢˊ)前煎			前煎	
3	(司ㄙˉ)司詞			司詞	
3	(單ㄉㄢˉ)單彈			單彈	
3	(尤ㄧㄡˊ)尤就			尤就	
3	(或ㄏㄨㄛˋ)或國			或國	
3	(早ㄗㄠˇ)早草			早草	
3	(黃ㄏㄨㄤˊ)黃廣礦			黃廣礦	
3	(齊ㄑㄧˊ)齊濟			齊濟	
3	(巠ㄐㄧㄥˉ)經輕			經輕	
3	(叚ㄐㄧㄚˇ)假蝦			假蝦	
3	(曷ㄏㄜˊ)喝渴			喝渴	
3	(肖ㄒㄧㄠˋ)稍趙			稍趙	
3	(幷ㄅㄧㄥˋ)餅拼瓶			餅拼瓶	
3	(屯ㄊㄨㄣˊ)飩頓			飩頓	
3	(喿ㄗㄠˋ)澡操			澡操	
3	(戠ㄓˊ)職識			職識	
3	(尌ㄕㄨˉ)樹廚			樹廚	
3	(扁ㄅㄧㄢˇ)遍篇			遍篇	
3	(雚ㄍㄨㄢˋ)觀歡			觀歡	
4	(不ㄅㄨˋˊ˙)不杯				不杯
4	(一ˇ乙)乙乾吃				乙乾吃
4	(二ㄦˋ)二仁				二仁
4	(你ㄋㄧˇ)你您				你您
4	(兄ㄒㄩㄥˉ)兄祝				兄祝

Type	The whole group with the same component	Analogy type 1=homophone	Analogy type 2=partial homophone (tonal)	Analogy type 3=same rhymes	Analogy type 4=different sounds
4	（先ㄒㄧㄢˉ）先洗贊				先洗贊
4	（免ㄇㄧㄢˇ）免晚				免晚
4	（共ㄍㄨㄥˋ）共港暴選散				共港暴選散
4	（冬ㄉㄨㄥˉ）冬疼				冬疼
4	（刀ㄉㄠˉ）刀初				刀初
4	（力ㄌㄧˋ）力男				力男
4	（區ㄑㄩˉ）區歐				區歐
4	（午ㄨˇ）午許				午許
4	（即ㄐㄧˊ）即節				即節
4	（厭ㄧㄢˋ）厭壓				厭壓
4	（去ㄑㄩˋ）去丟法				去丟法
4	（另ㄌㄧㄥˋ）另別捌拐				另別捌拐
4	（右ㄧㄡˋ）右若				右若
4	（同ㄊㄨㄥˊ）同興				同興
4	（回ㄏㄨㄟˊ）回圖牆				回圖牆
4	（士ㄕˋ）士壺壽				士壺壽
4	（多ㄉㄨㄛˉ）多夥				多夥
4	（女ㄋㄩˇ）女努如好安				女努如好安
4	（巾ㄐㄧㄣˉ）巾幣布帶				巾幣布帶
4	（必ㄅㄧˋ）必秘				必秘
4	（戶ㄏㄨˋ）戶淚				戶淚
4	（才ㄘㄞˊ）才閉				才閉
4	（易ㄧˋ）易踢				易踢
4	（更ㄍㄥˋ）更便				更便
4	（月ㄩㄝˋ）月明朋				月明朋
4	（束ㄕㄨˋ）束嗽速懶辣				束嗽速懶辣
4	（東ㄉㄨㄥˉ）東陳				東陳
4	（林ㄌㄧㄣˊ）林禁森麻麼摩				林禁森麻麼摩
4	（樂ㄌㄜˋㄩㄝˋ）樂藥				樂藥
4	（此ㄘˇ）此些紫嘴				此些紫嘴
4	（步ㄅㄨˋ）步[頻]蘋				步（頻）蘋
4	（比ㄅㄧˇ）比餵				比餵
4	（泉ㄑㄩㄢˊ）泉線				泉線
4	（火ㄏㄨㄛˇ）火灰秋				火灰秋
4	（牛ㄋㄧㄡˊ）牛件				牛件

