
SINO-PLATONIC PAPERS

Number 116

May, 2002

On the Problems of Chinese Lettered Words

by
LIU Yongquan

Victor H. Mair, Editor
Sino-Platonic Papers
Department of East Asian Languages and Civilizations
University of Pennsylvania
Philadelphia, PA 19104-6305 USA
vmair@sas.upenn.edu
www.sino-platonic.org

SINO-PLATONIC PAPERS is an occasional series edited by Victor H. Mair. The purpose of the series is to make available to specialists and the interested public the results of research that, because of its unconventional or controversial nature, might otherwise go unpublished. The editor actively encourages younger, not yet well established, scholars and independent authors to submit manuscripts for consideration. Contributions in any of the major scholarly languages of the world, including Romanized Modern Standard Mandarin (MSM) and Japanese, are acceptable. In special circumstances, papers written in one of the Sinitic topolects (*fangyan*) may be considered for publication.

Although the chief focus of *Sino-Platonic Papers* is on the intercultural relations of China with other peoples, challenging and creative studies on a wide variety of philological subjects will be entertained. This series is **not** the place for safe, sober, and stodgy presentations. *Sino-Platonic Papers* prefers lively work that, while taking reasonable risks to advance the field, capitalizes on brilliant new insights into the development of civilization.

The only style-sheet we honor is that of consistency. Where possible, we prefer the usages of the *Journal of Asian Studies*. Sinographs (*hanzi*, also called tetragraphs [*fangkuaizi*]) and other unusual symbols should be kept to an absolute minimum. *Sino-Platonic Papers* emphasizes substance over form.

Submissions are regularly sent out to be refereed and extensive editorial suggestions for revision may be offered. Manuscripts should be double-spaced with wide margins and submitted in duplicate. A set of "Instructions for Authors" may be obtained by contacting the editor.

Ideally, the final draft should be a neat, clear camera-ready copy with high black-and-white contrast.

Sino-Platonic Papers is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

Please note: When the editor goes on an expedition or research trip, all operations (including filling orders) may temporarily cease for up to two or three months at a time. In such circumstances, those who wish to purchase various issues of *SPP* are requested to wait patiently until he returns. If issues are urgently needed while the editor is away, they may be requested through Interlibrary Loan.

N.B.: Beginning with issue no. 171, *Sino-Platonic Papers* has been published electronically on the Web. Issues from no. 1 to no. 170, however, will continue to be sold as paper copies until our stock runs out, after which they too will be made available on the Web at www.sino-platonic.org.

On the Problems of Chinese Lettered Words*

LIU Yongquan

(Institute of Linguistics, CASS)

Abstract: The author notes in the development of the Chinese lexicon that there are two kinds of words in Modern Chinese. One of these is Chinese character words, which constitute the majority of the Chinese lexicon. The other is lettered words, which occupy only a small part of the lexicon and are mostly loanwords. The author also notes that emergence of lettered words is a result of the Policy of Reformation and Opening, and of the needs of today's information society. Moreover, this is a new and promising method for borrowing words from foreign languages. This paper discusses in detail the word-formation of lettered words as well as relevant problems, such as pronunciation, style of writing, and indexing system. The indexing system proposed here is computer-based and totally automated, and may be a worthy reference for the automation of dictionary compilation.

1. Arguments on Lettered Words

Under the guidance of the Reformation and Opening Policy, China's economy, culture, science, and technology have rapidly developed. New things and new concepts are introduced unceasingly; in accordance with this trend, there has emerged a large batch of new terms in which such lettered words as MTV 音乐电视 and BP 机 ("beeper") are included. How should these lettered words be dealt with? Shall we welcome or dispute them? There have been many arguments on this topic. Perhaps even today there are still two sharply different opinions. And there are also dissenters even among the author's friends. However, after the publication of my paper "Talk about Lettered Words" in 1994¹, a great many colleagues have come to support my point of view^{2, 3, 4, 5}. Particularly valuable is their description of the application of

