

LEARNING EFFICIENCIES FOR DIFFERENT ORTHOGRAPHIES:
A COMPARATIVE STUDY OF HAN CHARACTERS
AND VIETNAMESE ROMANIZATION

by

WI-VUN TAIFFALO CHIUNG

Presented to the faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2003

Copyright © by Wi-vun Taiffalo Chiung 2003

All Rights Reserved

ACKNOWLEDGMENTS

I appreciate the many contributions that people have made for this dissertation. First of all, I would like to express my deepest appreciation to my committee members, Dr. Jerold Edmondson, Dr. David Silva, Dr. Jinny Choi, Dr. Mary Morgan, and Dr. Richard Watson for their helpful guidance and insightful comments on this research.

Special thanks to the students who were involved in the experiments. This study would not have been carried out without their participation. Thanks also goes to their teachers, such as Hàn-chhoan Lí 李漢村, Dr. Chèng-hong Si 施正鋒, cô Thảo, Dr. Phùng Thị Huệ, and Dr. Nguyễn Văn Hồng, who kindly offered me their classes for conducting the survey.

I am indebted to all my friends who offered me big favors while I was collecting data in Taiwan and Vietnam. They are: Phoè-hun Chng 莊佩芬, Bèng-téng Chhòa 蔡明庭, Lē-koan Tân 陳麗娟, Kiat-tông Tēⁿ 典仔, Kok-cheng Ngô 吳國禎 and Tek-hòa Tân 陳德樺 in Taiwan; Dr. Đoàn, Thiên Thuật, Dr. Lý Toàn Thắng, Dr. Nguyễn Quang Hồng, Tú Sót, Thế Anh, Hoàng Tiến, Nguyễn Thanh Xuân, Nguyễn Thị Thu Phương, Nguyễn Nhật Nam, Nguyễn Thị Mai Lan, Nguyễn Hoàng Yén, Cao Lê Anh, Nguyễn Công Hoàng, Nguyễn Thu Hà, Nguyễn Thu Thủy, Ngô Thị Thanh Dung, Nguyễn Thanh Nga, Nguyễn Thị Bích Ngọc, Vũ Thanh Tùng, Hoàng Minh Thắng, Nguyễn Thanh Nhạn, Hoàng Kim Loan, Nguyễn Kim Thanh, Nguyễn Thanh Lâm, and Trần Hoàng Khánh Ly in Vietnam. I am also thankful for the following friends' assistance while I was reviewing and digitizing dictation tests of the Taiwanese subjects: Sunny Chiam 詹惠雯, Gi-téng Iúⁿ 楊貽婷, Anny Ho 何韻如, Jenny Hong 洪欣玉, Fanny Hong 洪錦芳, Sarah Yao 姚崇閔, Vera Yao 姚崇瑤, Chun-Shih Han 韓君實, Joy Lin 林蒼汝, Jessica Kuo 郭立潔, Anita Lin 林星吟 and Sammi

Lin 林亭慧. In addition, I do appreciate Jeff Barnett, Ruel Macaraeg, and Ian Connally for their kind proofreading of my dissertation. I wish to express my gratitude to Dr. Pete Unseth and his librarian wife Carole, who often offered me helpful information for my research. My gratitude also goes to the local Kongsì members, who are the most Taiwan-loving people I have ever met in the USA.

Fundings from the Taiwanese-American Foundation of Boston, Dr. Li Yan-Yen Foundation, and Dr. Chen Wen Chen Memorial Foundation are helpful and I am grateful to their support.

Finally, I owe all my achievements to my parents Giók-ki Sun 孫玉枝 and Hoat-thài Chiúⁿ 蔣發太, and my brothers Êng-khim Chiúⁿ 蔣榮欽, Êng-hok Chiúⁿ 蔣榮福, Êng-hong Chiúⁿ 蔣榮豐, and Ûi-chì Chiúⁿ 蔣爲志. Without their support, I would not have been able to complete this dissertation.

April 4, 2003

ABSTRACT

LEARNING EFFICIENCIES FOR DIFFERENT ORTHOGRAPHIES: A COMPARATIVE STUDY OF HAN CHARACTERS AND VIETNAMESE ROMANIZATION

Publication No. _____

Wi-vun Taiffalo Chiung, Ph.D.

The University of Texas at Arlington, 2003

Supervising Professor: Jerold A. Edmondson

In order to address the question of whether or not to abandon Han characters (Hanji), it is important to evaluate empirically the efficiency of Han writing. The purpose of this study is to compare the efficiency of learning to read and write in Hanji versus learning to read and write in phonemic writing systems, such as Vietnamese Chu Quoc Ngu (CQN) or Mandarin Bopomo.

Three experiments were conducted in this study. The first experiment focused on a study of reading comprehension; the second one focused on a study of accuracy of writing dictation; and the last was a study of oral reading. A total of 453 subjects from Taiwan and 350 subjects from Vietnam were involved in the experiments. Subjects consisted of elementary school and college students.

The reading comprehension tests were divided into groups Hanji, Bopomo, and CQN, in which subjects were examined with reading texts in Hanji, Bopomo, and CQN, respectively. The results of the reading comprehension tests reveal no statistically significant difference between Hanji and CQN groups. However, students from the second

to fifth grades in the Bopomo group had significantly lower scores than students in the other groups.

In dictation tests, subjects were divided into groups Taiwanese and Vietnamese. Tests in each group were given in soft and hard articles. The statistical results of tests on soft article reveal that students in both Taiwanese and Vietnamese groups significantly increased their score each year until the fourth grade, by which time they had the same statistical score as college students. As for tests on hard article, Taiwanese students spent more years in the acquisition of Hanji, and even the sixth graders' scores do not statistically reach the same level as college students. However, Vietnamese students had reached a college level at the fifth grade. Errors in the dictation tests were also analyzed, and twelve error types were found in the Taiwanese group. The major errors were made due to similarity in sound between correct and incorrect Han characters. The phonetic similarity errors account for 85.70% in the dictation test two.

In addition to dictation tests, CQN also showed superiority in oral reading tests. The results indicate that CQN beginners are able to produce about 90% accuracy in oral reading after three or four months of learning, and reach nearly 100% accuracy a year later.

In short, these results lead to the conclusion that Vietnamese CQN is more efficient than Chinese characters in learning to read and write.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
ABSTRACT	vi
LIST OF FIGURES	xiv
LIST OF TABLES	xvi
Chapter	
1. INTRODUCTION	1
1.1 Motivation and purpose of the study	1
1.2 Research questions and hypotheses	3
1.3 Overview of the study	5
2. LANGUAGE, LITERACY AND NATIONALISM IN HAN SPHERE	6
2.1 Language planning and nationalism	7
2.2 Diglossia and digraphia	8
2.2.1 Digraphia with biliteracy	11
2.2.2 Digraphia without biliteracy	13
2.3 Socio-cultural background in the Han sphere	14
2.4 Orthographic reform in Korea	17
2.5 Orthographic reform in Japan	20
2.6 Orthographic reform in Vietnam	25
2.6.1 Colonial background	25
2.6.2 Orthographic tradition	29

2.6.3	Development of Romanization	32
2.7	Orthographic reform in Taiwan	39
2.7.1	Colonial background	40
2.7.2	Orthographic tradition	43
2.7.3	Romanization prior to 1945	45
2.7.3.1	Sinkang Romanization (1624-early nineteenth century)	45
2.7.3.2	Peh-oe-ji Romanization (1865-present)	47
2.7.4	Romanization after 1945	51
2.7.4.1	Romanization for Mandarin Chinese	52
2.7.4.2	Romanization for Taiwanese	58
2.7.5	Contemporary language movement	61
2.7.6	The future of Romanization in Taiwan	66
3.	WRITING SYSTEMS AND LEARNING EFFICIENCIES	70
3.1	Development of writing systems	70
3.1.1	Criteria of classification	70
3.1.2	Evolution of writing	73
3.1.3	New proposed criteria for the typology of writing	76
3.1.4	Orthography and Learning efficiency	79
3.1.5	Criteria for a new orthography	81
3.1.6	Psycholinguistic view of reading	82
3.2	Han characters	83
3.2.1	Formation of Han characters	84
3.2.2	Evolution of Han characters	87

3.2.3	Influence of Han characters on word perception	95
3.3	Roman script	99
3.3.1	How Romanization works	99
4.	SOUND AND WRITING SYSTEMS IN TAIWANESE, CHINESE AND VIETNAMESE	101
4.1	Taiwanese	101
4.1.1	Sound system	101
4.1.2	Written form in Romanized Peh-oe-ji	105
4.1.3	Current Taiwanese language education in Taiwan	111
4.2	Chinese	112
4.2.1	Sound system	112
4.2.2	Written form in Romanization	114
4.2.3	Written form in Bopomo	118
4.2.4	Current Chinese language education in Taiwan	121
4.3	Vietnamese	123
4.3.1	Sound system	124
4.3.1.1	Consonants	125
4.3.1.2	Vowels	126
4.3.1.3	Tones	127
4.3.2	Written form in Romanized Chu Quoc Ngu	131
4.3.3	Current Vietnamese language education in Vietnam	139
5.	METHODOLOGY	141
5.1	Reading comprehension tests	141
5.1.1	Research design	141
5.1.2	Selection of reading texts	142

5.1.3	Selection of subjects	144
5.1.4	Conducting procedure	146
5.1.5	Data analysis	147
5.2	Dictation tests	149
5.2.1	Research design	150
5.2.2	Preparation for passages	150
5.2.3	Selection of subjects	153
5.2.4	Procedure	153
5.2.5	Data analysis	154
5.3	Oral reading tests	154
5.3.1	Research design	155
5.3.2	Selection of oral reading texts	155
5.3.3	Selection of subjects	155
5.3.4	Conducting procedure	155
5.3.5	Data analysis	156
6.	RESULTS AND DISCUSSION	157
6.1	Results of reading comprehension tests	157
6.1.1	Hanji group	157
6.1.2	Bopomo group	160
6.1.3	Chu Quoc Ngu group	163
6.1.4	Three scripts in comparison	166
6.2	Results of dictation tests in Taiwanese group	173
6.2.1	Analysis on Hanji in dictation one	173
6.2.2	Analysis on Bopomo in dictation one	180
6.2.3	Analysis on Hanji in dictation two	184

6.2.4	Analysis on Bopomo in dictation two	191
6.3	Results of dictation tests in Vietnamese group	195
6.3.1	Analysis on CQN in dictation one	195
6.3.2	Analysis on CQN in dictation two	205
6.4	Results of oral reading tests in Vietnamese group	215
6.4.1	Analysis on CQN in oral reading one	215
6.4.2	Analysis on CQN in oral reading two	221
7.	CONCLUSIONS	228
7.1	Summary of the findings	228
7.2	Conclusions and implications	232
7.3	Recommendation for further studies	237

Appendix

A.	VIETNAMESE PHONEMES AND THEIR CORRESPONDING CHU QUOC NGU.....	240
B.	ORIGINAL TEXTS FOR READING COMPREHENSION TESTS IN HANJI GROUP	246
C.	ORIGINAL TEXTS FOR READING COMPREHENSION TESTS IN BOPOMO GROUP	255
D.	ORIGINAL TEXTS FOR READING COMPREHENSION TESTS IN VIETNAMESE GROUP	265
E.	ORIGINAL TEXTS FOR DICTATION TESTS IN TAIWANESE GROUP	277
F.	ORIGINAL TEXTS FOR DICTATION TESTS IN VIETNAMESE GROUP	279
G.	ORIGINAL TEXTS FOR ORAL READING TESTS IN VIETNAMESE GROUP	281

REFERENCES 283
BIOGRAPHICAL INFORMATION 302

LIST OF FIGURES

Figure	Page
1. Example of political cartoon in Taiwanese	14
2. Relationship among Chinese nationalism-nationism, Taiwanese nationalism, and Taiwanese nationism	64
3. Descriptive-representational picture	74
4. Identifying-mnemonic device	75
5. Singlish example of ‘sea’ and ‘see’	86
6. Each Han character has only one syllable	97
7. Each syllable was misperceived as having meaning and character	97
8. Process of lexical Sinification	98
9. Example of two-dimensional placement by Roman alphabet	100
10. Tone Circle in Taiwanese	104
11. Samples of Han characters with Bopomo	120
12. Scores received by grades in Hanji group	160
13. Scores received by grades in Bopomo group	162
14. Scores received by grades in CQN group	165
15. Mean scores received by grades and scripts	167
16. Percentage of correct Hanji in dictation one	175
17. Percentage of correct Bopomo in dictation one	182
18. Percentage of correct Hanji in dictation two	186
19. Percentage of correct Hanji in soft and hard articles	187
20. Percentage of correct Bopomo in dictation two	192

21. Percentage of Bopomo used in soft and hard articles	193
22. Percentage of correct CQN segments in dictation one	197
23. Percentage of correct CQN segments in dictation two	206
24. Percentage of correct CQN in soft and hard articles	207
25. Percentage of correct Hanji in soft and hard articles	208
26. Mean scores received by grades and scripts	229
27. Mean score of Taiwanese students in dictation tests	230
28. Mean score of Vietnamese students in dictation tests	231

LIST OF TABLES

Table	Page
1. The relationship between bilingualism and diglossia	9
2. The relationship between digraphia and biliteracy	10
3. Double-nested digraphia in Taiwan	11
4. Relation between languages and political status in Vietnam	25
5. Numbers of literate in Vietnam	37
6. Relation between language and political status in Taiwan	40
7. Mandarin consonants represented by IPA, HP, TYP, and Jhuyin Fuhao	58
8. Number of each category of non-Han schemes	59
9. English example of Universal Orthography	79
10. Hanji example of Universal Orthography	87
11. Evolutional steps and representational levels of Han writing system	88
12. Percentage of Han characters by different principles and periods	91
13. Example of ‘punji’	96
14. Taiwanese consonants in IPA	102
15. Taiwanese vowels in IPA	102
16. Tonal categories in Taiwanese (Cheng 1997)	102
17. Examples of spelling variations among Medhurst, Douglas, and Campbell	106
18. Symbols for Taiwanese consonants in the spelling of Peh-oe-ji	107
19. Symbols for Taiwanese vowels in the spelling of Peh-oe-ji	107
20. Mandarin consonants in IPA	113
21. Mandarin vowels	113

22. Tonal categories in Mandarin	114
23. Symbols for Mandarin consonants in the spelling of Pinyin	116
24. Symbols for Mandarin vowels in the spelling of Pinyin	117
25. Bopomo with equivalent Pinyin and IPA	119
26. Bopomo example of Universal Orthography	121
27. Periods for learning Mandarin in elementary school	122
28. Percentage of colloquial and classical writings in high school textbooks	123
29. Vietnamese consonants in IPA	125
30. Vietnamese simple vowels in IPA	126
31. Vietnamese short vowels in IPA	126
32. Vietnamese diphthongs in IPA	127
33. Traditional categories of Vietnamese tones	128
34. Vietnamese tone system (Thompson 1987)	128
35. Vietnamese example of Universal Orthography	132
36. Symbols for Vietnamese consonants in the spelling of Chu Quoc Ngu	135
37. Vietnamese vowels in CQN	136
38. Vietnamese tone marks in CQN	137
39. Periods for learning Vietnamese in elementary school	139
40. Number of frequently used characters	144
41. Subjects from Taiwan	145
42. Subjects from Vietnam	146
43. Characters contained in dictation one for Taiwanese group	151
44. Characters contained in dictation two for Taiwanese group	152
45. Scores received by students in Hanji group	158
46. Homogeneous subsets by grades in Hanji group	159

47.	Time spent on reading Hanji texts by collegians	160
48.	Scores received by students in Bopomo group	161
49.	Homogeneous subsets by grades in Bopomo group	161
50.	Time spent on reading Bopomo texts by collegians	162
51.	Scores received by students in CQN group	163
52.	Homogeneous subsets by grades in CQN group	164
53.	Time spent on reading CQN texts by collegians	165
54.	Mean scores received by grades and scripts	166
55.	Homogeneous subsets by scripts (2 nd to 5 th graders)	169
56.	Homogeneous subsets by scripts among literacy beginners	169
57.	Homogeneous subsets by scripts among collegians	171
58.	The rates in reading text-based Hanji and CQN by collegians	172
59.	Correct Han characters in dictation one	174
60.	Homogeneous subsets of correct Hanji by grades in dictation one	175
61.	Incorrect Han characters in dictation one	176
62.	Original error types of Hanji in dictation one	178
63.	Factor loadings of twelve error categories	179
64.	Basic error types of Hanji in dictation one	180
65.	Correct Bopomo in dictation one	181
66.	Homogeneous subsets of correct Bopomo by grades in dictation one	182
67.	Original error types of Bopomo in dictation one	183
68.	Factor loadings of eight error types	184
69.	Correct Han characters in dictation two	185
70.	Homogeneous subsets of correct Hanji by grades in dictation two	186
71.	Incorrect Han characters in dictation two	188

72. Original error types of Hanji in dictation two	189
73. Factor loadings of ten error categories	190
74. Basic error types of Hanji in dictation two	190
75. Correct Bopomo in dictation two	191
76. Homogeneous subjects of correct Bopomo by grades in dictation two	192
77. Error types of Bopomo in dictation two	194
78. Correct segments of CQN in dictation one	196
79. Homogeneous subset of correct CQN segments in dictation one	197
80. Error analysis on CQN by all grades in dictation one	199
81. Sorted probability of error types among all grades in dictation one	201
82. Sorted probability of error types among first graders in dictation one	203
83. Correct segments of CQN in dictation two	205
84. Homogeneous subset of correct CQN segments in dictation two	206
85. Error analysis on CQN by all grades in dictation two	209
86. Sorted probability of error types among all grades in dictation two	211
87. Sorted probability of error types among first graders in dictation two	213
88. Correct segments of CQN in oral reading one	216
89. Homogeneous subsets of correct CQN segments in oral reading one	216
90. Time spent by subjects on oral reading one	217
91. Time spent by collegians on oral reading one	217
92. Error analysis on CQN in oral reading one	218
93. Sorted probability of error types in oral reading one	220
94. Correct segments of CQN in oral reading two	222
95. Homogeneous subsets of correct CQN segments in oral reading two	222
96. Time spent by subjects on oral reading two	223

97. Error analysis on CQN in oral reading two	224
98. Sorted probability of error types in oral reading two	226
99. Orthographic types by placement and correspondence	239

CHAPTER 1

INTRODUCTION

The fundamental purpose of this study is to compare the efficiency of learning to read and write in Han characters (Hanji 漢字) versus learning to read and write in an alphabetic writing system. As representative of alphabet writing, Vietnamese Chu Quoc Ngu and Mandarin Bopomo were chosen as the contrastive orthographies in comparison to Han characters. This chapter is divided into three sections. Section 1.1 lays out the motivation and purpose of the study. Section 1.2 describes the research questions and hypotheses, and section 1.3 provides an overview of the study.

1.1 Motivation and purpose of the study

Debates on the standardization of national language, and on the use of Han characters in the Han cultural sphere have been going on for more than a hundred years (Chen 1999; Hannas 1997). The citizenry of Vietnam and Korea eventually shifted from the use of Hanji to phonemic writing of their own, i.e., Romanized *Chữ Quốc Ngữ*¹ in Vietnam, and *Hangul*² 한글 in Korea. However, writing in Hanji is still the parallel dominant writing system in Taiwan and China though the two places use traditional and simplified characters, respectively (Chen 1990, 1989; Chiung 2001; Norman 1988; DeFrancis 1950). In Taiwan and China, many people have refused to abandon Hanji because they regard Han characters

¹ *Chữ Quốc Ngữ* literally means 國語字 or the National Orthography. It was derived from missionary orthography in the seventeenth century, and is currently the official Romanized writing system for the Vietnamese language (DeFrancis 1977).

² *Hangul* 한글, the Korean alphabet was originally invented by King Sejong in the 15th century. *Hangul* is a pure Korean word, thus it does not have corresponding Han characters (Lee 1957). In Taiwan, the Han characters 諺文 are usually used to refer to *Hangul*.

as the best suited orthography for the Chinese spoken language (DeFrancis 1990). But others have argued that the high number of Han characters constitute a burden on the learner, and may cause further hindrance to national modernization. For example, Chen (1994:367) points Han characters are largely “responsible for the country’s high illiteracy and low efficiency, and hence an impediment to the process of modernization.” Although the controversy about the continued use or abandonment of Hanji has been going on over a century, little comparative research has been done with regard to the learning efficiencies of Hanji and other phonemic writing systems. Lack of true empirical research reveals that the source for people’s insistence on maintaining or abandoning Han characters is merely based in their socialization or ideological preference rather than on a scientific study of the alternative orthographies.

In order to address the question of whether or not to abandon Han characters, it is important to evaluate the efficiency of Han writing empirically. While there are several aspects and different approaches with regard to such a study, this dissertation compares the efficiency of learning to read and write in Hanji versus learning to read and write in alphabetic writing systems, such as Vietnamese Chu Quoc Ngu and Mandarin Bopomo. More specifically, the primary concern of this comparative study is to determine how long it takes for a student to be able develop an ability to read and write textual material of general level³ written in Han characters, Chu Quoc Ngu and Bopomo. It is assumed that the more school years needed by students to achieve a certain level of reading and writing capacity, the more difficult the writing system is to master.

³ The term “general level” is defined as articles appearing in the selected popular newspapers in this investigation.

1.2 Research questions and hypotheses

Because the terms “efficient” or “inefficient” will be significant only if we compare Hanji to other writing systems, we need to include other orthographies in the comparison. Theoretically, we should have contrastive groups, such as Hanji groups vs. Roman groups. Subjects in these two groups would to have identical backgrounds (e.g. mother tongue, intelligence quotient, educational level, and social class) differing only by having been equally educated in different writing systems, i.e., Hanji in the Hanji groups, and Roman script in the Roman groups. However, in practice, it would be extremely difficult to find groups that meet these criteria as well as to find volunteers willing to join the experimental groups. Thus, we have to examine the learning efficiency in an indirect way. For this reason, Vietnamese Chu Quoc Ngu (CQN) was chosen in this study as a contrastive orthography. The main reasons for choosing Vietnamese are: 1) Vietnamese is typologically closer to Mandarin Chinese than other languages using Romanized writing systems; 2) Vietnamese though much of its written history has used Han characters, and it completely shifted to Romanization only in 1945. Nowadays, Romanized CQN is taught through Vietnam’s national education system; and 3) This researcher has easier access to get subjects and data from Vietnam than other countries.

Three experiments were conducted to examine learning efficiency. The first experiment focused on the study of reading comprehension; the second one focused on the study of accuracy of writing dictation; and the last was the study of oral reading. A total of 453 students from Taiwan and 350 students from Vietnam were involved in the experiments.

The first experiment, Reading Comprehension Tests (RCT), was designed to determine the average number of years of typical educational experience needed for students in Taiwan and Vietnam to be able to read and comprehend general newspapers in

their language using their standard orthographies, i.e., Mandarin (Hanji vs. Bopomo⁴) and Vietnamese (Chu Quoc Ngu), respectively. Why were students from Taiwan divided into groups Hanji and Bopomo? Mainly because Mandarin and Vietnamese are two different languages. Even if the results reveal that the Vietnamese Romanization is more efficient than the Han system, Romanization for Mandarin may not be as practicable or as efficient as it is in Vietnamese since 1) different languages have different phonological characteristics and they may need specially-adapted writing systems, and 2) Taiwan and Vietnam may have different backgrounds and resources for literacy education. Thus, the Bopomo group was proposed to solve this potential problem. Therefore, Taiwanese students were divided into two groups, i.e., Hanji and Bopomo. Students in the Hanji group were given reading comprehension tests in Hanji version, and students in the Bopomo group were tested in Bopomo version. The score students received are treated as an index of the efficiency of different writing systems. The research questions in this experiment are: 1) Do students achieve different scores in the tests? If so, what factors contribute to their different proficiencies? 2) Comparing Hanji to Bopomo and CQN, which one is learned faster? 3) Do students have different learning processes with regard to Hanji, Bopomo, and CQN?

The second experiment consisted of a dictation task administered to Taiwanese and Vietnamese students to examine the accuracy of writing with comparison to students' reading ability. The research questions in this experiment are: 1) Is performance in writing significantly different from reading? If so, what are the factors? 2) Is the performance of Taiwanese students (in Hanji) significantly different from Vietnamese (in CQN)? If so,

⁴ *Bopomo* or ㄅㄆㄇ注音符号, the National Phonetic Symbols, which is used as a supplementary tool to the learning of Mandarin. It is taught along with Han characters through the national education system in Taiwan.

what are the factors? 3) What are the easiest and most difficult parts of writing in Hanji and CQN? And how accurate they are? 4) What errors are the students more likely to made?

The third experiment included oral reading tests of school students in Vietnam. In this experiment, subjects read prepared texts aloud. The third experiment was found necessary to be added after the first two experiments were completed. The purpose of oral reading tests was to examine the accuracy of speaking with comparison to students' reading and writing abilities.

1.3 Overview of the study

This study is divided into seven chapters. The first chapter provides an overall introduction to this study. Chapters two, three, and four discuss various aspects of literacy. In chapter two, we examine the sociolinguistic context with regard to language, literacy and nationalism in the Han sphere. Chapter three re-clarifies traditional classification of writing systems, and examines the role different type of writing systems play in relation to learning efficiency. In chapter four, the writing schemes Taiwanese Peh-oe-ji, Mandarin Bopomo, and Vietnamese Chu Quoc Ngu are surveyed, so readers will have better understanding of the three writing systems involved in the experiments of this study. The methodology for quantitative experiments of this study is demonstrated in chapter five, and the experimental results and discussion are arranged in chapter six. Finally, conclusions and recommendations for further studies are provided in chapter seven.

CHAPTER 2

LANGUAGE, LITERACY AND NATIONALISM IN HAN SPHERE

The Han character sphere, including Vietnam, Korea, Japan, Taiwan and China, adopted Han characters and classical Han writing as the official written language long before the twentieth century. However, great changes transpired with the advent of the twentieth century. After World War II, Han characters in Vietnam and Korea were officially replaced by the Romanized *Chu Quoc Ngu* and phonemic Hangeul 한글, respectively. In Japan, the number of Han characters in use was decreased and the syllabary *Kana* system was promoted to national status. In Taiwan, although Romanization was developed centuries ago, Han characters remained the dominant orthography in current Taiwanese society. As for China, the simplification of Han characters seems the only result of China's efforts at reforming characters for over a century.

This chapter examines the orthographic transition within the Han sphere in terms of their linguistic and sociolinguistic features. It centers on two issues: 1) the development and trend of linguistic structure of the later devised scripts, and 2) the factors which contributed to the transition.

From the perspective of linguistics, an orthographic structure in evolution tends to represent smaller sound units, that is, from morphosyllabic to syllabic to phonemic writing, and from two dimensions to a single dimension.

Both internal and external factors have contributed to the different outcomes of orthographic reform in these countries. Internal factors include the general public's demand for literacy and protest against feudal social hierarchy; external factors include the political

relationships between these countries and the state that is origin of Han characters (i.e., China).

2.1 Language planning and nationalism

Language planning, as defined by Rubin (1971: xvi), “is deliberate language change; that is, changes in the systems of language code or speaking or both that are planned by organizations that are established for such purposes or given a mandate to fulfill such purposes.” Language planning is a complex job, and whether or not it will be successful is usually determined by various factors, such as political, economic, and sociocultural factors (Davis 1994: xiii). Two large categories of language planning activities are usually distinguished as status planning and corpus planning by Kloss (1969), or language determination and language development in terms of Jernudd (1973).

Status planning or language determination refers to the choices of languages to be used for specific purposes, such as the selections of official language or medium of instruction in school. Corpus planning or language development refers to the selection and promotion of variants within a language, such activities as reforming existing spelling, adopting a new script, and coining new terms. Ferguson (1968: 28) sees three dimensions relevant for measuring language development: 1) graphization, which means an adoption of a writing system, 2) standardization “the development of a norm which overrides regional and social dialects,” and 3) modernization “the development of intertranslatability with other languages in a range of topics and forms of discourse characteristic of industrial, secularized structurally differentiated, “modern” societies.”

Fasold (1984: 250) points out two approaches with regard to language planning. The first is instrumental consideration, which regards language fundamentally as a tool (e.g. for socio-economic advancement). The only criteria to be used in the standardization of mechanical tools are concerned with making them more suitable to the task they are used

for. For those languages being used or learned for instrumental purpose, they are considered inherently better than others. The second is socio-linguistic consideration, which takes the symbolic value of language into account, and regard language as a resource that can be used in improving one's social position.

Regarding the relationship between language and nationalism, Fishman (1968) classified nationalism into nationalism and nationism. He defines nationalism as the "process of transformation from fragmentary and tradition-bound ethnicity to unifying and ideologized nationality" (Fishman 1968: 41). The role of language in nationalism is sociolinguistic in that it serves as a link to the glorious past and with authenticity. A language is not only a vehicle for the history of a nationality, but also a part of history itself (Fasold 1984: 3).

Fishman (1968: 42) describes nationism as "wherever politico-geographic momentum and consideration are in advance of sociocultural momentum and consideration." The role of language in nationism is that whatever language does the job best is the best choice (Fasold 1984: 3). For example, considering government administration and education, a language or languages, which do the job best, must be chosen.

2.2 Diglossia and digraphia

Before we go through the discussions of language and literacy in Han sphere, we need to clarify two key concepts, i.e., diglossia and digraphia, which would help readers understand the linguistic and orthographic situations in this sphere.

Charles Ferguson (1959) is usually referred to as the first scholar who used the term diglossia to denote situations where two varieties of the same language are used for different social functions. Of the two varieties, one is called High (or simply H) and the other the Low (or L). H dialect has higher prestige and more literary heritage. H is better-

standardized and used in formal and public domains. In contrast, L dialect is considered inferior and is used in informal and private domains.

Joshua Fishman (1967) later revised and expanded the concept of diglossia. Fishman places less emphasis on the importance of situations with only two language varieties. For Fishman, diglossia can refer to any degree of linguistic difference from the most subtle stylistic differences within a single language to the use of two totally unrelated languages as long as the differences are functionally distinguished within the society (Fasold 1984: 40). Fishman further clarified the relationship between bilingualism and diglossia, as illustrated in **Table 1**.

Table 1. The relationship between bilingualism and diglossia

		Diglossia	
		+	-
Bilingualism	+	(1) Both diglossia and bilingualism	(2) Bilingualism without diglossia
	-	(3) Diglossia without bilingualism	(4) Neither diglossia nor bilingualism

The term ‘digraphia’ was intentionally created in analogy to the concept of diglossia. The major difference is that digraphia refers to written language, while diglossia refers to spoken.

Digraphia is defined by Dale (1980: 5) as “the use of two (or more) writing systems for representing a single language,” or by DeFrancis (1984: 59) as “the use of two or more different systems of writing the same language.” Digraphia in this study is expanded to the use of more than one writing varieties to serve different communicational tasks within a society. In other words, the use of multiple writing varieties is not restricted to the cases

within a single language. Further, digraphia with/without biliteracy, as is shown in **Table 2**, is proposed as a parallel to Fishman's (1967) idea of diglossia with/without bilingualism.

Table 2. The relationship between digraphia and biliteracy

		Digraphia	
		+	-
Biliteracy*	+	(1) Both digraphia and biliteracy	(2) Biliteracy without digraphia
	-	(3) Digraphia without biliteracy	(4) Neither digraphia nor biliteracy

* biliteracy means control of both H and L writing systems.

Let me demonstrate in the case of Taiwan of **Table 2**. Discussions of digraphia in this section are limited to linguistic situation after 1945 when Chinese KMT occupied Taiwan. The digraphic situation in Taiwan can be regarded as a double-nested digraphia as shown in **Table 3**, in which "H" and "L" represent High and Low languages (or orthographies) with the digraphia among languages; and "h" and "l" refer to high and low languages (or orthographies) of digraphia within a single language. For example, Chinese is serving as High in contrast to Taiwanese Low. When examining orthographies of a single language, Hanji (Han characters) counts as high, and Roman script (or Bopomo) as low.

Table 3. Double-nested digraphia in Taiwan

	Classical Han	
h	-----	
	Mandarin Chinese (in Hanji)	H
l	-----	
	Mandarin Chinese (in Bopomo)	
h	Taiwanese (in Hanji)	
l	-----	L
	Taiwanese (in Roman script)	

High (or high) and Low (or low) are functionally distinguished within the society. The functional distribution for High and Low means that they are situations in which only High is appropriate, and others in which only Low can be used. There are very little overlap between High and Low situations. Generally speaking, H has higher prestige, and the functions calling for H are formal and guarded. In contrast, L has lower prestige, and it is informal and relaxed.

2.2.1 Digraphia with biliteracy

Digraphia with biliteracy means that people have control over both (or several) high and low writing systems, but they are functionally distinguished. This type of digraphia-biliteracy relationship refers to the cases of digraphia within a single language. They are: 1) Hanji vs. Roman script within written Taiwanese, and 2) classical Han vs. Hanji vs. Bopomo within Mandarin Chinese.

Cheng (1990: 219-237) and Tiu^N (1998: 230-241) have pointed out that there are currently three main writing schemes for writing Taiwanese. They are: (1) Han character only, which means the exclusive use of Hanji, (2) Han-Lo ‘Hanji with Roman script,’ which means a combination of Hanji with Roman script, and (3) Roman-only, or ‘exclusive

use of Roman script.’ Generally speaking, Han writing has a longer literary heritage. Han-Lo writing is a new proposal, in which about 15% of the Taiwanese words are proposed to be written in Roman script, and others in Hanji. Writing in Roman-only is usually limited to the older generations of church Peh-oe-ji⁵ users.

Hanji and Roman script are two different orthographies among the three writing schemes. In general, Hanji is more prestigious and dominant. Many people in the Taiwanese Writing Circle enjoy finding the so-called 本字 ‘pun-ji’ or original characters for Taiwanese words in order to prove that Taiwanese is a prestigious language. This phenomenon has shown that most people in Taiwan consider Hanji a classical and prestigious orthography. This phenomenon supports the results of Chiung’s (2001a) survey of 244 college students’ attitudes toward different writing systems of written Taiwanese. Chiung’s experimental results reveal that college students tend to prefer Hanji more than Roman script.

Digraphia with biliteracy occurs within Mandarin Chinese as well as within Taiwanese. Classical Han writing is the high language in contrast to colloquial Mandarin Chinese writing. Mandarin Chinese written in Hanji is relatively high comparing to the Mandarin in Bopomo.

In Taiwan, all students are taught Bopomo, modern Chinese writing (in Hanji), and classical Han writing through the national education system. Bopomo is first taught as an auxiliary tool to the learning of Mandarin. Thereafter, Hanji is taught as an official writing system for Mandarin. Students are inculcated that Hanji are 國字 the ‘National Characters,’ and to avoid using Bopomo in their compositions in higher grades. Later, classical Han

⁵ For details about Peh-oe-ji, see next sections and chapters.

writing is taught. Students are considered intelligent if they are able to read such literary works as 唐詩 ‘Tang Poetry’ and 論語 ‘The Analects of Confucius.’

2.2.2 Digraphia without biliteracy

Digraphia without biliteracy means that H and L are functionally distinguished, but they are not both controlled by the same linguistic group. This type of digraphia-biliteracy relationship refers to the cases of digraphia between Taiwanese (L) and Chinese (H).

In Taiwan, Mandarin Chinese is recognized as the only ‘National Language,’ but Taiwanese is deprecated as a 方言 or dialect. Both spoken and written Mandarin are taught through the national education system, but Taiwanese is excluded from the system. Consequently, most Taiwanese speakers do not know how to read and write in Taiwanese. Many of them do not even know that Taiwanese can serve as a written language.

As a High language, Mandarin Chinese is well standardized and used for administrative and educational purposes. All government documents and school textbooks are published in Mandarin Chinese. Compared to Chinese, written Taiwanese is less standardized, and publication in it is much more marginalized to themes of homeland and authors’ love for Taiwan. Occasionally, Taiwanese captions may appear on political cartoon on newspapers.



Figure 1. Example of political cartoon in Taiwanese.

2.3 Socio-cultural background in the Han sphere

The Chinese attitude towards their neighbors and foreigners can be expressed exactly in the old Chinese philosophy, the System of the Five Garments (五服制 *Wufuzhi*; *Ngou-hok-che*). The Chinese empire set up a world outlook: the capital is great, civilized, and the central point of the world. Further, the empire used the capital as the center of the circle, to draw five circles of 500 kilometers radius. The farther barbarians are from the central capital, the more barbaric they are. Chinese people call the barbarians from the east as “*Dong-yi*,” barbarians from the south as “*Nan-man*,” barbarians from west as “*Xi-rong*,” and barbarians from the north as “*Bei-di*.” All the words are different animal names.

In such control concept of System of the Five Garments, the Chinese empire always tried to subdue the “barbarians” and brought them under the domination of China in order to “civilize” them. As a consequence, the “barbarians” were either under China’s direct domination or were demanded to pay tributes each certain year in recognition of the empire’s suzerainty (i.e., become a vassal state under China).

In this pattern, Vietnam, Korea and Taiwan had been directly occupied by China for long periods. Although later on they were no longer under direct domination, they became vassal states of China effectively until modern times. For example, Vietnam was brought under China's direct domination in 111 BC by *Han Wu Di*, the Chinese emperor of *Han* dynasty. Vietnam could not liberate itself from China until AD 939, during the fall of the powerful Chinese *Tang* dynasty (Hodgkin 1981). Thereafter, although Vietnam had established its own independent dynastic tradition, the Vietnamese had to recognize the suzerainty of Imperial China to exchange a later millennium of freedom until the late nineteenth century (SarDesai 1992: 19).

Although Japan was not under direct domination of China, due to China's powerful regimes during the times of Han and Tang dynasties, China was the model of imitation for Japan until the nineteenth century. For example, Japan's *Taika* Reform in the seventh century "marked the first step in the direction of the formation of a Chinese-style centralized state" (Seeley 1991: 40).

In general, China's main influences on these sino-spheric countries included: 1) the adoption of Han characters and classical Han writing (*bun-gian*; *wenyan*) to write Vietnamese, Korean, Japanese, and Taiwanese; and 2) the importation of Buddhism, Confucianism, the civil service examination and the government official system.

According to the civil service examination system, the books of Confucius and Mencius, which were written in classical Han, were accorded the status of classics among scholars and mandarins who assisted the emperor or king in governing his people (Taylor and Taylor 1995: 144-152). Everyone who desired to become a scholar or mandarin had to learn to use Hanji and read these classics and pass the Imperial examination, unless he had a close relationship with the emperor. Consequently, as Coulmas (2000: 52) has pointed out such literacy skills functioned "as a crucial means of social control," and "the Mandarin

scholar-bureaucrat embodied this tradition, which perpetuated itself above all through the civil service examination system.” Han characters and classical Han writing thus became the orthodoxy of written language in the Sinosphere for over a thousand years. The influence of Han characters on these counties was so deeply entrenched that the first historical annals compiled by their governments to record their early history were all written in classical Han. They are *Kojiki* (AD 712) and *Nihon shoki* (AD 720) in Japan, *Samguk sagi* (AD 1145) in Korea, and *Dai Viet Su Ky* (AD 1253) in Vietnam.

From the perspective of literacy, the classics were not only difficult to read (i.e., Hanji), but also hard to understand (i.e., the texts), because the texts were written in classic literary style instead of colloquial speech (*peh-oe; baihua*).⁶ In other words, because most of the people were farmers who labored in the fields all day long, they had little interest in learning Hanji and classical writing. As a consequence, a literate noble class and an illiterate peasant class developed and the class system strengthened the feudal society. The complexity of Hanji could be well expressed by the old Taiwanese saying, “*Hanji na thak e-bat, chhui-chhiu to phah si-kat,*” which means that you cannot understand all the Han characters even if you studied until you could tie your beard into a knot. Another saying similarly describes its challenges, “*Si-su Ngou-keng thak thau-thau, m-bat ku pih chau,*” which means you still cannot distinguish the characters of tortoise, turtle, and cooking stove (because they look so similar in shape) even if you have studied all of the classics.

In short, as Chen (1994: 367) has pointed out, high illiteracy and low efficiency caused by the use of Han characters have hence become impediments to nation’s modernization, the demand for widespread literacy has become one of the advising factors pushing orthographic reform in the Han sphere.

