

FREQUENTLY-USED CHINESE CHARACTERS AND LANGUAGE COGNITION

Chin-Chuan Cheng
University of Illinois at Urbana-Champaign
City University of Hong Kong
c-cheng1@uiuc.edu

To facilitate the spread of literacy in China, several lists of frequently-used Chinese characters have been published in the last 70 years. The number of characters in each list does not exceed 8,000. The number of character types in each Chinese book such as the 25 Dynasty Histories published in the past 2,000 years also does not exceed this number. Even when synchronous dictionaries already accumulated over 30,000 distinct characters, writers of large volumes normally used only 4,000 to 8,000 characters. The range of 4,000 to 8,000 morphemes then is proposed as the optimal number of linguistic symbols for human manipulation.

1. Number of frequently-used characters

Comprehensive Chinese dictionaries list over 50,000 distinctive Chinese characters. Many of these characters are no longer in use. From the articles in J. Wang 1995 I collected the following counts of frequently-used characters in practical lexicons or pedagogically oriented frequency books. (These references are cited in J. Wang 1995 and are therefore not given in our own References.) These counts range from 2,000 to slightly over 7,000. The titles of the volumes are given to help the reader understand their orientation.

- | | |
|-----------|---|
| (1) 4,261 | Chen Heqin 陳鶴琴 (1928, <i>Practical Lexicon for Colloquial Style</i> 語體文應用字匯) |
| 2,000 | Sichuan Education College 四川省教育科學院 (1946, <i>Frequently Used Characters</i> 常用字選) |
| 3,500 | Chinese Comprehensive Dictionary Editorial Board
中國大詞典編纂處 (1954, <i>List of 3,500 Characters for Learning</i> 識字正音3500字表) |
| 3,000 | Education Department of Shantung Province
山東省教育廳 (1958, <i>List of Frequently-Used Characters in Putonghua</i> 普通話常用字表) |
| 3,100 | Secondary and Primary School Textbook Editorial Board
of Education Bureau of Beijing City
北京市教育局中小學教材編審處 (1965, <i>List of Frequently Used Characters</i> 常用字表) |
| 4,444 | Chinese Character Section of Committee on Chinese Lan- |

	guage Reform 中國文字改革委員會漢字組 (1975, <i>List of 4,500 Characters 4500字表</i>)
7,292	Ministry of Communication of People's Republic of China 中華人民共和國郵電部 (1983, <i>Standard Telegraphic Code 標準電碼本</i>)
4,574	Language Teaching Department of Beijing Language Institute 北京語言學院語言教學研究所 (1985, <i>Frequency list of Chinese Characters 漢字頻率表</i>)
3,500	Language Commission 國家語言文字工作委員會 (1988, <i>List of Frequently-Used Characters in Modern Chinese 現代漢語常用字表</i>)

The term 'frequently-used characters' may give an impression that the number of characters was drastically reduced for elementary learning and literacy promotion. However, T'sou et al. 1997 show that current Chinese newspapers in Taipei, Hong Kong, and Singapore used about 4,000 distinct characters. It is therefore reasonable to say that a few thousand characters will suffice for linguistic expressions of most matters and events in modern times.

When we examine books of earlier times, we still see a similar number of characters used. For example, the novel *Dream of the Red Mansion* (紅樓夢) from the 18th century used only slightly over 4,000 characters. As it is quite certain that the first 80 and the last 40 chapters were written by two different individuals, I used the electronic version offered by Yuanze University of Taiwan to tabulate the characters separately as follows:

(2)	CHARACTER	CHARACTER	
	TOKENS	TYPES	
	496,855	4,293	First 80 Chapters 紅樓夢前80回
	234,980	3,217	Last 40 Chapters 紅樓夢後40回
	731,835	4,501	All 120 Chapters 紅樓夢120回

The entire book had over 730,000 character tokens. But only 4,501 distinct characters were used. In the same 18th century, the dictionary *Kangxi Zidian* (康熙字典) published in 1716 already collected a total of 47,035 distinct characters. The fact that out of the available 47,000 distinct characters only a few thousand of them were employed in a novel full of poems and descriptions of social institutions and personal feelings should be regarded as something significant about human use of linguistic symbols. In the following sections I will show that during the past 2,000 years the number of distinct Chinese characters increased from 9,000 to over 56,000. But individual authors used only a few thousand characters. This small number has been a constant across historical stages.