Type	The whole group with the same component	Analogy type 1=homophone	Analogy type 2=partial homophone (tonal)	Analogy type 3=same rhymes	Analogy type 4=different sounds
4	（用ㄩㄥˋ）用備				用備
4	（田ㄊㄧㄢˊ）田思累細				田思累細
4	（米ㄇㄧˇ）米糕				米糕
4	（而ㄦˊ）而需				而需
4	（能ㄋㄥˊ）能（罷）擺				能擺
4	（自ㄗˋ）自咱息				自咱息
4	（舍ㄕㄜˋ）舍舒［捨］				舍舒
4	（角ㄐㄧㄠˇ）角解				角解
4	（言ㄧㄢˊ）言信這罰				言信這罰
4	（兑ㄉㄨㄟˋ）閱稅說				閱稅說
4	（豆ㄉㄡˋ）豆短壹				豆短壹
4	（貝ㄅㄟˋ）貝蹟噴價				貝蹟噴價
4	（車ㄔㄜˉ）車軍運連				車軍運連
4	（辛ㄒㄧㄣˉ）辛辦				辛辦
4	（魚ㄩˊ）魚魯蘇				魚魯蘇
4	（黑ㄏㄟˉ）黑墨				黑墨
4	（龍ㄌㄨㄥˊ）龍寵				龍寵
4	（祭ㄐㄧˋ）際察				際察
4	（欠ㄑㄧㄢˋ）歉吹次飲				歉吹次飲
4	（兼ㄐㄧㄢ）兼賺歉				兼賺歉
4	（旨ㄓˇ）指嘗				指嘗
4	（良ㄌㄧㄤˊ）朗娘				朗娘
4	（聿ㄌㄩˋ）律肆筆				律肆筆
4	（卑ㄅㄟˉ）牌啤				牌啤
4	（ㄙㄞˉ）賽寒				賽寒
4	（ㄐㄐㄧㄡˉ）叫收				叫收
4	（辰ㄔㄣˊ）晨農				晨農
4	（享ㄒㄧㄤˇ）敦郭				敦郭
4	（攸ㄧㄡˉ）修條				修條
4	（吉ㄐㄧˊ）喜結臺				喜結臺
4	（堇ㄑㄧㄥˊ）漢難灘				漢難灘
4	（枼ㄝˋ）葉碟				葉碟
4	（服?ㄈㄨˊ）服報				服 報
4	（卬ㄤˊ）仰迎				仰迎
4	（殳ㄕㄨˉ）没鍛聲醫				没鍛聲醫
4	（彔ㄌㄨˋ）錄綠				錄綠
4	（余ㄩˊ）途除斜				途除斜

Type	The whole group with the same component	Analogy type 1=homophone	Analogy type 2=partial homophone (tonal)	Analogy type 3=same rhymes	Analogy type 4=different sounds
4	(戈ㄍㄜ˙)划找				划找
4	(戌ㄒㄩˉ)威歲				威歲
4	(幵ㄐㄧㄢˉ)研形				研形
4	(幸ㄒㄧㄥˋ)譯執				譯執
4	(穴ㄒㄩㄝˋ)窗空究穿				窗空究穿
4	(婁ㄌㄡˊ)樓數				樓數
4	(勺ㄕㄠˊ)釣的約				釣的約
4	(罙ㄕㄣˉ)深探				深探
4	(坴ㄌㄧㄡˋ)陸熱				陸熱
4	(谷ㄍㄨˇ)容浴腳				容浴腳
4	(儿口ㄧㄢˇ)鉛船				鉛船
4	(尃ㄈㄨˋ)傅博				傅博
4	(周ㄓㄡˉ)週調				週調
4	(堯ㄧㄠˊ)曉燒				曉燒
4	(豊ㄌㄧˇ)禮體				禮體
4	(昔ㄒㄧˊ)籍借錯醋				籍借錯醋
4	(卯ㄇㄠˇ)貿留聊				貿留聊
4	(02ㄧㄥˊ)營勞				營勞
4	(斗ㄉㄡˇ)科料				科料
4	(疋ㄆㄧˇ, ㄕㄨˉ)楚蛋嚏				楚蛋嚏
4	(冉ㄖㄢˇ)再稱				再稱
4	(貫ㄍㄨㄢˋ)慣實				慣實
4	(丘ㄑ丨ㄡ)乒乓[兵]				乒乓
4	(予ㄩˇ)預舒野				預舒野
4	(矛ㄇㄠˊ)茅橘				茅橘
4	(尗ㄕㄨˉ)叔椒戚				叔椒戚
4	(虍ㄏㄨ)虎號處劇戲				虎號處劇戲
4	(兆ㄓㄠˋ)桃跳				桃跳
4	(俞ㄩˊ)瑜輸				瑜輸
4	(鬼ㄍㄨㄟˇ)瑰塊				瑰塊
4	(頁ㄧㄝˋ)類煩夏順頭顏顧				類煩夏順頭顏顧
4	(糸 言糸 ㄌㄧㄢˋ)變灣				變灣
4	(奧ㄠˋ)澳噢				澳噢
4	(雈ㄏㄨㄢˊ)舊護				舊護
4	(垂ㄔㄨㄟˊ)睡郵				睡郵
4	(冓ㄍㄡˋ)購講				購講
4	(业ㄧㄝˋ)業對普				業對普