⁶ For more details about classical Han writing, see Norman (1988: 83-132).

2.4 Orthographic reform in Korea

Han characters were probably introduced to Korea by Chinese immigrants fleeing China's civil wars during the Warring States Period (BC 403-221) (Ledyard 1966: 22-23). Han characters became institutionalized thereafter Han Wu Di brought northern Korea under direct Chinese domination in 108 BC (Ledyard 1966: 23). China's control of northern Korea lasted until the fourth century. Meanwhile, the deposed Korean people migrated south and split into three kingdoms: *Silla* (BC 57-668 AD), *Paekche* (BC 18-660 AD), and *Koguryo* (BC 37-668 AD), which were all unified by Silla in 668AD. In addition to the territory of Chinese domination, Han characters were also used among the elite in these three kingdoms (Taylor and Taylor 1995: 203). In 958AD, the Chinese-style civil service examination system was established by the *Koryo* kingdom, which had replaced the Unified Silla in 935. The state examination system lasted in Korea for a thousand years ending only in 1894 (Taylor and Taylor 1995: 255-259).

As Han characters were being adopted in Korea, the Koreans encountered difficulties in understanding the classical Han writing. They gradually developed their own remedial measures to make writing in Han characters more approachable to the Korean-speaking people. Beginning in the late sixth and early seventh centuries, two major remedies were developed, and they were later known as *Hyangch'al* and *Idu*. *Hyangch'al* was mainly applied in transcribing vernacular poetry. *Idu* served as a bureaucratic tool for the clarification of administrative documents written in literary Chinese, and it lasted until the end of the nineteenth century (Ledyard 1966: 34). Texts in *Hyangch'al* and *Idu* were both written in Han characters. The arrangements of word order in *Hyangch'al* were in accordance with the grammatical details of the Korean language. However, the text of *Idu* "wavers between Chinese and Korean syntax and is marked by the insertion of Korean grammatical forms intended to aid Korean readers" (Ledyard 1966: 33). In both types,

either “phonetic borrowings” or “semantic borrowings” were incorporated into Korean while still choosing Han characters to represent the Korean language (Taylor and Taylor 1995: 204-207).

Although the Korean elite had developed Hyangch'al and Idu, the demand for a more accessible writing system grew stronger as the fifteenth century progressed (Ledyard 1966: 70). During the fifteenth century, the Korean King *Sejong* and his scholars undertook a project of inventing a new script for writing the Korean vernacular. The project was carried out in 1443, and was officially proclaimed in the title of *Hun Min Jong Um*⁷ 訓民正音 in 1446 (Ledyard 1966: 91-99). The script of the *Hun Min Jong Um* is known today as Hangul, the Korean alphabets, consisting of 28 letters to write Korean in a phonemic way (Shin et al. 1990). King Sejong prefaced in *Hun Min Jong Um* that why he devised the new script:

國之語音 異乎中國 與文字不相流通 故愚民有所欲言而終不得伸其情者多矣 予為此憫然 新制二十八字 欲使人人易習 便於日用矣…有其聲而無其字 假中國之字以通其用 是猶柄鑿之鉏鋤也 豈能達而無礙乎 要皆各隨所處而安 不可強之使同也 吾東方禮樂文章 倖擬華夏 但方言俚語 不與之同 學書者患其有趣之難曉 治獄者病其曲折之難通 昔新羅薛驄 始作吏讀 官府民間 至今行之 然皆假字而用 或澀或窒 非但鄙陋無稽而已 至於言語之間 則不能達其萬一焉…(Lee 1957: 47)

The language of this country is different from that of China, so that it is impossible (for us Koreans) to communicate by means of Chinese characters. Therefore, among the unlearned people, there have been many who, having something to put into words, have in the end been unable to express themselves. Feeling sorry for this, I have newly made twenty-eight letters only because I wish them to be easy for everyone to learn and convenient for use in daily life...The feature and circumstances of various places are different, and so naturally sounds and breaths differ accordingly. The language of countries other than China have their own sounds, but not their own letters. These countries have borrowed the Chinese characters for their use. This absurdity is like putting a square peg in a round hole. How can it be freely used without

⁷ Literary, “Correct Sounds to Instruct People.”

hindrance? In fact, all things are safe in their proper places and should not be forced to be the same. The ceremonials, music and literature of Korea are comparable to and imitative of those of China, but the language of this country is not the same as that of the Chinese. Those who study books have been troubled by the difficulty of understanding the meaning [of Chinese characters] and those who enforce laws have been distressed at the difficulty of distinguishing right and wrong. In the olden times of Silla, Sol-chong originated the I-du characters which have been used in the government offices and among the people until now. But they use borrowed Chinese characters [for Korean sounds], and in some cases the I-du characters are awkward and in others hinder understanding of the language. They are not only rustic and crude and unreasonable; when it comes to the realm of actual speech, they are impossible to express the language in one out of ten thousand cases (translation by Lee Ton-Ju, in Shin 1999: 1-35).

Although the new system of Hangul was very efficient and could have made widespread literacy possible, it soon had opposition from the privileged bureaucratic and literate classes. For example, the most well known anti-Sejong faction was led by *Malli Choi*, the highest purely academic rank in the College of Assembled Worthies (Ledyard 1966: 99-114). In 1444, Choi presented Sejong a petition against the new orthographic invention, as follows:

我朝自祖宗以來 至誠事天 一遵華制 今當同文同軌之時 創作諺文 有該觀聽 儻曰諺文 皆本古字非新字也 則字形雖倣古之篆文 用音合字盡反於古 實無所據 若流中國 或有非議者 豈不有愧於事天慕華...

In the first place it is a violation of the principle of maintaining friendly relations with China, to invent and use letters, which do not exist in China...Those who seek position in the government will not seek to learn Chinese characters with patience, and consequently, Chinese literature, which is our only study and sole literature, will flourish no longer. The Vulgar Script [i.e., Hangul], which is a mere novelty will cause hindrance to study, disadvantage and inefficiency to administration...(Lee 1957: 30-31).

The opposition to the new script lasted decades even after the death of Sejong. Moreover, writing in Hangul was banned by the regent *Yonsan'gun* after the literati purge of 1504 (Ledyard 1966: 322). Consequently, Hangul was used in very limited circles and domains. For centuries after its creation, Hangul was variously called “*onmun*” (vulgar

script), “women’s letters,” “monks’ letters,” or “children’s letters” (Taylor and Taylor 1995: 212). For most of its history, Hangul was “regarded as a poor person’s substitute for real writing, which was either classical Chinese written in characters or stilted Korean written in Chinese characters” (Hannas 1997: 51).

The inferior development of Hangul reached a turning point with the advent of the twentieth century. During the Japanese occupation of Korea (1910-1945), Japan’s harsh policy to restrict the use of the Korean language had enhanced the Korean identity of Hangul (Coulmas 2000: 56). Moreover, the user-friendly characteristics of Hangul made it favorable to the Korean nationalists in the consideration of literacy. In other words, Hangul, corresponding to Chu Quoc Ngu in Vietnam, was chosen as the tool to eliminate illiteracy in order to fight against Japanese imperialism. As Hangul gained more recognition and had become wider spread than ever before, it was thus further promoted to the official national script when the Korean people built their modern nation-state(s) after the World War II.

After the war, Han characters in North Korea were officially abolished in favor of the exclusive use of Hangul. As for South Korea, although the policy abolishing Han characters has not been consistently executed, the use of Han characters has dramatically decline over the past decades (Taylor & Taylor 1995: 208-210). In short, Han characters have shifted from a dominant status to a supplementary use as a tool to the Hangul.

2.5 Orthographic reform in Japan

It is estimated that around the fifth century, Han characters were brought to Japan by Korean scholars (Seeley 1991: 6). Thereafter, due to an increasing cultural dependence on China, such as the Taika Reform (645-649), Han characters and classics written in it had become more prominent and prestigious in Japanese society by the seventh century (Seeley 1991: 40).

Once the Japanese embraced classical Han writing, they encountered difficulties in reading the Chinese classics as had occurred in the cases of Vietnam and Korea. Again, the Japanese utilized ‘phonetic borrowings’ and ‘semantic borrowings’ to overcome the problems. Those remedies were both well adopted in the famous *Man’yoshu*, Collection of the Myriad Leaves, a collection of Japanese poems compiled around 759AD. In the method of sound borrowings of the *Man’yoshu*, the original meaning of Han characters was disregarded, while their Chinese pronunciation or Japanese *kun* pronunciation in accordance with the characters was borrowed. The sub-methods based on sounds were called *shakuon* and *shakukun* respectively. Because *shakuon* and *shakukun* are the prominent features in the writing of *Man’yoshu*, they are generally called *man’yogana*, a combination of *man’yo*, from the title of *Man’yoshu*, and *kana*, the syllabary (Habein 1984: 12).

Because of complication and inconsistency, starting in the ninth century the *man’yogaga*-like systems were moving toward a process of simplification to Han characters used as phonograms (Seeley 1991: 59). Among the various simplified syllabaries, *Katakana* and *Hiragana*, which are currently in use after modern standardization, were well developed and widely used at least by the tenth century (Habein 1984: 22-35; Seeley 1991: 69-75). *Katakana* was called imperfect kana, which was developed by priests. *Hiragana* was called *onnade* “woman’s handwriting” or *onnamoji* “woman’s letters.” Because women were excluded from the study of literary Chinese, they were most likely to use *hiragana* (Habein 1984: 25).

The Heian period (794-1192) was a period that included an introduction of reading and writing to the noble class. In the later centuries, literacy was brought to the broader public, leading to diversification and complication of writing styles, which include literary Chinese, *Kana*, and a hybrid of Han characters with *Kana* (Habein 1984: 4).

From the later part of the nineteenth century onwards, the issue of script reform was raised, and people were highly concerned again with the opening of Japan to the West. After the imperial regime was restored in 1868, Emperor Meiji opened his door to foreign countries, which resulted in enormous changes in daily life. Among the changes was the increase of new words coined for the overwhelming unfamiliar concepts and objects from the West. In this situation, the intellectuals rose the issues of language reform in the interest of better literacy and education. There were three major proposals in such a reform: 1) to replace the current chaotic systems with a Kana-only system, 2) to replaced the existing systems with Romanization, and 3) to limit the number of Han characters in use (Seeley 1991: 136-142).

After the successful political reform of Emperor Meiji, manifested in its victories in the Sino-Japanese war of 1894-95 and the Russo-Japanese war of 1904-05, the Japanese began to believe that their nation could be mobilized through more effective education, to which script reform was considered important (Gottlieb 1995: 25). This belief eventually brought language reform into practical trials in the early part of the twentieth century. Because using Kana-only or Romanization was considered too radical, the orthographic reform thus, centered on restricting the number of commonly used Han characters and the standardization of the Kana usage (Seeley 1991: 142). Such efforts were reflected in the examples: new regulations aimed at simplifying the teaching of written Japanese at the primary schools were issued by the Education Ministry in 1900; *Kanji seirian* “Proposed Modifications to Han Characters” was published in 1919 by *Hoshina et al.*; *Kanazukai no Kaitei An* “Proposal for the Revision of Kana Usage” was released by the Interim Committee on the National Language in 1924; *Toyo kanjihyo* “List of Characters for general Use” was proposed by the Interim Committee in 1923, and a later revised *Toyo kanjihyo* in 1931, which consisted of 1856 characters.

As time went on, Japan's language policy was driven by imperatives from modernization to imperialism in the first half of the twentieth century (Gottlieb 1995: 21). The influence of the military and the ultranationalists became more and more powerful as Japan became more aggressive in preparation for conquering China. The influence was substantial especially after the Manchuria Incident of 1931, in which three northeast provinces of China were under Japan's occupation. From the perspective of the military and ultranationalists, Han characters and historical Kana usage were *kotodama*, the "spirit of the Japanese language," which constitutes the essence of the Japanese national spirit. Therefore, reform proposals, such as abolition of Han characters, Romanization, or new Kana usage, were considered attempts to tamper with Japan's spirit, culture, and history. For example, the Interim Committee's proposals of 1931 to restrict characters and to carry out a new Kana usage were dismissed due to fierce oppositions from the conservatives. In another case in 1939, a number of Romanization advocates were arrested on the charge of anti-nationalist sympathies (Gottlieb 1995: 75-88; Seeley 1991: 147-148).

Although many efforts were brought to the script reform, wider adoption of reform proposals would not become reality until the end of World War II, when the Japanese army surrendered to the Allied Forces (Seeley 1991: 151; Hannas 1997: 43). After Japan's defeat in 1945, the arrogant military and ultranationalists were suppressed. As Eastman (1983: 23) has pointed out, without any social, cultural, or political changes, orthography reform is not likely to succeed. Japan's dramatic changes after the war thus created the atmosphere and conditions to carry out script reform. In 1946, under the supervision of the Supreme Command for the Allied Powers (SCAP), Japan's cabinet promulgated *Toyo kanjihyo*, the list of 1850 characters for daily use, and *Gendai kanazukai*, the new modern Kana usage, as the first step of script reform after the war (Unger 1996: 58; Seeley 1991: 152).

At present, Han characters and Kana syllabary all serve as the official scripts in the hybrid Japanese writing system. This fact makes Japan the only case, among Vietnam, Korea, and Japan, in which Han characters were not officially abolished after domestic scripts were promoted to national status. Why were Han characters not abolished in Japan? Both internal and external factors have contributed to this outcome. From the perspective of literacy and anti-feudal hierarchy, by the early twentieth century Japan reached a much higher degree of literacy and modernization in comparison to other Asian countries.⁸ This achievement gave the conservatives the impression that Han characters need not be abolished as long as Kana syllabary was in actual use. Furthermore, although Han characters were originally imported from China, they were converted from a pure foreign invention to an indigenized writing system after more than a thousand years of adoption. In other words, Hanji was regarded by the Japanese as part of their language, in which situation it was totally different from the case of Vietnamese, where they considered Han characters as Chinese scripts and Chu Nom as their own. Why did Japan and Vietnam have reverse perceptions of Han characters? Recall that Japan historically never was under China's direct control. On the contrary, Japan's imperialism and militarism became a fateful threat to China in modern history. However, conflicts with China frequently occurred in the history of Vietnam. That is to say, the Japanese did not consider the use of Han characters as a cultural link to a potential invader (i.e., China). As a matter of fact, the use of Han characters was even considered necessary once Japan launched an invasion into China. For example, the Interim Committee's proposal, *Toyo kanjihyo* of 1931, was

⁸ For example, Koji Taira estimated that "male and female literacy rates rose from about 35 and 8 percent, respectively, to about 75 and 68 percent between the beginning and end of the Meiji period (1868-1912)" (quoted in Unger 1996: 35). Aso and Amano reported that 86.9 percent of Japanese children attended four-year compulsory schooling in 1905 (quoted in Okano and Tsuchiya 1999: 19).

strongly opposed by the military for the practical need to write a large number of Chinese personal and place names of the newly occupied Chinese territories (Seeley 1991: 147).

2.6 Orthographic reform in Vietnam

Hanji was first employed in the writing system of Vietnam. Later *Chữ Nôm* (Chu Nom) appeared in the tenth century, and Romanized Vietnamese devised in the seventeenth century. The relation between languages and political status since 111 B.C. in Vietnam is shown in **Table 4**:

Table 4. Relation between languages and political status in Vietnam

Period	Political Status	Spoken Languages	Writing Systems
111B.C.-939A.D.	Chinese colonialism	Vietnamese/Chinese	Chinese (Han characters)
939-1651	Monarchical independence	Vietnamese/Chinese	Chinese/Nom
1651-1861	Monarchical independence	Vietnamese/Chinese	Chinese/Nom/pre-Quoc Ngu
1861-1945	French colonialism	Vietnamese/Chinese/French	Chinese/Nom/Quoc Ngu /French
1945-	National independence	Vietnamese	Quoc Ngu

*Based on John DeFrancis 1977.

2.6.1 Colonial background

After the first Chinese emperor *Qin Shih Huang* 秦始皇, who built the Great Wall, annexed six countries (221 B.C.), he continued to suppress South of the Mountain Passes (嶺南 present southern China). In 207 B.C., Trieu Da,⁹ a Chinese general who commanded the *Kwantung* and *Kwangsi* provinces of present day China, brought the Red River Delta under his jurisdiction and built up an autonomous state called *Nam Viet* 南越. In 111 B.C., the Chinese emperor of the *Han* dynasty, *Han Wu Ti* 漢武帝, sent his forces against Nam Viet and annexed Nam Viet which remained under the direct domination of China until 939.

⁹ Trieu Da is written as Triêu Đà in Vietnamese, and Zhao Tuo 趙佗 in Chinese.

In that year, Nam Viet¹⁰ separated from China at the moment of the fall of the powerful *Tang* dynasty, and then became an independent monarchy. This government was ancestor of the current Socialist Republic of Vietnam of today (Chavan 1987; Hodgkin 1981; Holmgren 19??; SarDesai 1992; Tran 1992).

Although the Vietnamese established their own independent monarchy, Vietnamese had to recognize the suzerainty of the Chinese Emperor to exchange a later millennium of freedom until the late nineteenth century (SarDesai 1992: 19). As SarDesai (1992: 21) describes “despite strong political hostility toward the Chinese, the Vietnamese rulers deliberately set their nation on a course of sinicization.” China’s influence on Vietnam was never dismissed even though Vietnam had achieved independent political status. For example, during the early feudal period of Vietnam, *Ly* and *Tran* dynasties (1010 A.D.-1428 A.D.), the Vietnamese government established a Confucian Temple of Literature and the Han-Lin Academy for study of Confucianism¹¹, and imported many Chinese administrative practices including the civil service examination and the hierarchical system of bureaucracy from China (SarDesai 1992: 21). Consequently, Chinese classics such as Four Classical Books 四書 and Five Canonical Books 五經 became the textbooks and resources for Vietnamese scholars and officials to study to obtain office (Nguyen 19??: 2; Pham 1980). In short, although Vietnam was not under China’s direct domination in the second millennium, there was also great influence on Vietnam from China, as the late Vietnamese historian, Tran Trong Kim (1882-1953) described (Tran 1921: 8):

No matter adult or child, the Vietnamese only learned Chinese history instead of Vietnamese history when they went to school. They had to obtain materials from Chinese literature when they wrote poems or articles and they never mentioned

¹⁰ I keep using the term *Nam Viet* for readers' convenience, although its name changed over different periods. See Hodgkin (1981:349) for names given to Vietnam at different periods of history.

¹¹ For details of Confucianism in Vietnam, see Nguyen 1979.

their own country, Vietnam. Besides, Vietnamese always looked down on their own history and thought that it was not useful to know Vietnamese history. This was because the Vietnamese did not have their own Vietnamese writing, the Vietnamese had to acquire knowledge through other people's language and other people's characters.

After Chinese domination, Vietnamese faced the French imperial power from 1862, in which year Vietnam ceded three provinces of Cochin China to France, until the end of First Indochina War in 1954 (Hodgkin 1981; SarDesai 1992; McLeod 1991).

In the end of the fifteenth century, new technology made west European traveling by sea easier and led to the discovery and control of new sea routes; as such Europeans as the Portuguese, the Dutch and the Spanish, gradually appeared in South China Sea for trade, mission work, or colonization. Before French imperial power entered Vietnam, there had been some missionary activities there. For example, in 1624 a French missionary named *Alexandre de Rhodes*, who is usually referred to as the inventor of chữ Quốc ngữ, a method of writing the Vietnamese language in Roman script instead of the traditional Han characters, arrived in Vietnam to begin his mission of four decades there. In many colonized countries, missionary activities resulted in some conflicts between missionaries and local people, and that was no exception in Vietnam. Due to religious conflicts, there was a marked increase in hostility toward Catholics and forward foreign influences. Consequently, large scale persecution of converts and missionaries began in the 1820s under Emperor *Minh Mang* (SarDesai 1992: 32). Religious conflicts became an excuse for the French to invade Vietnam and finally take the whole of the country.

In many colonies, intellectuals usually have different thoughts on the relationship between locals and immigrants, i.e., they may choose to collaborate, to resist, or to retreat. That was also no exception in the case of Vietnam. According to SarDesai (1992: 44), the Vietnamese mandarin class in the transition from the nineteenth century to the twentieth century was divided into three groups: 1) those who had collaborated with the French, 2)

those who retreated to the villages in a kind of passive non-cooperation, and 3) those who battled to bring new meaning and ethnic salvation to their country. Prior to the twentieth century, many Vietnamese mandarins were under the illusion that Vietnamese would maintain cultural and spiritual independence even though they had lost their land and political control to French. But a new generation of mandarins became aware of the pervasive educational and cultural impact of colonial rule, thus they devoted themselves to nationalist resistance movements.

In the second decade of the twentieth century, the Vietnamese nationalism gradually gained strength. SarDesai (1992: 46-47) attributes the result to two primary reasons. First, the result of French education. Although it was a colonial education, it provided Vietnamese a chance to gain knowledge and ideas from abroad such as nationalism, democracy and the concept of nation-state. Second, the early twentieth century was a period characterized by the rise of nationalism. More than 100,000 Vietnamese soldiers and workers in France had experienced nationalism during World War I (1914-1918). Besides, the pronouncement of the right of self-determination of nations (1918) by the U.S. President, Woodrow Wilson, inspired the nationalist movement.

On September 2, 1945, the Vietnamese communist leader, Ho Chi Minh, declared the birth of Democratic Republic of Viet Nam. However, the new Republic was not soon recognized by any country, and it caused the First Indochina War (1946-1954), in which the French power attempted to suppress the independence of Vietnam. Ultimately, this French power failed to maintain control of Vietnam, and Vietnam was divided into two zones. The Second Indochina War (1964-1975) expelled all alien forces including the United States, and the country was reunited as the Socialist Republic of Viet Nam (Chavan 1987; Hodgkin 1981; SarDesai 1992).

2.6.2 Orthographic tradition

In Vietnam, Han characters were employed since 207 BC during the *Nam Viet* (South Viet) period (Nguyen 1999: 2). Of course, by the Han occupiers and perhaps others. Thereafter, Han characters retained their orthodoxy status during the millennium of Chinese occupation. Not until the tenth century when Vietnam liberated itself from the Chinese Empire, could the domestic scripts Chu Nom have been prominently developed (DeFrancis 1977: 21). Chu Nom, or Nom scripts, means southern writing or southern orthography in contrast to Chu Han, Han writing or Han characters. Chu Nom in the early period was used as an auxiliary tool of classical Han to record personal or geographical names and local specialties (Nguyen 1999: 2). The thirteenth century was marked the first literary writing in Chu Nom (DeFrancis 1977: 23). Literary works in Chu Nom achieved popularity from the sixteenth century to the eighteenth century, and reached their peak at the end of the eighteenth century (DeFrancis 1977: 44). For example, *Truyen Kieu*, a novel in Chu Nom, considered a masterpiece of Vietnamese literature, was published at the end of the eighteenth century.

Generally speaking, ordinary people, monks, relegated mandarins, and very limited ultra-anti-Chinese nationalist elites favored Chu Nom. In contrast to official domains dominated by Han characters, such as governmental administration, education, academia, and classic literary works, Nom scripts were mainly used for recording folktales, composing literary works in pure Vietnamese, translating the Buddhist Scripture since access the masses was its goal, and being used as an auxiliary tool to read classics in Han characters (Nguyen 1999).

The orthographic structure of Nom scripts consists of two main categories (DeFrancis 1977: 24-26). The first one may be called 'simple borrowings,' which is in accordance with the existing Han characters in shape, but different in sound or semantic meaning. In other

words, Han characters were borrowed for their phonetic or semantic value to represent Vietnamese words. For example, in the case of phonetic borrowings, 昆 originally meant ‘insect’ in Chinese and had a Sino-Vietnamese pronunciation /kon/, was borrowed to represent the Vietnamese word *con* (pronounced as /kon/ and means ‘child’ or ‘offspring’). In such a case, the original Chinese meaning was ignored and only their Sino-Vietnamese sound was preserved to indicate the pronunciation of the corresponding Vietnamese words. In other situations, the pronunciation of characters was ignored and only their semantic meaning was preserved in the case of semantic borrowings.

The second category may be called “composite creations,” which were developed relatively later than the first type.¹² In this category, Nom scripts were made by combining two Han characters, usually where one was taken over for its meaning and the other for its pronunciation (DeFrancis 1977: 25). For example, the Vietnamese word *con* was also written as 子昆 at a later time. It consisted of two Han characters 子 (with Chinese meaning ‘child’ or ‘offspring’ and Sino-Vietnamese pronunciation /tu/) and 昆. In this case, 子 refers to the meaning and 昆 indicates the pronunciation.

Although Nom scripts experienced a beginning and development period at least for a thousand years, they are still far from being standardized (The Anh 1999: 5; DeFrancis 1977: 24-30). Because of inconsistency, a Vietnamese word may be written in different Nom scripts, such as 字字字喃, 字喃, and 字宁字喃, all referring to the same word ‘Chu Nom.’ The major causes of inconsistency are 1) lack of institutional support since the mandarin and scholar class as a whole looked with disdain on the Nom literature, 2) the fact that Nom scripts were not devised under explicit linguistic planning as was the case with Hangul was

¹² Personal communication with Nguyen Quang Hong.

in Korea; instead, Nom scripts were created by individual authors in different time and places, and 3) the inconsistency was inherited from Han characters.

Although the domestic Nom scripts have been around since the tenth century, they neither reached the same prestige as Han characters, nor replaced the classical Han writing. In contrast, Chu Nom was generally regarded as a vulgar written form, which refers to the low language in digraphia. Moreover, Nom scripts were eventually forced to yield themselves to the *Chu Quoc Ngu*, a Romanized writing system originally devised in the early seventeenth century, which finally became the only official orthography of Vietnam in 1945. Three are three factors that contributed to the fate of Chu Nom.

First, the Vietnamese were deeply influenced by the Chinese values with regard to Han characters. Since Hanji was highly regarded as the only official orthography in China, to which was pledged in fief Vietnam, the Vietnamese people had no choice but to follow this traditional value. As a consequence, the Vietnamese rulers in all dynasties, except a few short-lived strongly anti-Chinese rulers, such as *Ho Quy Ly* (1400-1407) and *Quang Trung* (1788-1792), had to recognize Han characters as the institutionalized writing of the realm.

Second, literacy in Nom scripts was restricted by those who had passed the civil service examination. Because the examination system tested knowledge of the Chinese classics written in Hanji, all the literati who wished to pass the exam had to study the classics. Once they passed the exam and became bureaucrats, they had to maintain the examination system to ensure their monopoly of power and knowledge in the Chinese style feudal hierarchy (DeFrancis 1977: 47).

Third, the development of Nom scripts was highly restricted by the nature of their orthographic structure. Because Chu Nom were composed of one or two Han characters to form a new Nom graph, the process inherited all the defects of Han characters (DeFrancis

1977: 25). The much more complicated structure has even caused Nom scripts more problems in such aspects as efficiency, accuracy, and consistency. Normally, one has to learn Han characters first before s/he could fully master Nom scripts.¹³ Consequently, learning to read and write in Nom scripts was even more difficult than writing in Han characters.

2.6.3 Development of Romanization

In the late sixteenth century and the early seventeenth century, European missionaries from countries including Portugal, Italy, and France, gradually came to preach in Vietnam. To get their ideas across to the local people, it was recognized by missionaries that knowledge of spoken Vietnamese was essential. A Romanized writing system was thus devised to assist missionary-learners to acquire the Vietnamese language (Do 1972). Among the variants of Vietnamese Romanization, Alexandre de Rhodes¹⁴ is usually referred to as the person who provided the first systematic work of Vietnamese Romanization (DeFrancis 1977: 54). In 1651, de Rhodes published the first Romanized dictionary, *Dictionarium Annamiticum, Lusitanum et Latinum* (Vietnamese-Portuguese-Latin), and a Vietnamese catechism *Cathechismus*. De Rhodes' Romanized system, with some later changes, became the present Chu Quoc Ngu.¹⁵

The development of Romanized writing in Vietnam can be divided into four periods in terms of its spread: 1) Church period, from the early seventeenth century to the first half of the nineteenth century. Roman scripts were mainly used in church and among religious followers. 2) French promotion period during the second half of the nineteenth century after

¹³ Personal communication with Nguyen Quang Hong.

¹⁴ For more information about Alexandre de Rhodes, see Tran, Phan, Hong, and Do (1998) and Chu (1996)

¹⁵ For the historical changes of Chu Quoc Ngu, refers to Nguyen (1994).

the French invaded Vietnam in 1858 (Vien Van Hoc 1961: 21-23). In this period, Romanized Vietnamese was intentionally promoted by the French aiming to replace the classical Chinese ultimately with French (DeFrancis 1977: 129-134). 3) Nationalist promotion period during the first half of the twentieth century. Vietnamese Romanization was promoted by anti-colonialism organizations, such as the *Dong Kinh Nghia Thuc* 東京義塾 or 'Dong Kinh Free School' and *Hoi Truyen Ba Quoc Ngu* 'Association for Promoting Chu Quoc Ngu' (Vien Van Hoc 1961: 24; Hoang 1994: 94; Nguyen, Pham, and Tran 1998: 44-54). Because Roman scripts were no longer closely associated with the French colonialist, but considered as an efficient literacy tool, Romanization thus received much greater recognition by the Vietnamese people than in the period of French promotion (DeFrancis 1977: 159). 4) National status period after 1945, when Ho Chi Minh declared the exclusive use of Chu Quoc Ngu (Ho Chi Minh 1994: 64-65).

From the perspective of literacy, Roman script was much easier to acquire than Han character or Chu Nom. However, Vietnamese Romanization was not widespread until the early twentieth century. There are two primary reasons. First, the use of Romanized Vietnamese was primarily limited to the Catholic community prior to the twentieth century. DeFrancis (1977: 64) has pointed out that most missionaries "looked upon it [Romanization] chiefly as a tool in working with the Vietnamese language and were not greatly concerned with urging its use in other areas." Moreover, even if people outside the Catholic community wanted to learn the Romanization, they were afraid of being treated as collaborators with foreign missionaries since there were conflicts between local people and foreign missionaries. Consequently, there was no general and secular public usage outside the Catholic community (DeFrancis 1977: 61). Second, it was the reflection of people's socialized preference of Han characters since Han characters had reached the status of orthodoxy since the Ly dynasty. This phenomenon of preference is especially true of

traditional scholars and officials. For example, it was reported that Confucian schools, which were essential to acquire a knowledge of Han writing and Chinese classics, continued to exist and attract students as late as the first decade or two of the twentieth century (DeFrancis 1977: 124).

Since French colonization was involved in the colonial history of Vietnam, what role have the French (1861-1945) played in the orthographic transition of the Vietnamese language? First of all, the French weakened or even replaced the role played by the Chinese in Vietnam. In the nineteenth century, China was losing dominance over Asia since the Opium war in 1842. In addition, Japan's successful Westernization, shown in such wars as her victories over China in 1895 and over Tsarist Russia in 1904-1905, had impressed the Vietnamese. The appearance of the French power in Indochina¹⁶ caused the Vietnamese people to experience the new political power from Western society, and further reconsidering their relationship with the traditional feudal China.

Second, the French's antagonism toward Chinese had strengthened the promotion of the Romanized system. Their hostile attitudes toward Chinese was summed up in a letter of 15 January 1866 by a French administrator, Paulin Vial, who held the position of *Directeur du Cabinet du Gouverneur de la Cochinchine*: "From the first days it was recognized that the Chinese language was a barrier between us and the natives...; it is the only one which can bring close to us the Annamites of the colony by inculcating in them the principles of European civilization and isolating them from the hostile influence of our neighbors" (quoted in DeFrancis 1977: 77). Thus, the actions taken by the French colonialists included termination of the traditional civil service examination, and promotion of the Romanized Vietnamese,¹⁷ which was regarded as a closer connection to French since both French and

¹⁶ Indochina includes present Vietnam, Laos, and Cambodia.

¹⁷ For example, *Gia Định Báo*, the first Romanized newspaper was published by the colonial

discussion on primary education in 1931. Furthermore, Ho even concluded that it is going to take Vietnamese more than five hundred years to improve their “patois” to the level of French and Chinese (Marr 1981: 136).

Immediately after Ho Chi Minh declared the establishment of Democratic Republic of Viet Nam on September 2, 1945, he issued a decree to promote Vietnamese and Chu Quoc Ngu on September 8 (DeFrancis 1977: 239). Concurrently, Ho Chi Minh launched an “Appeal to Fight Illiteracy” in October as follows (Ho, 1994: 64-65):

Citizens of Viet Nam!

Formerly, when they ruled over our country, the French colonialists carried out a policy of obscurantism. They limited the number of schools; they did not want us to get an education so that they could deceive and exploit us all the more easily.

Ninety-five percent of the total population received no schooling, which means that nearly all Vietnamese were illiterate. How could we have progressed in such conditions?

Now that we have won back independence, one of the most urgent tasks at present is to raise the people's cultural level.

The Government has decided that before a year has passed, every Vietnamese will have learnt *quoc ngu*, the national Romanized script. A Popular Education Department has been set up to that effect.

People of Viet Nam!

If you want to safeguard national independence, if you want our nation to grow strong and our country prosperous, every one of you must know his rights and duties. He must possess knowledge so as to be able to participate in the building of the country. First of all he must learn to read and write *quoc ngu*.

Let the literates teach the illiterates; let them take part in mass education.

Let the illiterates study hard. The husband will teach his wife, the elder brother his junior, the children their parents, the master his servants; the rich will open classes for illiterates in their own houses.

The women should study even harder for up to now many obstacles have stood in their way. It is high time now for them to catch up with the men and be worthy of their status of citizens with full electoral rights.

I hope that young people of both sexes will eagerly participate in this work.

The number of people who acquired literacy in Quoc Ngu after the achievement of independence was reported by Le Thanh Khoi (quoted in DeFrancis 1977: 240) to have

risen from 20 percent in the year 1945 to 70 percent in 1953. Similar statistics were reported by Huang (1953: 20) as shown in **Table 5**.

Table 5. Numbers of literate in Vietnam

Year	Numbers of Literate	Percentage of Total Population
1945	2,520,678	14%
1946	4,680,000	27%
1947	6,880,000	39%
1948	9,680,000	55%
1949	11,580,000	66%
1950	12,000,000	68%
1953	14,000,000	79%

*Percentage was calculated based on the data of Huang (1953: 20)

How could Vietnam successfully replace Han characters and Chu Nom with Romanized Chu Quoc Ngu? Hannas (1997: 88-92) stated twelve factors, and concluded “the compelling factor behind this success is that Vietnam never had a top-down, coordinated, state-backed movement to effect the reform” (1997: 84). Although it is true that bottom-up grass root movement played an important role in Vietnam’s orthographic transition, one can attribute the success to two crucial factors: 1) the internal factors resulting from the social demand for literacy and anti-feudal hierarchy, and 2) the external factors resulting from political friction between Vietnam and China in the international arena during the first half of the twentieth century.

The internal factors of social demand for literacy and anti-feudal hierarchy are understandable. Recall that China was the only major threat to the traditional feudal society of Vietnam prior to the nineteenth century. In that situation, although the adoption of Han characters could cause the majority of Vietnamese to be illiterate, it could, on the other hand, minimize a potential invasion from China, and more importantly, preserve the vested interests of the Vietnamese bureaucrats in the Chinese style feudal hierarchy. However,

with the advent of the twentieth century, Vietnam has faced a train of international colonialism. Since Ho Chi Minh claimed that 95 percent of Vietnam's total population were illiterate, it was important to equip the people with primary education, which was considered essential to modernization in order to fight against imperialism (Ho Chi Minh 1994: 64-65). Although the domestically-created Nom scripts, to some extent, represented the Vietnamese spirit, their fatal weakness in blocking literacy removed them from being candidates for a national writing system. Thus, the efficient and easily learned Romanization was the best choice for literacy and stood in contrast to the complexity of Han characters and Nom scripts. Since the majority of Vietnamese were illiterate, and only a few elites were skilled in Han writing or French during the promotion of Chu Quoc Ngu, it was clear that Romanized Vietnamese would be favored by the majority, and thus won any literacy contest.

The external factor involves the complexity of the international situation in the 1940s, as Hodgkin (1981: 288) states it, the Vietnamese "faced with a varying combination of partly competing, partly collaborating imperialisms, French, Japanese, British and American, with *Kuomintang* China." At that time, Vietnam was considered an important base to attack southern China¹⁹ after Japan became more aggressive and invaded China in the 1930s (Hodgkin 1981: 288). The Japanese military eventually entered Vietnam and shared with the French the control of Vietnam in the early 1940s. From the perspective of China, suppression of Japan's military activities in Vietnam was desired. However, in the viewpoint of the French, they were afraid that China would take over Vietnam again if Chinese troops were to enter Vietnam on the excuse of suppressing Japanese forces (Jiang

¹⁹ In the view point of Japan, domination of Vietnam and its northern trade-route was essential for effective control of southern China since the *Tonkin* Railway from *Haiphong* to *Yunnan* was a vital source of supplies for *Kuomintang* China (Hodgkin 1981: 288).

1971: 181; Chen 1969: 134). For the Vietnamese people, the question was how to maintain their national identity and achieve national independence from the imperialisms. These were the issues considered priority by their leaders such as Ho Chi Minh. Ho's Chinese strategy was to keep Chinese forces away from Vietnam and minimize the possibility of a Chinese comeback in Indochina. Politically speaking, Ho refused to allow the Chinese army to enter Vietnam (Jiang 1971: 107; Chen 1969: 146) as well as instigating an anti-Chinese movement (Jiang 1971: 228-240; Chen 1969: 121); culturally, Romanized Vietnamese was considered a distinctive feature of cultural boundary between Vietnam and China. These considerations have impelled Ho in favor of Romanization rather than Han characters which are used in China.

2.7 Orthographic reform in Taiwan

The first written language in Taiwan was the so-called *Sinkang* Manuscripts 新港文書, a Romanized system used to write the vernacular of indigenous Siraya tribes during the Dutch occupation of Taiwan in the seventeenth century. Thereafter, classical Han writing was adopted as an official language by the government, and *Koa-a-chheh* was treated as the popular writing for the public during the Koxinga and the Qing occupations. In addition to those two written forms, other Romanized systems have been developed to write Taiwanese²⁰ and Hakka languages since the nineteenth century. After Taiwan became a part of Japan (1895-1945), Japanese writing became the official written language in Taiwan. After World War II, Mandarin Chinese became the standard of writing under Chiang Kai-shek's occupation of Taiwan. The relation between language and political status in Taiwan is shown in **Table 6**.

²⁰ Taiwanese is also called Taigi, Tai-yu, Holoee, Southern Min, or Min-nan. The broad definition of Taiwanese includes all the indigenous languages, Hakfa, and Holoee. Occasionally, Taiwanese refers to Holoee only. Holoee speakers account for 73.3% of Taiwan's population, Hakfa 12%, indigenous 1.7%, and Mandarin speakers who came to Taiwan with KMT account 13% (Huang 1993:21).

Table 6. Relation between language and political status in Taiwan

Period	Political status	Spoken Languages	Writing Systems**
-1624	Tribal society	Aboriginal	Tribal totem
1624-1661	Dutch colonialism	Aboriginal/Taiwanese*	Sinkang (新港文) Classical Han (文言文)
1661-1683	Koxinga colonialism	Aboriginal/Taiwanese	Classical Han Sinkang
1683-1895	Qing colonialism	Aboriginal/Taiwanese	Classical Han Koa-a-chheh (歌仔冊) Peh-oe-ji Sinkang
1895-1945	Japanese colonialism	Aboriginal/Taiwanese/Japanese	Japanese Classical Han Colloquial Han (in Taiwanese) Colloquial Han (in Mandarin) Peh-oe-ji Kana-Taiwanese (臺式假名)
1945-2000	KMT colonialism	Aboriginal/Taiwanese/Mandarin	Chinese (Mandarin) Taiwanese Aboriginal

* Taiwanese means Hakka-Taiwanese and Holo-Taiwanese here.

** The order of listed writing systems in each cell of this column do not indicate the year of occurrences. The first listed orthography refers to the official written language adopted by its relevant governor.