lettered words in Singapore, Hong Kong, and Taiwan. It seems that this problem is being considered among all Chinese communities. It is worth noting that there is a new appendix of lettered words in the 185th printing of the *Xiandai Hanyu Cidian (Modern Chinese Dictionary)*, i.e., its first revised edition. Since then, lettered words have also appeared in many other dictionaries, such as *the Xiandai Hanyu Xiao Cidian (Pocket Chinese Dictionary)* (Commercial Publishing House, 1999), *Xiandai Hanyu Da Cidian (The Great Dictionary of Modern Chinese)* (The Great Dictionary of Modern Chinese Publishing House, 2000), *Xinhua Cidian (Xinhua Lexicon)* (Commercial Publishing House, 2001), *Zhongxuesheng Guifan Cidian (The Dictionary for Middle School Students)* (Chinese Youth Publishing House, 2001). Moreover, a special *Zimuci Cidian (Dictionary of Lettered Words)*⁶ has recently been published by the Shanghai Lexicon Publishing House. All of these new publications illustrate that lettered words have become a part of the Chinese lexicon, a new form of loanwords, and indispensable entries in dictionaries.

2. Name and Definition of Lettered Word

In 1994, the lettered word in my paper "Talk on Lettered Words" was defined as "a Chinese word partly combined with letters (mainly Latin letters) or completely consisting of letters, the former like B 超 ("Bi-scan") and 卡拉 OK ("karaoke"), and the latter including CD and UFO. They are a new form for borrowing words from other languages." Now it seems better to call them Chinese lettered words straightforwardly. The reason is very simple: words in English, French, German, and Russian languages consist of letters only, so there is no need for the term "lettered word," while the majority of the Chinese lexicon consists of Chinese characters and only a small part of the lexicon consists of pure letters (such as DNA, WTO, APEC, CDMA), or characters with letters (such as BP 机, 卡拉 OK, X 染色体 [X-chromosome], 维生素 E [vitamin E]). Therefore the term "lettered word" is specific only for languages like

Chinese that are written mainly with non-Latin scripts. There will be no need for the term "lettered word" if there will appear a Chinese variant---phoneticized (alphabetical) Chinese (i.e., one language and two scripts, or digraphia)---some day in the future. Then BP 机 will turn out as BP-ji, 卡拉 OK will be KALA-OKE, and so forth.

The majority of lettered words are loanwords, and the rest are Chinese words, for instance GB ("国家标准") and HSK ("汉语水平考试 Chinese Proficiency Test"). Therefore, it is not proper from this point of view that these lettered words be called "words initialized with Western letters," or "words initialized with foreign letters," because GB, HSK, PSC (Putonghua Shuiping Ceshi ["Mandarin Proficiency Exam"]), etc. are composed using Chinese Pinyin—the Chinese alphabet—instead of with a Western alphabet or another foreign alphabet. Besides, if we were to adopt the designation "words initialized with Western /foreign letters," not only would the above-mentioned acronyms composed of Chinese Pinyin (which tend to be produced more and more) be excluded, many words initialized with Chinese characters would also not be covered. Consequently, it is better to give them the clear-cut name "Chinese lettered words."

To sum up, there are two kinds of words in the Modern Chinese lexicon: one of them consists of Chinese characters, and constitutes the majority of the Modern Chinese lexicon; the other consists of lettered words, which occupy only a small part of the Chinese lexicon and are mostly loanwords. It should be sufficiently comprehensive to define Chinese lettered words as those composed using the Latin alphabet (including Pinyin) or the Greek alphabet, or composed of them along with symbols, numbers, or Chinese characters respectively.

3. Necessity and Possibility of Lettered Words

Lettered words are the products of the Reformation and Opening Policy, and of the rapid development of

science and technology. The aim of the reformation policy is to encourage development, and the opening policy offers the possibility to absorb new things. No great progress can be made in the field of science and technology, as well as in the economy and culture, if we adopt a closed-door policy. Under that condition, new things and new concepts could not be absorbed, let alone new terms.