Type	The whole group with the same component	Analogy type 1=homophone	Analogy type 2=partial homophone (tonal)	Analogy type 3=same rhymes	Analogy type 4=different sounds
12	(付ㄈㄨˋ)付府腐附	府腐/付附	府腐/付附		
12	(兩ㄌㄧㄤˇ)兩倆輛	兩倆	兩倆/輛		
12	(官ㄍㄨㄢˉ)官管館	管館	官/管館		
12	(己ㄐㄧˇ)己紀記	紀記	己/紀記		
12	(幾ㄐㄧˇˉ)幾機磯	機磯	幾/機磯		
12	(正ㄓㄥˋ)正整政	正政	正整政		
12	(長ㄓㄤˇㄔㄤˊ)長張漲帳賬	長漲/帳賬	長張漲帳賬		
12	(馬ㄇㄚˇ)馬嗎媽碼	馬碼	馬嗎媽碼		
13	(其ㄑㄧˊ)其期棋基	其期棋		(其ㄑㄧˊ)其期棋/基	
13	(奇ㄑㄧˊ)奇騎寄椅	奇騎		(奇ㄑㄧˊ)奇騎/寄/椅	
13	(弟ㄉㄧˋ)弟涕第	弟第		(弟ㄉㄧˋ)弟第/涕	
13	(成ㄔㄥˊ)成城盛	成城		(成ㄔㄥˊ)成城/盛	
13	(般ㄅㄢˉ)般搬盤	般搬		(般ㄅㄢˉ)般搬/盤	
13	(悤ㄘㄨㄥˉ)蔥聰總	蔥聰		(悤ㄘㄨㄥˉ)蔥聰/總	
14	(也ㄝˇ)也池地他她拖	他她			也池地他她拖
14	(五ㄨˇ)五伍語	五伍			伍語
14	(合ㄏㄜˊ)合盒答拿給拾	合盒			合盒答拿給拾
14	(因ㄧㄣˉ)因恩嗯	恩嗯			因/恩嗯
14	(是ㄕˋ)是提題	提題			是提題
14	(未ㄨㄟˋ)未味妹	未味			未味/妹
14	(止ㄓˇ)止址肯	止址			止址/肯
14	(生ㄕㄥˉ)生姓性	姓性			生/姓性
14	(登ㄉㄥˉ)登燈證	登燈			登燈證
14	(白ㄅㄞˊ, ㄅㄛˊ)白伯碧怕帕拍習	怕帕			(白ㄅㄞˊ, ㄅㄛˊ)伯/碧/怕帕/拍/習
14	(百ㄅㄞˇ)百佰宿	百佰			百佰/宿
14	(系ㄒㄧˋ)系係縣	系係			系係/縣
14	(羊ㄧㄤˊ)羊洋鮮	羊洋			羊洋/鮮
14	(禺ㄩˋ)萬遇寓	遇寓			萬/遇寓
14	(里ㄌㄧˇ)里裡理裹裹	里裡裹理			里裡理裹/裹
14	(非ㄈㄟˉ)非啡菲排	非啡菲			非啡菲/排

Type	The whole group with the same component	Analogy type 1=homophone	Analogy type 2=partial homophone (tonal)	Analogy type 3=same rhymes	Analogy type 4=different sounds
14	(音一ㄣˉ)音竟意億章	意億			音/竟/意億/章
14	(亡ㄨㄤˊ)望忘忙盲贏	望忘/忙盲			望忘/忙盲/贏
14	(德?ㄉㄜˊ)德聽廳	聽廳			德/聽廳
14	(夬ㄍㄨㄞˋ)快筷決	快筷			快筷/決
14	(倝ㄍㄢˋ)韓朝潮	朝潮			韓/朝潮
14	(柬ㄐ一ㄢˇ)蘭練煉鍊	練煉鍊			蘭/練煉鍊
14	(隹ㄓㄨㄟˉ)準准誰推蘿雜蕉進雙售應雖	準准			準准/誰推蘿雜蕉進雙售應雖
14	(甲ㄐ一ㄚˇ)甲押鴨	押鴨			甲/押鴨
23	(交ㄐ一ㄠˉ)交餃較校		交餃較	交餃較校	
23	(包ㄅㄠˉ)包飽抱跑		包飽抱	包飽抱/跑	
23	(方ㄈㄤˉ)方房訪放旁謗		方房訪放	方旁謗房/訪/放	
23	(重ㄓㄨㄥˋ, ㄔㄨㄥˊ)重種懂動		(重ㄓㄨㄥˋ, ㄔㄨㄥˊ)重種/懂動	(重ㄓㄨㄥˋ, ㄔㄨㄥˊ)重種/懂動	
23	(召ㄓㄠˋ)招照超紹		招照	招照超紹	
23	(甫ㄈㄨˇ)輔葡鋪		葡鋪	輔葡鋪	
24	(丁ㄉ一ㄥˉ)丁訂頂打停		丁訂頂/停		丁訂頂/打/停
24	(化ㄏㄨㄚˋ)化花貨		化花		化花貨
24	(又一ㄡˋ)又友叉怪		又友		又友/叉怪
24	(句ㄐㄩˋ)句狗夠		狗夠		句/狗夠
24	(子ㄗ˙ˇ)子字仔季李孫		子字		子字仔季李孫
24	(寸ㄘㄨㄣˋ)寸村討		寸村		寸村討
24	(斤ㄐ一ㄣˉ)斤近折所		斤近		斤近/折/所
24	(每ㄇㄟˇ)每莓敏海		每莓		每莓敏海
24	(炎一ㄢˊ)炎談毯		談毯		炎/談毯
24	(真ㄓㄣˉ)真鎮填		真鎮		鎮填
24	(買ㄇㄞˇ)買賣續		買賣		買賣/續
24	(首ㄕㄡˇ)首道導		道導		首/道導