2.7.1 Colonial background

Taiwan is a multilingual and multiethnic island country. There are currently more than twenty languages in Taiwan, including indigenous languages, Hakka, Holo Taiwanese, and Mandarin Chinese (Grimes 2000). Generally speaking, Taiwan's population can be divided into four primary ethnic groups: indigenous (1.7%), Hakka (12%), Holo (73.3%), and Mainlanders (13%) (Huang 1993: 21). Hakka and Holo are the so-called Han people. In fact, many of them are the descendants of intermarriage between sinitic immigrants and local Taiwanese aboriginals during the *Koxinga* and *Qing* periods. Mainlanders are the latest immigrants from China, who came to Taiwan with the Chiang Kai-shek's KMT regime in the late 1940s. Although Hakka, Holo, and Mainlanders are all immigrants originally from China, they do have different national identities. For example, most of the

Table 6. Relation between language and political status in Taiwan

Period	Political status	Spoken Languages	Writing Systems**
-1624	Tribal society	Aboriginal	Tribal totem
1624-1661	Dutch colonialism	Aboriginal/Taiwanese*	Sinkang (新港文) Classical Han (文言文)
1661-1683	Koxinga colonialism	Aboriginal/Taiwanese	Classical Han Sinkang
1683-1895	Qing colonialism	Aboriginal/Taiwanese	Classical Han Koa-a-chheh (歌仔冊) Peh-oe-ji Sinkang
1895-1945	Japanese colonialism	Aboriginal/Taiwanese/Japanese	Japanese Classical Han Colloquial Han (in Taiwanese) Colloquial Han (in Mandarin) Peh-oe-ji Kana-Taiwanese (臺式假名)
1945-2000	KMT colonialism	Aboriginal/Taiwanese/Mandarin	Chinese (Mandarin) Taiwanese Aboriginal

* Taiwanese means Hakka-Taiwanese and Holo-Taiwanese here.

** The order of listed writing systems in each cell of this column do not indicate the year of occurrences. The first listed orthography refers to the official written language adopted by its relevant governor.

2.7.1 Colonial background

Taiwan is a multilingual and multiethnic island country. There are currently more than twenty languages in Taiwan, including indigenous languages, Hakka, Holo Taiwanese, and Mandarin Chinese (Grimes 2000). Generally speaking, Taiwan's population can be divided into four primary ethnic groups: indigenous (1.7%), Hakka (12%), Holo (73.3%), and Mainlanders (13%) (Huang 1993: 21). Hakka and Holo are the so-called Han people. In fact, many of them are the descendants of intermarriage between sinitic immigrants and local Taiwanese aboriginals during the *Koxinga* and *Qing* periods. Mainlanders are the latest immigrants from China, who came to Taiwan with the Chiang Kai-shek's KMT regime in the late 1940s. Although Hakka, Holo, and Mainlanders are all immigrants originally from China, they do have different national identities. For example, most of the

Holo and Hakka people identified themselves as Taiwanese. However, according to Ong's investigation, 54% of the surveyed Mainlanders still identified themselves as Chinese. Only 7.3% identified themselves as Taiwanese, and the rest were neutral (Ong 1993: 87). Their divergent opinions about identity on Taiwan is also a factor influencing the promotion of Taiwanese language(s).

In addition to being a multiethnic society, Taiwan has been colonized by several foreign regimes since the seventeenth century. Prior to foreign occupation, Taiwan was a primitive society with many different indigenous tribes, which did not belong to any countries, not China nor Japan. In 1624, the Dutch occupied Taiwan and established the first alien regime in Taiwan, which led to the first use of Roman scripts. The first Romanization was used to write the now-extinct indigenous Siraya language. In 1661, Koxinga, leader of a remnant force of the former Chinese *Ming* Dynasty, failed to restore the Ming Dynasty against the new Qing Dynasty; therefore, he retreated to Taiwan. Koxinga expelled the Dutch and established a sinitic regime in Taiwan as a base for retaking the Mainland. Confucianism and civil service examination were thus imposed in Taiwan during Koxinga's regime and at a later time Qing Dynasty. The Koxinga regime was later annexed by the Chinese Qing Dynasty in 1683. During the late Qing period, *Peh-oe-ji*, or Scripts of Vernacular Speech, the second Romanization in Taiwan, was introduced by western missionaries in the second half of the nineteenth century. Peh-oe-ji is mainly used for Holo Taiwanese, which constitutes the majority of Taiwan's current population. Two centuries later after Qing's occupation, the sovereignty of Taiwan was transferred from China to Japan as a consequence of the Sino-Japanese war of 1895. At the end of World War II, Japanese forces surrendered to the Allied Forces. *Chiang Kai-shek*, the leader of the Chinese Nationalist (KMT²¹ or *Kuomintang*) took over Taiwan on behalf of

²¹ KMT was the ruling party in Taiwan since 1945 until 2000, in which year Chen Shui-bian, the

the Allied Powers under General Order No.1 of September 2, 1945 (Peng and Ng 1995: 60-61). Simultaneously, Chiang Kai-shek was fighting against the Chinese Communist Party in Mainland China. In 1949, Chiang's troops were completely defeated and then pursued by the Chinese Communists. At that time, Taiwan's national status was supposed to be dealt with by a peace treaty among the fighting nations. However, because of Chiang's defeat on the Mainland China, Chiang decided to occupy Taiwan as a base from which he would fight to retake the Mainland (Kerr 1992; Peng and Ng 1995; Su 1980; Ong 1993). Consequently, Chiang's political regime, Republic of China (ROC) was renewed in Taiwan and has remained there since 1949.²²

In comparing the Taiwanese and Vietnamese histories, although both of them were dominated by alien forces, there were distinctive differences which explain why the Taiwanese did not establish an independent country at the end of World War II as did the Vietnamese. First of all, Taiwan was never established as an independent state. In fact, the idea of establishing a modern nation-state did not dawn on the Taiwanese people until 1947,²³ when the February 28 Massacre occurred (Ng 1994: 202). The concept of nation, as Anderson defines it, is an "imagined political community" (Anderson 1983: 6). It indicates that nations are invented rather than being naturally born. Thus, having collective historical experience is quite important for the members of a group to restore their collective memory

presidential candidate of opposition party Democratic Progressive Party, was elected the new president.

²² Republic of China was formerly the official name of the Chinese government (1912-1949) in China, but was replaced by the People's Republic of China (PRC) in 1949. Once the ROC was renewed in Taiwan, the ruling party KMT claimed that ROC has sovereignty over Mainland China and is the only legal government, which represents all of China. This extravagant claim was not changed until 2000, when the opposition party DPP won the presidential election.

²³ The exact date for the origin of Taiwanese independence movement may vary from scholar to scholar. But, the February 28 Massacre of 1947, in which over twenty thousand of Taiwanese people were killed by Chiang's troops (Kerr 1992:303), is usually considered the origin of current Taiwanese independence movement (Ng 1994).

and thus take further actions to achieve their collective objectives. For example, many revolts during the Qing occupation of Taiwan gathered masses by restoring their historical experience of “Anti-Qing and restoring the Ming dynasty”反清復明. By the end of the nineteenth century, the collective memory that the Taiwanese had was to propose a restoration and renaissance of the Ming, which would have put in place an alien historic Chinese dynasty rather than a localized modern Taiwanese nation-state. It indicates that creation of a Taiwanese nation-state was not yet a mature idea. Second, the re-occupation of Taiwan after 1945 by the Chinese KMT was a crucial point in explaining why the Taiwanese have not yet form their own nation-state. Would Vietnam have become independent if it was still under the control of China at the end of World War II? During the occupation of Taiwan, the Chinese ROC regime reconstructed a Chinese identity for the Taiwanese people in the way which will be detailed in the following sections.

2.7.2 Orthographic tradition

The classic writing system of Han characters²⁴ 文言文 was the official written language before twentieth century in all Hanji cultural areas, including Taiwan, China, Vietnam, Korea, and Japan (Hannas 1997; Chiung 1997). Classical Han writing had become the official written language, much as Latin was in pre-modern Europe (Norman 1991: 78). In addition to those standard Han characters used for classical Han writing, there were some local or vernacular characters, which were popularly used by local people for the purpose of vernacular writing. One of these was *Chu-Nom* 字喃 in Vietnam and *Koa-a-chheh* 歌仔冊 in Taiwan.

²⁴ The classical writing of Han characters was also called classical Chinese or literary Chinese.

The Koa-a-chheh²⁵ (literally, song book) orthography was named because many traditional song books were written in this system. Each sentence was composed of either five or seven characters. Koa-a-chheh texts were not regarded as classical Han writing because they were written close to the spoken form. Even so, they were distinct from modern colloquial forms. Therefore, they might be classified as pre-modern colloquial writings.

Although Koa-a-chheh writing was regarded as a writing system, the usage of Han characters could vary from user to user. Different writers might choose different characters to express the same spoken form. In general, people chose characters from an inventory of available characters or created new characters. There are three main principles for choosing from available characters.

First, the same etymon is written with the same Han characters. Such as 想 ‘think’ in the Koa-a-chheh sentence 蚊仔想著足怨切 (*bang-a siunn tiah chiok oan-chheh*: the mosquito become very sad while he thought about that).

Second, the original meaning of a character was ignored; only the sound was attended to. This idea is equivalent to rebus writing in English. For example, 足 ‘*chiok*’ possesses the meaning of ‘foot’ in classical Han writing, however, it is used to express ‘very’ in the Koa-a-chheh sentence above.

Third, the pronunciation of a character was ignored, and its meaning borrowed to express the same meaning in different languages. For example, the meaning of 蚊 was borrowed to express ‘mosquito’ in Taiwanese.

²⁵ For more details regarding Koa-a-chheh, see Ong (1993a: 169-215), and Ong Sun Liong 王順隆 at <http://plaza16.mbn.or.jp/~sunliong/kua-a-chheh.htm>

2.7.3 Romanization prior to 1945

Romanization in Taiwan prior to 1945 can be divided into two eras. The first era of Romanization is *Sinkang* writing, which was mainly devised for the indigenous languages, lasted from the first half of the seventeenth century during the Dutch occupation of Taiwan to the early nineteenth century. The second Romanization is *Peh-oe-ji* writing. It was devised for Holo and Hakka Taiwanese languages, and it has existed in Taiwan since the second half of nineteenth century.

2.7.3.1 Sinkang Romanization (1624-early nineteenth century)

Sinkang writing was the first Romanization and the first writing system in the history of Taiwan. It was devised by Dutch missionaries and applied mainly for writing Siraya, an indigenous language in southwest plain of Taiwan. Although Romanized writing in indigenous language had been mentioned in earlier historical materials such as *Chulo Koanchi* 諸羅縣志 ‘Topographical and Historical Description of Chulo’ (1717), and *E-tamsui-sia Kiagi* 下淡水社寄語 ‘A Glossary of the Lower Tamsui Dialect’ (1763), Romanization in Sinkang was not well known until the discovery of Sinkang manuscripts.

For the Dutch the main purposes for occupying Taiwan were to convert the locals to Christianity as well as to exploit resources. As Campbell has described it, “during that period they [i.e., Dutch] not only carried on a profitable trade, but made successful efforts in educating and Christianising the natives; one missionary alone having established a number of schools and received over five thousand adults into the membership of the Reformed Church” (Campbell, 1903: vii). The natives around Sinkang²⁶ were first taught

²⁶ Sinkang , originally spelled in *Sinkan*, was the place opposite to the *Tayouan* where the Dutch had settled in 1624. The present location is *Sin-chhi* 新市 of *Tainan* county.

Christianity through the learning of the Romanization of Sinkang.²⁷ There were some textbooks and testaments written in Romanized Sinkang, such as the *The Gospel of St. Matthew in Formosan Sinkang Dialect and Dutch (Het Heylige Euangelium Matthei en Jonannis Ofte Hagnau Ka D'llig Matiktik, Ka na Sasoulat ti Mattheus, ti Johannes appa. Overgefet inde Formosaansche tale, voor de Inwoonders van Soulang, Mattau, Sinckan, Bacloan, Tavokan, en Tevorang)*, which was translated and published by Daniel Gravius in 1661 (Campbell 1996; Lai 1990: 121-123).

After Koxinga drove the Dutch from Taiwan, the Roman scripts were still used by the plain tribes for some period. There were several manuscripts found after those native languages had disappeared. Those manuscripts were written either wholly in language(s) of native aborigines or they were bilingual texts with Romanization and Han characters. The majority of the manuscripts were either sale contracts, mortgage bonds, or leases (Murakami 1933: IV). Because most of those manuscripts were found in Sinkang areas and were written in Sinkang language, they were named Sinkang Manuscripts by scholars, or *Hoan-a-khe* 番仔契 (literally, the contract of barbarians) by the public (Lai 1990:125-127).

There are 141 examples of Sinkang Manuscripts reported in Murakami's Sinkan Manuscripts. The earliest dated from 1683, and the most recent one is dated 1813. In other words, the indigenous people continued to use the Romanization for over a century-and-a-half after the Dutch had left Taiwan (Murakami 1933: XV). In recent years, additional some forty manuscripts were found (cited in Li 2002). Among them, fifteen were reported by Li (2002), and were all written in the Siraya language.

²⁷ For detailed discussion on early Formosan literacy, see Heylen (2001a).

2.7.3.2 Peh-oe-ji Romanization (1865-present)

If Sinkang writing represents the first foreign missionary activities in Taiwan, then the development of *Peh-oe-ji*²⁸ reveals the return of missionary influences after the Dutch withdrawal from Taiwan.

More and more missionaries came to preach in China in the seventeenth century, even though there were many restrictions on foreign missionaries during the Qing Dynasty. The restrictions on foreign missionaries were continued until the Treaty of *Tientsin* was signed between the Qing Dynasty and Western countries in 1860. Taiwan, at that time, was under the control of the Qing Dynasty, therefore, foreign missionaries were allowed after that treaty. Consequently, the first mission after the Dutch, settled in *Taioan-hu*²⁹ by missionary James L. Maxwell and his assistants in 1865 (Hsu 1995:6-8; Lai 1990: 277-280).

Before missionaries arrived in Taiwan, there were already considerable missionary activities in southeast China. They had started developing Romanization of some languages such as Southern Min and Hakka. For instance, the first textbook for learning the Romanization of the Amoy³⁰ dialect, *Tngoe Hoan Ji Chho Hak* 'Amoy Spelling Book' was published by John Van Nest Talmage in 1852 in Amoy. That Romanization scheme was called *Poe-oe-ji* in Taiwan. It means the script of vernacular speech in contrast to the complicated Han characters of *wenyan*.

Peh-oe-ji was originally devised and promoted by missionaries for religious purposes. Consequently, most of its applications and publications are related to church activities. Those applications and publications of *Peh-oe-ji* since the nineteenth century can be

²⁸ For details about *Peh-oe-ji*, see Cheng 1977, Chiung 2000c, and Tiuⁿ 2001.

²⁹ Present *Tailam* or *Tainan* city.

³⁰ Amoy was a dialect of Southern Min, and was regarded as mixed *Chiang-chiu* and *Choan-chiu* dialects. The Amoy dialect was usually chosen by missionaries as a standard for Southern Min.

summarized into the following six categories: 1) textbooks, 2) dictionaries, 3) translation of the Bible, catechisms, and religious tracts, 4) newspaper, 5) private note-taking or writing letters, and 6) other publications, such as physiology, math, and novels.

Missionaries' linguistic efforts on the Romanization are reflected in various Romanized dictionaries.³¹ Medhurst's *A Dictionary of the Hok-keen Dialect of the Chinese Language* published in 1837 is considered the first existing Romanized dictionary of Southern Min compiled by western missionary (Ang 1996: 197-259; Heylen 2001: 146). Douglas' *Chinese-English Dictionary of the Vernacular or Spoken Language of Amoy* of 1873 is regarded as the remarkable dictionary of influence on the orthography of Peh-oe-ji (Ang 1993b: 1-9). After Douglas' dictionary, most Romanized dictionaries and publications followed his orthography with no or only minor changes. Generally speaking, missionary efforts on the linguistic features of Southern Min and Peh-oe-ji have reached a remarkable level since Douglas's dictionary (Ang 1993b: 5). William Campbell's dictionary *Ē-m̄ng Im ê Sin Jī-tián (A Dictionary of the Amoy Vernacular Spoken throughout the Prefectures of Chin-chiu, Chiang-chiu and Formosa* 1913), which was the first Peh-oe-ji dictionary published in Taiwan, is the most widespread Romanized dictionary in Taiwan. By 1987, this dictionary had been published in fourteen editions (Lai 1990).

The first New Testament in Romanized Amoy, *Lán ê Kìù-chú Iá-so' Ki-tok ê Sin-iok* was published in 1873, and the first Old Testament *Kū-iok ê Sèng Keng* in 1884. The wide use of Peh-oe-ji in Taiwan was promoted by the missionary Reverend Thomas Barclay while he published monthly newspaper *Tâi-oân-hú-siá' Kàu-hōe-pó*³² 'Taiwan Prefectural City Church

³¹ For more details about the development of early Romanized dictionaries in Southern Min, refer to Heylen (2001b).

³² Taiwan Prefectural City Church News has changed its title several times, and the recent title (1988) is *Taiwan Kau-hoe Kong-po* (Taiwan Church News). It was published in Peh-oe-ji until 1970, and thereafter it switched to Mandarin Chinese (Lai, 1990: 17-19).

News' in July 1885. In addition to publications related to Christianity, there were some other publications written in Peh-oe-ji, such as *Pit Sòan ê Chhò Hák* 'Fundamental Mathematics' by Ûi-lîm Gê in 1897, *Lāi Gōa Kho Khàn-hō-hák* 'The Principles and Practice of Nursing' by G. Gushue-Taylor in 1917, the novel *Chhut Sí-Sòaⁿ* 'Line between Life and Death' by Khe-phoàn Tēⁿ in 1926, and the collection of commentaries *Cháp-Hāng Koán-Kiàn* 'Opinions on Ten Issues' by Pêe-hóe Chhòa in 1925.

Usually, the religious believers apply Peh-oe-ji writing to their daily life after they acquire the skill of Romanization. For example, they may use Peh-oe-ji as a skill of note taking or writing letters to their sons or daughters or friends in addition to reading the Bible. Peh-oe-ji was widely used among church³³ people in Taiwan prior to the 1970s.³⁴ Among its users, women were among the majority. Most of those women did not command any literacy except in Peh-oe-ji. Today, there are still a few among the elder generations especially women who read only Peh-oe-ji.

From the perspective of literacy, it is not surprising that Peh-oe-ji would occur in Taiwan. Because Romanized Peh-oe-ji writing is much easier than the classical Han writing, it provides the general public a convenient tool for literacy. *Poe-hoe Chhoa* 蔡培火 points out that writing in Han characters is a heavy burden for most Taiwanese. He therefore advocates using Taiwanese Romanization to liberate the illiterate. He mentions the relationship between new Taiwan and Roman scripts in his book "Opinions on Ten Issues³⁵," which was published in 1925.

Pún-tó lāng lóng-kiōng ū saⁿ-pah lák-cháp-bān lāng, kīn-kīn chiah chha-put-to jī-cháp-bān lāng u hák-būn, kiám m̄-sī chin chió mah? Che sī sím-mih goân-in neh?

³³ Especially the Presbyterian Church in Taiwan.

³⁴ *Taiwan Kau-hoe Kong-po* (Taiwan Church News), which was originally published in Peh-oe-ji, switched to Mandarin Chinese in 1970. I use this year as an indicator to the change of Peh-oe-ji circulation.

³⁵ "Chap-Hang Koan-Kian" 十項管見 was entirely written in Peh-oe-ji

Chit hāng, sī lán ka-tī bē-hiáu khòⁿ hák-būn tāng; chit hāng, sī siat-hoat ê lāng bô ū cháp-hun ê sêng-sim. Iáu koh chit hāng, chiū-sī beh òh hák-būn ê bûn-jī giân-gú thài kan-kè hui-siōng oh-tit òh (Chhòa 1925: 14-15).

We Taiwanese have 3.6 millions of population, but only two hundred thousand of them are literate. Isn't it too few? What are the reasons? One is that we think little of literacy; another reason is that the ruler is not sincere to promote education; and the third is that the orthography [i.e., Hanji] and language are too difficult to learn literacy.

In a Hanji dominated society, having command of Hanji is considered a mark of intelligence and prestige. On the other hand, literacy other than Hanji is regarded as common and vulgar. Among Peh-oe-ji users, the majority were women who did not command any Han character or orthography except Peh-oe-ji. This phenomenon reflects the fact that the traditional women of lower social status were not likely to be educated with Han characters, and they had to choose the 'childlike' but easily learned Peh-oe-ji if they wished to be able to read and write. There was a widespread bias against Peh-oe-ji as observed by Rev. Thomas Barclay, the editor and publisher of TPCCN. He mentioned the Peh-oe-ji in the first issue of TPCCN, as follows:

Khó-sioh lín pún-kok ê jī chin oh, chió chió lāng khòⁿ ê hiáu-tit. Só-í goán ū siat pát-mih ê hoat-tō, ēng pēh-oē-jī lâi ìn-chheh, hō lín cheng-lāng khòⁿ khah khoài bat... Lāng m̄-thang phah-sng in-ūi i bat Khóng-chú-jī só-í m̄-bián òh chit-hō jī; iā m̄-thang khòⁿ-khin i, kóng sī gin-á só-thák-mê (Barclay 1885).

Because your traditional Han characters are too difficult to learn, only a few of you can read and write. That is why we have tried to print books in Peh-oe-ji, so you will be able to read easily...do not think you do not have to learn Peh-oe-ji if you already knew Hanji, neither look down the *Peh-oe-ji*, nor regard it as a childish writing.

Although Peh-oe-ji was originally devised for religious purposes, it is no longer limited to religious applications since the contemporary *Taibun*³⁶ 台文 movement was raised in the late 1980s. Peh-oe-ji has been adopted by many Taibun promoters as one of

³⁶ Taibun literally means Taiwanese literature or Taiwanese writing. It refers to the orthography issue in the Taiwanese language movement since 1980s. For details of the modern movement of written Taiwanese, see Chiung (1999:33-49).

the Romanized writing systems to write Taiwanese. For example, famous Taibun periodicals such as *Tâioân-jī*, *Tâi-bûn Thong-sin* and *Tâi Bûn Bóng Pò* adopt Peh-oe-ji as the preferred Romanization for writing Taiwanese. In addition, there were since 1996 a series of novels translated from world literature into Peh-oe-ji in a planned way by the members of 5% *Tai-ek Ke-oe*³⁷ ‘5% Project of Translation in Taiwanese’.

In short, the Peh-oe-ji was the basis for all forms of Romanization of modern Taiwanese colloquial writing. Even though several different schemes of Romanization for writing Taiwanese have been developed, many of them, such as TLPA were derived from Peh-oe-ji.³⁸ Peh-oe-ji and its derivatives are the most widely used Romanization even today.

2.7.4 Romanization after 1945

Romanization after 1945 can be categorized into Romanized Chinese and Romanized Taiwanese in terms of the language the Romanization is used for. The development of Chinese Romanization can be traced back to the KMT’s language planning in China in the first half of the twentieth century. Generally speaking, Chinese Romanization is not considered by the KMT to be an independent writing system, but rather as a set of phonetic symbols for transcribing Han characters. As for the Taiwanese Romanization, it is intentionally ignored (once forbidden) by the KMT regime, but it has become the main concern of the promoters of the Taibun movement. For most Taibun promoters,

³⁷ In November of 1995, some Taiwanese youths who were concerned about the writing of Taiwanese decided to deal with the Taiwanese modernization and loanwords through translation from foreign language into Taiwanese. The organization 5% Project of Translation in Taiwanese was then established on February 24, 1996. Its members have to contribute 5% of their income every month to the 5% fund. The first volume includes 7 books. They are *Lear Ong*, *Kui-a Be-chhia*, *Mi-hun-chhiuⁿ e Kui-a*, *Hoa-hak-phin e Hian-ki*, *Thiⁿ-kng Cheng e Loan-ai Ko⁻su*, *Pu-ho⁻-lang e Lek-su*, and *Opera Lai e Mo⁻-sin-a*, published by Tai-leh press in November 1996.

³⁸ For more information about different Romanized schemes, see Iuⁿ 1999.

Romanization is regarded as an independent orthography and thus has been proposed as an orthographic system for writing Taiwanese.

2.7.4.1 Romanization for Mandarin Chinese

In the late nineteenth and the early twentieth century, the language issues that concerned Chinese government and the general public were: 1) the unification of pronunciation (of Han characters) and the formation of a national language, and 2) the transition of written language from the classical standard ‘*wenyen*’ to colloquial writing ‘*baihua*.’³⁹

The pronunciation of Beijing was eventually chosen as the national language and the oral standard for reading Han characters. At that time, neither domestically created phonetic symbols nor western Roman scripts were considered independent orthographies, but rather were regarded as auxiliary tools for learning the national language (DeFrancis 1950: 221-236; Norman 1988: 257-263). *Jhuyin Zimu* 注音字母 or Phonetic Alphabet, a set of symbols derived from radicals of Han characters was devised and proposed by *Duyin Tongyihue* ‘the Committee on Unification of Pronunciation’ in 1913 and later officially adopted by the Chinese government in 1918 as a tool for learning the correct pronunciation of the national language. It was revised slightly in 1928 and renamed *Jhuyin Fuhao*⁴⁰ or Phonetic Symbols (henceforth NPS1) in 1930. This scheme was used in China until 1958 when *Hanyu Pinyin* (henceforth HP) was officially adopted and replaced it. *Jhuyin Fuhao* was brought to Taiwan by the KMT in 1945 and it has been taught through Taiwan’s national education system and has been in continuous use ever since.

³⁹ For details, see Chen 1999; DeFrancis 1950; Gao 1992; Jhou 1978; Norman 1988; Png 1965; and Tsao 1999.

⁴⁰ The purpose of using *Jhuyin Fuhao* 注音符號 ‘sound-annotating symbols’ is to “dispel any faint hope that they were to be used as bona fide writing systems” (Chen 1999: 189). This scheme was later called *Guoyu jhuyin fuhao di yi shih* or National Phonetic Symbols, 1st Scheme in Taiwan (henceforth NPS1).

The first Romanized phonetic scheme proposed and recognized by the Chinese government was the *Guoyu Luomazi* 國語羅馬字 or National Language Romanization, which was approved in 1928 (Chen 1999: 182). Although Guoyu Luomazi was approved by the government, in reality it was not promoted for practical use. It was even less widely used in comparison to another Romanized scheme *Latinxua sin wenz*⁴¹ (Norman, 1988: 259). Guoyu Luomazi was later brought together with Jhuyin Fuhao to Taiwan by the KMT during the Chiang Kai-shek occupation of Taiwan. The Guoyu Luomazi scheme was later revised and renamed *Guoyu jhuyin fuhao di er shih*⁴² or National Phonetic Symbols, 2nd Scheme (henceforth NPS2) and promulgated by Taiwan's Ministry of Education (MOE) in 1986.

Although both NPS1 and NPS2 were officially promulgated by the KMT regime in Taiwan, only NPS1 is taught in schools and is actually used as an auxiliary tool for learning to pronounce Mandarin. In contrast, NPS2 is excluded from school curriculum and is simply used to transliterate Chinese names into other languages (Chen 1999: 189). As a matter of fact, not only NPS2 but also other traditional Romanized schemes devised by foreigners, such as Wade-Giles and Postal schemes are also used for Mandarin transliteration.⁴³ Moreover, the majority of Taiwanese people who are not educated in the Romanized schemes, simply adapt the English K.K. phonetic symbols⁴⁴ to transliterate as

⁴¹ *Latinxua sin wenz* was first published in 1929 and employed among the 10,000 Chinese living in the USSR. It was considered an autonomous writing system and later introduced to China. This scheme was very popular especially in the Northwestern part of China where were under the control of the Chinese Communist Party at that time (Chen 1999:184-186).

⁴² Guoyu Luomazi was renamed National Phonetic Symbols 2nd Scheme, to distinguish it from the 1st scheme of Jhuyin Fuhao.

⁴³ Even for the government, different departments and different counties may use different Romanized schemes.

⁴⁴ In Taiwan, the Kenyon and Knott (K.K.) phonetic symbols are taught in schools serving as instructions of pronunciations in learning English.

they saw fit (Yu 1999). Consequently, the Romanization of Han characters in Taiwan is in a seriously chaotic situation. For example, 曹 may be transliterated *tsao*, *tsau*, *ts'ao*, *ts'au*, *chao*, *chau*, *chhao*, *chhau*, *c'ao*, *c'au*, and so on.

As a result of this chaos, much attention was paid to transliteration issues, with the government trying to unify the Romanized schemes in the late 1990s. In April 1999, a national conference on transliteration schemes was held by the MOE, focusing on the review of the four existing Romanized schemes, i.e. Wade-Giles, NPS2, HP, and *Tongyong Pinyin* (TYP).⁴⁵ In July of the same year, the Executive Yuan (*Heng-cheng-i'*ⁿ; similar to the presidential cabinet in western countries) announced that HP would be adopted as the standardization for future transliteration. However, this announcement soon aroused opposition and protests against the HP system in August (Chiang, Luo, Tiuⁿ, and Yu, 2000). Consequently, the final decision on a transliteration scheme was intentionally left until after the presidential election of March 2000. However, the result of the 2000 presidential election fell short of the KMT's expectation. The pro-Taiwanese Independence Democratic Progressive Party (DPP) won the election and the KMT lost power for the first time since 1945 after ruling Taiwan for fifty-five years.

Since the 2000 presidential election, the Mandarin transliteration issue has remained unresolved and it has even engendered more heated controversy and conflict between the new government and the pro-Chinese opposition parties, i.e., KMT, People First Party (*Cinmindang*), and New Party (*Sindang*). On September 16 of that year, the Mandarin Promotion Council (*Guoyu Tuesing Ueyuanhoe*) under the MOE of the new government approved the TYP for Mandarin transliteration. In October that also aroused criticism and protests from the opposition parties. *Ma Yingjiu*, the KMT mayor of Taipei started a

⁴⁵ For more discussion on these schemes, see Cheng 2000; Tsao 1999.

boycott against the new government on the pinyin issue. He criticized the TYP saying that it was not an international standard for Mandarin Chinese; it would create an obstacle for Taiwan to achieve globalization. He further asserted that Hanyu Pinyin would have to be adopted to achieve this objective (Jhongshih 2000; Jhongyangse 2000; Mingrihbao 2000).

This ‘pinyin controversy,’ or dispute over Mandarin transliteration schemes has been generally considered the biggest crisis for the new government aside from the ‘anti-nuclear power plant’ event.⁴⁶ In fact, the current pinyin controversy is probably the most widely broadcasted dispute over the issues of transliteration that has ever occurred in Taiwan. One may be curious as to why a linguistic issue could result in such heated debate and political crisis. There are two contributing factors: 1) the different national identity possessed by different parties, 2) the fact that the ruling DPP was a minority party in the Legislative Yuan (*Lip-hoat-i⁴⁷*; similar to congress).

The conflicts between TYP and HP fundamentally resulted from different perspectives of national identity rather than different linguistic designs. From the point of view of Chinese nationalism, it was important to avoid contributing to pro-Taiwanese independence activities. During the old days while the pro-unification KMT was a ruling party, there was no doubt or problem in using Guoyu Luomazi with regard to the nationalism issue. However, the pro-unification support has been flagging since the late 1980s when the native political movement began its ascendancy flourishing (Chiung 1999a). Moreover, the pro-Taiwanese independence DPP became the ruling party after the 2000 presidential election. In this strong pro-Taiwanese independence atmosphere, using a transliteration scheme different from China was suddenly perceived in a new way as an

⁴⁶ The 4th nuclear power plant in Taiwan was approved and under construction in the 1990s during the period of KMT government. After the DPP became the ruling party, the new government stopped its construction. Consequently, it aroused protests and boycott from the opposition parties, which proposed to unseat the new president Chen Shui-bian.

attempt of the new government to move toward Taiwanese independence.⁴⁷ Although Mayer Ma criticized the TYP scheme of not being an internationally recognized system, what he really implied was that TYP was distinct from the ‘Chinese PRC standard.’⁴⁸ What really concerned Ma was that TYP would lead further de facto division between Taiwan and China (Kang 2000; Te 2000).

Although the DPP won the 2000 presidential election, the new government could do little until the next election of legislative representatives at the end of 2001. The fact that the KMT and PFP together are the majority in the Legislative Yuan has inflated the pinyin controversy. To some extent, what mostly interested the KMT was putting up fronts to justify boycotting the new government rather than arriving at a finding on a transliteration scheme. In this case, to unseat the new president was probably the first priority, and the adoption of HP was simply the second. For example, those who accused the new government of not adopting the HP did not accuse the KMT of promulgating the NPS2.

In order to better understand the pinyin controversy, the history and differences between TYP and HP are briefly described in the following. TYP (Tongyong Pinyin), literally means general or common transliteration scheme. TYP was proposed and devised by a research fellow at Academia Sinica, *Yu Buocuyan* and his associates in the late 1990s. The fundamental purpose of this new design was to find the maximum transferability between the Hanyu Pinyin scheme and Taiwanese vernacular scheme. In other words, Yu tried to devise a transliteration scheme, which could be used for both Mandarin and Taiwanese languages without lethal conflicts in learning. There were two proposals for

⁴⁷ For example, in a press conference on November 29, 2000 the *Guotaiban* (Office for Taiwan Affairs) of the PRC claimed that someone was trying to promote Taiwanese independence in the areas of culture and education through using a different transliteration scheme from Hanyu pinyin.

⁴⁸ For example, if Ma really was concerned about the international standardization and globalization, he should also abandon the Jhuyin Fuhao, which is used in Taiwan only.

TYP, i.e., TYP1 (*Ka-sek*) and TYP2 (*It-sek*) (Cheng 2000; Yu 2000). In the scheme of TYP1, the letter p represents [p] in IPA; however, in TYP2, the letter p represents [p^h], and b represents [p]. TYP2 was the scheme involved in the pinyin controversy. Generally speaking, TYP2 is considered to be the revised version of Hanyu Pinyin, with minor change such as the initial symbols q, x, and zh (see Table 7). It was estimated that there were around 15% differences between transliterations using TYP2 and HP (Chiang & Huang, 2000).

HP (Hanyu Pinyin) literally means transliteration scheme for the Han standard language (That is to say, Mandarin). HP was designed during the mid-1950s in China and officially promulgated in 1958 by the government of the People's Republic of China. HP is currently considered the only legal transliteration scheme in China for the transcription of Modern Standard Chinese (Wenzi 1983). It was also adopted by the International Standardization Organization in 1982 as the standard form for transcribing Chinese words (Chen 1999: 187). Although the original design of HP was to lead ultimately to an autonomous orthography, it has been continuously claimed by the Chinese government that HP was intended for learners as an aid in learning standard Chinese (Chen 1999: 188-189; Hannas 1997: 24-25; Norman 1988: 263; Wenzi 1983: 6-21). In fact, not only HP, but also other phonemic writing schemes, such as Guoyu Luomazi and Jhuyin Fuhao have always been prevented from serving as independent writing systems. From the point of view of Chinese nationalism, Han characters embody the functionality of linguistic uniformity. In contrast, alphabetic writing would result in linguistic polycentrism and further be harmful to national unity (DeFrancis 1950: 221-236; Norman, 1988: 263). Apparently, national and political unity was considered to have priority over literacy by the Chinese government.

Table 7. Mandarin consonants represented by IPA, HP, TYP, and Jhuyin Fuhao

IPA

p ph m f t th n l k kh h tɕ tɕh ɕ tʂ tʂh ʂ ɿ ts tsh s

Hanyu Pinyin 漢語拼音

b p m f d t n l g k h j q x zh ch sh r z c s

Tongyong Pinyin 通用拼音

b p m f d t n l g k h ji ci si jh ch sh r z c s

Jhuyin Fuhao 注音符號

ㄅ ㄆ ㄇ ㄈ ㄊ ㄊㄥ ㄋ ㄌ ㄍ ㄍㄥ ㄏ ㄐ ㄑ ㄒ ㄓ ㄔ ㄕ ㄖ ㄗ ㄘ ㄙ

2.7.4.2 Romanization for Taiwanese

At present, because spoken Taiwanese is not well standardized, there are correspondingly many proposals for writing Taiwanese. Those proposals may be generally divided into two groups based on their scripts: Han character scripts⁴⁹ and non-Han character scripts. Non-Han characters may be further divided into two subtypes: A new alphabet, such as *Ganbun* (Hangul-like scheme) designed by Ang Ui-jin, or a ready-made alphabet, which makes use of the present Roman letters or Jhuyin Fuhao to write Taiwanese. To better understand the development of non-Han schemes, the number of each category is listed in Table 8 based on the 64 collections by Iuⁿ and Tiuⁿ (1999).

⁴⁹ This is the traditional way to write Taiwanese in classical style, as Hancha in classical Korean prior to the invention of Hangul. There are several problems encountered when writing colloquial speech by using Han characters. For more details in relation to this issue, see Chiung (1999a: 50-51, 1998).

Table 8. Number of each category of non-Han schemes

	Roman script	Revised Jhuyin Fuhao	Revised Kana	Hangul -like	Total
-1895	1	0	0	0	1
1985-1945	4	0	2	0	6
1945-1987	15	3	0	1	19
1987-	30	6	0	2	38
Total	50	9	2	3	64

Owing to the wide use of the personal computer and electronic networks in Taiwan from the 1990s onward, most orthographic designs that need extra technical support other than regular Mandarin Chinese software could not survive. Therefore, the majority of recent Taiwanese writing schemes were either in Han characters-only, Roman script-only or mixed scripts with Roman and Han.⁵⁰ At present, there are mainly three competing Romanized schemes in relation to the Taiwanese language, i.e., Peh-oe-ji, TLPA, and TYP. Among the Romanization proposals, Peh-oe-ji is definitely regarded as an independent orthography rather than just a transliteration scheme (Cheng 1999; Chiung 2001c). However, so far there is no common agreement of whether TLPA and TYP would be treated as writing systems or simply transliteration schemes.⁵¹

As introduced in the previous section, Peh-oe-ji is the traditional Romanization for writing Taiwanese (Holo and Hakka). Prior to the Taibun movement in the 1980s, Peh-oe-ji was the only Romanized scheme in practical use for writing Taiwanese. Compared to other Romanized schemes, Peh-oe-ji is still the Romanization with the richest inventory of

⁵⁰ Roman and Han mixed scheme was proposed mainly to solve the problem that some native Taiwanese words do not have appropriate Han characters (Cheng 1990, 1989). To some extent, it is like the mixture style of Korean Hancha plus Hangul or Japanese Kanji plus Kana. For more discussion on these three Taiwanese schemes, see Chiung (1999a) and Tiuⁿ (1998).

⁵¹ For comparisons and contrasts between Peh-oe-ji and TLPA, see Cheng (1999) and Khou (1999).

written work, including dictionaries, textbooks, literature works, and other publications in many areas (Iuⁿ, 1999).

TLPA or Taiwanese Language Phonetic Alphabet was proposed in the early 1990's by the Taiwan Languages and Literature Society.⁵² The major motivation for the TLPA designers to modify Peh-oe-ji is to overcome the inconvenience of typing and transmitting some special symbols on modern computing networks. TLPA has been revised several times, and the latest version was finalized in 1997. In January 1998, the MOE announced that TLPA would be adopted as the official Romanized scheme for Hakka and Holo Taiwanese. The hasty decision adopting TLPA immediately aroused fierce opposition from Peh-oe-ji users and Taibun-promoting groups.⁵³ Based on the petition proposed by the Taibun groups against TLPA, we can summarize three factors initiating the controversy. First, the MOE's procedure for determining the Romanized scheme for Taiwanese was considered insufficiently detailed. Taibun groups object, moreover, that TLPA was approved without public hearings and discussions. The protestors even considered the whole event a strategy of the MOE to polarize Taibun groups. Second, the TLPA was simply a theoretical design that had never seen practical use. However, Peh-oe-ji has been used since the early nineteenth century, and thus had a long history of literacy convention. Third, Peh-oe-ji is a definite orthography rather than a transliteration. However, the designers of TLPA have never clarified whether or not TLPA is intended to be a writing system.⁵⁴ The ambiguity about orthographic status of TLPA was a further weakness pointed to by the protestors.

⁵² Taiwan Languages and Literature Society (*Tai-oan Gi-bun Hak-hoe* 台灣語文學會) was established in 1991. For more information, visit its website at < <http://www.tlls.org.tw> >

⁵³ For example, see Ngou (1998), Lu (1998), Teⁿ (1998), and the "Petition against the MOE's adoption of TLPA" (March 14, 1998).