2. Accumulation of Chinese characters

The oracle writing of 3,300 years ago had about 5,000 distinct characters. It is generally agreed that 3,000 of them are now recognizable. Over time, additional characters were created, and the 1986 edition of the *Hanyu Da Zidian* (漢語大字典) had over 56,000 entries. The linear increase of distinct characters

can be clearly seen in the following list of dictionaries. The publication dates and the number of characters covered were gathered from Hsieh et al. 1992 and Yin & Rohrenow 1994:

(3) DATE	CHARACTERS	
1300BC	3,000	Oracle Writing 甲骨文
5AD	5,340	<i>Xuncuanpian</i> 訓纂篇
100	9,353	<i>Shuowen Jiezi</i> 說文解字
4C.	12,824	<i>Zilin</i> 字林
543	16,917	<i>Yupian</i> 玉篇
1008	26,000	<i>Guangyun</i> 廣韻
1039	31,000	<i>Leipian</i> 類篇
1067	30,000	<i>Jiyun</i> 集韻
1615	33,179	<i>Zihui</i> 字匯
1670	33,000	<i>Zhengzitong</i> 正字通
1716	47,035	<i>Kangxi Zidian</i> 康熙字典
1915	48,000	<i>Zhonghua Da Zidian</i> 中華大字典
1959	49,965	<i>Da Han-He Cidian</i> 大漢和辭典
1968	49,905	<i>Zhongwen Da Cidian</i> 中文大辭典
1986	56,000	<i>Hanyu Da Zidian</i> 漢語大字典

The sudden jump within a short span, for example, from a total of 48,000 characters in the year 1915 to 49,965 in 1959 and from 49,905 in 1968 to 56,000 in 1986 could not be attributed to the additional characters introduced during the intervals. The differences have to do with the comprehensiveness of the coverage of the dictionaries. Some editors had more resources to collect a larger number of characters. However, over the past 2,000 years, the gradual increase was obvious in this listing.

As users of dictionaries, we often feel that many of the characters or words in them are totally unfamiliar. Unabridged dictionaries are supposed to collect all words, frequent as well as rare, for users to look up. So it is natural for us to encounter unfamiliar characters. As we know, all aspects of language change in the course of history. Dictionary compilers are scholars who work on words unfamiliar to them; they have to look up words in other dictionaries in the compilation process. Why do we have unfamiliar words? Over a century ago, Darwin attributed the reason to a limit to the powers of the memory: 'We see variability in every tongue, and new words are continually cropping up; but as there is a limit to the powers of the memory, single words, like whole languages, gradually become extinct' (Darwin 1871:94-5). We will come back to this point later.

Not all words or characters in a dictionary are unknown to us. We may be able to derive the senses of some characters in contexts, but we have never used them before. We may say that these characters are in our passive vocabulary. There are familiar ones, of course. They are in our active vocabulary. How many are there of familiar ones? That is, how many characters do we know? This question has to do with human language cognition. Earlier we saw that 4,000 may be considered as an optimal number for language manipulation. We will pursue this

issue further in the following sections.

3. How many words do you know?

Traditionally a way to find the answer to the question 'how many words do you know?' is through experiments. In such experiments subjects were asked to identify known words from randomly selected pages of a dictionary, and the statistics were projected to cover the entire repertoire of the lexicon (Crystal 1995). Crystal 1995 states that English speakers can have 31,500 to 56,250 words in their active vocabulary and 38,300 to 76,350 words in their passive vocabulary. Miller & Gildea 1991 state that in the United States high school graduates at age 17 normally have 80,000 words in their vocabulary. Thus in the first 16 years of life they acquired 5,000 words per year or 13 words per day on average.

Do we indeed have 31,000 or 80,000 words in our active vocabulary? Are we actually able to use that many words? I have proposed to use the number of words in various books to help find the answer (Cheng 1997, 1998). First of all, in English the inflectional endings change a word into several graphic forms. For example, the nine words below are derived from the basic forms of 'write' and 'kick':

- (4) write, writes, wrote, writing, written, kick, kicks, kicked, kicking

We can say that these nine words are 'graphic words' and the two basic forms are the 'concept words'. I wrote a computer program to lemmatize words by returning words ending in -s, -ed, -ing, -ly, -er, -est and various forms of pronouns to the basic forms. Following is a list of words that show the inflected and basic forms:

- (5) ABILITIES: ABILITY
 ABRASIVELY: ABRASIVE
 ABRASIVES: ABRASIVE
 BUILDERS: BUILDER
 BUILDING: BUILD
 BUILDS: BUILD
 BUILT: BUILD
 COOLED: COOL
 COOLER: COOL
 COOLEST: COOL
 COOLING: COOL
 ME: I

I ran the program on scores of electronic English texts obtained from the Web. The numbers show a consistent pattern. No matter how long the books are, the concept words used range from about 4,000 to 8,000 in number. Some examples are given below:

(6) TOKEN	GRAPHIC	CONCEPT	TITLE
32,361	4,727	3,431	<i>Call of the Wild</i>
74,038	7,427	5,316	<i>Tom Sawyer</i>

87,044	8,630	6,046	<i>Beauty and the Beast</i>
161,751	9,281	6,433	<i>Dracula</i>
137,060	8,877	6,218	<i>The American</i>
39,631	4,236	3,190	<i>Aspern Papers</i>
80,493	8,976	6,377	<i>Paradise Lost</i>
161,974	7,097	4,647	<i>Emma</i>
120,735	6,288	4,199	<i>Sense and Sensibility</i>
123,270	6,288	4,146	<i>Pride and Prejudice</i>
84,128	5,741	3,934	<i>Persuasion</i>
729,792	13,765	8,641	Austen's 6 Books

The last set of numbers is for Jane Austen's six books combined. The books are *Emma*, *Sense and Sensibility*, *Pride and Prejudice*, *Persuasion*, *Mansfield Park*, *Northanger Abbey*. The novels by Austen covered different areas of human activities, and yet the total number of distinct words used is just over 8,000.

In Chinese, we can also examine individual authors' use of characters to determine the number of morphemes that we can actively control. Earlier we showed that the Chinese novel *Dream of the Red Mansion* used only 4,501 characters. Here we will examine the 25 Dynasty Histories. These Histories recorded all the activities of the emperors, ministers, and local officials, conditions of economy and society, and others. Generally speaking they are not skewed to one particular aspect of human life. The numbers of characters used in the 25 Dynasty Histories are taken from the tabulation by Hsieh et al. 1992. The publication dates, character tokens, and character types are given below. We see that the number of characters used rarely exceeds 8,000:

(7)	DATE	TOKENS	TYPES	TITLE
	93BC	533,505	5,122	<i>Shi Ji</i> 史記
	83AD	742,298	5,833	<i>Han Shu</i> 漢書
	289	377,807	4,388	<i>Sanguo Zhi</i> 三國志
	445	894,020	6,161	<i>Houhan Shu</i> 後漢書
	488	811,893	5,842	<i>Song Shu</i> 宋書
	514	299,257	4,962	<i>Nanqi Shu</i> 南齊書
	554	998,329	5,417	<i>Wei Shu</i> 魏書
	636	294,438	4,973	<i>Liang Shu</i> 梁書
	636	163,382	4,033	<i>Chen Shu</i> 陳書
	636	212,506	4,032	<i>Beiqi Shu</i> 北齊書
	636	262,659	4,161	<i>Zhou Shu</i> 周書
	648	1,158,126	5,997	<i>Jin Shu</i> 晉書
	656	701,698	5,592	<i>Sui Shu</i> 隋書
	659	677,624	5,376	<i>Nan Shi</i> 南史
	659	1,106,543	5,572	<i>Bei Shi</i> 北史
	945	2,002,600	6,346	<i>Jiu Tang Shu</i> 舊唐書
	974	790,879	5,109	<i>Jiu Wudai Shi</i> 舊五代史
	1060	1,694,794	6,771	<i>Xin Tang Shu</i> 新唐書
	1072	291,476	3,909	<i>Xin Wudai Shi</i> 新五代史
	1344	296,254	4,071	<i>Liao Shi</i> 遼史

1344	931,070	5,264	<i>Jin Shi</i> 金史
1345	3,980,123	7,389	<i>Song Shi</i> 宋史
1370	1,611,849	5,854	<i>Yuan Shi</i> 元史
1739	2,802,544	7,124	<i>Ming Shi</i> 明史
1927	4,514,567	8,080	<i>Qing Shi Gao</i> 清史稿