In order to absorb advanced culture promptly, a new method for borrowing words from other languages has emerged: lettered words based on the Chinese Pinyin alphabet (Latin alphabet). There have been two ways to absorb loanwords through the ages: phonetic translation and semantic translation. It was unsuccessful to translate foreign words phonetically by means of Chinese characters before the appearance of Chinese Pinyin. For instance, FAIRPLAY was clumsily transcribed into Chinese as FEI'APOLAI 费阿泼赖 and PRESIDENT as BOLIXITIANDE 伯理玺天德. Therefore, there are very few phonetic loans in Chinese. TANKE 坦克 ("tank"), KAFEI 咖啡 ("coffee"), XIMENGSI 席梦思 ("Siemens") are representatives of the few survivors. Why is this so? The reason lies in the facts that: 1) There are no spaces between words in Chinese texts –phonetic loanwords are obscured in their midst. It goes without saying that they have not been well received. 2) Alphabetic words are phonetic symbols, while Chinese characters are semantic symbols. To transcribe phonetic symbols by means of semantically-based ones is not only uneconomic but also easily ambiguous. Each Chinese character is a syllable. Consequently, syllables would be multiplied when transcription occurs, such that, for example, the trisyllabic PRESIDENT is converted into five characters, namely 伯理玺天德 (BOLIXITIANDE). Ambiguity occurs because almost every Chinese character carries lexical meaning; for example, 伯 (BO) means "uncle," 理 (LI) means "principle," etc. Confusion will inevitably occur if Chinese characters are used as phonetic symbols.

Now the situation has improved. These two questions have been solved by Chinese Pinyin. VCD, MBA, MPA, ISBN, and ADSL are very eye-catching among Chinese characters; they are not only economical but

are also easy to use (VCD has only four strokes while 激光视盘[JIGUANG SHIPAN] has 41 strokes), and no ambiguity occurs. Because of Chinese Pinyin, the Chinese language is able to borrow loanwords by means of phonetic transcription or direct letter-for-letter rendering, so that a large number of lettered words are able to emerge. Therefore, the argument that the Chinese language does not tolerate phonetic loanwords is wrong. Many facts prove that it is not the Chinese language but the Chinese characters which do not tolerate loanwords. Chinese would be able to absorb a large number of phonetic loanwords just as other languages do if an appropriate tool were available.

4. Classification of Lettered Words

1) Classifying from the viewpoint of alphabets used, lettered words are mainly composed using the Latin alphabet and only a few of them are composed using the Greek alphabet.

2) Classifying from the viewpoint of purity, there are two groups. One of them is pure lettered words, such as CT ("computed tomography"), and the other is words with letters, such as BP机 ("beeper"). Moreover, words with letters may be subdivided into letter + symbol (.net), letter + Arabic number (3D) and letter + Chinese character (X光 ["X-ray"]).

3) Classifying by the form of letters: (1) all in capitals, such as DNA and ISO which are always written in capitals; (2) all in lowercase, such as ml and .com which are always written in lowercase (the overwhelming majority of lettered words composed of Greek letters are also written in lowercase, for instance α 粒子["alpha particles"], γ 射线["gamma rays"]); (3) blends (capital with lowercase letters) such as dBASE, vCommerce, Cobol, Internet (the last two are often written as COBOL, INTERNET too); (4) either in capitals or in lowercase, for instance EMAIL, email, Email.

4) Classifying by the amount of letters, they can be subdivided into: (1) a single letter, generally combining with Chinese characters, e.g., A股 ("A stocks"), B超, C调 ("key of C") (the majority of lettered words

formed using the Greek alphabet have only a single letter); (2) with two letters, such as CD, PC, TV, AA 制(everybody pays his own expense at a dinner party),卡拉 OK; (3) with three or more letters, such as OCR, UNIX,TOEFL, PROLOG, CHINANET, MOS 工艺("MOS technique"), BASIC 语言("BASIC language"), 低压 CVD ("low pressure CVD").