⁵⁴ For example, in the design of TLPA, Taiwanese tones are represented by Arabic numerals, such as *hun5* (cloud) representing the fifth tone. People criticized that numerals should not be used in an orthography.

In brief, the major differences between Peh-oe-ji and TLPA are phonetic symbols, tone marks and spelling rules. In regard to phonetic symbols, there are three differences. That is, *ch* and *chh* in Peh-oe-ji were modified and became *c* and *ch* in TLPA; back vowel *o* was represented by *oo* in TLPA; and superscript *n* was replaced with regular letters *nn*, such as in the case of *tinn* ‘sweet’. In TLPA, tones were represented by Arabic numerals. For example, *tai5* ‘platform’ represents tone five. As for the spelling, some Peh-oe-ji spellings such as *eng*, *ek*, *oa*, and *oe* were revised to *ing*, *ik*, *ua*, and *ue*.

2.7.5 Contemporary language movement

The National Language Policy or monolingual policy was adopted both during the Japanese and KMT occupations of Taiwan (Huang 1993; Tsao 1999; Png 1965; Tiuⁿ 1974). In the case of KMT’s monolingual policy, the Taiwanese people were not allowed to speak their vernaculars in public. Moreover, they were forced to learn Modern Standard Chinese and to identify themselves as Chinese through the national education system (Cheng 1996; Tiuⁿ 1996). As Hsiao (1997: 307) has pointed out, “the usage of Mandarin as a national language becomes a testimony of the Chineseness of the KMT state,” the Chinese KMT regime is trying to convert the Taiwanese into Chinese through Mandarin monolingualism. Consequently, research by scholars such as Chan (1994) and Young (1988) has revealed that a language shift toward Mandarin is in progress. Huang (1993: 160) goes so far as to suggest that the indigenous languages of Taiwan are all endangered. In addition, the monolingual policy has shown strong impact on three-generational relationships among Taiwanese families (Chuang 2000).

In response to the KMT’s National Language Policy, Taiwanese promoters have protested against the monolingual policy, and have demanded bilingual education in schools. This is the so-called *Taigibun Untong* ‘Taiwanese language movement’ that has substantially arisen since the second half of the 1980s (Hsiao 1997; Erbaugh 1995; Huang

1993; Li 1999; Lim 1996; Chiung 1999a). There are two core issues for the Taiwanese language movement. First, the movement wishes to promote spoken Taiwanese⁵⁵ to allow speakers to maintain their vernacular speech. Second, the movement aims to promote and standardize written Taiwanese in order to develop Taiwanese (vernacular) literature.⁵⁶ Because written Taiwanese is not well standardized and not taught in the national education system, Taiwanese speakers have to write in Modern Written Chinese (MWC) instead of in Written Taiwanese (WT). In other words, people speak in Taiwanese, but write in MWC. Although more than a hundred orthographies have been proposed by different enthusiasts for the standardization of written Taiwanese, most of the designs have probably been accepted and used only by their own designers. Moreover, many of the designs were never applied to practical writing after they were devised. Because of the wide use of the personal computer and electronic networks in Taiwan since the 1990s, most orthographic designs, which require extra technical support other than regular Mandarin software, are unable to survive. Therefore, the majority of recent Taiwanese writing systems are either in Han characters, Roman alphabet or a mixed system combining Roman and Han, as Cheng (1990) and Tiuⁿ (1998) have documented.

The orthographic situation in Taiwan is as complicated as Taiwan's political status and people's national identity. Linguistically, people in Taiwan have to face the issue whether to use MWC or WT as their written language. Further, people who choose WT, have to decide which scripts will be adopted while they are writing in Taiwanese.

⁵⁵ The broad definition of Taiwanese includes all the indigenous languages, Hakka, and Holoee. Occasionally, Taiwanese refers to Holoee only, which is the language spoken by the Holo people. Holoee is also called Holo Taiwanese, Taigi, Tai-yu, Holoee, Southern Min, or Min-nan.

⁵⁶ Although the issues of written Taiwanese include Hakka and indigenous languages, most literary works are written in the Holo language. This fact makes the Holo language the focus of the written Taiwanese. Therefore, the term "written Taiwanese" in this paper refers only to the written form of the Holo language, if not specified.

Politically, Taiwan is currently in an ambiguous political status, i.e., neither nominally an independent Republic of Taiwan nor substantially a province of the People's Republic of China (Peng and Ng 1995). This political ambiguity mirrors people's divergent national identity, which is usually categorized as 1) Taiwanese-only, 2) Chinese-only, and 3) both Taiwanese and Chinese.⁵⁷ Consequently, the diversity of the public's national identity led to different political claims, i.e., independence, unification with China, or maintaining the status quo.⁵⁸

The contemporary Taiwanese language movement since the 1980s reflects Taiwan's socio-political complexity and its colonial background. In terms of Fishman's (1968) nationalism and nationism, it reveals the controversial relationship among Chinese nationalism-nationism,⁵⁹ Taiwanese nationalism and Taiwanese nationism as illustrated in **Figure 2**.

⁵⁷ Taiwanese-only means that people identify themselves as Taiwanese rather than Chinese. "Both Taiwanese and Chinese" refers to people who identify themselves as both Taiwanese and Chinese. For more information about the identity issue, readers may refer to Chang (1993), Si (1994, 1996, 1997, 1998), Huang (2000), and Tse (2000).

⁵⁸ Their proportion of supporters may vary slightly from poll to poll, but in general, less than 20% of Taiwan's populations in recent years are in favor of unification with China (Huang 2000; Tse 2000).

⁵⁹ At the beginning of Chinese KMT's occupation of Taiwan, Chinese nationalists may have held different opinions from Chinese nationalists. However, later on when the use of Mandarin by people in Taiwan dramatically increased, the objects of Chinese nationalism and Chinese nationism became the same. That is, to keep using Mandarin since it has dominated educational and governmental functions in Taiwan. Therefore, I do not distinguish Chinese nationalism from Chinese nationism here.

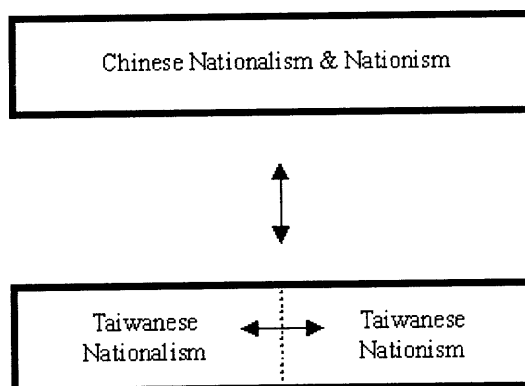


Figure 2. Relationship among Chinese nationalism-nationism, Taiwanese nationalism, and Taiwanese nationism.

In the dimension of nationalism and nationism, it reveals the political tensions between Chinese and Taiwanese. Chinese nationalism can be inherited from the internal Chinese KMT and as well as external People's Republic of China. The strong conflicts between KMT's Chinese nationalism and Taiwanese nationalism were overt in the anti-KMT movement⁶⁰ during the second half of the 1980s and the entirety of the 1990s. The conflicts between PRC Chinese nationalism and Taiwanese nationalism started in the late 1980s⁶¹ and reached the climax in 1999 when the former president Teng-hui Lee claimed that Taiwan and China hold "special state to state" relationship.

In the dimension of Taiwanese, it shows the expanding tension between Taiwanese nationalism and Taiwanese nationism. Some Taiwanese politicians and intellectuals who lead socio-political movements, such as *Hong-Beng Tan*, *Sui-kim Phenn* and *Chhun-Beng*

⁶⁰ In this paper, I consider 1986, when the first native opposition party Democratic Progressive Party was born, the beginning of anti-KMT movement though its origin can be traced back to the 1970s. KMT lost its ruling status in the 2000 presidential election; therefore, 2000 was considered the end of the anti-KMT movement.

⁶¹ For example, Iu-choan Chhoa, Cho-tek Khou, and Lam-iong Tenn claimed the independence of Taiwan to the public in 1987.

Ng, do not valorate the Taiwanese language movement as a necessary step even though they identify themselves as Taiwanese rather than Chinese. In their ideology, they disapprove of the KMT's strict national language policy; however, they have come to accept the results of the national language policy. In other words, they recognize the legitimate status of the colonial language, i.e., Mandarin Chinese as the official language since it has become widespread in Taiwan after more than fifty years of promotion. However, they are criticized by Taiwanese nationalists who say they have ignored the threat of Chinese nationalism from China. From the perspective of Taiwanese nationalism, the Taiwanese language is not only a medium for communication, but also a part of history and spirit of Taiwan. Moreover, it is considered a national defense against Chinese nationalism of the PRC and the ROC (Lim 1996, 1997, 1998; Li 1999; Chiuⁿ 1996).

The complexity of this social-political background has prevented any of Taiwan's domestic scripts from being promoted to national status. Therefore, in contrast to Vietnam, Korea, and Japan, Taiwan has uniquely a system of vernacular writing still under development. Both internal and external factors, as I have outlined, have contributed to the stalemate in regard to the development of Taiwanese orthography.

In terms of the internal demands for literacy and an anti-feudal hierarchy, written Taiwanese was not effectively promoted at the right time when the public met their demands in the early twentieth century during the Japanese occupation.⁶² Nowadays, Taiwan has shifted from a traditional feudal society to a modern one, in which a minimum of nine years of compulsory education has been required since 1968. Taiwan's Minister of Interior (2002) claimed that Taiwan's current population has reached a literacy rate of 96%,

⁶² The causes are complicated. On one hand, it was because of the opposition from the Japanese colonialist; on the other hand, the elites' preference for Han characters was caused by their internalized socialization and misunderstanding of the nature and function of Han characters.

indicating that the majority of people in Taiwan, to some extent, have acquired literacy skills in Han characters and Modern Written Chinese. This fact has thus reduced the public's literacy demands for a new orthography.

From the perspective of external factors, because of the complexity and ambiguity of the political relationship between Taiwan and China, Han characters are not substantially regarded as a foreign script by the people in Taiwan. In contrast, Roman scripts are generally considered foreign even though Romanized writing has been used in Taiwan for hundreds of years (Chiung 2001a). As Gelb (1952: 196) has pointed out, "in all cases it was the foreigners who were not afraid to break away from sacred traditions and were thus able to introduce reforms which led to new and revolutionary developments." The weak awareness of national identity by Taiwanese people from Chinese people has thus shaken the promotion of Roman scripts and written Taiwanese.

2.7.6 The future of Romanization in Taiwan

Despite the uncontested evidence that any Romanization is much more efficient than Han characters, Romanizations are currently not widely accepted by people in Taiwan (Chiung 2001a). Writing in Roman script is regarded as the low language in a digraphic situation. There are several reasons for this phenomenon:

First, the preference of the people in Taiwan for Han characters is caused by internalized socialization. Because Han characters have been adopted as the official orthography for two thousand years, being able to master Han characters well is the mark of an educated adult in the Han cultural areas. Writing in scripts other than Han characters may be regarded as childlike writing (Chiung 2001c). For example, when *Tai-oan-hu-sia*" *Kau-hoe-po*, the first Romanized Taiwanese newspaper, was published in 1885, the editor and publisher Rev. Thomas Barclay exhorted readers of the newspaper not to "look down at Peh-pe-ji; do not regard it as childish writing" (Barclay 1885).

Second, misunderstanding of the nature and function of Han characters has confirmed people's preference for Han characters. Many people believe that Han characters are ideally suited for all forms of the Han language family, which includes Holo and Hakka Taiwanese. They believe that Taiwanese cannot be expressed well without Han characters, because Han characters are logographs and each character expresses a distinctive semantic function. In addition, many people believe Lian Heng's (1987) claim that "there are no Taiwanese words which do not have corresponding characters." However, DeFrancis (1996: 40) has pointed out that Han characters are "primarily sound-based and only secondarily semantically oriented." In DeFrancis' opinion, it is a myth to regard Han characters as logographic (DeFrancis 1990). He even concludes that "the inefficiency of the system stems precisely from its clumsy method of sound representation and the added complication of an even more clumsy system of semantic determinatives" (DeFrancis 1996: 40). If Han characters are logographs, the process involved in reading them should be different from phonological or phonetic writings. However, research conducted by Hung et al. has pointed out that "the phonological effect in the reading of the Chinese characters is real and its nature seems to be similar to that generated in an alphabetic script" (Hung, Tzeng and Tzeng 1992: 128). Their research reveals that the reading process of Han characters is similar to that of phonetic writing. In short, there is no sufficient evidence to support the view that the Han characters are logographs.

The third reason that Romanization is not widespread in Taiwan is because of political factors. Symbolically, writing in Han characters was regarded as a symbol of Chinese culture by Taiwan's ruling Chinese KMT regime. Writing in scripts other than Han characters was forbidden because it was perceived as a challenge to Chinese culture and Chinese nationalism. For example, the Romanized New Testament *Sin Iok* was once seized in 1975 because writing in Roman script was regarded as a challenge to the orthodoxy

status of Han characters (Li 1996). In addition, it was reported by New York Times in 1974 that a Romanized Taiwanese-English dictionary was officially banned and the Romanization was cited as the reason.⁶³

Usually, many factors are involved in the choice and shift of orthography. From the perspective of social demand, most people in current Taiwan have already attained the reading and writing skills in Han characters to a certain high level. It seems not easy for them to abandon their literacy conventions and shift to a completely new orthography. However, for the younger generation who are at the threshold of literacy, a new orthography may be attractive to them if it is much easier to learn to read and write. If education in Romanized writing could be included in schools and taught to the beginners, Romanization could quickly be a competing orthography to Han writing.

From the perspective of politics, political transitions usually affect the language situation (Si 1996). In the case of Taiwan, the current ambiguous national status and the diversity of national identity reflect people's uncertain determinations on the issues of written Taiwanese. On the other hand, people's uncertain determinations on the Taibun issues also reflect the political controversy on national issues of Taiwan. My research (Chiung 2001a) on the attitudes of Taiwanese college students toward written Taiwanese reveals that national identity is one of the most significant factors that affect students' attitudes toward Taiwanese writing. It is true that national identity played an important role in the orthographic transition of Vietnam, where Romanization eventually replaced Han characters and became the official orthography (Chiung 2002a, 2002b; DeFrancis, 1977). Will this replacement happen to the case of Taiwan? Whether or not Roman script will replace Han characters and Taiwanese replace Chinese depends on people's orthographic

⁶³ New York Times. 1974. Guide to dialect barred in Taiwan: dictionary tried to render local Chinese sounds. September 15, section 1, p.15. Available at <<http://hoklo.org/HokloCulture/Articles/?show=1#1>>

demands and their attitudes toward written Taiwanese. Moreover, people's national identity will play a crucial role in the transition. From my point of view, Han characters, at least, will retain their dominant status until the Taiwanese people are released from their ambiguity in regard to national identity.

CHAPTER 3

WRITING SYSTEMS AND LEARNING EFFICIENCIES

This chapter provides general discussion on the relationship between writing systems and learning efficiencies, and it consists of three sections. Section 3.1 discusses the overall concepts of writing systems, including criteria for classification, evolution of writing, orthography and learning efficiency, and script reform. Section 3.2 provides discussion on the specific orthography of Han characters, including the formation of Han characters and their evolution toward a syllabic writing system. The last section demonstrates how Roman alphabet could be adopted to compete with Han characters.

3.1 Development of writing systems

3.1.1 Criteria of classification

Traditionally, people regard orthography as either logographic or phonographic writing systems. For example, Han characters are considered logographic or ideographic system, and English alphabet is phonographic or phonetic according to traditional ideas about writing systems. However, this kind of distinction between logographic and phonographic writing systems is not always accurate and appropriate because neither consists of purely logographic or phonographic symbols in its writing. For instance, ‘semi-’ and ‘-er’ in English refer to the semantic meaning ‘half’ and ‘agent.’ In contrast, none of the characters in the Chinese word 雷射 ‘laser’ is related to ‘laser.’ Characters have been chosen simply to represent the sound of English loanword ‘laser.’

If Han characters are logographs, the process involved in reading them should be different from phonological or phonetic writing. However, research conducted by Hung et al. has pointed out that “the phonological effect in the reading of the Chinese characters is real and its nature seems to be similar to that generated in an alphabetic script” (Hung et al. 1992: 128). Their research reveals that phonological activation is just as involved in the reading process of Han characters as in phonetic writing system. In addition, Li (1992: 21) has pointed out that 形聲 or the radical-phonetic principle,⁶⁴ which employ both semantic and phonetic⁶⁵ components in the structure of most Han characters, increased from 27 % (eleventh century B.C.) to 90 % (twelfth century A.D.). In other words, most of the existing Han characters are decomposable into semantic and phonographic components and are not exclusively logographic. DeFrancis (1996: 40) has pointed out that Han characters are “primarily sound-based and only secondarily semantically oriented.” In DeFrancis’ opinion, it is just a myth to regard Han characters as logographic. He even asserts that “the inefficiency of the system stems precisely from its clumsy method of sound representation and the added complication of an even more clumsy system of semantic determinatives” (DeFrancis 1996: 40).

Regarding the orthographic issues, Gelb (1952) and Smalley (1963) have developed a remarkable classification of the world’s writing systems. That is, orthographic systems should be classified based on the category of sound units they represent. According to this norm, orthographies can be grouped into distinct types based on four linguistic levels, i.e., word (or morphemic) orthographies, syllabic orthographies, phonemic orthographies, and phonetic writing systems, corresponding to the sound units of words (or morphemes),

⁶⁴ For example, the left side or radical of 江(/kang/, river) refers to the semantic meaning “water,” and the right side or phonetic 工 represents the sound /kang/.

⁶⁵ To be exact, it is a syllabic sound unit.

syllables, phonemes, and phonetic features. Han characters are the best example of morphemic⁶⁶ writing, because almost every Han character can symbolize a (bound) morpheme or an (unbound) word and can combine with other characters to form new words. Japanese *Kana* is an example of syllabic writing (syllabary). Examples of phonemic⁶⁷ systems are Taiwanese *Peh-oe-ji*, Vietnamese, English, and many other languages using Latin scripts. In phonemic writing, the symbols correspond to phonemes.⁶⁸ Phonemic or alphabetic⁶⁹ writing systems are more widespread than others among the world's existing writing systems. As for phonetic writing, all the detailed features of sound difference are reflected in the transcription. This kind of script is usually the tool of a phonetician who wishes to transcribe spoken language data into written form for linguistic analysis. Although it transcribes smaller sound units than phonemes, it does not increase, and may even decrease the degree of efficiency. The primary reasons are that a phonetic transcription is more complex than a phonemic one, and a native speaker may not be aware of the different phonetic features, which require highly trained ears to detect. Consequently, as Smalley (1963: 5) says, "a genuinely phonetic writing system can never be the basis for a popular orthography." Therefore, phonetic writing is not widespread except among linguists.

⁶⁶ Gelb (1952) uses the term 'word-syllabic' in his book. In terms of DeFrancis (1990), the Han writing system is a form of morphosyllabic writing.

⁶⁷ Phonemic writing should be distinguished from phonetic writing. Many people confuse phonemic with phonetic writing, and treat them as identical. In fact, in most cases what people call phonetic writing is actually a phonemic writing system (Smalley 1963: 6).

⁶⁸ Whether or not symbols and phonemes have a one-to-one relationship depends on various languages. In *Peh-oe-ji*, the Taiwanese Romanization, each symbol generally represents only one phoneme. In contrast, English symbols have more than one corresponding relationship.

⁶⁹ Alphabetic system is defined here in terms of Smalley's "writing systems based directly upon the individual phonemes of language" (Smalley 1963: 6).

3.1.2 Evolution of writing

After Gelb posited the new criteria of classification, he further pointed out that writing has evolved from larger representative sound units to smaller ones, that there has been a tendency to develop from word-syllabic systems to syllabic and finally to alphabetic systems (Gelb 1952: 191).

Why is the evolution moving from larger sound units toward smaller ones? In this author's view, it is based on the human abilities to analyze speech utterances. In other words, writing systems are not likely to move toward smaller representative units until people have been able better to analyze speech utterance into smaller sound units. For example, the Japanese syllabic writing system could not occur until its devisers became aware of the fact that their word forms consisted of syllables. The Korean phonemic system had to wait until Korean scholars could analyze syllables into phonemes.

In ancient times prior to the advent of full writing systems, the most primitive ways of communication using visible symbols made use of the descriptive-representational and the identifying-mnemonic devices (Gelb 1952: 191). In the descriptive-representational device, simple pictures that contain elements, which are important for the transmission of communication, are drawn. Those simple pictures are similar to the drawings of artists, but lack aesthetic embellishments. In the identifying-mnemonic device, pictures are drawn with symbols that are used to assist in recording or identifying persons or subjects. The desire to record things through similar symbols constituted an important factor leading to the development of real writing (Gelb 1952: 192).

Suppose we lived ten thousand years ago before the creation of any written language. How could I put into written form the events that happened to me "one day, I saw a lion while I was walking in the forest. The lion was eager to eat me with its widely open mouth. I was so afraid. I had no choice but to pick up a rock and fight against the lion..."? I might

draw a picture as shown in **Figure 3**. In this case, the picture is considered a kind of special ‘writing,’ which corresponded to a large holistic sound sequence, i.e., a full story passage. In other words, this descriptive-representational picture encoded many hidden semantic meanings the drawer tried to express. Although this type of device is simple and compact, it is not easy for readers to decode. When someone else see this picture, for instance, they may infer that “the drawer or a third person wanted to show his bravery, so he went to a forest and challenged a lion...” or that “the guy wanted to show his kindness, so he fed a hungry lion a slice of meat.” In other words, different readers may have different or even conflicting ideas about the meaning of the events expressed by this drawer. These many interpretations indicate that descriptive-representational devices may be easy to encode, but they are neither easy nor precise to decode.

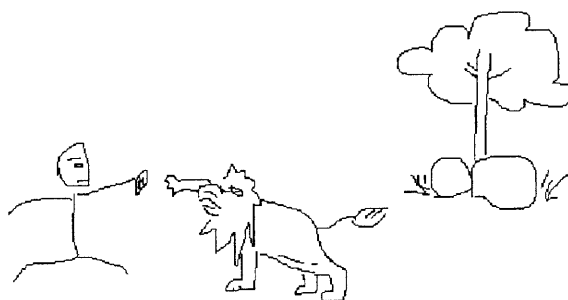


Figure 3. Descriptive-representational picture

After having only mitigated success with descriptive-representational pictures, the drawer might evolve to a series of symbols or simpler pictures, as shown in **Figure 4**, to express “one day, I saw a lion while I was walking...” **Figure 4** is an example of identifying-mnemonic devices, in which the drawer tries to segment a full story, and each segment or symbol is used to help remember and identify an object or a being. For example, the eye symbol was used to refer to “see.” Thus, the correspondence between the written

symbols and their spoken counterparts are further established. Once it was discovered that words could be expressed in written form, the word or logographic writing system was nearly established. One must say “nearly” because a process of conventionalization is still necessary for the device to become a mature word writing system, in which an individual word is always and conventionally expressed by an identical sign, so that misunderstanding among readers and writers will not occur. For example, in Chinese ‘fish’ and the symbol or graph 魚 always correspond. So, people understand that its meaning is ‘fish’ whenever they see the symbol 魚.



Figure 4. Identifying-mnemonic device

Although in the history of writing word writing systems represent great progress over purely descriptive-representational systems, they are not likely to survive and further develop to a full writing system without the process of phonetization that involves “attaching to a sign a phonetic value independent of the meaning which this sign has as a word” (Gelb 1952: 193-194). As Gelb (1952: 193) pointed out that “to create and memorize thousands of signs for thousands of words and names existing in a language and to invent new signs for newly acquired words and names is so impracticable that either a logographic writing can be used as a limited system or it must find new ways to overcome the difficulties,” phonetization is thus an important principle to simplify word writing system in order to deal with a large and potentially open lexical inventory. Another modern term for phonetization is ‘rebus writing.’ For example, the drawing of a bee and a leaf expresses the

word ‘belief,’ and ‘4 U’ means ‘for you’ in English. Another example of rebus writing in Taiwanese, 七 /ts^hit/ ‘seven’ and 桃 /tho/ ‘peach’ together means *chhit-tho* ‘to play.’ Yet another example is 𠄎 seen by Jerold A. Edmondson in Beijing, where 𠄎 ‘water’ and 七 ‘seven’ was used for *qi* [tɕ^hi] ‘paint.’

Because pure word writing systems could hardly exist, users have to either move toward syllabic or phonemic writing. Even Chinese, the most logographic of all the existing systems, is not a purely so. Recall that Li (1992: 21) has pointed out that by the twelfth century, the radical-phonetic principle had grown to 90% of the inventory of Chinese characters. Although Chinese characters have tended to move somewhat towards syllabary, these have not fully developed due mainly to social-cultural factors. Consequently, Chinese characters have fossilized on the roadway toward a syllabic writing. That is why Gelb considered them to be a word-syllabic writing system. We will have more detailed discussion of Chinese characters in the next sections.

3.1.3 New proposed criteria for the typology of writing

In addition to the popular tripartite schemes of logographies, syllabaries, and alphabets, a variety of other typologies have been proposed as summarized by Daniels (1996: 8-10). Nevertheless, none of them could provide sufficient descriptions of any of the various world’s writing systems. Therefore, this author is proposing the following scheme for describing the structure of any given orthographies in details.

Before one can demonstrate the scheme, one needs to clarify a couple of key terms, i.e., writing and grapheme.

‘Writing’ has various definitions varying from investigator to investigator. In this dissertation, this author cites Gelb’s definition “a system of human intercommunication by means of conventional visible marks” as the broad definition of writing. On the other hand,

a narrow definition of writing might be “a system of more or less permanent marks used to represent an utterance in such a way that it can be recovered more or less exactly without the intervention of the utterer” (Daniels 1996: 3). A full writing system is further defined on the basis of this narrow definition as “a system of more or less consistent symbols that represent speech sounds within lexical level.”

The term ‘grapheme’ is defined as “the smallest contrastive unit in the writing system of a language” (Crystal 1992: 161). In other words, a grapheme is the minimal graphic unit that can represent a change in semantic meaning. In the English example, letters *p* and *t* are considered two graphemes because they make distinctions between *pin* and *tin*.

This scheme I am proposing may be called ‘Universal Orthography’ and is intended to describe precisely all full writing systems found in the world. The Universal Orthography (UO) is arranged by the space of placement (columns) and manners of correspondence (rows) as illustrated with the English example in **Table 9**.

The space of placement includes direction category, which consists of rightward (R), leftward (L), upward (U), and downward (D). What I meant by ‘direction’ here refers only to the placement of graphemes within a lexical item. The direction of sentences is not considered since it is, to some extent, a matter of typesetting. For example, Han characters in sentences were and, to some extent still can be, horizontally arranged from right to left instead of the modern usage of left-to-right. Although there are two different typesettings, they have the same orthographic structure. In the direction category, rightward and leftward mean that graphemes are arranged horizontally from left to right or from right to left, respectively. Upward and downward mean that graphemes are arranged vertically from bottom to top or from top to bottom, respectively. Their properties of distinctive features are binary, either ‘+,’ which means ‘yes,’ or ‘-,’ meaning ‘no.’ For example, the direction of English is R(+), L(-), U(-), D(-). The reason we need the pairs leftward vs. rightward and

upward vs. downward instead of only leftward and upward is that in some cases, such as Chinese characters, they have the characteristics R(+), L(+), U(+), D(+). When featuring with R(+) and L(+), it means that some characters are arranged from left to right and the others are from right to left.

The manners of correspondence include three major categories: 1) representational levels, 2) added information, and 3) correspondences between graphemes/grapheme sets and speech sounds. The representational levels can be further divided into morpheme, syllable, phoneme, and phonetic feature levels. The added information includes supra-segmental and extra-information. The category ‘supra-segmental’ means whether or not the supra-segmental features of a language, such as tones or voice quality, are expressed in its writing. The category ‘extra information’ refers to any other information added to orthographic structure. For example, some semantic features of Han characters, which were constructed under the radical-phonetic principle, are regarded as extra information.

The category of correspondences between graphemes/grapheme sets and speech sounds is divided into 1) one-to-one relationship, 2) one grapheme/grapheme set corresponds to multiple sounds, 3) multiple graphemes/grapheme sets correspond to one sound, and 4) ambiguity. The subcategory ‘ambiguity’ means whether or not graphemes/grapheme sets provide precise correspondence, where ‘+’ means the orthographic symbol and sound do not correspond to each other. For example, most cases of phonetic components of Han characters, such as 虹, provide only a hint to its corresponding sound of its character. In the case of 虹 (pronounced as /hong/), the phonetic component 工 (individually pronounced as /kong/) does not exactly match its actual pronunciation of 虹.

For better understanding of this new proposal, the principle of Universal Orthography applied to English are seen in **Table 9**.

Table 9. English example of Universal Orthography

		directions			
		R	L	U	D
Morphemes	Stem	-	-	-	-
	Prefix	-	-	-	-
	Infix	-	-	-	-
	Suffix	-	-	-	-
Syllables	Syllables	-	-	-	-
	Initial	-	-	-	-
	Medial	-	-	-	-
	Final	-	-	-	-
Phonemes	Consonants	+	-	-	-
	Vowels	+	-	-	-
Phonetic features		-	-	-	-
Supra-segmental		-	-	-	-
Extra information		-	-	-	-
One-to-one		+	-	-	-
One-to-multiple		+	-	-	-
Multiple-to-one		+	-	-	-
Ambiguity		-	-	-	-

3.1.4 Orthography and Learning efficiency

Having clarified the classification of different orthographic systems, this author now turns to the relationship between learning efficiency and these orthographies. The efficiency issue can be examined through the perspective of the Universal Orthography, i.e., manners of correspondence and space of replacement.

Generally speaking, a smaller sound unit represented by a unique symbol will be more efficient than a larger representational unit. Among the types of phonemic, syllabic, and morphemic (or morpho-syllabic) writing systems, phonemic writing are the most efficient systems because they require the learning of the fewest number of symbols to represent the fullest range of speech. In contrast, the least efficient systems are those of the morphemic type, since in that writing system every morpheme has to be individually

learned (Smalley 1963: 7). This is because human languages always have a limited number of consonants and vowels (these sounds can be regarded as phonemic units), and a higher number of phonemic combinations (to form syllables) and an even greater number of syllabic combinations (to form multi-syllabic words).

In addition to the number of symbols to be learned, how these symbols are arranged over the space can also affect the efficiency of learning. In other words, the spelling rules in a phonemic system or the placement rules in the Han characters will also determine how easy or difficult they will be to learn. In general, the easier rules a system has, the easier it is to learn.

Consider a couple of examples to demonstrate the learning efficiency of different type of writing systems. In English, all lexical items can be expressed by the 26 alphabet letters. In contrast, the number of Hanji learned by elementary students in Taiwan is about 2,669 (Chiung 1999: 52). Norman (1988: 72-73) has pointed out that an ordinary literate Chinese person knows and uses somewhere between 3,000 and 4,000 Han characters. In other words, literacy learners have to learn minimally some 3,000 characters to be able to achieve norm levels of literacy in Chinese. Moreover, if readers wish to be able to have comprehensive knowledge of every characters in the cannon of Chinese literature, they have to learn as many as ca 100,000 characters, which were collected in the dictionary *Hanyu Da Zidian* (1999). Even if the readers have learned all 100,000 characters, they might not be able to read characters in such cases when writers add a stroke or radical to the existing characters to form a new one.

Someone may say that Han character learners do not have to learn all the characters. Rather, they just need to learn the limited number of basic components or the so-called radicals which constitute Han characters. Once they have learned the components, they will naturally comprehend all characters. That might sound reasonable but, in fact, is not

reliable. The total inventory of Han character components varies from dictionary to dictionary. According to the Chinese Character Component Standard of GB 13000.1 Character Set for Information Processing published in 1997, there are a total of 560 components. Even if we do not count the number of characters, but the number of components composing the characters, still numbers 560 and that is still a big burden compared to the 26 alphabet letters. Moreover, the placement of components and their corresponding sound and semantic interpretation in Han characters are not always consistent. This inconsistency usually raises further troubles for readers decoding the exact sound and meaning of the characters. This inconsistency issue will be further discussed in the next sections.

3.1.5 Criteria for a new orthography

Script reform can be divided into two aspects: 1) adjustment of the existing orthography, such as spelling reform or simplification of characters, and 2) replacement of the entire existing writing system. Any attempts to reform existing script usually raise heated debates. As Coulmas (1989: 241) notes, “Once written norms are established, they attract emotional attachment, and hence discussions about the reform of a given orthography or script often resemble a religious war more than a rational discourse, generating more heat than light.” Whether or not a new orthography will be accepted by the public in a society depends on various linguistic and non-linguistic factors. Smalley (1963: 34-52) has proposed five criteria of an adequate writing system as follows (in order of importance):

First, maximum motivation for the learner, acceptance by its society, and controlling groups such as the government. Second, maximum representation of speech. Third, maximum ease of learning. Fourth, maximum transfer. By ‘transfer,’ Smalley means whether or not the users of a new scripts can apply their skills in new scripts to the trade or

national language. Fifth, maximum ease of reproduction. Smalley's criteria for a new orthography will be adopted to examine Peh-oe-ji, the Romanized Taiwanese writing system in the next chapter.

3.1.6 Psycholinguistic view of reading

Goodman (1967: 33) referred to reading as “a psycholinguistic guessing game” and defined reading as “a selective process. It involves the partial use of available minimal language cues selected from perceptual input on the basis of reader's expectation. As this partial information is processed, tentative decisions are made to be confirmed, rejected, or refined as reading progress.” In other words, reading involves prediction, selection, and sampling of cues, instead of a precise process of letter-by-letter or word-by-word identification.

Reading requires visual and nonvisual information (Smith 1994: 66). Visual information is the printed material in front of reader's eyes, and it goes away when the lights go out. Nonvisual information or prior knowledge is the information you already have behind the eyeballs, such as knowledge of the language, and subject matter. In general, the more nonvisual information a reader has, the less visual information the reader needs (Smith 1994: 66). A proficient reader is someone skilled in anticipating upcoming print from contextual cues and prior knowledge (Robeck and Wallace 1990: 14). S/he needs neither a letter-by-letter nor a word-by-word recognition process, but a process of forward and backward saccades. On the contrary, inefficient readers rely too heavily on grapho/phonetic information.

Letter and word identifications are two of basic concerns in the study of reading process in alphabetic writing systems. Letter identification is “the process by which any two visual configurations are cognized to be the same” (Smith 1994: 106). Letters further constitute words. There are three major theories about word identification: whole-word

identification, letter-by-letter identification, and an intermediate position involving the identification of letter clusters (Smith 1994: 119). The frequent controversy over word identification is the argument that whether or not phonological awareness plays an obligatory role in its process. If so, to what extent phonological awareness is involved in word recognition. Many experimental results, such as Claudia, Turvey, and Lukatela (1992), Perfetti, Zhang, and Berent (1992), and Orden, Stone, Garlington, Markson, Pinnt, Simonfy, and Bricchetto (1992) have revealed that phonology plays an obligatory part in the word identification process. In short, phonics may not be the only strategy available for mediated word identification, but, it indeed frequently plays a central role in reading process (Smith 1994: 134).

Although there is a consensus that, to some extent, phonological awareness involves word identification in alphabetic writing systems, does that automatically mean that phonological information is also involved in reading Han characters, which are usually misidentified as a “logographic” system? In fact, this issue has aroused even more controversy than alphabetic writing system. Nevertheless, much research, such as Tien (1983), Horodeck (1987), Hung, Tzeng; and Tzeng (1992), Flores d’Arcais (1992), Cheng (1992), Shu and Anderson (1999), and Li (2000), have reported that phonological information does play a crucial role in reading Han characters just as in alphabetic writing systems. It seems that the remaining question is does sound play the only role in reading Han characters? If not, what other information is involved in Han character recognition.

3.2 Han characters

After this survey of the development of the world’s writing systems, attention is now turned to Han characters. A couple of key terms need to be clarified before discussion of Han characters.

First, the term 字 *zi* or *ji* ‘character’ refers to the minimal meaningful writing unit in the Han writing system, and each character has its own shape, sound, and meaning. A character can usually be further divided into two or more components (the so-called radical and phonetic). They can be either ‘bound components’ or ‘free components.’ Bound components cannot be used as independent characters if removed from a composite character, such as 氵 and 礻 in the examples 江 and 裡. By contrast, free components are the ones, which can function as independent characters, such as 工 and 里 in the previous examples. Usually, free components serve as firstly phonetic and secondly as semantic functions. By contrast, bound components serve firstly semantic and secondly phonetic purposes. There are several levels of components. The minimal components, which cannot be further divided, are called basic components.

The distinction between characters and components is not always clear. In general, independent characters in printed materials are adjusted in size and configuration to assure a overall square shape similar to each other. For example, 理 嫫 賄 are the samples of three characters instead of six, whereas 王 里 女 宗 貝 有 would represent six characters. What we have to be aware of is that a character, in many cases, is in fact a composite character made up of two element instead of a single character. For example, the character 明 ‘tomorrow’ consists of two unbound characters 日 ‘sun’ and 月 ‘moon.’ Thus, a Han character should be regarded as a grapheme set instead of a single grapheme in terms this author has defined in the previous section.

3.2.1 Formation of Han characters

The formation of Han characters is first categorized by Xu Shen in his etymological dictionary *Shuowen Jiezi* 說文解字, which was published about A.D. 121. Based on his statistical review of 9,353 characters, 六書 six categories or principles were identified as the basic methods of forming characters. Xu’s six principles were soon accepted by most

Chinese scholars as the norm for explaining how Han characters were constructed. Six principles are summarized as follows (DeFrancis 1990: 78-82; Liu 2001: 66-79):

First is 象形 or the “pictographic principle.” This principle is the most primitive method, and consists of characters coined on the basis of their shape, such as 日 ‘sun,’ 月 ‘moon,’ 象 ‘elephant,’ 魚 ‘fish,’ and 羊 ‘sheep.’ Characters coined with this principle are mostly words signifying concrete objects.

Second is 指事 or the “simple indicative principle.” This principle is also a primitive way to create characters. Instead of referring to concrete things, this principle is applied to the formation of abstract idea. For example, the characters 一 二 and 三, which have one, two, and three lines respectively, and refer to the numerical idea of the cardinalities one, two, and three. The characters 上 and 下 with strokes on either side of horizontal line, indicate the semantic meaning ‘above’ and ‘below,’ respectively.

Third is 會意 or the “compound indicative principle.” Characters build by this principle consist of two or more existing characters, and their meaning is derived by combining the meanings of constituent characters. For example, the character 明 ‘tomorrow’ consists of two parts 日 ‘sun’ and 月 ‘moon.’ The association between sun and moon to symbolize the passage of a day and a night is utilized to indicate the new meaning of 明.

Fourth is 形聲 or “radical-phonetic principle” or the “semantic-phonetic principle.” Characters in this principle consist of two components. The semantic components are mostly bound components and occasionally free components, and they provide information with regard to the derived meaning of the constituted character. The phonetic components usually are free components, and they provide clues to the pronunciation of the character. For example, 河 ‘river’ pronounced in Mandarin as [xɤ¹], consists of semantic component 氵, which means ‘water,’ and a phonetic component 可, pronouncing [k^hxɤ¹] with original

meaning ‘be able.’ In this case, the semantic meaning of the phonetic component 可 is ignored, while its sound is treated as a cue to the character’s pronunciation. We may take the following *Singlish* example to see how the semantic-phonetic principle could be applied to the English language.

Assume the English lexical items ‘sea’ and ‘see’ were spelled in the same way as ‘SI.’ How could they be differentiated by the semantic-phonetic principle? We may add semantic components 氵 ‘water’ and 目 ‘eyes’ to SI, so the Singlish characters look like the ones in **Figure 5**.

氵SI ‘sea’ 目SI ‘to see’

**Figure 5. Singlish example
of ‘sea’ and ‘see.’**

Fifth is 假借 or the “phonetic loan principle.” As the name of this principle suggests, the phonetic value of a character is borrowed to refer to another lexical item, which has identical or similar sound. For example, 無, originally means ‘dancing’ with pronunciation [wu˥], is borrowed to refer to another lexicon *wu* [wu˥], which means ‘nothing.’ In English, the equivalent term is rebus writing, such usage as ‘2’ in ‘back 2 school.’