The token column shows the length of each book. In the past 20 centuries, even though some of the books could be as short as 163,000 or as long as 4,514,000 characters, the total number of distinct characters used was between 4,000 and 8,000. This constant range existed when the dictionaries had a total number of character types of 5,000 in the 1st century, 12,000 in the 4th century, 16,000 in the 6th century, 30,000 in the 11th century, 33,000 in the 17th century, 47,000 in the 18th century, and 56,000 in the 20th century. The entries in the lists in (3) and (7) are now combined and arranged in chronological order to show the increase in the number of characters collected in dictionaries and the constant number of characters used in Dynasty Histories over time:

(8) -1300	3,000	Oracle Writing 甲骨文
-93	5,122	<i>Shi Ji</i> 史記
5	5,340	<i>Xuncuanpian</i> 訓纂篇
83	5,833	<i>Han Shu</i> 漢書
100	9,353	<i>Shuowen Jiezi</i> 說文解字
289	4,388	<i>Sanguo Zhi</i> 三國志
4C	12,824	<i>Zilin</i> 字林
445	6,161	<i>Houhan Shu</i> 後漢書
488	5,842	<i>Song Shu</i> 宋書
514	4,962	<i>Nanqi Shu</i> 南齊書
543	16,917	<i>Yupian</i> 玉篇
554	5,417	<i>Wei Shu</i> 魏書
636	4,973	<i>Liang Shu</i> 梁書
636	4,033	<i>Chen Shu</i> 陳書
636	4,032	<i>Beiqi Shu</i> 北齊書
636	4,161	<i>Zhou Shu</i> 周書
648	5,997	<i>Jin Shu</i> 晉書
656	5,592	<i>Sui Shu</i> 隋書
659	5,376	<i>Nan Shi</i> 南史
659	5,572	<i>Bei Shi</i> 北史
945	6,346	<i>Jiu Tang Shu</i> 舊唐書
974	5,109	<i>Jiu Wudai Shi</i> 舊五代史
1008	26,000	<i>Guangyun</i> 廣韻
1039	31,000	<i>Leipian</i> 類篇
1060	6,771	<i>Xin Tang Shu</i> 新唐書
1067	30,000	<i>Jiyun</i> 集韻
1072	3,909	<i>Xin Wudai Shi</i> 新五代史
1344	4,071	<i>Liao Shi</i> 遼史
1344	5,264	<i>Jin Shi</i> 金史
1345	7,389	<i>Song Shi</i> 宋史

1370	5,854	<i>Yuan Shi</i> 元史
1615	33,179	<i>Zihui</i> 字匯
1670	33,000	<i>Zhengzitong</i> 正字通
1716	47,035	<i>Kangxi Zidian</i> 康熙字典
1739	7,124	<i>Ming Shi</i> 明史
1915	48,000	<i>Zhonghua Da Zidian</i> 中華大字典
1927	8,080	<i>Qing Shi Gao</i> 清史稿
1959	49,965	<i>Da Han-He Cidian</i> 大漢和辭典
1968	49,905	<i>Zhongwen Da Cidian</i> 中文大辭典
1986	56,000	<i>Hanyu Da Zidian</i> 漢語大字典

The numbers were plotted in Figure 1 (see Appendix) to show the small range of variation in the number of characters used in the Dynasty Histories in contrast with the large numbers of characters in dictionaries. While dictionary compilers collected known as well as unfamiliar words in a volume, authors of books used familiar words to express ideas or to describe events. The small range of variation in the number of characters used in the Chinese writings and in the number of words used by English writers compels us to conclude that the optimal number of linguistic symbols a person can handle is between 4,000 and 8,000.

I have made a cross-language comparison of English and Chinese and concluded that the optimal number of linguistic symbols for manipulation is between 4,000 and 8,000. However, English 'words' and Chinese 'characters' are not identical linguistic units. We will discuss this issue below.

4. Characters, and words, and linguistic symbols

As is well known, a Chinese character represents a morpheme, and a morpheme can be analyzed as bound or free. A free morpheme is a word. A large number of Chinese characters represent words. But Chinese words may consist of more than one morpheme and therefore more than one character in writing. The English words that we tabulated above are lemmatized basic forms. They may be single morphemes. Some may consist of stems and affixes.

In terms of internal structure some English words may not be identical to Chinese characters. We now examine occurrences of 'words'.