- 5) Classifying by the position of letters and Chinese characters they can be subdivided into: (1) in front of Chinese characters (the overwhelming majority), such as T 恤 ("T-shirt"), AB 型血 ("AB-type blood"), APC 大夫 ("unqualified doctor"); (2) behind Chinese characters (these are fairly few), such as 卡拉 OK, 维生素 E, 新柏式 TCT("a new method for examining women's disease"). (3) in the middle of Chinese characters (very few), such as 三 C 革命 ("revolution of communications, computers, and control"), 中国 C 网 ("national information net," China's own Internet), 后 PC 时代 ("post- PC age").

5. Formation Patterns of Lettered Words

Firstly, we must note that almost all lettered words are acronyms, and the shortening methods vary.

- 1) Lettered words derived from Chinese are shortened in two ways: (1) lettered words are composed of the initial letters of each syllable, such as ZL from ZhuanLi ("patent"), RMB from RenMinBi ("Chinese currency"); (2) lettered words are composed of the initial letters of each original word, such as HSK from Hanyu Shuiping Kaoshi (Chinese Proficiency Test), and PSC from Putonghua Shuiping Ceshi (Putonghua Proficiency Exam).
- 2) Lettered words derived from Western languages are more complicated in the ways they are shortened. They may be composed of: (1) the initial letters of the prefix and suffix of the original word, such as kg from kilogram, cm from centimeter; (2) the initial letter and some other letter(s) of

the original word, such as BP 机 from BeePer, EPO from ErythroPOintin; (3) some other combination of letters from the original word, such as EXPWY from EXPressWaY (this is the written form in China, but it is written as EXPY abroad; (4) the initial letters of each original word, such as GRE from Graduate Record Examination; (5) the initial letters of each notional word, such as GMDSS from Global Maritime Distress and Safety System, in which "and" is omitted; and GATT from General Agreement on Tariffs and Trade, in which "on" and "and" are omitted (when shortening, conjunctions and prepositions are often omitted); (6) the initial letter of the first word and the whole second word, such as E-BANK from Electronic BANK; (7) several letters from the original wordgroup (generally the initial letter of the first word is indispensable), such as when DKK is selected from Denmark Krone (DKK is the written form used by the Bank of China, but DKr is used in English) and ETX from End of TeXt character; (8) when shortening, EX in the original word is often converted into X, such as XO (wine) from EXTRA OLD, PAX from Private Automatic eXchange, and Windows XP from Windows experience; (9) some letters are changed, for example, the CS in FACSIMILE is changed into X (FAX); (10) a letter's form is used to create lettered words, such as in T-shirt (it is worthwhile mentioning here that T 型台 [stage for model's show] in Chinese is also created by means of a letter's form); (11) lettered words are formed by the use of euphony, such as ICQ from "I seek you"; (12) lettered words are appended by "semantic marks made of Chinese characters" which are specific to many words partly combined with letters, for instance COBOL 语言 (L in COBOL, in fact, is a short form of "language", and the semantic marker 语言 is not really necessary, so it is often omitted); API 指数 is similar, since API (Air Pollution Index) already involves the word Index 指数.

6. Ways of Arranging Lettered Words

At the present time, there are two ways to arrange lettered words: arrangement in the appendix of an

ordinary dictionary, and arrangement in a specialized dictionary of lettered words.

1) Arrangement in appendix:

(1) Firstly, to arrange lettered words with Greek letters, just as ordinary dictionaries do.

These kinds of words are very few--- the frequently used ones are around 10 in number. After the arrangement of these words, one or two lines are left empty and are then arranged with Latin letters. However, it is also possible to emulate the arrangement in the dictionary of lettered words, i.e., to put these words behind all the words with Latin letters.