The last principle is 轉注 or the “synonymous principle.” What Xu Shen meant by synonymous is that characters in this group have same meaning. For example, the characters 考 and 老 are the in same group. In fact, whether or not the “synonymous principle” is a principle of forming characters remains in doubt. Many would say that the synonymous principle is simply a term of classification of Han characters rather than an approach for creating characters. Those synonymous characters could be considered the consequence of formation by different devisers in different time and space.

In short, the structure of Han characters in terms of my proposed Universal Orthography chart can be described as in **Table 10**.

Table 10. Hanji example of Universal Orthography

		directions			
		R	L	U	D
Morphemes	Stem	+	+	+	+
	Prefix	-	-	-	-
	Infix	-	-	-	-
	Suffix	-	-	-	-
Syllables	Syllables	+	+	+	+
	Initial	-	-	-	-
	Medial	-	-	-	-
	Final	-	-	-	-
Phonemes	Consonants	-	-	-	-
	Vowels	-	-	-	-
Phonetic features		-	-	-	-
Supra-segmental		-	-	-	-
Extra-information		+	+	+	+
One-to-one		-	-	-	-
One-to-multiple		+	+	+	+
Multiple-to-one		+	+	+	+
Ambiguity		+	+	+	+

3.2.2 Evolution of Han characters

Although most Chinese scholars regard the six principles as an adequate method to form characters, this author would consider them ‘steps’ rather than ‘approaches’ in terms of Gelb’s idea of the evolution of writing systems. The synonymous principle is first excluded since it is merely the products of other principles. The other five principles are divided into three steps. The major principles in each step and their corresponding representational level are illustrated in **Table 11**.

Table 11. Evolutionary steps and representational levels of the Han writing system

Steps Rep. levels	Pictographic step	Ideographic step		Phonographic step	
Word	Pictographic principle	Simple indicative principle	Compound indicative principle		
Logosyllable				Phonetic loan principle	
Morphosyllable					Semantic-phonetic principle

The pictographic step is the earliest one, where conventional simplified pictures or symbols were used to refer to concrete objects. Those pictures or symbols eventually became the current shape of Han characters. Each character corresponds to a grapheme of word representational level, thus characters created in this period can be regarded as a word writing system.

The simple and compound indicative principles are the second step. While the pictographic principle was adopted to describe concrete objects, the simple indicative principle was employed to express abstract ideas. Because the simple indicative principle was unable to bear the increasing number of lexical items with abstract content, the compound indicative principle was developed to meet the demand. In this period, characters were mostly grapheme sets instead of a single grapheme since most of them were combinations of two or more independent characters. The representational level of characters created in this step is also lexical, thus they are a system of word writing. They are the so-called ideograph.

As I mentioned in previous sections, a word writing system could hardly survive and further develop to a full writing system without the process of phonetization. Thus, the phonetic loan and semantic-phonetic principles were devised as a way of phonetization in the third step of the Han writing system. They are the way to create new characters on the basis of existing inventory. When the phonetic loan principle was first employed, however, it soon was found that pure phonetic-loan could hardly deal with the considerable amount of homophony in the spoken language. Therefore, a semantic component was added to phonetic-loan characters to distinguish homophones. In other words, the semantic-phonetic principle was devised to meet the demand of differentiation. Chinese thus might be said to be a language in which the written form often distinguishes where the spoken language does not. So it is that the function of semantic components in Han characters is analogous to those in phonemic systems when homophones are differentiated by using different spelling rules. Characters in the phonetic loan principle could be considered logosyllabary,⁷⁰ and characters in the semantic-phonetic principle are morphosyllabary. Both logosyllabary and morphosyllabary characters are in the transition from a word writing to a system of smaller representational level.

Then, what level might they be said to be progressing towards? As a consequence of syllable and meaning(s) have become the major features of characters in previous steps one and two, the smaller level is referring to syllables or morphemes. We say ‘syllables or morphemes’ because they possess both morphemic and syllabic feature. They are a kind of hybrid system. For example, ‘yogurt’ and ‘outstanding’ in Taiwan Mandarin are both

⁷⁰ Logosyllabary was originally defined by Daniels (1996: 4) as “the characters of a script denote individual words (or morphemes) as well as particular syllables.” In this study, the definition of logosyllabary is modified by ignoring its original semantic meaning while representing syllables. The major difference between logosyllabary and syllabary is simplification. In syllabary, a syllable is represent by a grapheme. But, in logosyllabary, syllables are represented by the same sound of characters which normally denote phonemes or words.

disyllabic words, written in two characters as 優格 [jouŋ kɤʅ] and 優秀 [jouŋ ɕiouʅ], respectively. They both share the same first character 優, which semantically means ‘excellence.’ However, 優 in ‘yogurt’ is simply a phonetic-loan character, which has nothing to do with yogurt. In contrast, 優 in ‘outstanding’ preserves both its semantic and phonetic functions. There is no doubt that individual Han characters correspond to a syllable and meaning; however, whether or not they still maintain their original meaning and become part of the newly created word largely depends on the arbitrary decision of their first users. This feature indicates that whether Han writing is moving towards a system of syllable or morpheme depends on how characters are used.

As shown in **Table 11**, the evolution of the Han writing system is moving from pictographic or ideographic towards a phonographic writing. To be more precise, it is moving from a word toward a syllabic system though a pure syllabary is not achieved. What we mean by ‘evolution’ here is that a later-devised principle is developed on the basis of principles in previous steps. It does not necessarily mean an earlier-devised principle will not be employed again in a later step. For example, pictographic characters are the basis of compound-indicative-characters. That does not necessary mean that the pictographic has principle completely disappeared in the ideographic step.

To show proof of moving toward a syllabic system, Li’s (1986) statistical data on analysis of Han characters based on the six principles are listed in **Table 12**. Li’s data consist of characters collected from three major time spans. The first period contains 1,225 inscriptions incised on bones and shells, which appeared between the fourteenth and eleventh centuries B.C. Those inscriptions are usually considered to represent the earliest Han characters (Ji 1992: 223). Characters in the second period are collected from the etymological dictionary *Liushu Yaolie* 六書爻列, which was published in the second century A.D. Characters collected in another etymological dictionary 六書略 *Liushulue*,

which was published in the twelfth century, provide the data of the third period. The statistical results in **Table 12** reveal that percentages of pictographic characters, simple-indicative characters, and compound-indicative characters have been decreasing for over two thousand years, while semantic-phonetic characters have been increasing and have reached 90% of character inventory in the twelfth century. It has even achieved a value of 97% of the 47,035 characters of the *Kangxi* dictionary in the eighteenth century (DeFrancis 1990: 84). This phenomenon has shown that the semantic-phonetic principle has played a role of increasing importance in the formation of Han characters as time has gone on.

Table 12. Percentage of Han characters by different principles and periods

Types		Picto-graphic	Simple indicative	Compound indicative	Semantic-phonetic	Phonetic loan	Synonym-ous	Unclear	Total
BC 14 th -11 th Century	Number	276	20	396	334	129	0	70	1225
	Percent	22.53%	1.63%	32.33%	27.27%	10.53%	0%	5.71%	100%
2 nd Century AD	Number	364	125	1167	7697	115	7	0	9475
	Percent	3.84%	1.32%	12.31%	81.24%	1.21%	0.07%	0%	100%
12 th Century AD	Number	608	107	740	21810	598	372	0	24235
	Percent	2.50%	0.44%	3.05%	90.00%	2.47%	1.53%	0%	100%

Although Han characters have shown a tendency towards a syllabic system, they did not successfully transfer into a full syllabic system. What inhibited them in transition? There are both linguistic and non-linguistic factors. The linguistics factors are deadly due to the inconsistency of utilizing components of the semantic-phonetic principle, as DeFrancis (1996: 40) has pointed out, “the inefficiency of the system stems precisely from its clumsy method of sound representation and the added complication of an even more clumsy system of semantic determinatives.”

We may examine two aspects of the inconsistency, 1) correspondence between characters and sound and meaning, and 2) placement of all components within characters. On the correspondence between characters and sound, an identical phonetic value of

characters could be represented by different phonetic components. For example, all the following semantic-phonetic characters 惛 ‘beating heart,’ 牯 ‘bull,’ and 瞽 ‘blind’ 鵠 ‘swan’ have the same pronunciation [ku˥]. However, they are represented by different phonetic components 骨[ku˥], 古[ku˥], 鼓[ku˥], and 告[kau˥], respectively. In regard to the correspondence between characters and meaning, a semantic component does not always correspond to the same semantic category. For instance, all the characters 蚶 ‘oyster,’ 蟻 ‘ant,’ and 虹 ‘rainbow’ consist of the same semantic component 虫 ‘insect’ even though they are semantically different. What semantic category a character has assigned to it is, to some extent, arbitrary. Moreover, the boundary between semantic and phonetic components is not always definite. In other words, a component could serve either phonetic or semantic functions in different situations. For example, 骨 serves a phonetic component as in 惛, but has semantic roles as in 骸 ‘skeleton’ and 髀 ‘shinbone.’ Regarding the placement of components, both semantic and phonetic components can be placed on either at the left, right, top, bottom, or center. For example, the phonetic component 古 is normally placed on the right such as 牯, while occasionally placed on the left 故, on the top 辜, on the bottom 罟, and in the central 固.

All of this inconsistency has prevented Han characters from becoming an efficient and accurate writing system. Moreover, the inconsistency could even mislead readers to a wrong or even opposite meaning and pronunciation of characters, such as 虹, which could be misunderstood as a “insect.” We may wonder why the problems of inconsistency were not solved or overcome over the thousands of years? I would attribute this to two major factors. First, it is due to the limitation of linguistics knowledge. The second, even more important factor, is the cultural conservatism of the Chinese people.

In comparison to the Japanese and the Korean, the ancient Chinese scholars were not well aware of the existence of syllables and phonemes. This is probably because the

Chinese scholars were too much attached to the properties of Han characters. In earlier times, a character learner had to learn its pronunciation from another character. For example, in Xu Shen's dictionary, it listed “*句讀若鳩*,” which means the character *句* is pronounced as *鳩*. Another “advanced” way to record sound is the so-called *反切* ‘*fanqie*,’ which literally means ‘turn and cut,’ which was devised sometime around the second and third centuries A.D., and continued to be employed in most dictionaries until the early twentieth century (Ji 1992: 71-73). In *fanqie*, the sound of a character was always represented by two other characters, the initial was represented by the first character and the final by the second. For example, *東* [toŋ] is represented by characters *德* [tɤʅ] and *洪* [hoŋ]. To learn the pronunciation of *東*, readers have to gather [t] and [hoŋ] from *德* and *洪*, respectively, and then combine them together. Although the ancient Chinese seems to become aware of initials and finals through the device of *fanqie*, it was not successfully simplified and refined into syllabary or alphabets as the Japanese and the Korean did. Among several attempts to improve *fanqie*, *三十字母* or Thirty Basic Characters⁷¹ proposed by a monk *Shouwen* 守溫 in the tenth century, is considered the forerunner of Chinese alphabets (Zhou 1987: 87). Based on *Shouwen*'s remaining manuscripts, thirty consonants were identified and each of them was represented by a unique Han character. In other words, characters were treated as a kind of alphabet to represent sounds. As for vowels, they were not mentioned at all. Although *Shouwen*'s scheme shows the first step towards having one-to-one correspondence between characters and sounds, it did not draw much attention from the Chinese people. Consequently, *Shouwen*'s scheme did not progress, as Zhou (1987: 87) comments, “it just steps halfway on the way to alphabets.”

⁷¹ Six characters were added by other scholars to represent additional consonants at a later time. It was thus also called *三十六字母* Thirty Six Basic Characters.

Why did Shouwen's Thirty Basic Characters not draw much attention? Basically, it could be attributed to cultural conservatism, which was further entrenched by the civil service examination, as I have detailed in chapter two. As Confucianism became the dominant philosophy in feudal hierarchy since the Han dynasty in the second century B.C., Confucius' books in Han characters gradually became the orthodox canon for officials, scholars, and even literacy beginners to follow. To what extent the Chinese people follow their ancestors is well expressed by the old Chinese saying, 貴古賤今 'the ancient is dear; the contemporary cheap,' which means 'those ancestors are noble and perfect while the contemporaries are of little value.' This saying indicates that it is rather difficult for later generations to make changes in regard to the legacy of previous generations. Although literacy beginners might not find the sounds of characters in a fanqie dictionary, their teachers would never consider improving the approach. Instead, they just blamed the students for not studying hard enough. Once people were attached to Han writing, it was hard for them to reform or reject Han characters. There is another Chinese saying 一字值千金 'a character is worth a thousand gold pieces,' which means 'Han characters are so valuable that you cannot neglect even one of them.' Indeed, Han characters might be more valuable than gold since the acquisition of Han characters was the fundamental requirement to pass the civil service examination and to become a government official, who were regarded as the most prestigious people in the Han sphere. As a matter of fact, not only did the ancient Chinese pay respect to Han characters, but this idea continues in the general public of modern times. For example, as a member of Taiwanese society under Chinese influence, I was often told by my grandmother "do not leave papers with Confucius characters on the ground or people will step on them," though she could not read any of these characters.

3.2.3 Influence of Han characters on word perception

Because sound and semantic representations in Han characters are not always consistent, this fact has further resulted in psychological reading problems. The prominent problem is that components in Han characters usually mislead to a wrong meaning and sound. For example, there is a place called 三貂角 ‘Sam-tiau-kak’ in Taipei county, Taiwan. This name was originated from Spanish word ‘San Diego’ when Spanish came to northern Taiwan in the seventeenth century. Because San Diego was written in Han characters as 三, meaning ‘three,’ 貂 ‘marten,’ and 角 ‘angle,’ these characters have mislead readers to the connection between San Diego and “three marten.” Another famous example of misleading is the origin of 基隆 ‘Kelang,’ which is a county near Taipei. There used to be indigenous Ketagalan tribes in the great Taipei area. For some reason, Ketagalan was written in Han characters as 雞籠 (means ‘chicken cage’ and pronounced as /ke lang/), and at a later time 基隆, to refer to the current Kelang county. Consequently, many people thought Kelang was named because the mountains in Kelang look like chicken cages.

This misleading phenomenon not only occurred with general public, but also with scholars, particularly those etymologists in Taiwan. This phenomenon can be observed from their action to find the so-called ‘*punji*’ 本字 or ‘original character’ or ‘the etymologically correct character.’ For example, many traditional etymologists in the Taiwanese linguistics circle tried very hard to find ‘*punji*’ for those Taiwanese words, which do not have appropriate or corresponding characters. They, such as Lian (1987), claimed that all Taiwanese lexical items have their own ‘*punji*.’ However, the characters they found as *punji*, usually vary from scholar to scholar. For example, 个, 其, 個, 格, 兮, 的, and 𠄎 were claimed by different scholars to be the *punji* of the Taiwanese word *e* ‘of.’ More such examples are given in **Table 13**. Apparently, their identification of ‘*punji*’ is more likely to be arbitrary.

Table 13. Example of ‘punji’ by different scholars

Gloss Author	e ‘of’	e ‘classifier’	beh ‘want’	m ‘no’	che ‘many’	khah ‘more’
Ang 洪惟仁	个	个	卜	毋	濟, 儕	卡
Khou 許成章	其, 個, 格	員, 個, 箇, 个, 負	望, 卜, 媒, 謀, 薄, 甫, 俾	不, 無, 否	濟, 侈, 庶, 孳, 滋, 眾, 鼓, ...	較, 校
Lian 連橫	兮	兮	憐	恠	多	較
Tan 陳修	的	個	必	不	較, 濟	較
IuN 楊青矗	个, 兮	个, 個	要	咁, 唔	濟	卡, 忒, 較

According to 《台語文摘》《台灣漢語辭典》《台灣語典》《台灣話大詞典》《國台雙語詞典》

Although some lexical items may have their own punji, it does not mean that punji can be found for all lexical items. Those etymologists have paid too much attention to punji, and ignored the fact that 1) Han characters were not originally created for representing “colloquial” speech, and 2) many lexical items in Taiwanese have non-Sinitic origin.

Chiung’s (1998) study has revealed that the perception of words possessed by Hanji educated Taiwanese speakers are more likely to be misled by Han characters. In his study, a total of sixteen Taiwanese speakers, including Hanji and non-Hanji educated background, were interviewed. The subjects were told to reveal their perception of syllables in words based on a prepared word list. The word list consisted of polysyllabic words and compound words of monosyllabic morphemes. The results showed that Hanji educated subjects were more likely to perceive polysyllabic words as compound words, and attempt to assign a meaning to each syllable. In other words, they were misled by the monosyllabic feature of Han characters, and misperceived each syllable as having corresponding meaning and character. This phenomenon is called ‘syllabic sinification’ 音節漢字化, in which any syllable is arbitrarily assigned a meaning and Han character, as illustrated in **Figure 6** and **Figure 7**.

Chiung (1998) further points out that ‘syllabic sinification’ is the consequence of lexical sinification, in which process syllables of words are assigned an arbitrary character (see **Figure 8**). Whenever anyone is not satisfied with the character attached to syllable, s/he may start over the process of lexical sinification again and replace the existing characters with another one. This process of lexical sinification is exactly what those etymologists do when they are searching for punji.

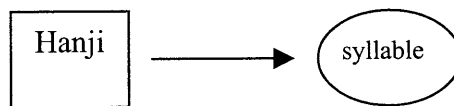


Figure 6. Each Han character has only one syllable.



Figure 7. Each syllable was misperceived as having meaning and character.

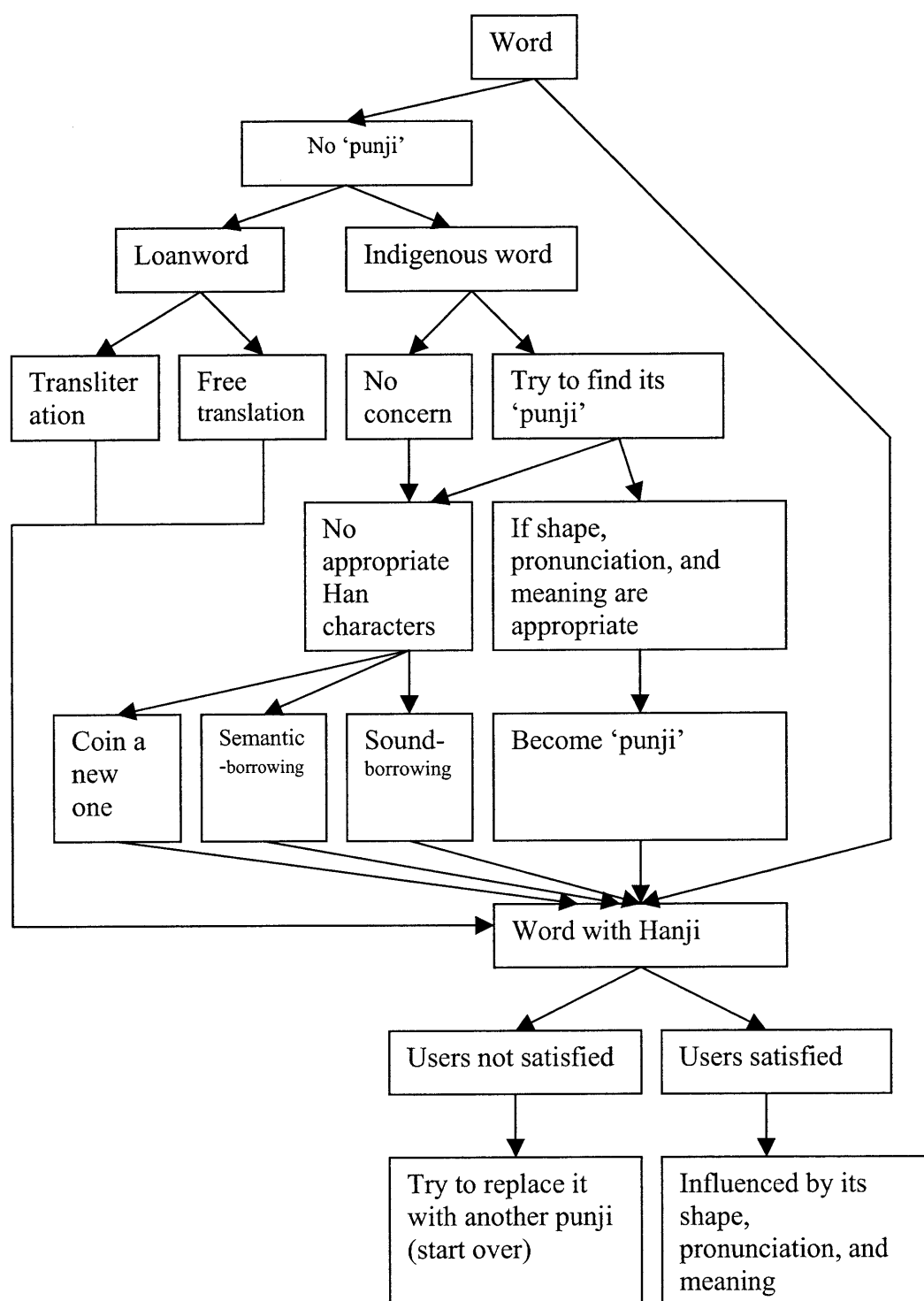


Figure 8. Process of lexical Sinification.

3.3 Roman script

The Roman script or Latin script is thought to derive from the Etruscan alphabet, which originated from a western Greek alphabet around eighth century B.C. (Bonfante 1996: 297; Hyatt 1990: 20-21). When the Romans came into power in the Italian peninsula, they took over the alphabet of the Etruscans. With some modifications, the revised Etruscan alphabet was later adopted by the Romans for writing Latin. This revised alphabet was later called the Roman script/alphabet or the Latin script/alphabet, and it became widespread through the promotion of Roman Catholicism in western Europe. Nowadays, the Roman alphabet is the most widely used script around the world, including most European and African nations, some nations in Asia, and all nations of the Americas and Oceania (Nakanishi 1990: 14).

3.3.1 How Romanization works

Although Romanized writing systems have proven themselves as writing systems for many of the world's major languages, many people doubt the capacity of Romanization for Han languages. They think that Romanization is incapable of differentiating the pernicious homophony in the Han languages. Such questions about the Romanization of Asian languages have been raised in the Hanji cultural sphere ever since the nineteenth century (cf. DeFrancis 1990; Hannas 1997; Chen 1999). As matter of fact, Romanization can differentiate homophonous morphemes as well as Han characters. It just depends on how the spelling of the Romanization is devised. For example, in English, *see* and *sea* are spelled in different ways to refer to different things despite identical pronunciation. The case of *to*, *too*, and *two* is another example from English. As for Taiwanese, for example, 科根 *crkunl* or radicals is proposed by Cengjiu Dan (1998) as a system to write Taiwanese. Basically, Dan defines 60 categories with 60 simple symbols, or the so-called *crkunl*, to refer to different semantic categories of words. He adds a *crkunl* to each Romanized

Taiwanese word so readers can distinguish the different meaning even when the pronunciation of the words is identical. For example, he defines the symbol <z> as animal and <%> as numeral. So, the homophones ‘dog’ and ‘nine’ in Taiwanese are differentiated by different spellings as ‘kuz’ and ‘kau%,’ respectively. Apparently, Dan has tried to solve the inconsistency problems of the semantic-phonetic principle, and he further simplified the symbols of the semantic components through his idea of crkunl.

Another way to differentiate homophony within the Roman alphabet is to expand its linear spelling to a two-dimensional placement, something like what the Korean alphabet has done. Suppose one is dealing with the English homophones ‘date,’ meaning time, and ‘date,’ meaning date palm, by applying this approach. Their new look is as illustrated in **Figure 9**.

$\frac{DA}{TE}$	$\frac{DT}{AE}$
‘date,’ meaning ‘time’	‘date,’ meaning date palm

Figure 9. Example of two-dimensional placement by Roman alphabet.

Although adding rules or affixes to spelling may increase the capacity of Roman alphabet to differentiate homophones, it can also increase the degree of difficulty in spelling, and thus reduce the efficiency and ease of learning. To what extent these accommodations will be applied to a Romanized writing system depends on how the designers evaluate the various costs and benefits.

CHAPTER 4

SOUND AND WRITING SYSTEMS IN TAIWANESE, CHINESE AND VIETNAMESE

In this chapter, specific writing systems, which are deeply involved in this study, are examined in detail so readers will have a better understanding of their correspondence between orthographic symbols and speech sounds. These systems are Taiwanese Romanization Peh-oe-ji, Chinese Romanization Hanyu Pinyin, Chinese Bopomo, and Vietnamese Romanization Chu Quoc Ngu. As for Han characters, readers may refer back to chapter three.

4.1 Taiwanese

4.1.1 Sound system

The Taiwanese language is also called Southern Min, Minnan, Holo, or Taigi. The most accepted phonological system for Taiwanese is as shown in **Table 14**, **Table 15**, and **Table 16**. In general, there are seventeen consonants, six vowels, and seven tones, though may vary from variety to variety. Among the consonants, the phoneme /l/ is in fact pronounced as [d] or [r] in most circumstances (Tiuⁿ 2001: 31-32). Nevertheless, we follow the traditional description of listing /l/ as a phoneme.

Table 14. Taiwanese consonants in IPA

	bi-labial	alveolar	velar	glottal
	-asp/+asp	-asp/+asp	-asp/+asp	-asp/+asp
voiceless stop	p / p ^h	t / t ^h	k / k ^h	
voiced stop	b		g	
voiceless fricative		s		h
voiceless affricate		ts / ts ^h		
voiced affricate		dz		
voiced lateral		l		
voiced nasal	m	n	ŋ	

Table 15. Taiwanese vowels in IPA

	front	central	back
high	i		u
mid	e	ə	o
low		a	

Table 16. Tonal categories in Taiwanese (Cheng 1997)

Categories	君 [kun˧˥] gentle	滾 [kun˨˥] boil	棍 [kun˨˨˥] stick	骨 [kut˧˥] bone	裙 [kun˨˥] skirt	-	近 [kun˨˥] near	滑 [kut˧˥] glide
Numerical categories	1	2	3	4	5	6	7	8
Tone marks in Peh-oe-ji	unmarked	˨	˨˨	unmarked	˨˥		˨˥	˧˥
Peh-oe-ji samples	kun	kún	kùn	kut	kùn		kūn	kút
Numerical tone values	44	53	31	3	12		22	8
IPA tone values	˧˥	˨˥	˨˨˥	˧˥	˨˥		˨˥	˧˥

There are currently seven tonal categories in the Taiwanese language. They are traditionally called tones 1, 2, 3, 4, 5, 7, and 8. The missing element, tone 6, has merged with tone 2 or tone 7 (Ang 1985: 2-3), therefore, nowadays there are only seven tones, as listed in **Table 16**.

Tone 5 is traditionally described as a low rising tone (12); tone 4 and tone 8 are abrupt tones with low and high contrasts, respectively (Cheng 1977, 1997; Ang 1985; Weingartner 1970). However, Ong (1993) has shown that tone 5 is a low falling and then rising tone based on acoustic measurement of his own pronunciation in 1945. Chiung's (2001f) acoustic measurement of 22 Taiwanese speakers reveals that young generation is more likely to possess falling-rising feature. Tseng (1995: 32) points out that the tone shape of tone 5 can be a rising contour (/) if the starting point is low; or a dipping contour (\) if the starting point is slightly higher. Both contours share the low-rising pattern. As for tone 4 and tone 8, it is said that distinctions between tone 4 and tone 8 are not apparent in some areas such as in Tai-tiong city of central Taiwan (Cheng 1977: 97; Khou 1990: 89).

There is a Tone Sandhi Rule which has the effect of converting all but the last full tone within a tone group into their corresponding sandhi tones (Chen 1987).

Tone Sandhi Rule(TSR)

$T \rightarrow T' / _ T$ within a tone group

Key: T= base tone, T'= sandhi tone

For example, consider 土地 *tho'-te* 'land', a compound word of two monosyllabic words *tho*. 'soil' and *te* 'ground.'

tho'-te

Base tones → 2 - 7

After TSR → 1 - 7

If we add 公 *kong* 'grandpa' to *tho.-te*, it becomes *tho.-te-kong* or 'God of Land.'

tho'-te-kong

Base tones → 2 - 7 - 1

After TSR → 1 - 3 - 1

How sandhi tones occur can be described by the Tone Circle or Tone Square⁷². In the Tone square, we assume sandhi tones and base tones share the same seven tone categories.

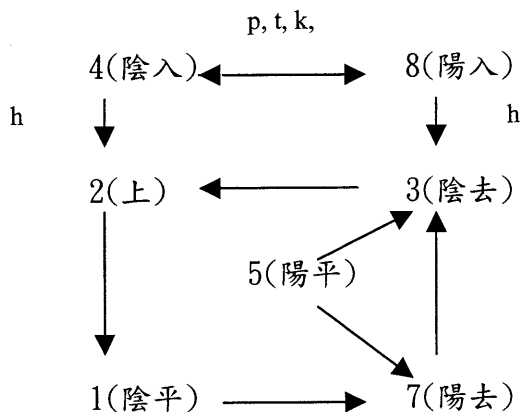


Figure 10. Tone Circle in Taiwanese.

1 st tone 陰平	→	7 th tone 陽去	e.g. hoe-hoe ‘confused’ 1-1→7-1
2 nd tone 上聲	→	1 st tone 陰平	e.g. te-te ‘short’ 2-2→1-2
3 rd tone 陰去	→	2 nd tone 上聲	e.g. cho-cho ‘make’ 3-3→2-3
4 th tone 陰入	→	2 nd tone 上聲	if with h ‘glottal stop’ in the final
			e.g. tah-tah ‘attach’ 4-4→2-4
		→ 8 th tone 陽入	elsewhere
			e.g. tap-tap ‘answer’ 4-4→8-4
5 th tone 陽平	→	7 th tone 陽去	in Southern Taiwan
			e.g. Tai-pak ‘Taipei’ 5-4→7-4
		→ 3 rd tone (陰去)	in Northern Taiwan
			e.g. Tai-pak ‘Taipei’ 5-4→3-4

⁷² This square is based on Robert Cheng (1977: 123) with some modification. Taiwanese Tone Square is a little bit different from Xiamen Tone Circle.

7 th tone 陽去	→ 3 rd tone 陰去	e.g. chai-chai(steady) 7-7→3-7
8 th tone 陽入	→ 3 rd tone 陰去	if with h ‘glottal stop’ in the final
		e.g. peh-peh ‘white’ 8-8→3-8
	→ 4 th tone(陰入)	elsewhere
		e.g. pak-pak ‘tie’ 8-8→4-8

In addition to following the Tone Square, there are some other special tonal sandhi rules such as Three Syllables Tonal Sandhi 三連音變調, Tone Sandhi Preceding /a/仔前變調, and Tone Sandhi Based on Preceding Tone 隨前變調. Since tone sandhi is not the major concern in this study, further elaboration must be left for another time.

4.1.2 Written form in Romanized Peh-oe-ji

Adoption of the Roman alphabet for writing Southern Min dates back to the sixteenth century when the Spanish Dominicans worked together with their translators among the Chinese community in Manila (Van der Loon 1966, 1967; Kloter 2002). Their systems differ in many aspects from Romanization developed in the nineteenth century (Kloter 2002).

Missionary linguistic efforts on the Romanization are reflected in various Romanized dictionaries. Medhurst’s *A Dictionary of the Hok-keen Dialect of the Chinese Language* 福建方言字典 published in 1837 is considered the first existing Romanized dictionary of Southern Min compiled by western missionary (Ang 1996: 197-259; Heylen 2001: 146). Douglas’ *Chinese-English Dictionary of the Vernacular or Spoken Language of Amoy* 廈英大辭典 of 1873 is regarded as a dictionary of significant influence on the orthography of Peh-oe-ji (Ang 1993b: 1-9). After Douglas’ dictionary, most Romanized dictionaries and publications followed his orthography with only minor changes. Generally speaking, the missionary linguistic efforts on Southern Min and Peh-oe-ji (POJ) have made considerable progress since Douglas’s dictionary (Ang 1993b: 5). William Campbell’s dictionary *E-*

mng-im Sin Ji-tian or *A Dictionary of the Amoy Vernacular* (1913), which was the first Peh-oe-ji dictionary published in Taiwan, is the most widespread Romanized dictionary in Taiwan and had been published in fourteen editions by 1987.

The following list consists of some examples of the variations of spelling among these three dictionaries.

Table 17. Examples of spelling variations among Medhurst, Douglas, and Campbell

Medhurst	Douglas	Campbell	Hanji	IPA
ee ^{ng}	i ⁿ	i ⁿ	嬰	[i]
ëen	ien	ian	煙	[en]
wa	oa	oa	蛙	[wa]
oe	o	o	烏	[o]

Since E-mng-im Sin Ji-tian is the most widespread Romanized dictionary in Taiwan, it will be briefly demonstrated how Peh-oe-ji works in the E-mng-im Sin Ji-tian. For more details on how the Roman alphabet is employed in Peh-oe-ji, readers may refer to Jiuⁿ (2001) and Cheng and Cheng (1977).

Table 18. Symbols for Taiwanese consonants in the spelling of Peh-oe-ji

Consonants	Peh-oe-ji	Conditions	Examples
/p/	p		pí 'compare'
/p ^h /	ph	initial only	phoe 'letter'
/t/	t		tê 'tea'
/t ^h /	th	initial only	thâi 'to kill'
/k/	k		ka 'add'
/k ^h /	kh	initial only	kha 'foot'
/b/	b	initial only	bûn 'literature'
/g/	g	initial only	gí 'language'
/h/	h		hí 'glad'
/s/	s	initial only	sì 'four'
/ts/	ch	before i, e	chi 'of'
	ts	elsewhere	tsa 'investigate'
/ts ^h /	chh	initial only	chha 'differ'
/dz/	j	initial only	jít 'sun'
/l/	l	initial only	lí 'you'
/m/	m		mī 'noodle'
/n/	n		ni 'milk'
/ŋ/	ng		âng 'red'

Table 19. Symbols for Taiwanese vowels in the spelling of Peh-oe-ji

Consonants	Peh-oe-ji	Conditions	Examples
/i/	i		ti 'pig'
/e/	e	Elsewhere	tê 'tea'
	ia	Followed by n or t	tīan 'electric' kiat 'to form'
/a/	a		ta 'dry'
/u/	u		tú 'meet'
/ə/	o		to 'knife'
/o/	ȯ	Elsewhere	ȯ 'black'
	o	Followed by finals, except /ŋ/	tong 'east' kok 'state'

The spelling rules of Peh-oe-ji are easier than those for Vietnamese Chu Quoc Ngu. In general, there is a one-to-one relationship between orthographic symbols and phonemes as

shown in **Table 18** and **Table 19**. The only exception is the pair of <ch> and <ts> that both refer to the phoneme /ts/ (nowadays, ts has been replaced by ch). The different usages between ts and ch are based on vowel position. That is, ts precedes back vowels such as ‘tso,’ and ch precedes front vowels such as ‘chi.’ This rule was influenced by the phenomenon of palatalization of /ts/, where [ts] become palatalized [tɕ] when followed by front vowels. In other words, the missionary devisors treated [ts] and [tɕ] as two different sounds in terms of phonetics instead of phonemics, though most spelling rules in POJ are made from the viewpoint of phonemics.

After phonemes are represented, tone marks are superimposed on the nuclei of syllables and a hyphen ‘-’ is added between syllables, such as in thó-tē-kong, ‘the God of Land.’ Because Taiwanese is a tone language with rich tone sandhi, there can be several ways to represent tones. In the design of POJ, the base tone or underlying tone of each syllable is chosen and represented by its tone mark. For example, ‘the God of Land’ must be represented by its underlying form thó-tē-kong rather than surface form tho-tè-kong (this is the form of actual pronunciation). Why is underlying form instead of surface form chosen to represent tones in POJ? This is an issue often challenged point by POJ reformers. Indeed, in some cases, especially polysyllabic loanwords, such as o-tó-bái ‘motorcycle,’ it seems ridiculous to represent motorcycle by the underlying form o-tó-bái rather than the actual pronunciation o-tò-bái. The major reason for the missionaries to choose underlying tone is probably the influence of the monosyllabic feature of Han Characters.

We may examine the invention of POJ in terms of Smalley’s criteria for a new orthography as mentioned in chapter three. All the strengths and weaknesses of Peh-oe-ji come from its nature as phonemic writing. In terms of efficiency, the relative ease of learning reading and writing in POJ over Hanji gives higher motivation to its learners. In a former agricultural society, most people were peasants who labored in the fields all day

long, and they had little interest in learning complicated Han characters. In contrast to Han characters, the ease of learning POJ provided those farmers a good opportunity to acquire literacy. This is one of the reasons why there are a number of people who have no command of Han characters, only POJ.⁷³ Although Peh-oe-ji might provide maximum motivation to individual learners, it may not have the same motivation in a Hanji dominated society and government. Chiung (2001a) studied empirically 244 college students' attitudes toward various contemporary Taiwanese writing schemes. That study revealed that college students educated in Hanji and Mandarin tend to favor Han characters over Roman scripts. As for the attitudes of the Chinese KMT government (1945-2000), POJ was not only excluded from the national education system but was also restricted in its daily use. For instance, the Romanized *Sin Iok* (New Testament) was once seized by KMT in 1975 because Hanji was considered the only national orthography, while Romanization was regarded as a challenge to KMT's Chinese nationalism.

To have maximum representation of speech usually requires a good linguistic analysis of the language before devising an orthography. Campbell's (1913) *E-mng-im Sin Ji-tian* demonstrated the achievement of missionary knowledge about Amoy and Taiwanese. Campbell's choices of symbols for representing Taiwanese consonants and vowels are listed in **Table 18** and **Table 19**. In Campbell's dictionary, 24 symbols are used to represent 23 Taiwanese phonemes (i.e., 17 consonants and 6 vowels), and those symbols consist of only 17 Roman letters. Campbell's linguistic analysis and choice of symbols are quite accurate and efficient and agree with modern linguistic treatments. For example, he primarily assigns one grapheme/grapheme set (except /ts/) to each phoneme. The letters he

⁷³ Huang (quoted in Xu 1992: 70) estimated that by 1955 a total of 115,500 people in all Southern Min speaking areas such as Hokkien, Malaysia, and Taiwan could use Peh-oe-ji. 32,000 of them were Peh-oe-ji users in Taiwan.

assigns to sound segments are very close to the IPA system (International Phonetic Alphabet), which is adopted by many contemporary linguists for transcribing linguistic data. If two letters for a phoneme are inevitable, he tries to make the symbols easy and rule based. For example, <h> indicates ‘aspiration’ when it is attached to *p*, *t*, *k*, or *ch*; and it represents glottal stop when it occurs in the final position of a syllable. Other than these situations, <h> refers to a glottal fricative.

Overall, Campbell’s choice of phonemic symbols is fairly standard. The only controversial point is the alveolar voiceless affricate sounds. While Campbell does distinguish between *ch* and *ts*, this difference is actually ‘phonetic’ rather than ‘phonemic.’ That is to say, he could choose either *ch* or *ts* to represent the segmental contrast.

In addition to the choice of phonemic symbols, the spelling in Campbell’s dictionary is also pretty straightforward. His fundamental principle of spelling is to do phonemic transcriptions of spoken language. That is, write down phonemically what you hear. His second principle is to treat POJ as an independent orthography once the spelling of words is confirmed, rather than a supplementary phonetic tool to the learning of Han characters. In Campbell’s opinion, the spelling of the Romanized Bible (1873) was considered the official orthography of POJ. Therefore, as Campbell described in the preface of his dictionary, “none of the current words whose spelling differs from that standard were taken in.” He made efforts to maintain the existing POJ orthography. The issue of spelling of POJ is still controversial among some of its users today. For example, people have tried to replace the forms *ian*, *oa*, and *eng*, with *en*, *ua*, and *ing*, respectively.

Although Romanized Peh-oe-ji has the strengths of maximum representation and efficiency, many people doubt its capacity for Romanizing the members of the Han language family because they think that the Roman alphabet is too inadequate to differentiate homophones. As I have demonstrated in chapter three, this doubt is not well

founded. If Romanization were incapable of serving as an adequate orthography, how could it have survived for more than a hundred years in Taiwan and four hundred years in Vietnam?