Francis & Kucera 1982 show that the Brown University's English corpus built in the 1960s had over 30,000 distinct basic words. However, they give only 5,996 words with a running frequency of five or above in their listing and consider the others rare words. It is interesting to note here that this number 5,996 is within our optimal range.

There are numerous words in Chinese. 'Cidian' (word dictionaries), in contrast to 'Zidian' (character dictionaries), normally give a separate count of monosyllabic (single character) and polysyllabic (multiple character) entries. Some of the following 'word' dictionaries have more than 90,000 distinct entries:

(9)	MONO-SYLLABIC	POLY-SYLLABIC	TITLE
		100,000	Chinese Comprehensive Dictionary Editorial Board 中國大辭典編纂處 (1937, <i>Guoyu Cidian</i> 國語辭典)
	10,000	30,000	He 何容 (1976, <i>Guoyu Ribao Zidian</i> 國語日報字典)
	6,000	50,000	Beijing Foreign Language College English Department 北京外國語學院英語系 (1978, <i>Chinese-English Dictionary</i> 漢英詞典)
		56,000	Chinese Academy of Social Science Linguistics Institute 中國社會科學院語言所 (1980, <i>Modern Chinese Dictionary</i> 現代漢語詞典)
	3,994	90,000	Fu and Chen 傅興嶺、陳章煥 (1982, <i>Frequently-Used Word Dictionary</i> 常用構詞字典)
	2,116	90,000	Liu 劉源 (1984, <i>Modern Chinese Word List</i> 現代漢語詞表)
	4,000	58,000	Zhang 張聿忠 (1986, <i>Words</i> 詞林)
	9,700	48,000	Li et al. 李國炎等 (1988, <i>New Chinese Dictionary</i> 新編漢語詞典)
	13,000	80,000	Wu 吳昌恒 (1988, <i>Old and Modern Chinese Practical Dictionary</i> 古今漢語實用詞典)
		60,400	Chinese Pinyin Lexicon Editing Group 漢語拼音詞匯編寫組 (1991, <i>Chinese Pinyin Lexicon</i> 漢語拼音詞匯)
	800	80,000	Beijing Foreign Language University English Department 北京外國語大學英語系 (1995, <i>Chinese-English Dictionary</i> 漢英詞典)
	13,000	36,000	Ye and Huang 葉立群、黃成德 (1996, <i>Longman New Advanced Chinese Dictionary</i> 郎文中文高級新辭典)
	11,000	28,000	Min 閔龍華 (1997, <i>Modern Chinese Usage Dictionary</i> 現代漢語用法詞典)

Many of the words here are combinations of morphemes and other words. They do not have to be memorized. Moreover, the commonly used ones are not as numerous as the figures would show. For example, T'sou et al. 1997 found over 40,000 'words' in the newspapers in Singapore, Hong Kong, and Taipei. But the highest occurring 5,000 to 7,000 words cover 90% of the texts:

(10) COVERAGE	SINGAPORE	HONG KONG	TAIPEI
90%	5,043	7,477	5,005

Thus we can say that even in terms of words, the optimal number is not very large. We will have to study the morphological organization of words in Chinese in detail to be able to differentiate those that have to be memorized and those that can be derived from morphological rules and patterns.

5. Language cognition

We have used 'characters', 'graphic words', and 'concept words' in printed matters to argue for the view that the human capacity for manipulation of linguistic symbols has a limit around 8,000 units. We see no problem in extending this view to the capacity of those who are not literate. As we know, with effort and specialization, some people may have a larger vocabulary. But literate or not, mature native speakers of any language possess similar powers of language use.

Earlier I cited Darwin's view that 'as there is a limit to the powers of the memory, single words, like whole languages, gradually become extinct'. The phrase 'powers of the memory' is perhaps the right word. We use a finite set of elements to make up other words. The words are combined to make up phrases and sentences. The powers for combination are fairly high. But I have shown that there is a limit to morphemes that we can actively manipulate. This limit cannot be due to the finiteness of physical memory locations or cells for the reason that bilinguals or multilinguals generally have the same powers for each language. If physical memory locations were the determining factor, then those who speak another language would have to replace the memory of what had been occupied by the first language. Therefore the limit of powers is a matter of memory function rather than memory location.