Type	The whole group with the same component	Analogy type 1=homophone	Analogy type 2=partial homophone (tonal)	Analogy type 3=same rhymes	Analogy type 4=different sounds
24	（乍ㄓㄚˋ）炸昨作怎		昨作		炸/昨作/怎
24	（寺ㄕˋ）詩時待等特		詩時		詩時待等特
24	（禾ㄏㄜˊ）利梨和香		利梨		利梨和香
34	（今ㄐㄧㄣˉ）今陰琴念			今陰琴	今陰琴念
34	（告ㄍㄠˋ）告靠酷			告靠	告靠/酷
34	（土ㄊㄨˇ）土肚社			土肚	土肚社
34	（我ㄨㄛˇ）我餓俄			餓俄	我/餓俄
34	（果ㄍㄨㄛˇ）果課夥			果夥	果夥/課
34	（至ㄓˋ）至室屋臺			至室	至室屋臺
34	（見ㄐㄧㄢˋ）見現視親襯			見現	見現/視親襯
34	（僉ㄑㄧㄢˉ）簽檢險臉驗			簽檢險	簽檢險/臉/驗
34	（學ㄒㄩㄝˊ）學覺覺			學覺ㄐㄩㄝˊ	學覺/覺ㄐㄧㄠˋ
34	（咸ㄒㄧㄢˊ）鹹減感			鹹減	鹹減感
34	（亲ㄔㄣˊ）新親襯			新親	新親襯
34	（尚ㄕㄤˋ）廠躺常當堂			廠躺/當堂	廠躺/常/當堂
34	（舌ㄕㄜˊ）颳話活			颳話	颳話/活
34	（圭ㄍㄨㄟˉ）桂掛鞋街封幫			鞋街	鞋街/桂掛封幫
34	（呈ㄔㄥˊ）程聖鐵			程聖	程聖鐵
123	（及ㄐㄧˊ）及級圾吸	及級	及級/圾	及級/圾/吸	
123	（工ㄍㄨㄥˉ）工功紅空恐	工功	空恐	工功紅空恐	
123	（反ㄈㄢˇ）返飯板版	板版	返飯	返飯/板版	
124	（主ㄓㄨˇ）主住注往	住注	主/住注		主/住注/往
124	（加ㄐㄧㄚˉ）加珈架駕茄咖賀	加珈/架駕	加珈/架駕		加珈/架駕/茄/咖/賀
124	（者ㄓㄜˇ）者著豬煮暑薯都	暑薯	者著/豬煮		者著/豬煮/暑薯/都
124	（門ㄇㄣˊ）門們悶間簡開聞問	門們悶	門們悶/間簡/聞問		門們悶/間簡/開/聞問
124	（占ㄓㄢˉ）佔站點店貼	佔站	點店		佔站/點店/貼