Maximum of transfer is another virtue of POJ. Since POJ consists of Roman letters, and Roman script is the most widespread orthography among the world's writing systems, its users will have a better starting point to the orthographies of other Romanized languages such as English.

From the perspective of the reproduction of orthography, reproducing Romanized POJ is even easier and more efficient than Han characters (recalling that there are a total of 47,035 characters in the *Kangxi* Dictionary). Compared to the small amount of Roman letters and diacritic marks in the POJ writing, Han characters are much more difficult to reproduce in typographic composition (DeFrancis 1996: 19-21). In the information age, although personal computers can easily reproduce Han characters, dealing with Han characters still involves more trouble than dealing with Roman scripts, such challenge as compatibility, OCR, and machine translation.

4.1.3 Current Taiwanese language education in Taiwan

There was no nation-wide Taiwanese language(s) education in Taiwan under the administration of the Chinese KMT. This situation was not changed until 2001 when the opposition party Democratic Progress Party came in power.

Starting in Fall semester 2001, all elementary schools are allowed to offer at least one period (40 minutes) of Taiwanese language(s) to their students. Those selective courses are called 鄉土語言課程 or "Subject on Homeland Languages." The subject includes the teaching of Holo, Hakka, and indigenous languages, depending on the availability offered by school. Schools have right to choose their own textbooks for the courses. Students have

to select one of the languages as their learning subject. However, their studying results are neither assessed nor included in records. According to the statistics by Department of Elementary and Junior High School Education⁷⁴ under the Ministry of Education, Holo language is offered to students in 2,098 schools, Hakka in 532 schools, and indigenous languages in 264 schools. Among the schools, 99% of them spend one period per week for this subject. Only 35 schools offer more than one period.

4.2 Chinese

4.2.1 Sound system

Mandarin Chinese is the official language in China and Taiwan. Its consonants, vowels, and tones are listed in **Table 20**, **Table 21** and **Table 22**. Taiwan Mandarin⁷⁵ is slightly different phonetically from Beijing Mandarin in some of its details. In general, the retroflex feature in Taiwan Mandarin is not as prominent as it is in Beijing Mandarin; [x] and [ʃ] of Beijing Mandarin are more likely to be pronounced by Taiwanese speakers as [h] and [ə], respectively. Tone values of Taiwan Mandarin (TM) also differ from Beijing Mandarin (BM).

Four base tones in Mandarin are usually described as 55, 35, 214, and 51, based on Chao's (1968) five-point scale. However, Fon and Chiang (1999) have pointed out that the four tones in TM should be considered 44, 323, 312, and 42, respectively, based on their acoustical study of a bilingual subject of Taiwanese and Mandarin. Chu's (1998) measurements of 24 Mandarin speakers from Taipei reveal that TM tone 2 falls before rising, and TM tone 3 is a low level or low falling. Chiung's (1999) measurements of 22

⁷⁴ For details about the statistics, visit their website at <http://www.eje.edu.tw/ejedata/kying/200210141849/911014.htm>

⁷⁵ Taiwan Mandarin in this study is defined as the language acquired by Taiwanese speakers through KMT's promotion of Mandarin in Taiwan.

bilinguals of Mandarin and Taiwanese also indicate that TM tone is a falling-rising tone, and TM tone is a low falling one. Based on these findings, I would suggest a modification of the four tones in TM as 44, 212, 31, and 53, as shown in **Table 22**.

Table 20. Mandarin consonants in IPA

		bi-labial	labio-dental	alveolar	retroflex	palatal	velar
		-asp/+asp		-asp/+asp	-asp/+asp	-asp/+asp	-asp/+asp
voiceless	stop	p / p ^h		t / t ^h			k / k ^h
voiceless	fricative		f	s	ʂ	ç	x
voiced	fricative				ʐ		
voiceless	affricate			ts / ts ^h	tʂ / tʂ ^h	tç / tç ^h	
voiced	lateral			l			
voiced	nasal	m		n			

Table 21. Mandarin vowels

	apical		laminal		
	front	back	front	central	back
			-rd/+rd		-rd/+rd
high	i	ɯ	i / y		u
mid		ɤ	e	ə	ɤ / o
low				a	

Table 22. Tonal categories in Mandarin

Categories	媽 [ma] mother	麻 [ma] sesame	馬 [ma] horse	罵 [ma] blame	嗎 [ma] particle
Traditional categories	陰平	陽平	上聲	去聲	輕聲 neutral t.
Numerical categories	1	2	3	4	
Tone marks in Pinyin	-	/	∨	∖	unmarked
Numerical tone values (Beijin Mandarin)	55	35	214	51	
IPA tone values (Beijin Mandarin)	┐	┌	√	∨	
Tone marks in Bopomo	unmarked	/	∨	∖	•
Numerical tone values (Taiwan Mandarin)	44	212	31	53	
IPA tone values (Taiwan Mandarin)	┐	√	∨	∨	

4.2.2 Written form in Romanization

Although there are several schemes for Chinese Romanization, none of them except Hanyu Pinyin, which was promulgated by the People's Republic of China in 1958, is taught through national education system either in Taiwan or in China. Since Hanyu Pinyin is the only Romanization included in current schooling of China, how it works is examined in this section.⁷⁶

The correspondences between orthographic symbols and speech sounds in Pinyin are demonstrated in **Table 23** and **Table 24**. Diacritic marks for tones are listed in **Table 21**. In general, there is a one-to-one correspondence between symbols and consonants. As for the vowels, seven out of the eleven vowels are represented by more than a symbol, which should be regulated by spelling rules. In either consonants or vowels, orthographic symbols

⁷⁶ For more information about Pinyin, readers may refer to Wenzhi (1983), Ingulsrud and Allen (1999), and Killingley (1998).

consist of a graph, such as , or a digraph, such as <sh>. After phonemes are represented by their symbols, a graph <'> is added between syllables if necessary. For example, <'> is added to show that *pi'ao* 'leather' is a disyllabic word rather than a monosyllabic word *piao* 'float.'

Although Pinyin could be considered an independent writing system, it is not recognized as an orthography but as a supplementary tool for learning the standard pronunciation of Mandarin and Chinese characters. As former Chinese premier Zhou Enlai addressed, "we should be clear that Hanyu Pinyin is to indicate the pronunciation of Chinese characters and to spread the use of standard Mandarin; it is not to substitute a phonetic writing system for the Chinese characters" (Wenzi 1983: 6; Ingulsrud and Allen 1999: 38).

Table 23. Symbols for Mandarin consonants in the spelling of Pinyin

Consonants	Pinyin	Conditions	Examples
/p/	b		bā 八 'eight'
/p ^h /	p		pī 披 'to wear'
/t/	d		dà 大 'big'
/t ^h /	t		tī 踢 'kick'
/k/	g		guó 國 'state'
/k ^h /	k		kāfēi 咖啡 'coffee'
/f/	f		fēi 飛 'fly'
/s/	s		sān 三 'three'
/ʃ/	sh		shān 山 'mountain'
/ç/	x		xī 西 'west'
/x/	h		hào 號 'number'
/z/	r		rào 繞 'to circle'
/ts/	z		zá 雜 'complex'
/ts ^h /	c		cā 擦 'wipe'
/tʃ/	zh		zhà 炸 'to bomb'
/tʃ ^h /	ch		chá 茶 'tea'
/tç/	j		jí 及 'and'
/tç ^h /	q		qī 七 'seven'
/l/	l		lā 拉 'to pull'
/m/	m		mā 媽 'mother'
/n/	n		ná 拿 'to hold'

Table 24. Symbols for Mandarin vowels in the spelling of Pinyin

Vowels	Pinyin	Conditions	Examples
/i/	i	elsewhere	Bīn 賓 'guest'
	y	without initials, followed by vowels other than /i/	yā 鴨 'duck'
	yi	without initials	yīng 鷹 'eagle'
/y/	ü	elsewhere	nǚ 女 'female'
	u	preceded by j, q, x	qū 區 'area' xūn 薰 'smoke'
	yu	without initials	yuē 約 'to date'
/ɿ/	i		zì 字 'characters'
/ʅ/	i		zhī 之 'of'
/u/	u	elsewhere	lù 路 'road'
	w	without initials, followed by vowels other than /u/	wān 灣 'bay'
	wu	without initials	wū 烏 'dark'
	o	Before [ŋ] or after [a]	dōng 東 'east' bāo 包 'to warp'
/ɛr/	er	elsewhere	èr 二 'two'
	r	final only	huār 花兒 'flower'
/e/	e	elsewhere	tiē 貼 'to post'
	ê	in isolation	ê
	unmarked	after [Cu]	gūi 規 'a rule'
/ə/	e	elsewhere	bēn 奔 'running' mēng 蒙 'blind'
	unmarked	after [Cu]	dūn 蹲 'squat'
/ɤ/	e		é 鵝 'goose' hē 喝 'to drink'
	o	elsewhere	bó 博 'knowledgeably' duō 多 'many'
/a/	unmarked	after [Ci]	niú 牛 'cow'
	a		dá 大 'big'

4.2.3 Written form in Bopomo

Bopomo ㄅㄆㄇ注音符號 or Phonetic Symbols for Mandarin was originally devised in China in the early twentieth century. It was brought to Taiwan in 1945 by the Chinese KMT and has been serving ever since as a supplementary tool for students' learning of Mandarin and Chinese characters. In contrast, Bopomo in China was replaced with Hanyu Pinyin in 1958.

Bopomo is an advanced invention based on the concept of Shouwen's Thirty Basic Characters, which was mentioned in chapter three. Bopomo divides syllables into three parts, i.e., initials, medials, and finals (Zhou 1987: 88). Their symbols and equivalent Pinyin and IPA are listed in **Table 25**, and tone marks are listed in **Table 22**. Because Bopomo represents not only phonemes, but also syllables, it is a hybrid system of phonemic and syllabic writing.

Table 25. Bopomo with equivalent Pinyin and IPA

Types	Bopomo	Pinyin	IPA	
Initials	ㄅ	b	[p]	
	ㄆ	p	[p ^h]	
	ㄇ	m	[m]	
	ㄈ	f	[f]	
	ㄉ	d	[t]	
	ㄊ	t	[t ^h]	
	ㄋ	n	[n]	
	ㄌ	l	[l]	
	ㄍ	g	[k]	
	ㄎ	k	[k ^h]	
	ㄏ	h	[x]	
	ㄐ	j	[tɕ]	
	ㄑ	q	[tɕ ^h]	
	ㄒ	x	[ç]	
	ㄗ	zh	[tʂ]	
		zhi	[tʂʅ]	
	ㄘ	ch	[tʂ ^h]	
		chi	[tʂ ^h ʅ]	
	ㄙ	sh	[ʂ]	
		shi	[ʂʅ]	
	ㄨ	r	[ʐ]	
		ri	[ʐʅ]	
	ㄗ	z	[ts]	
		zi	[tsʅ]	
	ㄘ	c	[tʂ ^h]	
		ci	[tʂ ^h ʅ]	
	ㄙ	s	[s]	
		si	[sʅ]	
	Medials	ㄟ	i	[i], [ɪ]
		ㄨ	u	[u]
		ㄩ	ü	[y]
	Finals	ㄚ	a	[a]
		ㄛ	o	[o]
ㄜ		e	[ɤ]	
ㄝ		ê	[e]	
ㄞ		er	[ɛ̯]	
ㄟ		ai	[aj]	
ㄠ		ei	[ej]	
ㄡ		ao	[aw]	
ㄣ		ou	[ow]	

Table 25. --Continued.

	ㄛ	an	[an]
	ㄛˊ	en	[ən]
	ㄛˊˊ	ang	[aŋ]
	ㄛˊˊˊ	eng	[eŋ]

In Taiwan, textbooks for elementary school students are mostly written in Han characters with Bopomo, arranging text from top to bottom and from right to left. **Figure 11** shows an example of how Han characters and Bopomo are arranged. **Table 26** shows the description of Bopomo in terms of Universal Orthography.

本^{ㄅㄨㄣˊ} 音^{ㄩㄢ} 這^{ㄓㄥˋ}
 。 符^{ㄈㄨˊ} 個^{ㄍㄜˊ}
 號^{ㄏㄠˋ} 句^{ㄐㄩˋ}
 的^{ㄉㄜˊ} 子^{ㄗㄩˋ}
 樣^{ㄩㄤˋ} 是^{ㄕㄟˋ}
 注^{ㄓㄨˋ}

Figure 11. Samples of Han characters with Bopomo.

Table 26. Bopomo example of Universal Orthography

		directions			
		R	L	U	D
Morphemes	Stem	-	-	-	-
	Prefix	-	-	-	-
	Infix	-	-	-	-
	Suffix	-	-	-	-
Syllables	Syllables	+	-	-	+
	Initial	+	-	-	+
	Medial	+	-	-	+
	Final	+	-	-	+
Phonemes	Consonants	-	-	-	-
	Vowels	-	-	-	-
Phonetic features		-	-	-	-
Supra-segmental		+	-	-	+
Extra information		-	-	-	-
One-to-one		+	-	-	+
One-to-multiple		-	-	-	-
Multiple-to-one		+	-	-	+
Ambiguity		-	-	-	-

4.2.4 Current Chinese language education in Taiwan

Mandarin Chinese is currently the only official language in Taiwan. Mandarin and Han characters are taught through the national education system. Moreover, they are included in the entrance examinations for senior high and college schools. As for the Taiwanese languages, they have never been included in any institutional examinations since Mandarin was adopted by KMT as the so-called National Language.

A total of nine years, including elementary and junior high school, are offered to all Taiwanese citizens as compulsory education. According the education statistics of Ministry of Education (MOE 2000), 99.68% of school-age children in 1999 is attending school. In

the same year, 99.89% of elementary school graduates is attending junior high school. According to the statistics of Ministry of Internal Affairs (2002), 96% of the population older than 15 is literate.⁷⁷

Prior to the execution of the new system 九年一貫教育 or “Consistent Curriculum for Nine Years of Compulsory Education” in 2001, all elementary and junior high schools had to follow a standard curriculum promulgated by MOE. The latest Standard Curriculum for Elementary Education (SCEE) and Standard Curriculum for Junior High School Education (SCJHE) were promulgated in 1993 and 1994, respectively. According to the SCEE, elementary students have to learn the subject ‘*Kuoyu*’ 國語 or National Language for 56 periods (40 minutes per period) out of a total of 188 periods in six years. For details, see **Table 27**, in which number in each cell refers to the periods taught each week. According to the same SCEE, Bopomo is regarded as an auxiliary tool to the learning of Han characters and Mandarin; and all 37 symbols of the Bopomo have to be taught to pupils in the first ten weeks of first semester. According to the latest (1995) elementary textbooks compiled by the National Compilation Agent, the number of Han characters learned by students at each grade is 328 for the first grade, 479 for second grade, 455 for third grade, 529 for fourth grade, 493 for fifth grade, and 385 for sixth grade.

Table 27. Periods for learning Mandarin in elementary school

Subjects	Grades						total
	1	2	3	4	5	6	
Mandarin	10	10	9	9	9	9	56
Others	16	16	24	24	26	26	132
total	26	26	33	33	35	35	188

⁷⁷ Statistics available at <<http://www.moi.gov.tw/W3/stat/home.asp>>

The subject ‘Kuoyu’ is replaced by ‘Kuowen’ 國文 or National Writing and Literature in high school. According to SCJHE, students have to learn Kuowen for 5 periods/per week through graduation. Textbooks for Kuowen consist of colloquial and classical Han writings. The percentage of texts in each type of writing is summarized in **Table 28**. It shows that classical Han writing is getting more weight through years.

Table 28. Percentage of colloquial and classical writings in high school textbooks

	Semesters					
	1	2	3	4	5	6
Colloquial	80%	70%	70%	60%	50%	40%
Classical	20%	30%	30%	40%	50%	60%

4.3 Vietnamese

Vietnam is a country with a rich diversity of ethnicities, including such language groups as Mon-Khmer (94% of total population), Kadai (3.7%; also called Daic or Tai-Kadai), Miao-Yao (1.1%), Austronesian (0.8%), Tibeto-Burman, and Han (Grimes 2000). It is reported that there are 54 official ethnic groups, 86 living languages, and 1 extinct language (Grimes 2000). Recently, *Nung Ven* and *Xapho* (or *Laghuu*), two languages in North Vietnam were discovered and reported by Edmondson’s research team. Among the ethnic groups, *Viet* or *Kinh* is the majority, and it accounts for 87% of Vietnam’s total population (Dang 2000:1). The mother tongue of Viet is called the Vietnamese language. The Vietnamese language is known to its native speakers as *Tieng Viet*, and formerly known as Annamese or Annamite. Vietnamese is currently the official language of Vietnam, and its speakers account for 87% of Vietnam’s population, which was reported to be 75 million in 1995 (Grimes 2000).

4.3.1 Sound system

The classification of the Vietnamese language has been disputed for a long time. However, at present it is widely believed that Vietnamese belongs to the Mon-Khmer language family, which is spoken throughout much of Southeast Asia, primarily in Laos, Vietnam, and Cambodia, but also in Thailand, Burma, the Malay Peninsula, and the Nicobar Islands in the Andaman Sea (Ruhlen 1987: 148). At present, there are about 156 Mon-Khmer languages (Grimes 2000). Among the Mon-Khmer languages, Vietnamese is the most well known, largest, and most widely spoken.

Vietnamese is an isolating language, that is, one in which the words are mostly monosyllables, there is no overt morphological alternation, and syntactic relationships are shown by word order, just as in the cases of Taiwanese and Chinese. Traditionally, Vietnamese was regarded as monosyllabic because most of Vietnamese words consist of single syllables. However, recent statistical studies have shown that there is a clear tendency toward poly-syllabicity in modern Vietnamese (Nguyen 1997: 35). In addition, Vietnamese is a tonal language. Modern Vietnamese possesses six tones, which distinguish different lexical meanings of words. Tone sandhi in Vietnamese is neither as substantial nor as rich as in Taiwanese.

Various foreign influences have influenced the development of the Vietnamese language because of the contacts in the past between the Vietnamese and other peoples. Among them, Chinese is probably the strongest and most lasting one donor language (Nguyen 1971: 153).

The modern Vietnamese language is based on the varieties spoken in Vietnam's capital city of Hanoi and surround Red River basin. Traditionally, Vietnamese vernacular types were proposed by Henri Maspero (1912) dividing the language into two main groups: 1) the Haut-Annam group, which comprehended numerous local speech types of the small

villages stretching from the north of Nghe-An province to the south of Tha-Thien province, and 2) Tonkinese-Cochinchinese, which covers all the remaining territory (Thompson 1987: 78). Generally speaking, Vietnamese variation form a continuum from north to south, each pattern somewhat different from a neighboring one on either side. Hanoi, Hue, and Saigon, located respectively in north, central, and south parts, represent three major remarkable dialects in Vietnam (Nguyen 1997: 10). In this dissertation, Vietnamese refers to the standard Hanoi dialect unless otherwise specified.

4.3.1.1 Consonants

The identification of Vietnamese phonemes may vary from scholar to scholar.⁷⁸ According to Doan (1999: 166), there are 19 consonants in the Hanoi variety of Vietnamese. These consonants were listed in IPA format in **Table 29**. In addition to the 19 consonants, other forms may contain retroflex consonants /tʂ/, /ʂ/, and /zʂ/ (Nguyen 1997: 20).

Table 29. Vietnamese consonants in IPA

		bi-labial	labial-dental	alveolar	palatal	velar	glottal
		-asp/+asp					
voiceless	stop			t / t ^h	c	k	ʔ
voiced	stop	b		d			
voiceless	fricative		f	s		x	h
voiced	fricative		v	z		ɣ	
voiced	lateral			l			
voiced	nasal	m		n	ɲ	ŋ	

⁷⁸ For example, /ʔ/ was not recognized as a phoneme by Nguyen (1997: 20), rather, he recognized /p/ as a phoneme because /p/ nowadays can also occur at the beginning of several loanwords from French, such as *pin* ‘battery,’ and *po-ke* ‘poker.’ However, according to Thompson (1987: 21), the glottal stop could be recognized as a phoneme. The voicing of [b] and [d] are predictable allophones of /p/ and /t/ respectively, following initial /ʔ/ (Thompson 1987: 21). For example, [b] occurs in initial only, and [p] in final only.

4.3.1.2 Vowels

Compared to Taiwanese, Vietnamese vowels are much more complex and difficult. The identification of the vowels varies from scholar to scholar. The Vietnamese vocalic system was divided into upper and lower vocalics (Thompson 1987: 19). The upper vocalics include six vowels, /i u u e ɤ o/. They are formed relatively high in the mouth and characterized by a three-way position (front, back unrounded, and back rounded). Lower vocalics include five vowels, /ɛ ɔ ɐ a ɑ/. They are formed relatively low and characterized by a two-way position distinction (front, back). However, according to Doan (1999), Vietnamese vowels may be categorized into nine simple vowels, four short vowels, and three diphthongs. The vowel /ɐ/ identified by Thompson is considered /ɤ/ by Doan. Despite different opinions on the Vietnamese vowel system, the vowels based on Doan (1999) are listed in **Table 30**, **Table 31** and **Table 32**.

Table 30. Vietnamese simple vowels in IPA

	front	central	back (-rd)	back (+rd)
upper ^{high}	i		u	u
upper mid	e		ɤ	o
lower ^{lower mid}	ɛ			ɔ
low			a	

Table 31. Vietnamese short vowels in IPA

	front	central	back (-rd)	back (+rd)
upper ^{high}				
upper mid			ɤ̃	
lower ^{lower mid}	ɛ̃			ɔ̃
low			ã	

Table 32. Vietnamese diphthongs in IPA

	front	central	back (-rd)	back (+rd)
upper	i_e		u_ɣ	u_o
lower				

4.3.1.3 Tones

Mon-Khmer languages have usually been noteworthy for the linguistic category of register, which most prominently includes voice quality as a contrastive feature. Although Vietnamese is not a classic register language, voice quality as well as pitch phenomena are both important in the tone system of Vietnamese (Nguyen and Edmondson 1997: 1).

Although the modern Vietnamese language is usually described as having six tones, the number of Vietnamese tones could be two, four, six, or eight, based on different criteria of classification (Doan Thien Thuat, p.c.). For example, eight tones were identified in traditional Vietnamese phonology, as listed in **Table 33**. The major reason that modern Vietnamese is recognized as six tones is mainly because the misleading of current Chu Quoc Ngu writing system. The missionaries were not aware of the differences between Falling and Entering categories while they were devising the system (Doan Thien Thuat, p.c.). The major different feature between these two categories is the length of duration: Entering tones have a slightly shorter duration than Falling tones. For example, *nát* ‘broken’ and *nạt* ‘scold’ are pronounced relatively shorter than *nán* ‘linger’ and *nạn* ‘disaster,’ respectively. Entering tones in Vietnamese always require a final *p t c* or *ch*. They are similar to tones 4 and 8 in Taiwanese.

Table 33. Traditional categories of Vietnamese tones

Traditional Categories	平(Level)		上(Uprising)		去(Falling)		入(Entering)	
	浮(F)	沉(S)	浮(F)	沉(S)	浮(F)	沉(S)	浮(F)	沉(S)
Traditional categories ⁷⁹	浮(F)	沉(S)	浮(F)	沉(S)	浮(F)	沉(S)	浮(F)	沉(S)
Modern categories	ngang	huyền	hỏi	ngã	sắc	nặng	sắc	nặng
Numerical tone values	33	21	313	435	35	3	5	3
Tone values in IPA	˥	˩	˨˨˩	˨˨˩	˥˩	˩˩	˥˩	˩˩
Conditions							With finals p t c ch	With finals p t c ch

Tones in modern northern Vietnamese are categorized as *sac*, *nga*, *ngang*, *huyen*, *hoi*, and *nang*. They are composed of contours of pitch combined with certain other features of voice production as given by Thompson (1987: 20) in **Table 34**.

Table 34. Vietnamese tone system (Thompson 1987)

TONE NAME	SYMBOL	PITCH LEVEL	CONTOUR	OTHER FEATURES
sắc	˥	High	Rising	Tenseness
ngã	˥˩	High	Rising	Glottalization
ngang	(unmarked)	High-Mid	Trailing-Falling	Laxness
huyền	˩	Low	Trailing	Laxness, breathiness
hỏi	˨˨˩	Mid-low	Dropping	Tenseness
nặng	˩˩	Low	Dropping	Glottalization or tenseness

Sắc tone

Sac tone is high and rising and tense. Examples are found in the pronunciation of cá ‘fish’ and khó ‘be difficult.’ According to Nguyen and Edmondson’s (1997: 8) acoustic measurements, sac tone of his informant began at a level of 42 semitones and rose to a value of about 48. Thus he assigns sac tone a value of 35 on the Chao’s scale-of-five

⁷⁹ 浮 literally means ‘float’ or ‘up’; 沉 literally means ‘sink’ or ‘down.’ 浮 and 沉 were devised by Vietnamese scholars to refer to the same categories 陰(Yin) and 陽(Yang) of traditional Chinese phonology.

system for transcribing tones. When sac tones do not end up with p, t, k, or ch finals, their shapes are similar to tone 2 in Beijing Mandarin (not Taiwan Mandarin, since tone 2 in TM has become a low falling and then rising tone), such as 麻, 答, 拔, and close to the rising part of tone 5 in Taiwanese, but pitch in sac tone is much higher than in Taiwanese tone 5 (about the height of Taiwanese tone 8, such as 毒, 直 and 逐). When sac tones end up with p, t, k, or ch, they are similar to Taiwanese tone 8.

Ngā tone

Nga tone is also high and rising. Its contour is roughly the same as that of sac, but it is accompanied by the rasping voice quality occasioned by tense glottal stricture. In careful speech such syllables are sometimes interrupted completely by a glottal stop or a rapid series of glottal stops (Thompson 1987: 40). Examples are found in the pronunciation of *sūa* ‘milk’ and *cūng* ‘likewise.’ In Nguyen and Edmondson’s measurements, nga tone began at the level of 44 semitones and rose to the same top of sac tone. Its trajectory showed a characteristic break in the voicing at about 225 msec (about half of the total duration) into the syllable. In terms of Chiung’s (2001f) MOTTA analysis, the angle between falling and rising parts of nga tone is smaller than that of Hoi tone. This tone neither exists in Taiwanese nor in Mandarin.

Ngang tone

Ngang tone is modal; in contour it is nearly level in non-final syllables not accompanied by heavy stress, although even in these cases it probably trails downward slightly (Thompson 1987: 40). Examples are found in the pronunciation of *ba* ‘three’ and *xe* ‘vehicle.’ Nguyen and Edmondson’s measurements coincide with Thompson’s description that ngang tone has a slight fall nature from a value of 45 semitones falling to 44 semitones (Nguyen and Edmondson 1997: 7). Although ngang tone is phonetically a slight falling, it is phonemically regarded as a level tone with a value of 33 on the Chao

scale. It is similar to Mandarin tone 1, such as 媽, 搭, 都, and Taiwanese tone 1, such as 君, 雞, 花, but with relatively lower pitch.

Huyền tone

Huyen tone is lax, starts quite low and trails downward toward the bottom of the voice range. It is often accompanied by a kind of breathy voicing, reminiscent of a sigh (Thompson 1987: 40). Examples are found in the pronunciation of về ‘return home’ and làng ‘village.’ Edmondson points out that huyen tone is lower than ngang tone, beginning at 38 semitones and falling to 36 semitones. He assigns huyen tone a Chao scale value of 21. Huyen tone is very close to Taiwanese tone 3, such as 棍, 庫, 豹. It is also similar to tone 3 in Taiwan Mandarin 叻, 滾, 把 or the falling part of tone 3 in Beijing Mandarin.

Hỏi tone

Hoi tone is tense; it starts somewhat higher than huyen and drops rather abruptly. In final syllables, and especially in citation forms, this is followed by a sweeping rise at the end, and for this reason it is often called the ‘dipping’ tone (Thompson 1987: 41). Examples are found in the pronunciation of phải ‘must’ and ảnh ‘photograph.’ In Nguyen and Edmondson’s (1997: 8) measurements, hoi tone began at 42 semitones and fell to 36 semitones only to rise again to about the level of the beginning. Its trajectory could be a value of 212 or 313 on the Chao scale. Though hoi tone is usually described as low falling and then rising tone, not all Vietnamese speakers have the rising part. Among Nguyen and Edmondson’s six informants, all three Hanoi speakers failed to have the rise, whereas the three non-Hanoi Northerners all had it.

When hoi tone consists of falling and rising contour, it is close to Taiwanese tone 5, such as 群, 財, 猴, and similar to Beijing Mandarin tone 3 (馬, 打, 把) or Taiwan Mandarin tone 2 (文, 純, 陳). When hoi tone consists of only the falling part, it is similar to Taiwanese tone 3 (棍, 兔, 暈) or Taiwan Mandarin tone 3 (馬, 打, 把). The development of

hoi tone from falling-rising to falling seems to be the same as the change of tone 3 in Mandarin from Beijing (falling-rising) to southern forms, such as Taiwan Mandarin (falling).

Nặng tone

Nang tone is also tense; it starts somewhat lower than hoi. With syllables ending in a stop [p, t, c, k] it drops only a little more sharply than *huyen* tone, but it is never accompanied by the breathy quality of that tone (Thompson 1987: 41). For example, *đẹp* ‘be beautiful.’ Other syllables have the same rasping voice quality as *nga*, drop very sharply and are almost immediately cut off by a strong glottal stop. Examples are found in the pronunciation of *mạ* ‘rice seedling.’ According to Nguyen and Edmondson’s measurements, nang tone began at almost the identical height of 42 semitones and fell to about 38 semitones. Nang was much shorter than other tones, and it was assigned a tone value of 32, with a tendency to go lower. Nang tone is similar to Taiwanese tone 4 (闊, 骨, 角), but with relatively longer duration when it does not end up with p, t, k, or ch finals.

4.3.2 Written form in Romanized Chu Quoc Ngu

The spelling of current Vietnamese Romanization Chu Quoc Ngu (CQN) can be traced back to its early history of missionaries in the late sixteenth century and the early seventeenth century. Although Alexandre de Rhodes is usually regarded as the person who provided the first systematization of Vietnamese Romanization, it is apparent that the Vietnamese Romanization resulted from collective efforts, with the influences of diverse missionaries of different national origins (Thompson 1987: 54-55; Ly 1996: 5). For example, *gi* [z] is borrowed from Italian spelling (Thompson 1987: 62), *nh* [ɲ] from Portuguese,⁸⁰ *ph* [f] from ancient Greek (DeFrancis 1977: 58). In some cases, their spelling

⁸⁰ Personal communication with David Silva and Jerold Edmondson.

is not simply influenced by a single language, but could be resulted from multilingual influences. For example, the use of *c*, *k*, *q* [k] could be influenced by French, Portuguese, and Italian. In short, Chu Quoc Ngu possesses the features as described in **Table 35**, in terms of the Universal Orthography.

Table 35. Vietnamese example of Universal Orthography.

		directions			
		R	L	U	D
Morphemes	Stem	-	-	-	-
	Prefix	-	-	-	-
	Infix	-	-	-	-
	Suffix	-	-	-	-
Syllables	Syllables	-	-	-	-
	Initial	-	-	-	-
	Medial	-	-	-	-
	Final	-	-	-	-
Phonemes	Consonants	+	-	-	-
	Vowels	+	-	-	-
Phonetic features		+	-	-	-
Supra-segmental		+	-	-	-
Extra information		-	-	-	-
One-to-one		+	-	-	-
One-to-multiple		-	-	-	-
Multiple-to-one		+	-	-	-
Ambiguity		-	-	-	-

In comparison to Taiwanese Peh-oe-ji, the spelling scheme of CQN is much more complex. The major factors are as follows:

First, CQN requires more diacritics, graphemic sets, and spelling rules because the sound system in Vietnamese is more complicated than in Taiwanese. For example, because Vietnamese has more vowels, additional diacritics < ^ ' ˇ > have to be added to the existing Roman letters to distinguish more vowels.

Second, it resulted from the influences of the multilingual backgrounds of missionaries mentioned above.

The third factor is the influence of language variation. For example, although retroflex /tʂ z ʂ/ are not found in standard Hanoi Vietnamese, they are spoken in other forms and are reflected in the spelling of CQN. In the case of /tʂ/, *tr* was chosen to represent the retroflex sound, and *ch* was used to represent its counterpart of non-retroflex. Consequently, the Hanoi speakers may have difficulty in distinguishing the difference between *tr* and *ch*. For instance, *trông* ‘to plant’ and *chông* ‘husband’ are homophones in Hanoi but spelled differently.

Fourth, the complexity of CQN was deepened due to the historical change over the past 400 years since its invention. For example, the graphemic sets *d* and *gi* were very likely devised in the early period of the seventeenth century to distinguish [d] and [k], which gradually merged and became [z] in modern times (Doan 1999: 163-164). Because the letter *d* was adopted to represent [d], missionaries had to create additional symbol *đ* to represent the phoneme /d/, which still exists and has a pre-glottalized quality [ʔd] today.

Fifth, to some extent, the complexity of CQN was due to the limitation of linguistic analysis in the early period of its development. For example, *k* and *q* were adopted to represent the same phoneme /k/ because the sound [k] followed by a glide [w] was regarded as different from that which occurred elsewhere. Consequently, *q* was chosen to represent [k] followed by a glide [w]. Another example is the short vowel [ɤ̃], which was supposed to be represented with *õ*, so it could be consistent with its long vowel counterpart *o* (i.e., /ɤ̃/). However, it was represented with *â*, which may lead to confusion between /ɤ̃/ and /a/.⁸¹ In short, despite its complexity and inconsistency, the CQN system overall shows a rather

⁸¹ According to Edmondson (personal communication), *â* is actually [ɐ] not [ɤ̃].

good score for efficiency compared to other systems, as Thompson and Thomas commented.⁸²

The consonants, vowels, and tones and their corresponding symbols in CQN are summarized in **Table 36**, **Table 37** and **Table 38**. For a detailed survey on their correspondence, readers may refer to appendix A.

⁸² Cited in Hannas (1997: 86).

Table 36. Symbols for Vietnamese consonants in the spelling of Chu Quoc Ngu

Consonants	CQN	Conditions	Examples
/t/	t		tôi 'I'
/t ^h /	th		thu 'autumn'
/c/	ch		cho 'give'
/tʂ/	tr	Dialects	trông 'to plant'
/k/	k	Followed by i, y, e, ê,	kê 'chicken'
	q	Followed by /w/	quả 'fruit'
	c	Elsewhere	cá 'fish'
/ʔ/	unmarked		ăn 'eat'
/b/*	b		ba 'three'
/d/*	đ		đi 'go'
/f/	ph		Pháp 'French'
/s/	x		xa 'far'
/ʂ/	s	Dialects	so 'compare'
/x/	kh		khi 'while'
/h/	h		hỏi 'ask'
/v/	v		về 'go back'
/z/	d	Must be learned	da 'skin'
	gi	Must be learned	gia 'home'
	g	Followed by i	gì 'what'
/z/	r	Dialects	ra 'go out'
/ʎ/	gh	Followed by i, e, ê	ghi 'record'
	g	Elsewhere	gà 'chicken'
/l/	l		là 'is'
/m/	m		mẹ 'mother'
/n/	n		nam 'south'
/ɲ/	nh		nhớ 'miss'
/ŋ/	ngh	Followed by i, e, ê	ngủ 'rest'
	ng	Elsewhere	ngọc 'jade'

* They are considered /ʔb/ and /ʔd/, respectively (Edmondson, p.c.)

Table 37. Vietnamese vowels in CQN

Vowels	CQN	Conditions	Examples
/i/	i		khi 'when'
	y	Sino-Vietnamese	đồng ý 'agree'
/e/	ê		ghế 'chair'
/ɛ/	e		em 'you'
/ɛ̃/	a	Followed by /p/, /c/	thanh 'sound'
/u/	u		cũ 'old'
/ɯ/	ư		từ 'word'
/o/	ô		cô 'aunt'
/ɤ/	ơ		thơ 'poem'
/ɤ̃/	â		thấy 'see'
/ɔ/	o		co 'to bend'
/ɔ̃/	o	Followed by /ŋ/, /k/	cong 'curved'
/a/	a		ta 'we'
/ã/	ã		ăn 'eat'
	a		tay 'hand'
/i_e/	iê	Elsewhere	tiên 'fairy'
	yê	Preceded by glottal stop /ʔ/ or glide /w/	yêu 'love' truyện 'story'
	ia	Without glide /w/ and coda	bia 'beer'
	ya	Preceded by glide /w/, and without coda	khuya 'midnight'
/u_o/	uô	Elsewhere	chuông 'bell'
	ua	Without coda	vua 'king'
/ɯ_ɤ/	ươ	Elsewhere	được 'able'
	ưa	Without coda	mưa 'rain'

Table 38. Vietnamese tone marks in CQN

Categories	ngang	sắc	huyền	hỏi	ngã	nặng
Tone marks	unmarked	´	`	?	~	.
Tone values	33	35	21	313	435	3
IPA tone value	—	↗	↘	↘↗	↗	·

In CQN system, consonants are represented by one or up to three graphs/letters. For example, /t/ is represented by <t>, /f/ by <ph>, and /ŋ/ by <ng> or <ng>. All graphs for consonants are made from existing Roman letters, except <đ>. Vowels are represented by one or two graphs/letters with appropriate diacritics <^'ˇ>, if necessary. The diacritic <^> represents 'upper' or 'close-mid' vowels, such as <ê> and <ô>; but, occasionally it represents different vowel quality in the only case of <â>, which has a phonetic value [ǎ]; <ˇ> represents 'short' vowels; and <'> shows 'unrounded' feature.

In general, CQN is spelled according to phonemic principles. That is, phonemes, instead of phones, are represented. But, there are some exceptions. For example, final consonant [c] is represented by <ch> instead of <c>, such as *thich*. In this case, [c] is treated by CQN as a phoneme. In fact, [c] is just an allophone of /k/. The spelling was misled by a phenomenon of palatalization, where final /k/ become [c] when preceded by front vowels. Another example of palatalization is the case, in which allophone [ɲ] in final position is spelled as <nh>, rather than <ng>.⁸³

Although the spelling in CQN may be more complex than Taiwanese Peh-oe-ji, there are predictive rules. Most rules are in accordance with vowel positions. That is, front vs.

⁸³ According to Richard Watson (personal communication), the final nh and ch are historically and phonemically correct. In the whole Mon-Khmer area it is common for final palatals to be realized as alveolars with a preceding palatal onglide. Even though the palatal onglide has weakened in most dialects of Vietnamese, it is still distinct from the historical final /n/ and /t/, which have become velar ng and k in some dialects.

back, upper vs. lower, and rounded vs. unrounded. For example, initial /k/ is spelled as <k> if followed by front vowels, or <q> followed by the glide /w/, or <c> elsewhere.

Tone marks are added to the nucleus of the syllable after phonemes are spelled out. For polysyllabic words, a space is usually placed between syllables; but, in some cases, mostly in recent foreign loanwords, hyphen <-> takes this job. For examples, *Áp-ga-ni-xtan* ‘Afghan,’ *Ô-xa-ma Bin La-đen* ‘Osama Bin Laden.’ Occasionally, polysyllabic words are spelled without space or hyphen between syllables, such as *Ucraina* ‘Ukraine,’ and *photo copy* ‘photocopy.’ Some Vietnamese scholars attribute the monosyllabic characteristic to the accommodation to the monosyllabic structure in spoken Vietnamese. However, this author would say that this characteristic of monosyllable in CQN is mainly the consequence of influence of Han characters, and only secondly the accommodation to the spoken language.

The monosyllabic feature in CQN has further resulted in some troubles. The major problem is the lack of word division. Many Vietnamese people, even with college degree, cannot easily identify the boundary between words. They even do not have the concept of words and syllables, and simply regard syllables as words. This is the same common phenomenon in Taiwan and China, where “the word is by no means a clear and intuitive notion” by their speakers (Packard 2000: 14). This ambiguity of word division can be observed from speakers’ contradiction in syntax. For example, ‘socialism’ in Vietnamese could be spelled *xã hội chủ nghĩa* or *chủ nghĩa xã hội*. In the first spelling, it is considered a word of four syllables; and the second spelling is a compound word comprising *chủ nghĩa* ‘doctrine’ and *xã hội* ‘social.’ Another example is the pair of *Á châu* vs. *châu Á* ‘Asia,’ in which *Á* means Asia and *châu* means continental. The ambiguity of word boundary also can be observed from the inconsistent use of capitalization for proper nouns. For example, the proper nouns ‘Taiwan,’ ‘Vietnam,’ ‘Indo,’ and ‘National Language’ are all Sino-

Vietnamese words, but they are differently capitalized as in *Đài Loan*, *Việt Nam*, *Ấn độ*, and *Quốc ngữ*, respectively. In the case of *Quốc ngữ*, it is weird to write *chữ Quốc ngữ*, when *chữ* ‘orthography’ is added to refer to National Orthography. In addition, ‘The Socialist Republic of Vietnam’ could be written as *Cộng hoà xã hội chủ nghĩa Việt Nam*, *Cộng hoà Xã hội chủ nghĩa Việt Nam* and *Cộng hoà Xã hội Chủ nghĩa Việt Nam*.