New words appear, and old words become extinct. This is a fact of language. I have proposed a level of optimal manipulation of linguistic symbols for the maintenance of a constant 4,000-8,000 range of characters in the 25 Dynasty Histories over time. In daily life, names of acquaintances fade away with time. With effort we can retrieve them. So are words. Various linguistic functions and activities associated with the proposed optimal number of symbols for manipulation can be profitably studied in conjunction with the studies of memory and other aspects of cognition.

REFERENCES

- Beijing Foreign Language College English Department 北京外國語學院英語系. 1978. *Chinese-English Dictionary* 漢英詞典. Beijing: Commercial Press 商務印書館.
- CHENG, Chin-Chuan. 1997. In search of cognitive basis of common Chinese characters. Paper presented at The Ninth North American Conference on Chinese Linguistics. Victoria, Canada. May 2-4.
- . 1998. Learning words with many texts. *The Proceedings of the First Inter-*

- national Conference on Multimedia Language Education 1-12. Taipei: Crane Publishing House.
- Chinese Academy of Social Science Linguistics Institute 中國社會科學院語言所. 1980. *Modern Chinese Dictionary* 現代漢語詞典. Beijing: Commercial Press 商務印書館.
- Chinese Comprehensive Dictionary Editorial Board 中國大辭典編纂處. 1937. *Guoyu Cidian* 國語辭典. Taipei: Commercial Press 商務印書館.
- Chinese Pinyin Lexicon Editing Group 漢語拼音詞匯編寫組. 1991. *Chinese Pinyin Lexicon* 漢語拼音詞匯. Beijing: Language Press 語文出版社.
- CRYSTAL, David. 1995. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.
- DARWIN, Charles. 1871. *The Descent of Man*. Washington Square, New York: New York University Press. (1989 Edition).
- FRANCIS, W. Nelson, & Henry KUCERA 1982. *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin Company.
- FU, Xingling, & Zhanghuan CHEN 傅興嶺、陳章煥. 1982. *Frequently-Used Word Dictionary* 常用構詞字典. Beijing: Chinese People's University Press 中國人民大學出版社.
- HE, Rong 何容. (ed.) 1976. *Guoyu Ribao Zidian* 國語日報字典. Taipei: National Language Daily 國語日報社.
- HSIEH, Ching-Chun, LIN Hsi, HSU Chin-Ting, FU Wu-Chang, & CHANG Tsui-ling 謝清俊、林晰、許金定、傅武嫦、張翠玲. 1992. Statistics and analyses of the characters in Twenty-five Dynasty Histories 二十五史的文字統計與分析. Paper presented at the Third National Conference on Chinese Writing System 第三屆中國文字學全國學術研討會論文.
- LI, Guoyan, Heng MO, Yaohai SHAN, and Chongkang WU 李國炎、莫衡、單耀海、吳崇康. 1988. *New Chinese Dictionary* 新編漢語詞典. Changsha, Hunan: Hunan People Press 湖南人民出版社.
- LIU, Yuan 劉源. 1984. *Modern Chinese Word List* 現代漢語詞表. Beijing: Chinese Standards Press 中國標準出版社.
- MILLER, George A., & Patricia M. GILDEA. 1991. How children learn words. *The Emergence of Language Development and Evolution*, ed. by William S-Y. Wang 150-58. New York: W. H. Freeman.
- MIN, Longhua 閔龍華. 1997. *Modern Chinese Usage Dictionary* 現代漢語用法詞典. Taipei: Wenqiao Press 文橋出版社.
- T'SOU, Benjamin K., Hing-Lung LIN, Godfrey LIU, Terence CHAN, Jerome HU, Ching-hai CHEW, & John K.P. TSE. 1997. A synchronous Chinese language corpus from different speech communities: construction and applications. *Computational Linguistics and Chinese Language Processing* 2:1. 91-104.
- WANG, Jun 王均. (ed.) 1995. *Current Chinese Language Reform* 當代中國的文字改革. Beijing: Current China Publishers 當代中國出版社.
- WANG, William S-Y. (ed.) 1991. *The Emergence of Language Development and Evolution*. New York: W. H. Freeman.
- WU, Changheng 吳昌恒. 1988. *Old and Modern Chinese Practical Dictionary* 古今漢語實用詞典. Chengdu, Sichuan: Sichuan People Press 四川人民出版社.