Type	The whole group with the same component	Analogy type 1=homophone	Analogy type 2=partial homophone (tonal)	Analogy type 3=same rhymes	Analogy type 4=different sounds
124	(立ㄌㄧˋ)拉垃啦位部	拉垃	拉垃/啦		拉垃/啦/位/部
124	(令ㄌㄧㄥˋ)零齡領冷	零齡	零齡/領		零齡/領/冷
124	(目ㄇㄨˋ)目看相箱想	相箱	相箱想		目看/相箱想
124	(元ㄩㄢˊ)元玩完院冠	玩完	元院		元/玩完/院/冠
134	(且ㄑㄧㄝˇ)且姐粗租助宜誼	宜誼		且姐/粗租	且姐/粗租/助/宜誼
134	(十ㄕˊ)十汁針什什計	十什		針什	十什/汁/針什/計
134	(半ㄅㄢˋ)半拌判胖	半拌		半拌/判	半拌/判/胖
134	(可ㄎㄜˇ)可哥歌河阿啊	哥歌/阿啊		可/哥歌/河	可/哥歌/河/阿啊
134	(台ㄊㄞˊ)台颱始治	台颱		始治	台颱/始治
134	(各ㄍㄜˋ)各客酪路露洛落	路露/洛落		各客	各客/酪/路露/洛落
134	(少ㄕㄠˇ)少炒吵沙省	炒吵		少炒吵	少/炒吵/沙/省
134	(彡ㄕㄢˉ)衫杉參參須	衫杉		衫杉參參須	衫杉/參參/須
134	(昜ㄧㄤˊ)陽楊湯場傷	陽楊		場傷	陽楊/湯/場傷
234	(京ㄐㄧㄥˉ)京景影涼		京景	京景影	京景/影/涼
234	(取ㄑㄩˇ)取趣聚最		取趣	趣聚	取趣聚最
234	(專ㄓㄨㄢˉ)專轉傳團		專轉	專轉/傳	專轉/傳/團
234	(皮ㄆㄧˊ)皮坡婆波被		坡婆	坡婆/波	皮/坡婆/波/被
234	(亥ㄏㄞˋ)孩該咳刻		咳刻	孩該咳刻	孩該/咳刻
234	(艮ㄍㄣˋㄍㄣˇ)跟很銀眼退腿		退腿	跟很	跟很/銀眼/退腿
1234	(古ㄍㄨˇ)古姑故做個苦胡湖	胡湖	古姑故	古姑故/苦/胡湖	古姑故/做/個/苦/胡湖
1234	(巴ㄅㄚˉ)巴吧把爸爬肥	巴吧	巴吧把爸	巴吧把爸爬	巴吧把爸爬/肥
1234	(青ㄑㄧㄥˉ)青清晴情請精睛猜	青清/晴情/精睛	青清/晴情/請	青清/晴情/精睛	青清/晴情/請/精睛/猜

REFERENCES

References marked with an asterisk indicate textbooks included in this study.

Andrews, S. (1999). Why do L2 teachers need to 'know about language'? Teacher metalinguistic awareness and input for learning. *Language and Education, 13* (3), 161-177.

Andrews, S., & McNeil, A. (2005). Knowledge about language and the 'good language teacher'. In N. Bartels (Ed.), *Researching* a*pplied linguistic in language teacher education* (pp.221-234). New York: Springer.

*Beijing Language Institute (1995). *Practical Chinese Reader: Componentary Course* (Traditional Character Edition, Book one). 實用漢語課本. MA: Cheng & Tsui Company.

Bernhardt, E. B. (1986). Reading in the foreign language. In B. H. Wing (Ed.), *Listening, reading, and writing: analysis and application* (pp.93-115). Northeast Conference on the Teaching of Foreign Language. Middlebury, VT.

Brisbois, J. E. (1995). Connections between first and second-language reading. *Journal of Reading Behavior*, 27, 4.

Chen, G. H. 陳國弘 (Ed.). (1987). *New Chinese Dictionary* 最新國語新辭典. Taiwan: Jian Li. (In Chinese).

Chen, L., & Nunes, T. (1998). Children's understanding of the formal and functional characteristics of written Chinese. *Applied psycholinguistics*, 19, 115-131.

Chan, L., & Wang, L. (2003). Linguistic awareness in learning to read Chinese: a comparative study of Beijing and Hong Kong Children. In C. McBride-Chang and H. Chen (Eds.), *Reading development in Chinese children* (pp.51-71). London: Praeger.

Chen, L. (1997). *Chángyòng Hànzì zhòng de yǒuxiào biǎoyīn zìfú jí qí xiéshēng zìbiǎo* 常用漢字中的有效表音字符及其諧聲字表. *Yuwen Jianshe, 3*, 32-33.

Cheung, H. C. & Ng, L. K. H. (2003). Chinese reading development in some major Chinese societies: an introduction. In C. McBride-Chang and H. Chen (Eds.), *Reading development in Chinese children* (pp.3-17). London: Praeger.

Coltheart, M. Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,*589-608.

Cui, Y. 崔永华 (1997). 汉字部件和对外汉字教学 [*Hànzìbùjiàn hé duìwà iHànzì jiàoxué*], *语言文字应用* [*Yǔyán wénzì yìngyòng*], 3, 49-62 (In Chinese).

Cummins, J. (1981). Bilingualism and minority-language children: Language and literacy series. Toronto: Ontario Institute for Studies in Education. (ERIC Document Reproduction Service No. ED 215 557).

*DeFrancis, J. (1976). *Character Text for Beginning Chinese* (2nd ed.) 初級漢語課本漢字本. New Haven and London: Yale University Press.