In short, if spelling reform in CQN is favored, reformers may consider solving the monosyllabic issue as the first priority.

4.3.3 Current Vietnamese language education in Vietnam

The Vietnamese language is the only official language taught through Vietnam’s national education system. In Vietnam, elementary school comprises only five grades, and junior high school consists of four grades. All elementary and junior high schools are under the management of Ministry of Education and Training (MET). Curriculum in elementary has to follow MET’s Guildline to the Curriculum (*Huong dan phan phoi chuong trinh*). Based on the guildline for year 1997-1998, the number of period/per week for the subject ‘Vietnamese’ is summarized in **Table 39**, in which a period is equal to 40 minutes. According to the same guildline, CQN letters and spelling rules are taught in the first 26 weeks. In other words, first graders should have taught all letters and rules by the middle of second semester. Thereafter, drills in reading and writing are continued.

Table 39. Periods for learning Vietnamese in elementary school

Subjects	Grades					total
	1	2	3	4	5	
Vietnamese	13	10	10	8	8	49
Others	11	13.5	14.5	17	17	73
total	24	23.5	24.5	25	25	122

It is reported by the former minister of education, Pham Minh Hac (1998: 78) that in early 1990s only about 80% of school-age children are attending school. The percentage of drop-out in elementary school is between 5% and 10 % in the first half of 1990s. The percentage of students who complete five grades is low: only 50% in 1991-1992, and 60% in 1994-1995. These facts have shown that primary education in Vietnam is not well-conducted and universalized in comparison with Taiwan. Despite the deficiency of primary education, Pham (1998: 80) also reported that 91% of Vietnam's population is literate by 1996. If Pham's report is accurate, CQN must be easily learned. Otherwise, how could Vietnam achieve 91% of literacy with a relatively less developed education system?

CHAPTER 5

METHODOLOGY

This study consists of three sets of experiment: the first set consists of reading comprehension tests, the second one is dictation tests, and the last one is oral reading tests. Test set I compares learners' skill in reading comprehension of different orthographies (i.e., Hanji, Bopomo and CQN) across different languages (i.e., Mandarin and Vietnamese). Test set II compares learners' skill in writing dictation of different orthographies (i.e., Hanji and CQN) across different languages (i.e., Mandarin and Vietnamese). Test set III conducted with the Vietnamese group only, was designed to test CQN learners' skill in reading aloud.

5.1 Reading comprehension tests

The main purpose of this experiment is to determine the average number of years needed for literacy learners in Taiwan and Vietnam to be able to comprehend newspaper articles in their own language and orthography, i.e. Mandarin (Hanji vs. Bopomo) and Vietnamese (Chu Quoc Ngu). Subjects from Taiwan were divided into Hanji and Bopomo groups because both Hanji and Bopomo can be used in the writing of Mandarin Chinese. In other words, there were three contrastive groups for the reading comprehension tests.

5.1.1 Research design

In this experiment, timed reading comprehension tests were conducted with the presumption that subjects' relative reading proficiencies in their own language and orthography will be reflected on their scores. Reading texts were divided into types H(anji), B(opomo) and V(ietnam) based on the orthographic differences. That is, texts in type H are written in Hanji, texts B are written in Bopomo, and texts V are in Romanized Chu Quoc

Ngu. Each reading type consisted of four different texts, and each text was followed by five comprehension questions. The subjects were students from Taiwan and Vietnam, including students in elementary schools and colleges. The subjects were divided into H(anji), B(opomo) and V(ietnam) groups according to their language and orthography. The subjects in each group were told to read prepared texts and answer comprehension questions within 30 minutes. Students' scores were awarded based on the percentage of right answers, ranging from zero to 100. Thus, if a student receives a higher score, it indicates that he or she has better proficiency than others. Furthermore, the learning periods needed for students to reach the maximum proficiency is treated as an index of learning efficiency of students' corresponding orthography. For example, if we find that fifth graders in H group obtain scores statistically equivalent to those students in college, then we conclude that it takes about five years for students in Taiwan to achieve full capacity of reading comprehension in Hanji.

5.1.2 Selection of reading texts

Each orthographic type consisted of four texts. The contents of the texts in types H and B were exactly the same except they were written in the two different writing systems. Type V had the same themes as those in types H and B, so that the influence of different themes on readers' comprehension tests can be minimized. All these test types had articles of similar length, so as to minimize the potential influence of different lengths, where "length" is defined by number of syllables in texts rather than visual layout of texts. In addition, each text began from the top of a new page, and was followed by their questions. Each question is accompanied by five possible answers. One right answer is given among the five choices. The fifth choice is 'no idea,' which means the subject does not know the right answer. The fifth choice is given to avoid or decrease guessing. Each subject's answers were written on a separate answer sheet. The Han characters were printed out with

14 size NewMing fonts 新細明體 by using Taiwanese version of Microsoft Word 2000. Vietnamese letters were printed out with 14 size VnTime fonts. Texts were all printed on A4 size paper. For the original texts and layout for the reading comprehension tests, please refer to appendixes B, C, and D.

The four reading texts were adopted from ‘soft articles’ in newspapers. Soft articles in this investigation are defined as entertainment and sports news. Soft articles were adopted because they are easier to read and understand for readers, as compared to hard articles such as political analyses and editorials. Newspapers were selected from the popular and widely-circulated newspapers of metropolitan areas where the students live. ‘Popular’ was decided by a small preliminary questionnaire survey of the citizens in the metropolis. ‘Widespread’ was decided by its volume of circulation.

Texts for types H and B were chosen from four articles published in Liberty Times 自由時報, all dated May 19, 1998. These four texts with their following questions consisted of a total of 2,346 characters, excluding punctuation. Those 2,346 characters in fact comprise only 492 different characters. Among them, 391 characters are ranked in the first set of frequently used characters, and 93 characters are in the second set, and the others are the least frequently used, as shown in **Table 40**. The ranking of frequently used characters are based on the Report on Usage Frequency of Characters and Words among Elementary School Students,⁸⁴ published by Taiwan’s Mandarin Promotion Council in 2000. Because frequently used characters are usually taught in elementary education, students should have acquired most characters contained in the texts before their graduation from elementary school.

⁸⁴ 國小學童常用字詞調查報告書

Table 40. Number of frequently used characters

Order of Frequent use	Number
1-999	391
1000-2999	93
3000-5000	8

Texts for Vietnamese group were adopted from four articles published in the newspaper Nhan Dan or 'People.' Those articles were dated October 22 and 29, and November 8, 2001. The texts consist of a total of 2297 syllables, excluding punctuation. Syllables were calculated because two major reasons: 1) it is sometimes difficult to identify the boundary between words and morphemes in Vietnamese, and 2) by doing so, we can be in accordance with the Hanji and Bopomo texts, where syllables (i.e., characters) were counted for statistical purposes.

5.1.3 Selection of subjects

Students' backgrounds such as Intelligence Quotient (IQ), Educational Quotient (EQ), first language, parents' social-economic class, and school's ranking level may contribute to students' different scores in the comprehension tests. In this study, students were assumed to have the same IQ and EQ except in a few cases, in which some chosen students might turn out to show extreme difficulty in reading and answering (dyslexia, etc). In such cases, students' scores were not included in this study. It is also assumed that students' first language have only very limited influence since Mandarin has been promoted for more than fifty years and much research has shown that language shift from Taiwanese to Mandarin has taken place completely among the young generations (Young 1989; Huang 1993; Chan 1994).

In order to minimize the influence of background on test results, subjects in the same testing group were chosen from students in the same school. A total of 453 students from Taiwan and 350 students from Vietnam were involved in the experiments. Subjects from Taiwan were chosen from Au-ang Elementary School and Tamkang University. Two classes of each grade from first to sixth were chosen from Au-ang Elementary School in Kaohsiung; and one class from Department of Public Administration of Tamkang University in Taipei was chosen. These choices were made based on my easy access to these schools, and students' availability. Further, subjects in each class were randomly divided into two groups. Group H was conducted with Hanji texts, and group B was conducted with Bopomo texts. The subjects from Taiwan are summarized in **Table 41**.

Table 41. Subjects from Taiwan

GRADES	GENDER	N. in Hanji	N. in Bopomo
1st graders	male	22	15
	female	14	17
	Total	36	32
2nd graders	male	17	18
	female	13	16
	Total	30	34
3rd graders	male	18	16
	female	12	14
	Total	30	30
4th graders	male	18	11
	female	13	22
	Total	31	33
5th graders	male	10	21
	female	24	11
	Total	34	32
6th graders	male	24	23
	female	15	12
	Total	39	35
collegians	male	12	8
	female	22	15
	Total	34	23
Total	male	121	112
	female	113	107
	Total	234	219

Subjects from Vietnam were chosen from To Hien Thanh Elementary School and Department of Oriental Studies of Hanoi National University. Some volunteers from different universities also participated in this study. Because To Hien Thanh is a tiny school with only two classes in each grade, the experiments were conducted on all students there. In addition, only five grades were tested because there are only five grades in Vietnam's elementary schools. The number of subjects in each grade is summarized in **Table 42**.

Table 42. Subjects from Vietnam

GRADES	GENDER	N
1st graders	male	33
	female	33
	Total	66
2nd graders	male	35
	female	24
	Total	59
3rd graders	male	22
	female	36
	Total	58
4th graders	male	27
	female	33
	Total	60
5th graders	male	28
	female	29
	Total	57
collegians	male	8
	female	42
	Total	50
Total	male	153
	female	197
	Total	350

5.1.4 Testing procedure

The tests were conducted with Taiwanese students at different times between January and May 2002. The first, second, and third graders of Au-ang Elementary School were tested on January 17, 11, and 15, respectively. The fourth graders were examined on April

22. The fifth graders were tested on April 22 and May 2, and the sixth graders were tested on April 25 and May 1.

Vietnamese subjects from the To Hien Thanh school were examined in the first half of December, 2001. Twenty two volunteers from different universities were recruited to be the collegian subjects in late May, 2002. Because the number recruited from voluntary was not considered sufficient, one class from the Department of Oriental Studies of the Hanoi National University was borrowed in early June 2002 to conduct the experiments.

Subjects in groups H, B and V were all told to read the prepared texts and answer the multiple-choice questions within 30 minutes. They were told to choose the fifth choice (i.e., 'no idea') in each question in the case if they were not sure about the correct answer. During the tests, they were not allowed to discuss the questions with each other. When time was up, their answer sheets were collected even if they had not finished reading.

For the collegian subjects, their time in reading comprehension tests was measured. The time in progress during the tests was written down on a blackboard every half minute. The students were told to write down the time appearing on the blackboard when they finished.

5.1.5 Data analysis

Students' scores on reading comprehension tests were keyboarded and then analyzed by using statistics software SPSS version 10. The statistical techniques employed in the reading comprehension tests were t-tests, ANOVA (analysis of variance) and post hoc tests.

In this section, only general concepts of the statistical techniques are provided. For details about statistics in linguistic studies, readers may refer to Rietveld (1993), Butler (1985), or Woods, Fletcher, and Hughes (1996).

Before going through each specific technique, some important statistical terms need to be introduced. First, whenever we do statistics, we always set up hypotheses for testing.

The hypotheses come in pairs: a research or alternative hypothesis (denoted by H_1 or H_a) versus the null hypothesis (H_0). The null hypothesis always states that there is nothing special going on in here. As Rietveld (1993: 5) described, “The null hypothesis (H_0) is the hypothesis actually tested in a statistical testing procedure. A null hypothesis is formulated in such a way that we can calculate the probability that H_0 is true.” If a null hypothesis is rejected under a certain condition of alpha level, then the alternative hypothesis (research hypothesis) is automatically accepted. The alpha level (α) is the probability of rejecting a null hypothesis when this hypothesis is in fact true. It usually refers to the significance level of a test. In this study, the significance level was set at $\alpha = 0.05$. That is, if $p < 0.05$ (less than 5 chances in 100 that the null hypothesis is valid), then we accept the research hypothesis. In other words, there are something special going on between the tested groups. Usually, a null hypothesis is not stated, but is implicitly present whenever statistical hypothesis-testing is executed. So, throughout the dissertation, the detailed null hypotheses are not mentioned, but only the results of the statistical tests. Although the significance level in this study was set at 5%, most significant results are in fact with a p value less than 0.01.

T-test: “The t-test is a parametric statistical test that tests whether the means of sets of scores from two samples are significantly different from each other” (Fasold: 1993: 98). For instance, independent-samples t-tests were conducted with scores of sixth graders between groups H and B. The mean scores of sixth graders in groups H and B are 96.92 and 91.86, respectively. Does it mean that groups H and B are statistically different from each other? The null hypothesis is that there is nothing different between the means of groups H and B. The results of independent-samples t-test by using SPSS reveal that $p > 0.05$, so we cannot reject null hypothesis. Thus, we have to conclude that there is nothing different in sixth graders between groups H and B.

ANOVA: In the case of comparing mean scores of reading samples, a t-test can only test between two groups. If we need to compare more than two groups simultaneously, then ANOVA are required. There are several ways of doing analysis of variance. General Linear Model Univariate or Univariate ANOVA (hereafter, UANOVA) under SPSS was employed in this study. UANOVA provides regression analysis and analysis of variance for one dependent variable by one or more factors and/or variables. The factor variables divide the population into groups. Using UANOVA, we can test null hypotheses about the effects of other variables on the means of various groupings of a single dependent variable. However, the results of ANOVA only provide us the general information. They do not point out which two means are significantly different from each other. If we need to specify which pairs of mean scores are significantly different from each other, we may use another technique such as post-hoc tests.

Post-hoc tests or post-hoc comparisons: There are several choices of *post-hoc* comparisons. In this study, Tukey's Honest Significant Difference (HSD) tests were employed for the purpose of simultaneous comparisons among a set of groups (variables). The results of post-hoc tests usually generate homogeneous subsets. Mean scores in the same subset are considered to be no significant difference, while they are significantly different from each other if they are listed in different subsets.

5.2 Dictation tests

Although students may be able to comprehend the reading texts, they may not be able to write accurately. Thus, dictation tests were conducted to examine the accuracy of writing in comparison to students' reading comprehension ability. The subjects were divided into groups Taiwan and Vietnam according to their nationality background. In other words, groups H and B for the reading comprehension tests were not distinguished any more in the

Taiwanese group. Methodological details of the dictation tests are the same as those in reading comprehension tests, unless otherwise specified.

5.2.1 Research design

The criteria for dictation tests are the same in both the Taiwan and Vietnam groups. The dictation consisted of two texts: 1) soft article mainly containing frequently used Han characters and words, and 2) hard article mainly containing less frequently used Han characters and words. The texts for dictation were first tape-recorded, and later be played to the students. Students were told to listen to the passages and write down what they heard. After the students completed their dictation, errors in the dictation tests were marked at a later time for further analysis of the number and kinds of errors made; for example, words or characters left out, missed strokes in the Han characters and misspelled words in the Bopomo and Romanized scripts. The percentage of Han characters correctly written by a Taiwanese subject was regarded as an index or score to evaluate his/her proficiency in writing Han characters. On the other hand, the percentage of correct segments (including sound segments and suprasegmental tones) received by a Vietnamese subject was regarded as an index or score to evaluate his/her proficiency in writing CQN.

5.2.2 Preparation for passages

The dictation texts for Taiwanese group included a short funny story and a news item on economics, respectively. The funny story was translated from the dictation text for Vietnamese group, and it consisted of a total of 130 characters, excluding punctuation. This text is in fact composed of only 73 different characters, as shown in **Table 43**. Text two comprised a total of 98 characters or 78 different characters, as shown in

Table 44. The frequencies of characters are in accordance with the statistics described in the Report on Usage Frequency of Characters and Words among Elementary School

Students (RUF CWESS), published by Taiwan's Mandarin Promotion Council. In **Table 43** and

Table 44, column F is the order of usage frequency reported by RUF CWESS. For example, the character 的 is the first frequently used character, and 吞 is ranked the 1623 in frequently used. Column S is the semester the character appeared on Mandarin textbook the first time. In average, characters in text one have an average of 270 in frequency order and are taught by second semester. Characters in text two have an average of 494 in frequency order and are taught by third semester. For the original texts, readers may refer to appendix.

Table 43. Characters contained in dictation one for Taiwanese group

	F	S		F	S		F	S		F	S
的	1	1	時	31	1	之	119	3	怪	525	4
一	2	1	下	38	1	常	120	3	剛	614	3
是	3	1	得	43	1	果	126	2	急	676	2
了	4	1	裡	49	1	外	144	1	貴	692	4
不	5	1	去	51	1	題	148	3	偷	898	3
我	6	1	很	57	1	主	165	3	險	917	5
有	7	1	心	65	1	少	189	2	桃	929	6
在	8	1	把	66	1	吃	191	1	桌	966	2
來	10	1	面	67	2	聽	206	1	危	1098	4
大	11	1	如	77	3	答	225	3	核	1206	12
上	12	1	三	78	1	它	230	2	粒	1236	4
到	14	1	發	79	2	放	260	1	肚	1281	4
們	15	1	沒	82	1	覺	261	2	吞	1623	6
個	16	1	同	84	2	非	306	3			
你	18	1	而	90	4	孩	342	2			
子	19	1	回	91	1	啊	372	2			
就	24	1	回	91	1	奇	425	2			
要	26	1	媽	99	1	阿	490	3			
說	27	1	從	107	2	忙	513	2			
中	28	2	問	111	1	誰	523	1			

Table 44. Characters contained in dictation two for Taiwanese group

	F	S		F	S		F	S		F	S
的	1	1	最	102	1	業	290	3	需	658	7
一	2	1	行	110	2	百	294	4	費	700	3
在	8	1	之	119	3	今	319	1	增	744	3
大	11	1	美	131	1	品	352	4	商	808	3
上	12	1	等	134	3	市	366	4	復	826	3
子	19	1	電	167	4	場	369	3	彈	948	7
說	27	1	五	168	1	達	382	4	價	1005	9
天	29	1	將	178	2	包	409	2	括	1026	3
下	38	1	加	187	3	導	412	4	億	1195	6
出	39	1	少	189	2	求	433	2	股	1312	6
多	46	1	新	190	3	報	436	2	週	1343	10
年	48	2	情	197	2	爾	441	5	廠	1431	9
分	56	3	全	207	2	千	458	4	幅	1574	7
動	59	2	元	237	4	強	459	3	刊	1712	8
面	67	2	特	240	4	至	463	6	勁	1763	5
氣	88	2	六	241	1	半	499	1	銷	2260	8
回	91	1	球	244	1	反	504	5	售	2435	0
頭	95	1	帶	248	2	英	533	4	甦	3398	0
現	96	2	量	267	3	消	570	4			
體	100	3	期	284	2	景	604	3			

Dictation tests for the Vietnamese group also included two texts. Text one was a funny story and thus considered a soft article, which consisted of 119 syllables, including 308 sound segments and 119 tones. The second text was a news on economic development published by Nhan Dan, dated August 11, 2001. This text contains 113 syllables, including 318 sound segments and 113 tones. For the original texts, readers may refer to appendix F.

The texts for the dictation tests were tape recorded by young female native speakers of Mandarin and Vietnamese. In the passages for each text, the full text was completely

read first at a normal speed. After that, each sentence in the text was read again with three repetitions and then followed by the next sentence. The repeated sentences were read at a relatively slower speed. Finally, the full text was completely read again at a normal speed. Thereafter, the passages for second text were recorded after the first passages with a five-second pause.

In summary, the first and second texts for Taiwanese group were tape-recorded with a length of 14 minutes 55 seconds, and 7 minutes 55 seconds, respectively. As for Vietnamese, the text durations were 13 minutes 27 seconds, and 11 minutes 13 seconds.

5.2.3 Selection of subjects

The subjects involved in the dictation tests are the same as those in reading comprehension tests. Readers may refer back to section 5.1 for the details.

5.2.4 Testing procedure

Dictation tests were conducted after subjects' reading comprehension tests. A ten minute break was allowed before the subjects began the dictation tests. Subjects were told to write down whatever they heard from the audio cassette player. Subjects in the Taiwanese group were told to write down standard Han characters for the passages. They were told to write down Bopomo only if they have no idea about the standard characters. Subjects in the Vietnamese group were simply told to write down the passages in CQN. The cassette was first played for a few seconds to make sure everyone in the test was able to appropriately hear the passages. When everyone was ready, the cassette was played back from the beginning of passages.

Subjects' dictations were collected at the end of play back, and digitalized at a later time for further analysis. Error checks in the Vietnamese dictations were assisted by my Vietnamese assistants, who were recruited from the collegians.

5.2.5 Data analysis

In addition to t-tests, UANOVA, and post-hoc tests, factor analysis was adopted for analyzing the scores on the dictation tests.

In some situations where numerous variables are used to characterize objects, we may want to know to what extent these variables have something in common. In other words, we want to know to what extent they measure the same “underlying” variables. Regularly, it occurs that clusters of variables turn out to be intercorrelated. Then we may want to assess whether or not these clusters measure common aspects of the dimensionality or domain. In other words, we are trying to figure out whether or not it is possible to reduce whole variables to fewer sets. An example of its application is the error analysis on Han characters. In dictation test one of the Taiwanese group, twelve error types were first identified. Five factors were further extracted from the twelve error types by using factor analysis. This result reveals that those twelve error types may be regrouped into only five basic types. For detailed procedure of doing factor analysis and error analysis, readers may refer to chapter six.

5.3 Oral reading tests

The purpose of the oral reading tests was to examine the accuracy of oral reading with regard to orthography. Oral reading tests were only conducted with Vietnamese group because there is no way for the Hanji beginners to be able to read advanced Han characters unless they had acquired the characters in advance. On the contrary, the Vietnamese writing system has very limited letter inventory and spelling rules. Once Vietnamese students have learned the letters and rules, they are able to read Vietnamese words and sentences. Therefore, oral reading tests were proposed to see how well Vietnamese students could apply their spelling skill in oral reading. In other words, we want to know how long it would take for a literacy learner of Vietnamese to be able to read CQN aloud.

5.3.1 Research design

Subjects were told to read aloud two prepared texts in CQN, one of which was considered soft and the other hard. Subjects' oral reading was timed and tape-recorded. Later on, their oral reading was transcribed and keyboarded for further analysis. It was presumed that the more correct segments the subjects have, the more proficiency the subjects have in oral reading skill. So, the correct segments produced by each subject, including sound segments and tonal features, were calculated. The percentage of correct segments was regarded an index or score to evaluate subjects' oral reading skill.

5.3.2 Selection of oral reading texts

The first text is a daily life story, talking about relationship between the author and her grandma. It consists of 101 syllables, including 271 sound segments and 101 tones. The second text is a news report on economics, adopted from Nhan Dan newspaper, dated August 11, 2001. It comprises 104 syllables, including 296 sound segments and 104 tones. Both texts were printed on a A4 paper with 14 size VnTime fonts.

5.3.3 Selection of subjects

Because of time limitations, not all Vietnamese subjects in previous tests were examined with oral reading. Only one class in each grade from first to third was chosen for oral reading tests. As for collegian subjects, only thirteen were randomly chosen from previous tests.

5.3.4 Testing procedure

Oral reading tests were conducted after subjects were tested in reading comprehension and dictation tests. My Vietnamese assistants were divided into four teams to conduct tests. Each team consisted of at least two Vietnamese; one was in charge of machine operation and timing, and the other held the microphone and paper for the subjects. Subjects were

individually called to participate in each team. They were told to read aloud whatever they saw on the prepared paper. Their pronunciations were recorded with a cassette recorder through microphone. At a later time, subjects' oral reading was first transcribed into paper, and then digitalized by using Microsoft Excel. This job was conducted by my Vietnamese assistants under my supervision. All the assistants were recruited from collegians with a foreign language or relevant major, and had received ten hours training in linguistics offered by me.

5.3.5 Data analysis

Same statistical techniques in previous tests were adopted by using SPSS. For details of the statistical techniques, readers may refer back to previous sections.

CHAPTER 6

RESULTS AND DISCUSSION

Statistical results and their discussion are arranged in four sections. Section 6.1 describes the results of reading comprehension tests. Results of dictation tests are divided into two sections. Section 6.2 contains the results of dictation tests in the Taiwanese group, and section 6.3 comprises the results in the Vietnamese group. The last section in this chapter presents the results of oral reading tests in the Vietnamese group.

6.1 Results of reading comprehension tests

A total of 803 subjects for reading comprehension tests are divided into three groups, i.e., Hanji, Bopomo, and CQN. The experimental results of each group are presented in each of the following sub-sections.

6.1.1 Hanji group

There were a total of 234 students, consisting of 121 males and 113 females, involved in the Hanji group. Their scores on reading comprehension tests are shown in **Table 45**.

Table 45. Scores received by students in Hanji group

GRADES	GENDER	N	Mean	Sd.
1st graders	male	22	28.41	19.54
	female	14	39.29	33.27
	Total	36	32.64	25.87
2nd graders	male	17	49.41	34.95
	female	13	50.77	32.46
	Total	30	50.00	33.32
3rd graders	male	18	87.22	14.37
	female	12	95.42	5.82
	Total	30	90.50	12.27
4th graders	male	18	96.39	4.47
	female	13	97.31	2.59
	Total	31	96.77	3.77
5th graders	male	10	93.50	7.84
	female	24	93.33	17.42
	Total	34	93.38	15.11
6th graders	male	24	95.83	6.54
	female	15	98.67	3.99
	Total	39	96.92	5.81
collegians	male	12	95.83	6.69
	female	22	97.50	6.12
	Total	34	96.91	6.28
Total	male	121	75.66	31.92
	female	113	83.94	28.51
	Total	234	79.66	30.54

The statistical results of Univariate Analysis of Variance (hereafter, UANOVA) reveal that gender is not significantly different at the 5% significance level. However, grade level is a significant factor. Three homogeneous subsets out of the seven grade categories were suggested by *post hoc* comparisons of UANOVA, as shown in **Table 46**. This means that scores received by grades in each subset are not significantly different from each other. For example, the scores of third, fourth, fifth, sixth graders and collegians, are considered not statistically different though their means are slightly different. Three subsets indicate that we may reclassify all subjects into three groups. The first group contains the first graders, the second group comprises the second graders, and the third group consists of all other levels.

Table 46. Homogeneous subsets by grades in Hanji group

GRADES	N	Subset		
		1	2	3
1st graders	36	32.64		
2nd graders	30		50.00	
3rd graders	30			90.50
5th graders	34			93.38
4th graders	31			96.77
collegians	34			96.91
6th grader	39			96.92
Sig.		1	1	0.758

Alpha = 0.05

We may rearrange the mean scores by grades as shown in **Figure 12**, so we can get a better picture. **Figure 12** shows that scores received by elementary students significantly increase over the years, and they have statistically reached the same level as college students by the third grade (suggested by the results of *post hoc* tests). Because college students are considered well-educated literates in a society, this result indicates that literacy beginners in Han characters may have acquired the basic reading skills in reading soft articles after about three years of learning. In addition, we may divide learners' learning of skills in reading soft articles into three steps. They are early, intermediate, and advanced steps, each of which step requires about one year's study. "Soft articles" are specified because the reading comprehension tests are conducted with easy-to-read texts.

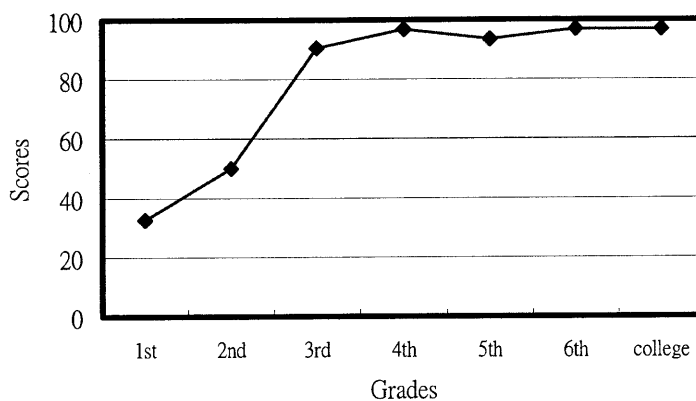


Figure 12. Scores received by grades in Hanji group.

In the reading comprehension tests, all subjects from first to sixth grades were given a fixed time of thirty minutes to complete their reading. The period of thirty minutes was not enough time for most pupils in early grades, but was more than enough for pupils in higher grades. As for the collegian subjects, they were given timed-reading tests. The statistical results of collegians' time spent on reading comprehension tests are listed in **Table 47**. It shows that collegians spent an average of 4.86 minutes to complete their reading tests.

Table 47. Time spent on reading Hanji texts by collegians

	N	Minimum	Maximum	Mean	Sd.
Mins.	34	1.0	8.0	4.86	1.63

6.1.2 Bopomo group

The Bopomo group consisted of 112 males and 107 females, and their descriptive statistics are listed in **Table 48**. Their statistical results of UANOVA reveal that grade, instead of gender, is a significant factor at the 5% significance level. Further *post hoc* tests suggest three homogeneous subsets by grade in Bopomo group, as shown in **Table 49**.

**Table 48. Scores received by students in
Bopomo group**

GRADES	GENDER	N	Mean	Sd.
1st graders	male	15	21.67	25.33
	female	17	23.53	23.17
	Total	32	22.66	23.83
2nd graders	male	18	27.50	21.02
	female	16	27.50	22.73
	Total	34	27.50	21.51
3rd graders	male	16	58.44	23.86
	female	14	61.79	28.26
	Total	30	60.00	25.60
4th graders	male	11	61.82	27.14
	female	22	75.00	22.31
	Total	33	70.61	24.42
5th graders	male	21	68.33	29.51
	female	11	61.82	22.61
	Total	32	66.09	27.14
6th graders	male	23	89.78	15.34
	female	12	95.83	7.33
	Total	35	91.86	13.34
collegians	male	8	95.63	5.63
	female	15	95.00	11.80
	Total	23	95.22	9.94
Total	male	112	59.82	34.00
	female	107	61.78	34.04
	Total	219	60.78	33.96

**Table 49. Homogeneous subsets by grades in
Bopomo group**

GRADES	N	Subset		
		1	2	3
1st graders	36	22.66		
2nd graders	30	27.50		
3rd graders	30		60.00	
5th graders	34		66.09	
4th graders	31		70.61	
6th grader	39			91.86
collegians	34			95.22
Sig.		0.978	0.492	0.997

Alpha = 0.05

Table 49 indicates three homogeneous groups among the subjects. The first group includes first and second graders. The second group consists of third, fourth, and fifth graders. The last group comprises sixth graders and collegians. As **Figure 12** has shown, scores of subjects in these groups increase over years. The first group is the lowest and the third group is the highest. It indicates that pupils are experiencing three steps toward full skills in reading Bopomo. Each Bopomo learning step requires two or three years before advancing to the next step. Full skills are defined as the same level as those collegians have.

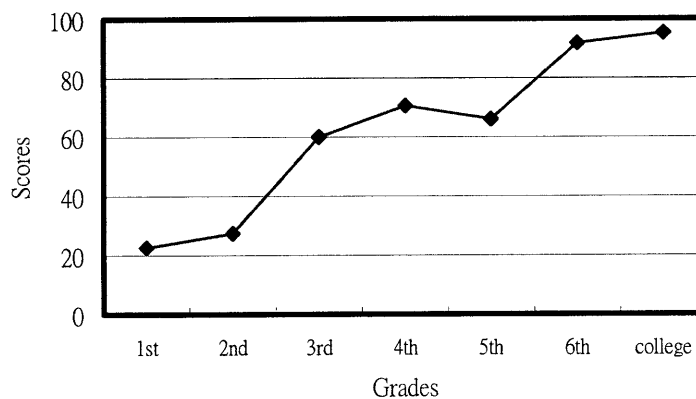


Figure 13. Scores received by grades in Bopomo group.

In contrast to a period of thirty minutes given to the pupils, the collegian subjects were given time-reading tests. The descriptive statistics of their tests are listed in **Table 50**, which reveals that collegian readers require a mean of 13.28 minutes to read those four Bopomo texts. As compared to the Han group's 4.86 minutes, Bopomo group spent more time on reading tests.

Table 50. Time spent on reading Bopomo texts by collegians

	N	Minimum	Maximum	Mean	Sd.
Mins.	23	5.0	27.0	13.28	6.41

6.1.3 Chu Quoc Ngu group

There were 153 male and 197 female subjects participating in the Vietnamese Chu Quoc Ngu group (CQN group). The reading comprehension tests were conducted on elementary students in early December 2001. It was found necessary to narrow the time span between the first grade and second grade down to make the learning process more precise. The first grade pupils were tested again with the same examination three months later. Thus, those re-tested pupils are marked as 1.5 graders in **Table 51**, **Table 52** and throughout the whole chapter. For some reason, nine students were missing and there were only a total of 57 students in the second examination. Scores received by students in the CQN group are described in **Table 51**.

Table 51. Scores received by students in CQN group

GRADES	GENDER	N	Mean	Sd.
1st graders	male	33	10.15	8.43
	female	33	18.03	22.15
	Total	66	14.09	17.09
1.5 graders	male	27	24.44	13.82
	female	30	25.33	15.48
	Total	57	24.91	14.59
2nd graders	male	35	53.14	20.37
	female	24	60.21	23.38
	Total	59	56.02	21.73
3rd graders	male	22	76.82	23.17
	female	36	82.36	17.67
	Total	58	80.26	19.92
4th graders	male	27	85.37	22.57
	female	33	88.48	16.42
	Total	60	87.08	19.32
5th graders	male	28	94.46	6.14
	female	29	95.34	4.21
	Total	57	94.91	5.22
collegians	male	8	91.25	12.75
	female	42	92.86	8.64
	Total	50	92.60	9.27
Total	male	180	56.81	35.74
	female	227	67.62	34.29
	Total	407	62.84	35.30

Statistical results of UANOVA reveal that both grade and gender are significant factors at the 5% significance level. However, if we increase the significance level to 1%, the gender factor is excluded. Since gender is not our main concern in this study, this factor is excluded from further analysis. Further *post hoc* tests were conducted, and five subsets were generated as shown in **Table 52**. Among the subsets, the fourth graders are listed twice, appearing both in subsets four and five. This means that the fourth graders are not significantly different from the fifth graders and collegians in the fifth subset. Neither are they significantly different from the third graders in the fourth subset. In other words, we could consider the fourth graders the component either in subsets four or five.

Table 52. Homogeneous subsets by grades in CQN group

GRADES	N	Subset				
		1	2	3	4	5
1st graders	66	14.09				
1.5 graders	57		24.91			
2nd graders	59			56.02		
3rd graders	58				80.26	
4th graders	60				87.08	87.08
collegians	50					92.60
5 th graders	57					94.91
Sig.		1	1	1	0.279	0.139

Alpha = 0.05

We may divide the subjects into five groups. The first, second, and third groups include the first, 1.5, and second graders, respectively. The fourth group contains the third graders, and the fifth group consists of fourth and fifth graders and collegians. To have a better picture of these groups, mean scores of grades are rearranged in **Figure 14**. It shows that mean scores are increasing from the first to the fifth groups. It implies that pupils' reading skills are improving and they have statistically achieved the same level as collegians in group five (i.e., since fourth graders). Although the mean score of the fourth

group (i.e., third graders) is significantly different from that of group five (i.e., the fourth, fifth graders and collegians are interpreted simultaneously), it is not different from the fourth graders. Because the boundary between the fourth and fifth groups is not statistically clear, we may consider the third grade the transition to achieve to collegian level. In short, literacy learners of Vietnamese CQN may have achieved collegian level at somewhere around the third or fourth grade.

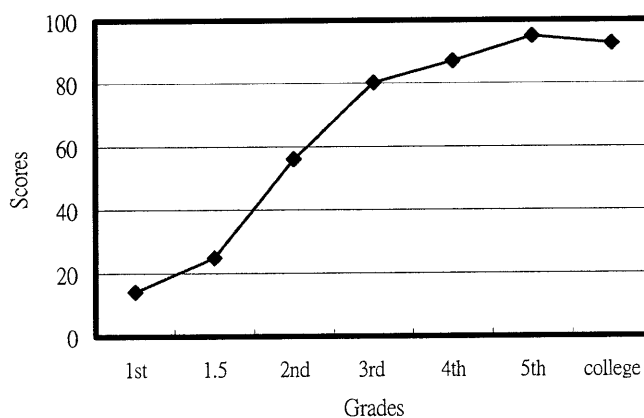


Figure 14. Scores received by grades in CQN group.

There were fifty collegians involved in the CQN group. The scores on the reading comprehension tests have a mean of 11.6 minutes, as summarized in **Table 53**.

Table 53. Time spent on reading CQN texts by collegians

	N	Minimum	Maximum	Mean	Sd.
Mins.	50	5.0	19.0	11.60	3.13

6.1.4 Three scripts in comparison

In this section, the three script groups, i.e., Hanji, Bopomo, and CQN, are examined simultaneously so we can have better understanding of their relationships to each other. To see these relationships, their descriptive statistics and graphs are first rearranged in **Table 54** and **Figure 15** for a better picture.

Table 54. Mean scores received by grades and scripts

GRADES	SCRIPTS	N	Mean	Sd.
1st graders	bopomo	32	22.66	23.83
	hanji	36	32.64	25.87
	CQN	66	14.09	17.09
	Total	134	21.12	22.63
1.5th graders	CQN	57	24.91	14.59
	Total	57	24.91	14.59
2nd graders	bopomo	34	27.50	21.51
	hanji	30	50.00	33.32
	CQN	59	56.02	21.73
	Total	123	46.67	27.58
3rd graders	bopomo	30	60.00	25.60
	hanji	30	90.50	12.27
	CQN	58	80.26	19.92
	Total	118	77.71	22.77
4th graders	bopomo	33	70.61	24.42
	hanji	31	96.77	3.77
	CQN	60	87.08	19.32
	Total	124	85.12	20.74
5th graders	bopomo	32	66.09	27.14
	hanji	34	93.38	15.11
	CQN	57	94.91	5.22
	Total	123	86.99	20.41
6th graders	bopomo	35	91.86	13.34
	hanji	39	96.92	5.81
	Total	74	94.53	10.34
collegians	bopomo	23	95.22	9.94
	hanji	34	96.91	6.28
	CQN	50	92.60	9.27
	Total	107	94.53	8.73
Total	bopomo	219	60.78	33.96
	hanji	234	79.66	30.54
	CQN	407	62.84	35.30
	Total	860	66.89	34.59

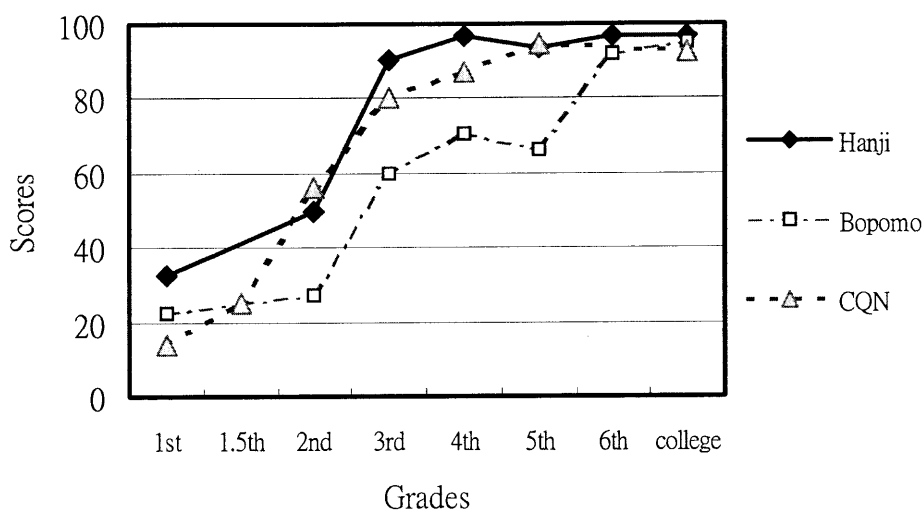


Figure 15. Mean scores received by grades and scripts.