- YE, Liqun and Chengde HUANG 葉立群、黃成德. 1996. *Longman New Advanced Chinese Dictionary* 朗文中文高級新辭典. Hong Kong: Longman Asia Publishers 朗文出版亞洲有限公司.
- YIN, Binyong and John ROHSENOW. 1994. *Modern Chinese Characters*. Beijing: Sinolingua.
- ZHANG, Yuzhong 張聿忠. 1986. *Words* 詞林. Beijing: Workers Press 工人出版社.

APPENDIX

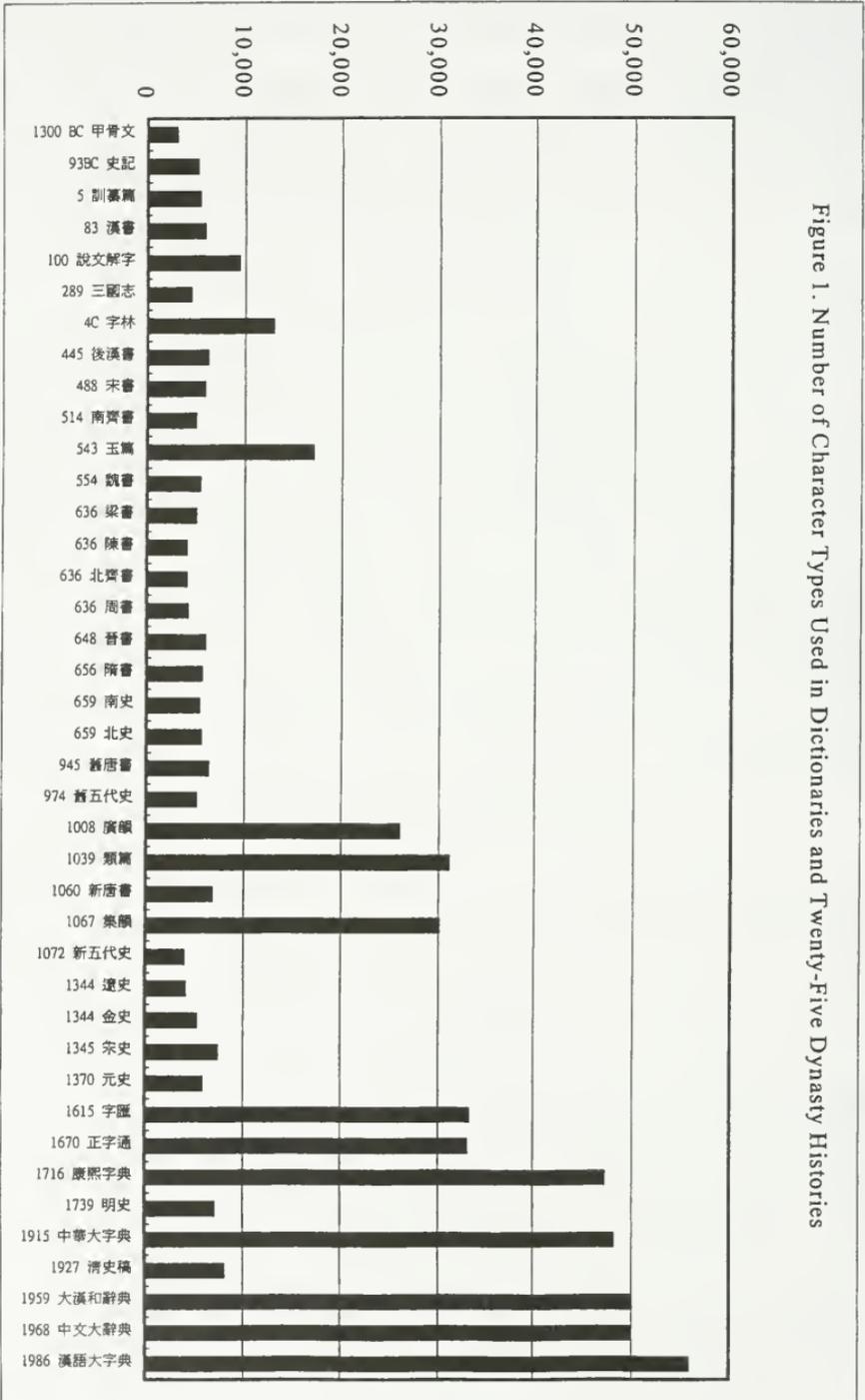


Figure 1. Number of Character Types Used in Dictionaries and Twenty-Five Dynasty Histories