*DeFrancis, J. (1977). *Beginning Chinese Reader* (2nd ed.). New Haven and London: Yale University Press.

Everson, M. (1986). The effect of word-unit spacing upon the reading strategies of native and non-native readers of Chinese: An eye-tracking study. Unpublished doctoral dissertation. Ohio State University, Columbus, OH.

Everson, M. E. (1988). Speed and comprehension in reading Chinese: Romanization vs. characters revisited. *Journal of Chinese Language Teachers Association*. 23(2), 1-16.

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, 194-204.

Everson, M. E. (2007). Developing orthographic awareness among CFL learners: what the research tells us. In A. Guder, X. Jiang, & Y. Wan (Eds.), *The cognition, learning and teaching of Chinese characters* (pp.33-50). China: Beijing Language & Culture University Press.

Everson, M. E., & Fan, H-M. M. (2008). *Which characters do beginning learners typically see?* Paper presented at the First International Symposium on Chinese Applied Linguistics, Iowa City, IA.

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.

Fan, H. Y. (1986). *Some discussion on the orthography of modern Chinese characters*. Paper presented at the National Conference of Chinese Character Research, China (In Chinese.)

Fei, J 费锦昌 (1996). 现代汉字部件探究 [*Xiàndài Hànzì bùjiàn tànjiū*], *语言文字应用* [*Yǔyán wénzì yìngyòng*], 2, 20-26 (In Chinese).

Feldman, L. B., & Siok, W. W. (1997). The role of component function in visual recognition of Chinese characters. *Journal of experimental psychology, 23*(3), 776-781.

Feldman, L. B. & Siok, W. W. (1999). Semantic radicals in phonetic compounds: implications for visual character recognition in Chinese. In J. Wang & A. W.

Inhoff, & H. C. Chen (Eds.), *Reading Chinese script: a cognitive analysis* (pp. 19-36). Mawah, NJ: Lawrence Erlbaum Associate.

Feng, Z. 冯志伟 (1990). 评 «汉字属性字典»[Review Hɑ̀nzì shǔxìng zìdiɑ̌n]. *語文建設[Yu wen jian she], 2*, 54-56 (In Chinese).

Fu, Y. 傅永和 (ed.) (1989). *Hànzì shǔxìng zìdiǎn* 汉字属性字典. Beijing: Yuwen Chubanshe (In Chinese).

Fu, Y. 傅永和 (1991). 汉字的结构 [Chinese character structures]. *语文建設* [*Yu wen jian she*], 9, 10-11 (In Chinese).

Fu, Y. 傅永和 (1992). 汉字部件出现的结构部位. *语言文字应用* [*Yǔyán wénzì yìngyòng*], *2*, 7-21 (In Chinese).

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 674-691.

Goodman, K. S. (1968). The psycholinguistic nature of the reading process. In K. Goodman (Ed.), *The psycholinguistic nature of reading process*. Detroit, MI: Wayne State University Press.

Gough, P. (1972). One second of reading. In J. F. Kavanaugh & I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, MA: MIT Press.

Gough, P.B., Juel, C., & Griffith, P. L. (1992). Reading, spelling, and orthographic cipher. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), Reading acquisition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Guder, A. (2007). The Chinese writing system as third dimension of foreign language learning. In A. Guder, X. Jiang, & Y. Wan (Eds.), The cognition, learning and teaching of Chinese characters (pp.33-50). China: Beijing Language & Culture University Press.

Guder-Manitius, A. (1999). Singographemdidaktik.: Aspekte einer systematischen vermittlung der chinesischen schrift im unterricht Chinesisch als fremdsprache: mit einer komonentenanalyse der hɑ̈ufigsten 3867 schriftzeichen. Heidelberg: Groos.

Harbaugh, R. (1998).中文字譜. New Haven, CT: Yale University Press and Taipei, Taiwan: Han Lu Book & Publishing Co.

Hatch, E., Brown, C. (1995). *Vocabulary, semantics, and language education*. New York: Cambridge University Press.

Henderson, L. (Ed.) (1984). *Orthographies and reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ho, C. S-H, Bryant, P. (1997). Learning to read Chinese beyond the logographic phase. *Reading research quarterly, 32*(3), 276-289.

Ho, C. S-H, Wong, W-L, & Chan, W-S. (1999). The use of orthographic analogies in learning to read Chinese. *Journal of child psychology and psychiatry, 40*(3), 393-403.

Ho, C. S-H., Yau, P. W-Y., & Au, A. (2003). Development of orthographic knowledge and its relationship with reading and spelling among Chinese kindergarten and primary school children. In C. McBride-Chang and H. Chen (Eds.), *Reading development in Chinese children* (pp.51-71). London: Praeger.