Furthermore, statistical analysis was conducted among groups in each grade. UANOVA statistical technique was employed if there were three groups in comparison; independent samples t-test was applied if there were only two groups. They were conducted by using SPSS 10 with the 5% significance level, and their statistical results are as follows:

First, there is no statistically significant difference among the mean scores of Hanji, Bopomo, and CQN groups among the students of sixth grade and college. This result reveals that orthography is not a significant factor in affecting students' scores on the reading comprehension tests once students are in sixth grade or higher. Because students in sixth grade are considered to have the same reading level as collegians, this result implies that Hanji, Bopomo, and CQN can all serve as independent writing systems without affecting readers' comprehension.

Second, among the students from the second to the fifth grades, there is no significant difference between the mean scores of Hanji and CQN groups. However, the Bopomo group is significantly different from the others. The statistical results of *post hoc* tests

among these three groups are listed in **Table 55**. Why is Bopomo group statistically different from the other two writing systems? There are two potential factors: 1) homophones in Bopomo may have reduced students' reading comprehension, and 2) fewer resources and less training in Bopomo may also have prevented their learners from becoming skilled as fast as learners in the Hanji group. If the first factor is true and in effect, then homophones in Vietnamese would be expected to affect the CQN group. Since this is not the case, it seems that factor two must be the major and dominant factor. Recall that Bopomo is only taught to pupils through the first ten weeks of the first grade. After that period, Han characters are taught exclusively and new characters are added every semester through sixth grade. Reading and writing in Han characters are repeatedly practiced semester by semester as the priority goal of literacy. On the contrary, Bopomo is used as an auxiliary tool in learning standard pronunciation of Han characters and Mandarin. Consequently, learners' acquisition of Bopomo is not as fast as Han characters. Here 'fast' here is defined based on the comparison of mean scores between Hanji and Bopomo groups. In fact, if we count the time students spend learning the basic writing unit, i.e., letters or characters, Han characters are probably not learned as fast as Bopomo. In the case of Taiwan, literacy beginners can acquire all thirty-seven letters of Bopomo in ten weeks and then start drills in reading and writing. However, they have to keep learning Han characters through six grades, and even during high school learning continues as they are taught classical Han writing. In this study, although third grade students in Han group have achieved the same score as collegians, it does not necessarily mean that students will have the same achievement when they are encountering texts other than 'soft articles.' Hard articles, in contrast to soft ones, require acquisition of more Han characters. In this situation, it might take more years for the pupils to achieve the same reading level as collegians. Further experiments on this issue are recommended.

**Table 55. Homogeneous subsets by scripts
(second to fifth graders)**

SCRIPTS	N		
		1	2
Bopomo	129	55.66	
CQN	234		79.47
Hanji	125		83.12
Sig.		1	0.261

Alpha = 0.05

The third set of statistical results comes from comparisons among different scripts in the early stage of literacy acquisition, i.e., first and 1.5th grades. To examine these beginners, the pupils were divided into four groups based on their orthographic background. Because pupils in CQN group were tested twice, they are regarded as two groups, i.e., CQN1 and CQN2, for the first and second time tests, respectively. Therefore, we have a total of four groups, Hanji, Bopomo, CQN1, and CQN2. UANOVA and *post hoc* tests were conducted to compare these groups simultaneously, and the results are shown in **Table 56**.

**Table 56. Homogeneous subsets by scripts
among literacy beginners**

SCRIPTS	N	subset	
		1	2
CQN1	66	14.09	
Bopomo	32	22.66	22.66
CQN2	57		24.91
Hanji	36		32.64
Sig.		0.174	0.082

Alpha = 0.05

Table 56 shows two homogeneous subsets are generated, and the Bopomo group is serving as the boundary between these two subsets. To further decide into which subset the Bopomo best fits, independent-samples t-tests were employed. The results of t-tests between Bopomo and Hanji groups showed no statistically significant difference. However,

Bopomo group is significantly different from CQN1 group at the 5% significance level. Therefore, Bopomo is better considered in the second subset. Thus, the first subset contains CQN1, and the second subset consists of Bopomo, CQN2, and Hanji groups. This result indicates that pupils in Bopomo and Hanji groups have the statistically same mean scores of reading comprehension tests on soft articles. As for the Vietnamese, although the mean score of pupils in CQN1 group is statistically lower than Bopomo and Hanji, they have caught up to the same level three months later in the CQN2 group. Recall that pupils in CQN1 and CQN2 groups were tested in December 2001 and March 2003, respectively; and Hanji and Bopomo groups were tested in January 2003. As for how many pupils and how deeply they have learned scripts in their pre-school days, we have no clues. All these facts indicate that we cannot simply assume pupils in the Hanji, Bopomo, and CQN groups have the same literacy starting point since the time span for the literacy beginners is so sensitive and any tiny increase in the amount of time can improve their orthographic skill greatly. In such a situation, I would suggest that literacy beginners in Hanji, Bopomo, and CQN have overall the same mean scores on reading comprehension tests before the learners are advanced to second grade. Even so, further study of the literacy beginners is needed and recommended to confirm this result.

The most surprising finding here is that the literacy beginners of Hanji are not significantly different from Bopomo and CQN. Under this situation, what does “inefficiency” have to say about Han characters? If the learning of Han characters is time-consuming, how could the first graders in the Hanji group have the statistically same mean scores as other groups? Recall that reading is neither a letter-by-letter nor word-by-word recognition process, but a process of forward and backward saccades (Smith 1994: 152). Yang and McConkie’s (1999: 212) study of thirteen Taiwanese subjects has shown a mean length of three characters in progressive saccades and 2.2 characters in regressive. In other

words, not all characters are actually read in the reading of Hanji texts. Therefore, the first graders could retrieve meanings from texts beyond their acquisition of limited number of Han characters. More research is needed to confirm this assumption.

In short, the mean scores of reading comprehension tests among Hanji, Bopomo, and CQN groups are statistically no different, except subjects in the Bopomo group from the second to fifth grades score significantly lower than subjects in other groups. Since collegian readers among the three scripts show no significant difference with regard to their scores on reading comprehension tests, is there a meaningful difference? The answer is yes. The major difference is the time the subjects spend on reading texts. UANOVA with *post hoc* tests are conducted among all the collegian subjects, and the results are shown in **Table 57**.

Table 57 shows two homogeneous subsets. The second subset consists of CQN and Bopomo groups. It means the reading time between CQN and Bopomo group has no statistically significant difference. On the other hand, Hanji is grouped in the first subset with a mean of 4.86 minutes. These results reveal that subjects in the Hanji group spend less time than CQN and Bopomo groups in reading their prepared texts.

Table 57. Homogeneous subsets by scripts among collegians

SCRIPTS	N	subset	
		1	2
Hanji	34	4.86	
CQN	50		11.60
Bopomo	23		13.28
Sig.		1	0.179

Alpha = 0.05

We may want to compare the rates in reading text-based Han characters and CQN. “Text-based” is emphasized here since reading individual characters is somewhat different

from reading characters arranged in a meaningful text. In this experiment, four prepared Hanji texts consist of a total of 2,346 characters, including multiple-choice questions, but punctuations are excluded. Those 2,346 characters divided by 4.86 minutes is equal to 482.72 characters per minute, and 4.86 minutes divided by 2,346 characters is equal to 0.124 seconds per character. As for the CQN texts, there are a total of 2,297 syllables,⁸⁵ and the same approach is applied to its calculation. The results of reading rates are shown in **Table 58**, which indicates that reading Han characters is 2.44 times faster than reading Vietnamese CQN.

Table 58. The rates in reading text-based Hanji and CQN by collegians

	Hanji	CQN
Total characters or syllables	2346	2297
Characters or syllables/per minute	482.72	198.02
Seconds/per character or syllable	0.124	0.303

Why is reading in Hanji faster than in CQN? There are two potential factors: 1) the influence of homophones in CQN, and 2) the influence of physical placement of orthographic symbols. Homophones in Vietnamese CQN are not as well distinguished as those in Chinese characters, so it could reduce the speed in reading CQN. As for the placement of orthographic symbols, the length of sentences in CQN is about twice that of sentences in Hanji when they both are arranged in a linear placement with 14 size fonts in a word processing file. This means that CQN readers have to spend more time with their eye

⁸⁵ Syllables instead of 'words' are counted so we can compare them with Han characters, which have a monosyllabic structure.

movement searching for words in a wider length. Consequently, the total amount of time spent on reading the prepared CQN texts is higher than Hanji. Whether or not these two factors are significant and which one is more dominant would require further study. This author would suggest applying the same approach in this experiment to other writing systems, particularly the Korean Hangul and English. Experiments in the Korean Hangul should be able to provide helpful information with regard to the relationship between length and reading speed since Korean Hangul is constructed with a monosyllabic shape and its text has roughly the same length as in Han character. On the other hand, experiments in English should be able to clarify the effect of homophones on reading, because English has the same linear structure as CQN, and possesses relatively fewer homophones compared to Vietnamese.

6.2 Results of dictation tests in Taiwanese group

Dictation tests were divided into Taiwanese and Vietnamese groups, each of which consisted of two dictation tests. Dictation tests one and two were conducted with a soft article and a hard article, respectively.

6.2.1 Analysis on Hanji in dictation one

There were a total of 415 subjects involved in the dictation tests. We had fewer subjects in dictation than in reading comprehension tests because some subjects did not return their dictation sheets. Their descriptive statistics are listed in **Table 59**, where ‘mean’ indicates the average number of correct characters students have written, and ‘%’ means the percentage of correct characters in the text. The percentage of correct Han characters is regarded as an index to evaluate students’ performance in dictation. ‘Maximum/minimum’ is the maximum/minimum number of correct characters among the students in a group. For example, the text for dictation one consists of 130 characters (excluding punctuation). The

male first graders have a mean of 25.24 correct characters, which constitute a percentage of 19.4 ($=25.24/130$). Among the male first graders, the maximum number of correct characters a student possesses is 57.

Table 59. Correct Han characters in dictation one

GRADES	GENDER	N	%	Mean	Sd.	Maximum	Minimum
1st graders	male	29	19.4	25.24	15.56	57	1
	female	26	20.2	26.31	15.98	59	2
	Total	55	19.8	25.75	15.62	59	1
2nd graders	male	26	42.1	54.73	26.13	118	6
	female	26	60.8	79.04	25.61	115	19
	Total	52	51.4	66.88	28.41	118	6
3rd graders	male	31	74.6	97.00	32.56	128	4
	female	26	78.1	101.50	35.06	128	13
	Total	57	76.2	99.05	33.49	128	4
4th graders	male	26	89.3	116.15	25.12	130	9
	female	34	93.9	122.09	8.04	130	99
	Total	60	91.9	119.52	17.67	130	9
5th graders	male	31	89.5	116.35	16.27	129	69
	female	35	93.9	122.03	11.21	130	69
	Total	66	91.8	119.36	14.00	130	69
6th graders	male	46	94.1	122.39	17.10	130	23
	female	26	97.6	126.92	3.32	130	117
	Total	72	95.4	124.03	13.93	130	23
collegians	male	19	98.6	128.16	3.88	130	119
	female	34	99.2	128.97	1.85	130	121
	Total	53	99.0	128.68	2.74	130	119
Total	male	208	73.4	95.45	41.55	130	1
	female	207	79.8	103.79	37.58	130	2
	Total	415	76.6	99.61	39.79	130	1

Statistical results of UANOVA reveal that gender and grade are significant factors at the 5% significance level. Gender factor is excluded for further analysis since it is not the main concern in this study. The results of *post hoc* tests generate four homogeneous subsets by grades in dictation one, as shown in **Table 60**.

Table 60. Homogeneous subsets of correct Hanji by grades in dictation one

GRADES	N	Subset			
		1	2	3	4
1st graders	55	25.75			
2 nd graders	52		66.88		
3 rd graders	57			99.05	
5 th graders	66				119.36
4 th graders	60				119.52
6 th graders	72				124.03
collegians	53				128.68
Sig.		1	1	1	0.132

Alpha = 0.05

To have a better picture, the percentage of correct characters achieved by students in each grade is reflected in **Figure 16**. It shows that students' percentage of correct characters significantly increases until the fourth grade. Because the results of *post hoc* tests indicate students from fourth grade to college are not statistically significantly different, we may conclude that the beginning learners of Han characters are improving their writing skills over years, and they have achieved the statistically same level as collegians at the stage of fourth grade. In other words, it takes about four years for a learner of Hanji to be able to write soft articles.

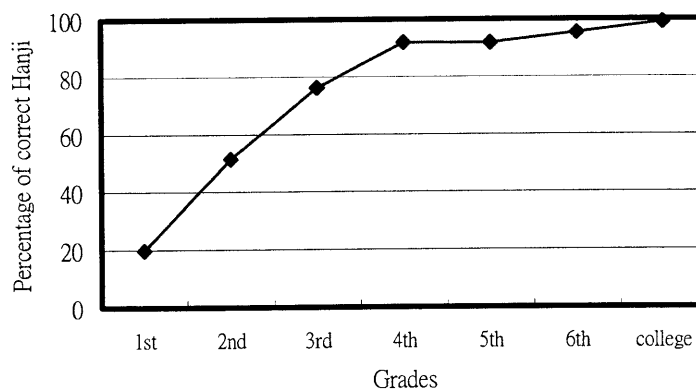


Figure 16. Percentage of correct Hanji in dictation one

Previous discussions focus on correctness in writing Han characters. How is the incorrectness measured? The descriptive statistics of incorrect characters are shown in **Table 61**. In **Table 61**, the incorrect characters are counted based on mistakes found in the characters the subjects have written. If a student did not write down anything, it is not considered incorrect characters. Percentage is the number of incorrect characters students produce divided by the number of all correct characters in dictation test one. Thus, in dictation one, collegian subjects have an average 0.84% of incorrect Han characters, with a maximum number of five. This says that, on average, a college student has 0.84 character out of 100 incorrect; and a student may have up to five characters incorrectly written.

Table 61. Incorrect Han characters in dictation one

GRADES	GENDER	N	%	Mean	Sd.	Maximum	Minimum
1st graders	male	29	1.56	2.03	3.91	21	0
	female	26	1.63	2.12	2.79	12	0
	Total	55	1.59	2.07	3.40	21	0
2nd graders	male	26	3.17	4.12	3.23	15	0
	female	26	2.81	3.65	3.25	13	0
	Total	52	2.98	3.88	3.22	15	0
3rd graders	male	31	1.98	2.58	2.22	9	0
	female	26	2.98	3.88	3.99	13	0
	Total	57	2.45	3.18	3.19	13	0
4th graders	male	26	1.63	2.12	2.94	14	0
	female	34	1.95	2.53	3.07	12	0
	Total	60	1.81	2.35	3.00	14	0
5th graders	male	31	3.58	4.65	4.77	18	0
	female	35	2.35	3.06	2.76	12	0
	Total	66	2.92	3.80	3.89	18	0
6th graders	male	46	2.15	2.80	3.38	14	0
	female	26	1.63	2.12	2.79	12	0
	Total	72	1.97	2.56	3.18	14	0
collegians	male	19	1.34	1.74	1.28	5	0
	female	34	0.57	0.74	1.21	4	0
	Total	53	0.84	1.09	1.32	5	0
Total	male	208	2.25	2.92	3.46	21	0
	female	207	1.95	2.53	3.02	13	0
	Total	415	2.10	2.73	3.25	21	0

Those errors made by all subjects in writing Han characters in the first dictation test are classified into the following twelve categories, and their descriptive statistics is listed in **Table 62**, where ‘%’ means the percentage of a specific error type out of all types.

HKI: Same sound but different meaning. That is, the characters written by subjects have identical sound but different meaning from the standard ones. For example, 兒 is considered HKI error of the standard character 而, and 河 is HKI error of 核.

HKL: Same sound with similar meaning. Here the written characters have identical sound and similar meaning with the standard ones. For example, 圓 is HKL error of 元.

HKP: Same sound, but different character component. That is, a component of characters is either missing, redundant, or incorrect. For example, the semantic component of 岡 is missing in contrast to the standard one 剛.

HLI: Similar sound and meaning. The incorrect characters have similar sound and meaning to those standard ones. For example, when the standard 期 is written as 集.

HL: Similar sound but different meaning. The incorrect characters have similar sound but different meaning to those standard ones. For example, 上 is HL error of 商.

HLH: Similar shape. In this type, the incorrect characters have similar shape as those standard ones. For example, 杲 is incorrect, in contrast to the standard 桌.

HLB: Similar meaning but different sound. For example, in the case when standard 很 is written as 好.

HKH: Strokes are missing. This is, some strokes in the characters are missing.

HCR: New characters are created in the semantic-phonetic principle. In this study, a few characters were created by students based on the semantic-phonetic principle.

HE: Variant forms of Han characters. Variant forms, including simplified, are used in the incorrect ones. For example, 仟 is the variant form of 千, and 发 is the simplified form of 發.

HO: All other types of errors.

HFL: A flip-flop phenomenon occurred in the incorrect character, such as 咲, which is the HFL error of 知.

Table 62. Original error types of Hanji in dictation one

GRADES		HKI	HKL	HKP	HLI	HL	HLH	HLB	HKH	HCR	HE	HO	HFL
1st grader	%	9.64	44.97	9.64	0.00	4.72	16.86	4.72	0.00	0.00	7.08	2.36	0.00
	Mean	0.12	0.56	0.12	0.00	5.8E-02	0.21	5.8E-02	0.00	0.00	8.8E-02	2.9E-02	0.00
	Sd.	0.41	0.50	0.41	0.00	0.34	0.77	0.24	0.00	0.00	0.29	0.17	0.00
	Max.	2	1	2	0	2	4	1	0	0	1	1	0
	Mini.	0	0	0	0	0	0	0	0	0	0	0	0
2nd grader	%	5.01	18.62	17.19	2.47	8.59	34.74	1.24	0.00	0.00	2.47	6.09	3.58
	Mean	0.14	0.52	0.48	6.9E-02	0.24	0.97	3.4E-02	0.00	0.00	6.9E-02	0.17	0.10
	Sd.	0.35	0.63	0.63	0.26	0.64	10.12	0.19	0.00	0.00	0.37	0.47	0.41
	Max.	1	2	2	1	3	3	1	0	0	2	2	2
	Mini.	0	0	0	0	0	0	0	0	0	0	0	0
3rd grader	%	17.77	8.14	17.77	0.00	2.74	48.13	1.37	0.00	0.00	0.00	4.07	0.00
	Mean	0.48	0.22	0.48	0.00	7.4E-02	1.30	3.7E-02	0.00	0.00	0.00	0.11	0.00
	Sd.	1.12	0.42	0.75	0.00	0.27	1.71	0.19	0.00	0.00	0.00	0.32	0.00
	Max.	5	1	2	0	1	6	1	0	0	0	1	0
	Mini.	0	0	0	0	0	0	0	0	0	0	0	0
4th grader	%	41.03	23.08	8.12	0.00	8.12	6.41	0.00	8.12	0.00	0.00	5.13	0.00
	Mean	0.96	0.54	0.19	0.00	0.19	0.15	0.00	0.19	0.00	0.00	0.12	0.00
	Sd.	1.64	0.58	0.49	0.00	0.63	0.37	0.00	0.49	0.00	0.00	0.33	0.00
	Max.	6	2	2	0	3	1	0	2	0	0	1	0
	Mini.	0	0	0	0	0	0	0	0	0	0	0	0
5th grader	%	14.57	12.14	6.07	6.07	6.07	38.04	4.86	0.00	3.68	0.00	8.50	0.00
	Mean	0.36	0.30	0.15	0.15	0.15	0.94	0.12	0.00	9.1E-02	0.00	0.21	0.00
	Sd.	0.78	0.59	0.51	0.36	0.71	1.87	0.55	0.00	0.38	0.00	0.55	0.00
	Max.	3	2	2	1	4	9	3	0	2	0	2	0
	Mini.	0	0	0	0	0	0	0	0	0	0	0	0
6th grader	%	23.82	6.54	2.53	11.21	6.54	48.11	1.26	0.00	0.00	0.00	0.00	0.00
	Mean	0.51	0.14	5.4E-02	0.24	0.14	1.03	2.7E-02	0.00	0.00	0.00	0.00	0.00
	Sd.	1.30	0.42	0.23	0.43	0.42	1.62	0.16	0.00	0.00	0.00	0.00	0.00
	Max.	5	2	1	1	2	7	1	0	0	0	0	0
	Mini.	0	0	0	0	0	0	0	0	0	0	0	0
Colleg-ian	%	11.84	13.81	2.37	4.73	2.37	31.39	0.00	0.00	0.00	26.37	7.11	0.00
	Mean	9.4E-02	0.11	1.9E-02	3.7E-02	1.9E-02	0.25	0.00	0.00	0.00	0.21	5.6E-02	0.00
	Sd.	0.35	0.42	0.14	0.19	0.14	0.62	0.00	0.00	0.00	0.57	0.23	0.00
	Max.	2	2	1	1	1	2	0	0	0	2	1	0
	Mini.	0	0	0	0	0	0	0	0	0	0	0	0
Total	%	17.82	16.25	9.43	3.95	5.76	34.07	1.98	1.10	0.66	3.51	4.83	0.66
	Mean	0.34	0.31	0.18	7.5E-02	0.11	0.65	3.7E-02	2.1E-02	1.3E-02	6.7E-02	9.2E-02	1.3E-02
	Sd.	0.95	0.53	0.48	0.26	0.47	1.30	0.25	0.17	0.14	0.32	0.33	0.14
	Max.	6	2	2	1	4	9	3	2	2	2	2	2
	Mini.	0	0	0	0	0	0	0	0	0	0	0	0
GRADES		HKI	HKL	HKP	HLI	HL	HLH	HLB	HKH	HCR	HE	HO	HFL

Further analysis of Hanji errors was conducted with statistical factor analysis by using SPSS. Five underlying factors (eigenvalues > 1) were extracted from the twelve error categories, as shown in **Table 63**. These factors explain 53.99% of the total variance.

Table 63. Factor loadings of twelve error categories

	Factor				
	1	2	3	4	5
HKI	0.602	0.129	-0.357	-0.222	0.165
HL	0.598	-0.448	0.231	-0.114	2.960E-02
HKL	0.583	0.233	-6.56E-02	0.107	0.244
HLH	0.574	0.273	-0.288	-4.50E-02	-6.89E-02
HO	0.482	-0.243	1.45E-02	0.211	-4.50E-02
HKP	0.252	0.539	0.463	0.117	-0.127
HKH	0.109	0.534	0.414	0.128	0.258
HFL	0.337	-0.506	0.579	-0.123	-9.23E-02
HCR	-3.57E-02	6.77E-03	-0.107	-0.738	0.348
HLI	0.201	-0.245	-0.343	0.479	0.102
HLB	0.109	9.09E-02	-0.270	-2.77E-02	-0.604
HE	-0.183	-0.167	-6.44E-02	0.316	0.576

The results of factor analysis reveal that those twelve error categories may be further regrouped into five basic types. Type one consists of components HKI, HL, HKL, HLH, and HO. Because the major common feature among these components is ‘similarity’ either in sound, meaning or shape, we may call type one ‘similarity’ errors. In dictation one, similarity errors account for 78.83% (i.e., 17.82 + 5.76 + 16.25 + 34.07 + 4.83) of total Hanji errors found in all subjects. The second type consists of components HKP and HKH. Because character components consist of strokes, we may call this ‘stroke’ error type, and it accounts for 10.53% of total errors. The third error type is HFL. Because only one component extracted in this type, we may just call this ‘flip-flop’ error type, and it accounts for 0.66%. In dictation one, flip-flop errors tend to occur among students in earlier grades.

The fourth error type includes components HCR and HLI and may be named ‘semantic-phonetic principle’ error type. This type accounts for 4.64% and tends to occur among students in higher grades. The last error type consists of components HLB and HE. ‘Semantic extension’ or ‘variant form’ is given to this error type, which accounts for 5.49%. In short, these five error types and their percentage are summarized in **Table 64**.

Table 64. Basic error types of Hanji in dictation one

Error types	%
Similarity	78.83
Stroke	10.53
Semantic extension	5.49
Semantic-phonetic principle	4.64
Flip-flop	0.66

6.2.2 Analysis on Bopomo in dictation one

In the dictation tests, students were told to write down Bopomo when they were encountering difficulties in writing Han characters. The purpose was to find out how students utilize Bopomo as an auxiliary tool in writing. The assumption was that the more Bopomo used in dictation, the more dependency the students have on its use. Only correct Bopomo was counted as an index to show students’ dependency of Bopomo. The descriptive statistics of correct Bopomo written by students is listed in **Table 65**. In the table, ‘%’ is the number of correct Bopomo students produce divided by the number of all standard Bopomo (i.e., 130) in dictation one. A correct Bopomo was counted based the correctness of syllables instead of individual letters. For example, 《 ㄅ is considered one correct Bopomo instead of two.

Table 65. Correct Bopomo in dictation one

GRADES	GENDER	N	%	Mean	Sd.	Maximum	Minimum
1st graders	male	29	39.92	51.90	24.39	97	5
	female	26	42.84	55.69	27.58	111	4
	Total	55	41.30	53.69	25.77	111	4
2nd graders	male	26	19.38	25.19	14.60	55	1
	female	26	18.82	24.46	13.80	67	4
	Total	52	19.10	24.83	14.07	67	1
3rd graders	male	31	10.15	13.19	13.65	51	0
	female	26	4.58	5.96	6.46	25	0
	Total	57	7.61	9.89	11.47	51	0
4th graders	male	26	5.06	6.58	16.42	84	0
	female	34	2.76	3.59	4.35	16	0
	Total	60	3.75	4.88	11.27	84	0
5th graders	male	31	2.13	2.77	3.01	11	0
	female	35	1.21	1.57	1.58	5	0
	Total	66	1.65	2.14	2.42	11	0
6th graders	male	46	0.42	0.54	1.33	7	0
	female	26	0.38	0.50	1.10	4	0
	Total	72	0.41	0.53	1.24	7	0
collegians	male	19	0.41	0.53	1.61	6	0
	female	34	0.00	0.00	0.00	0	0
	Total	53	0.15	0.19	0.98	6	0
Total	male	208	10.58	13.75	21.62	97	0
	female	207	9.02	11.73	21.44	111	0
	Total	415	9.81	12.75	21.53	111	0

Statistical results of UANOVA reveal that grade level is a significant factor at the 5% significance level. Four homogeneous subsets were generated as shown in **Table 66**. The data were rearranged and drawn as a graph in **Figure 17**. The results show that the percentage of correct Bopomo is significantly decreasing across the years until around the fourth or fifth grade. It implies that the dependency of the Hanji learners on Bopomo is decreasing while their skills in writing Hanji are increasing (compare **Figure 16**). In addition, if we assume the collegians are fully literate in Hanji without the assistance of Bopomo, then the pupils no longer need Bopomo at the grade of four or five.

Table 66. Homogeneous subsets of correct Bopomo by grades in dictation one

GRADES	N	Subset			
		1	2	3	4
collegians	53	0.19			
6 th graders	72	0.53			
5 th graders	66	2.14			
4 th graders	60	4.88	4.88		
3 rd graders	57		9.89		
2 nd graders	52			24.83	
1 st graders	55				53.69
Sig.		0.369	0.289	1	1

Alpha = 0.05

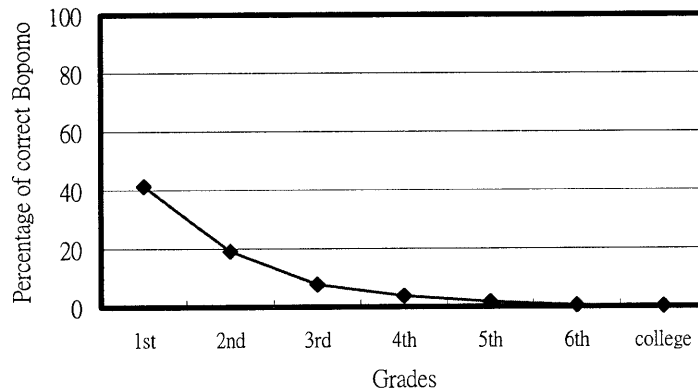


Figure 17. Percentage of correct Bopomo in dictation one.

There are eight error types found in Bopomo written by all subjects in dictation one. Recall that Bopomo is designed to represent the initial, medial, and final of syllables. Therefore, error types in Bopomo are counted based on its syllabic feature.

BW: It means initial, medial, and final are all incorrect.

BC: Initial is incorrect, but other parts are correct.

BM: Medial is incorrect, but other parts are correct.

BF: Final is incorrect, but other parts are correct.

BV: Everything is incorrect except initial.

BT: Tone is incorrect.

BL: Bopomo was written in similar sound. For example, 儿 was written as ㄛ.

BFL: A flip-flop phenomenon occurred in Bopomo. For example, ㄨ was incorrectly written as ㄩ and ㄨ as ㄩ.

Table 67. Original error types of Bopomo in dictation one

GRADES		BW	BC	BM	BF	BV	BT	BL	BFL
1st grader	%	4.37	29.75	5.97	15.67	6.93	32.63	4.37	0.31
	Mean	0.41	2.79	0.56	1.47	0.65	3.06	0.41	2.9E-02
	Sd.	0.70	2.28	0.70	1.60	1.01	3.18	0.61	0.17
	Max.	2	8	3	6	4	13	2	1
	Mini.	0	0	0	0	0	0	0	0
2nd grader	%	6.30	27.19	4.47	13.13	15.37	29.42	3.68	0.45
	Mean	0.48	2.07	0.34	1.00	1.17	2.24	0.28	3.4E-02
	Sd.	1.68	1.69	0.67	1.79	1.85	3.37	0.45	0.19
	Max.	7	5	2	6	7	13	1	1
	Mini.	0	0	0	0	0	0	0	0
3rd grader	%	16.18	48.00	2.00	5.93	2.00	19.96	5.93	0.00
	Mean	0.30	0.89	3.7E-02	0.11	3.7E-02	0.37	0.11	0.00
	Sd.	0.67	1.09	0.19	0.32	0.19	0.97	0.42	0.00
	Max.	2	3	1	1	1	4	2	0
	Mini.	0	0	0	0	0	0	0	0
4th grader	%	17.19	32.95	0.00	27.22	0.00	17.19	5.44	0.00
	Mean	0.12	0.23	0.00	0.19	0.00	0.12	3.8E-02	0.00
	Sd.	0.43	0.59	0.00	0.63	0.00	0.33	0.20	0.00
	Max.	2	2	0	3	0	1	1	0
	Mini.	0	0	0	0	0	0	0	0
5th grader	%	0.00	30.53	7.63	23.16	0.00	15.52	23.16	0.00
	Mean	0.00	0.12	3.0E-02	9.1E-02	0.00	6.1E-02	9.1E-02	0.00
	Sd.	0.00	0.42	0.17	0.38	0.00	0.35	0.29	0.00
	Max.	0	2	1	2	0	2	1	0
	Mini.	0	0	0	0	0	0	0	0
6th grader	%	0.00	0.00	0.00	66.67	33.33	0.00	0.00	0.00
	Mean	0.00	0.00	0.00	5.4E-02	2.7E-02	0.00	0.00	0.00
	Sd.	0.00	0.00	0.00	0.33	0.16	0.00	0.00	0.00
	Max.	0	0	0	2	1	0	0	0
	Mini.	0	0	0	0	0	0	0	0
Colleg-ian	%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Sd.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max.	0	0	0	0	0	0	0	0
	Mini.	0	0	0	0	0	0	0	0
Total	%	6.16	30.40	5.00	14.62	9.24	29.63	4.62	0.32
	Mean	0.16	0.79	0.13	0.38	0.24	0.77	0.12	8.4E-03
	Sd.	0.71	1.53	0.41	1.05	0.85	2.06	0.36	9.1E-02
	Max.	7	8	3	6	7	13	2	1
	Mini.	0	0	0	0	0	0	0	0
GRADES		BW	BC	BM	BF	BV	BT	BL	BFL

Further analysis of Bopomo errors were conducted with statistical factor analysis by using SPSS. Two underlying factors (eigenvalues > 1) were extracted from the original eight error types, as shown in **Table 68**. These factors explain 60.25% of the total variance. The first factor consists of BF, BC, BV, BT, BL, BM and BW. Therefore, type one may be called phonemic errors. In general (among all subjects), the percentage of errors in BF, BC, BV, BT, BL, BM and BW is 14.62%, 30.40%, 9.24%, 29.63%, 4.62%, 5%, and 6.16%, respectively. Among these errors, consonant and tonal errors are more frequent than others. The second factor covers BFL, which is a graphic flip-flop phenomenon. So, type two is called graphic errors.

Table 68. Factor loadings of eight error types

	Factor	
	1	2
BF	0.844	-0.105
BC	0.803	0.281
BV	0.748	-0.428
BT	0.710	0.338
BL	0.673	1.74E-02
BM	0.665	5.55E-02
BW	0.616	-0.394
BFL	0.185	0.738

6.2.3 Analysis on Hanji in dictation two

A total of 415 subjects were involved in the second dictation test. Among the subjects, 414 of them were the same subjects from dictation one. Unless otherwise specified, procedures in dictation one were also adopted to do analysis in dictation two. Tables and figures in dictation two are also arranged in the same formats as those in dictation one. So, we do not have to detail the same procedures and formats in the sections of dictation two.

The prepared text for dictation two consisted of a total of 89 Han characters. The number of correct characters students achieved is summarized in **Table 69**.

Table 69. Correct Han characters in dictation two

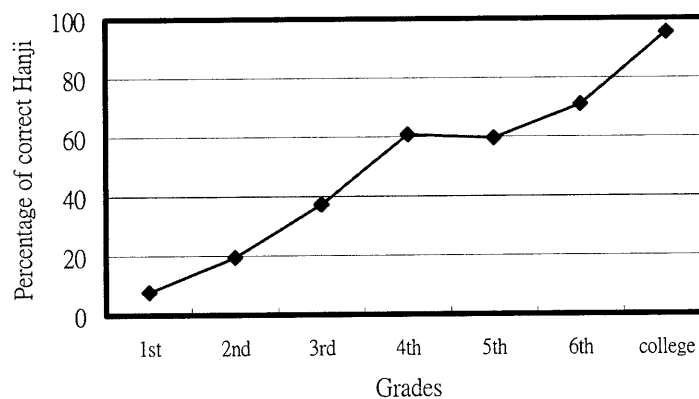
GRADES	GENDER	N	%	Mean	Sd.	Maximum	Minimum
1st graders	male	29	8.45	7.52	7.29	39	0
	female	25	7.06	6.28	4.10	16	0
	Total	54	7.80	6.94	6.01	39	0
2nd graders	male	26	17.42	15.50	10.47	46	0
	female	26	21.61	19.23	8.30	34	4
	Total	52	19.52	17.37	9.54	46	0
3rd graders	male	31	34.33	30.55	14.08	57	7
	female	26	40.92	36.42	18.07	66	7
	Total	57	37.34	33.23	16.14	66	7
4th graders	male	26	60.58	53.92	20.54	84	6
	female	34	59.91	53.32	18.59	89	20
	Total	60	60.20	53.58	19.29	89	6
5th graders	male	31	55.24	49.16	20.80	78	10
	female	35	63.37	56.40	15.79	77	22
	Total	66	59.55	53.00	18.53	78	10
6th graders	male	46	68.27	60.76	19.32	86	14
	female	26	75.42	67.12	12.07	83	37
	Total	72	70.85	63.06	17.24	86	14
collegians	male	19	94.80	84.37	4.49	89	72
	female	35	95.73	85.20	2.62	89	76
	Total	54	95.40	84.91	3.38	89	72
Total	male	208	48.03	42.75	28.07	89	0
	female	207	54.92	48.88	28.20	89	0
	Total	415	51.47	45.81	28.27	89	0

Statistical results of UANOVA reveal that gender and grade are two significant factors at the 5% significance level. Six homogeneous subsets of correct Hanji by grades are generated by *post hoc* tests. The data are rearranged in percentage by grades as shown in **Figure 18**.

Table 70. Homogeneous subsets of correct Hanji by grades in dictation two

GRADES	N	Subset					
		1	2	3	4	5	6
1st graders	54	6.94					
2 nd graders	52		17.37				
3 rd graders	57			33.23			
5 th graders	66				53.00		
4 th graders	60				53.58		
6 th graders	72					63.06	
collegians	54						84.91
Sig.		1	1	1	1	1	1

Alpha = 0.05

**Figure 18. Percentage of correct Hanji in dictation two.**

The results in **Table 70** and **Figure 18** reveal that students' skill in Han writing significantly improves over the years. The most important findings in the results is that the pupils, even in the sixth grade, have not achieved the same level as collegians. This fact implies that it takes more than six years for Hanji learners to be able to write hard articles at the collegian level. Terms "being able to write hard articles" is refers only to "being able to handle the required characters in writing hard articles;" it does not mean any other stylistic skills. The results indicate that the number of characters students have learned in

elementary school may be enough for soft articles, but it is insufficient for advanced hard articles, as illustrated in **Figure 19**.

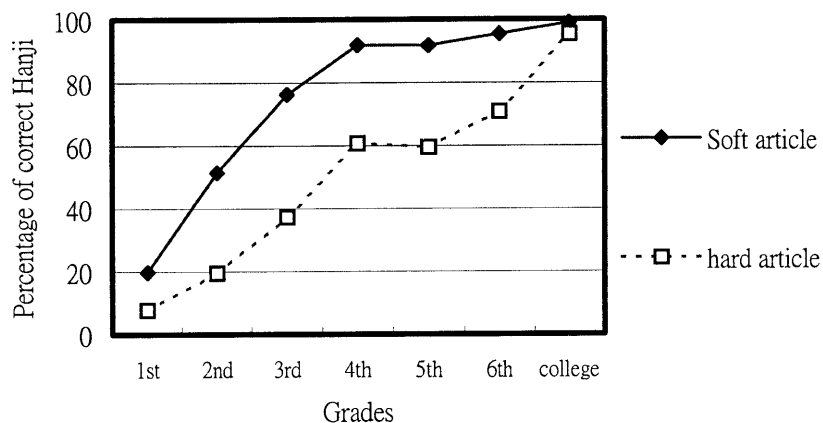


Figure 19. Percentage of correct Hanji in soft and hard articles.

The number of incorrect characters found in dictation two is statistically summarized in **Table 71**. Those incorrect characters belong to ten error types, as shown in **Table 72**. These ten error types are further analyzed by factor analysis, and four factors are extracted as shown in **Table 73**. These factors explain 56.18% of the total variance. Thus, these original ten types may be reclassified into four basic types.

Table 71. Incorrect Han characters in dictation two

GRADES	GENDER	N	%	Mean	Sd.	Maximum	Minimum
1st graders	male	29	2.09	1.86	2.29	10	0
	female	25	1.93	1.72	1.97	9	0
	Total	54	2.02	1.80	2.13	10	0
2nd graders	male	26	5.36	4.77	3.28	12	0
	female	26	6.91	6.15	3.96	17	0
	Total	52	6.13	5.46	3.66	17	0
3rd graders	male	31	8.12	7.23	4.35	19	0
	female	26	10.20	9.08	5.37	18	0
	Total	57	9.07	8.07	4.88	19	0
4th graders	male	26	8.82	7.85	4.82	19	0
	female	34	8.03	7.15	4.22	18	0
	Total	60	8.37	7.45	4.47	19	0
5th graders	male	31	10.40	9.26	4.26	18	2
	female	35	11.24	10.00	5.57	26	1
	Total	66	10.84	9.65	4.97	26	1
6th graders	male	46	8.51	7.57	4.62	23	0
	female	26	10.63	9.46	4.16	16	3
	Total	72	9.27	8.25	4.52	23	0
collegians	male	19	3.43	3.05	2.34	9	0
	female	35	3.49	3.11	1.84	7	0
	Total	54	3.47	3.09	2.01	9	0
Total	male	208	7.02	6.25	4.63	23	0
	female	207	7.53	6.70	5.03	26	0
	Total	415	7.27	6.47	4.83	26	0