Jackson, N. E., Everson, M. E., & Ke, C. (2003). Beginning readers' awareness of the orthographic structure of semantic-phonetic compounds: lessons from a study of learners of Chinese as a foreign language. In C. McBride-Chang and H. Chen (Eds.), *Reading development in Chinese children* (pp.3-17). London: Praeger.

Kane, D. (2006), The Chinese language: its history and current usage. Tuttle Publishing.

Kang, J. 康加深 (1993). 现代汉语形声字形符研究 [*Xiàndài Hànyǔxíngshēngzì xíng fúyánjiū*], In Chen Y. 陈原 (Ed.), *现代汉语用字信息分析* [*Xiàndài Hànyǔyòngzì xìnxī fēnxi*] (pp. 68-83), Shanghai, China: Shanghai Education Press.

Ke, C. (1996). *A model for Chinese orthographic awareness.* Unpublished manuscript.

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.

Ke, C., & Everson, E. E. (1999). Recent research in CFL reading and its pedeagogical implications. In M. Chu (Ed.), *Mapping the courese of the Chinese language field* (pp. 189-203). Kalamazoo, Michigan: Chinese Language Teachers Association.

Ke, C., Wen, X., & Kotenbeutel, C. (2001). Report on the 2000 CLTA articulation project. *Journal of the Chinese Language Teachers Association*, 36 (3), 25-60.

Ko, H., & Wu, C. F. (2003). The role of character components in reading Chiense. In C. McBride-Chang and H. Chen (Eds.), *Reading development in Chinese children* (pp.3-17). London: Praeger.

Koda, K. (1997). Orthographic knowledge in L2 lexical processing: a cross-linguistic perspective. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 35-52). New York: Cambridge University Press.

Koda, K. (2005). *Insights into second language reading: a cross-linguistic approach*. New York: Cambridge University Press.

Kupfer, P. (2007). Cloquent but blind-The problem of reading proficiency in Chinese as a foreign language. In A. Guder, X. Jiang, & Y. Wan (Eds.), The cognition, learning and teaching of Chinese characters (pp.1-16). China: Beijing Language & Culture University Press.

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293-323.

Leck, K. J., Weekes, B. S., Chen, M. J. (1995). Visual and phonological pathways to the lexicon: evidence from Chinese readers. *Memory & Cognitio. 23*(4), 468-476.

Li, H., & Chen, H-C. (1999). Radical processing in Chinese character recognition: evidence from lexical decision. *Psychologia, 42*, 199-208.

Li, J., Fu, X., & Lin, X. (2000). Study on the development of Chinese orthographic regularity in school children. *Xin li xue bao* 心理学报 *, 32*(2), 121-126.

Li, Y., & Kang, J. 李燕 & 康加深 (1993). 现代汉语形声字声符研究 [], In Chen Y. 陈原 (Ed.), *现代汉语用字信息分析* [*Xiàndài Hànyǔ yòngzì xìnxī fēnxi*] (pp. 84-98), Shanghai, China: Shanghai Education Press.

*Liu, Y., Yao, T.-C., Bi, N.-P., Ge, L., Shi, Y. (2008). *Integrated Chinese* (Traditional &Simplified, 3rd Edition, Level 1, Part 1&2) 中文聽說讀寫. MA: Cheng & Tsui Company Publisher & Distribution

Lu, T-H. (2003). *The relationship of phonological processing (phonological awareness, verbal short-term memory, and rapid naming) and visual short-term memory to reading disabilities in Chinese children*. Unpublished doctoral dissertation, University of Iowa, Iowa.

Marcel, A. J. (1980). Surface dyslexia and beginning reading: A revised hypothesis of the pronunciation of print and its impairments. In M. Coltheart, K. Patterson, & J. C. Marchsall (Eds.), *Deep dyslexia* (pp. 227-258). London: Routledge & Kegan Paul.

McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press

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

*National Taiwan Normal University. (2008). *Practical Audio-Visual Chinese* (2nd ed.). 新版實用視聽華語.Taiwan: Cheng Chung Book.

*Ning, C. (1994). *Communicating in Chinese: Reading and Writing* 漢語交流: 初級漢語閱讀書寫本. New Haven: Yale University Press

O'Dell, F. (1997). Incorporating vocabulary into the syllabus. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press.

Perfetti, C. A. & Tan, Li. H. (1999). The constituency model of Chinese word identification. In J. Wang & A. W. Inhoff, & H. C. Chen (Eds.), *Reading Chinese script: a cognitive analysis* (pp. 115-134). Mawah, NJ: Lawrence Erlbaum Associate.

中華人民共和國教育部國家語言文字工作委員會 [P.R.C. Department of Education National Linguistic Work Council]. (2009). *現代常用字部件及部件名稱規範 [Specification of common modern Chinese character components and component names]* (GF 0014-2009). China: 語文出版社 [Linguistic Press].

Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge: Cambridge University Press.

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

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

Shen, H. H. (2004). Level of cognitive processing: effects on character learning among non-native learners of Chinese as a foreign language. *Language and Education, 18*(2), 167-182.

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

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. 4). (pp. 45-66). Honolulu: University of Hawai'i, National Foreign Language Resource Center.

Shen, H. H., & Bear, D. R. (2000). Development of orthographic skills in Chinese children. *Reading and Writing: An Interdisciplinary Journal, 13*, 197-236.

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

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

Solomon, R. L., & Howes, D. H. (1951). Word frequency, personal values, and visual duration thresholds. *Psychological review, 58*, 256-270.

Stahl, N.A., Brozo, W.G., & Simpson, M.L. (1986). *Developing college vocabulary: a content analysis of instructional materials* (Tech. Rep. No. 86-07). Atlanta, Georgia: Georiga State University, College Reading and Learning Assistance.

Sternberg, R.J., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist 38*, 873-893.

Su, P. 苏培成 (1995). 现代汉字的部件切分 [Xiàndài Hànzì de bùjiàn qiē fēn],*语言文字应用* [*Yǔyán wénzì yìngyòng*], 3, 52-55 (In Chinese).

Taft, M., & Zhu, X. (1997). Submorphemic processing in reading Chinese. *Jouranl of Experimental Psychology: Learning, Memory, and Cognition, 23*, 761-775.

Taylor, I., & Taylor, M. M. (1995). *Writing and literacy in Chinese, Korean, and Japanese*. Amsterdam: John Benjamins Publishing.

Tomayo, J.M. (1987). Frequency of use as a measure of word difficulty in bilingual vocabulary test construction and translation. *Educational and Psychological Measurement, 47*, 893-902.

*Wang, F.-Y. 王方宇 (1961). *Read Chinese: A Beginning Text in the Chinese Character* (Expanded Edition, Book One) 華文讀本. New Haven: Yale University Press.

Wenlin Institute (2007). Wenlin software for learning Chinese 文林 (Version 3.4). [Computer software]. Eureka, CA: Wenlin Institute.

Xiao, D. 晓东 (1993). 也谈部件定义[*Yě tán bùjiàn dìngyì*].*語文建設[Yu wen jian she],* 6, 21-22.

Xiao, D. 晓东 (1994). 现代汉字独体与合体的再认识 [*Xiàndài Hànzì dútǐ yǔ héxiū de zài rènshi*],*語文建設[Yu wen jian she]*, 8, 28-31 (In Chinese).

Xiao, D. 晓东 (1995). 现代汉字部件分析的规范化 [*Xiàndài Hànzì bǔjiàn fēnxi de gu īfànhuà*],*语言文字应用* [*Yǔyán wénzì yìngyòng*], 3, 56-59 (In Chinese).

Xiao, R., Rayson, P., & McEnery, T. (2009). *A frequency dictionary of Mandarin Chinese: core vocabulary for learners*. NY: Routledge.

Xiao, Y. (2005). Raising the orthographic awareness of teachers of Chinese. In N. Bartels (Ed.), *Researching* a*pplied linguistic in language teacher education* (pp.221-234). New York: Springer.

*Yao, D. (2005). *Integrated Chinese*. (Traditional, 2nd Edition, Level 1, Part 1&2) 中文聽說讀寫. MA: Cheng & Tsui Company.

*****Yeh, T,-H. (Ed.). (2008). *Far East Everyday Chinese: Book I* 遠東生活華語. Taiwan: The Far East Book.

Yin, W. G., & Butterworth, B. (1992). Deep and surface dyslexia. In H.-C. Chen and O.J.L. Tzeng (Eds.), *Language processing in Chinese* (pp. 349-366). Amsterdam: Elsevier Science.

*Zhang, K., Liu, S., Chen, X., Zuo, S. , Shi, J., Liu, X. (2008). *New Practical Chinese Reader 1* 新實用漢語課本 1. China: Beijing Language & Culture Univ. Press/ Cheng & Tsui Company Publisher & Distribution

Zhu, Y. 朱一之(1993). 漢語水平考試: 漢字大綱所 2905 個漢字分析結果 Hanyu shuiping kaoshi-Hanzi dagang suoshou 2905 ge hanzi fenxi jieguo [The Chinese character analysis result of 2905 characters in HSK level]. Yu yan jiao xue yu yan jiu [language teaching and studies], 3, 98-111. (In Chinese.)

Zhu, Y. P. (1987). *Analysis of cueing function s of the phonetic in modern China*. Unpublished paper, East China Normal University. (In Chinese.)