(2) Words with Latin letters are large in quantity---collecting the frequently used words, there may be upwards of one hundred. So it has been suggested that they be arranged in pure alphabetic order automatically by computer. That is, to arrange them in the following way:

- A. First, to arrange lettered words initialized with symbols, such as @一代 (a new generation that has grown up with the appearance of the Internet), and .COM 经济(".com economics").
- B. Words initialized with Arabic numbers, such as 3+ X 模式 (a mode of examination used for entrance into university), 3A 级 (first class, i.e., AAA), and 5W 新闻 (all-sided news report).
- C. Words initialized with letters (in alphabetic order), such as A 型, B 股, DNA, and WTO.
- D. Words initialized with Chinese characters (in the order of Chinese Pinyin), such as 三 C 革命, 维生素 B, and 五 E 概念 (E-government, E-business,

E-communication, E-education, and E-family; E=Electronic).

2) Arrangement in Dictionary

- (1) Entries should be classified into different letters, i.e., divided by the initial letters into different areas, just as ordinary dictionaries do. Words not initialized with letters are also arranged into correspondent areas, for instance, 阿 Q is placed in the letter Q area, 3C into C, and 3+X into X. When referring to these words, one should look them up in the Q, C, X areas, respectively.
- (2) All entries are arranged in alphabetic order, ignoring the difference of capital and lowercase letters, for instance, CA (Certification Authority) and Ca (Calcium) may be arranged in any way with respect to their order.
- (3) Entries within the same area are arranged in the following sequence: first arrange the initial element with symbols (\$, &, (), ., @, \, <>), then those with Arabic numbers, then with letters, and finally with Chinese characters. If the entries have the same initial element, arrange them according to the second element in the same sequence, and if the second element is also the same, arrange them according to the third element. The rest may be deduced by analogy. N.B.: a hyphen between letters is ignored.
- (4) Lettered words in the text are automatically arranged according to the above-mentioned rules by computer. It must be noted in particular that we should use computers to do arrangement, retrieval, and other tasks as much as possible in order to modernize dictionary compilation. In the beginning, this effort may bring some inconvenience to those unfamiliar with computers. However, they may realize this system's many merits within a very short time. People are sure to accept it, especially when considering the fact that electronic products will adopt this convenient and accurate retrieval tool.

7. Style of Writing of Lettered Words

Some of the lettered words may be written either in capitals or in lowercase. However, some of them are written only in capitals or only in lowercase (most frequently all in capitals). For example, all lettered words in section 2 must be written in capitals, as analysis tells us that APEC, WTO, and the like are proper names and of course must be in capitals. Many technical terms, such as B 超, DNA, and CDMA must be in capitals too. Furthermore, the symbols of prefix names for marking decimal multiple units (over 10^6), such as M (Mega), G (Giga), and T (Tera) must also be in capitals. On the contrary, the symbols of prefix names for marking decimal fraction units, such as μ (micro), n (nano), and p (pico) must be written in lowercase. Names of measuring units such as kg, cc, and mg must also be in lowercase. Either is acceptable for lettered words like Internet or INTERNET, Fortran or FORTRAN. Either will also do for a word such as Email/ EMAIL/ email. In addition, there are some blends like dBASE, and vCommerce. All of these are conventions.

In order to avoid mistakes, one should bear firmly in mind when to use capitals and when to use lowercase. One must know that CD, Cd and cd represent different things.

8. Pronunciation of Lettered Words

- 1) Lettered words originating from foreign languages are pronounced according to the English alphabet without exception, as all of these words are imported through English. As for the pronunciation of lettered words originating from Chinese language, there is no final conclusion for the time being. However, people tend to pronounce them according to the English alphabet. For example, they pronounce 维生素 C as [si] instead of [ts'e]; and B in RMB as [bi] instead of [be].
- 2) Linkage: It is easy to read those lettered words with linkage, especially long words with vowels inserted, such as .com, JAVA, SOHO 族 (people working at home; SOHO=Small Office/ Home

Office), COBOL, PROLOG. The word ISO is pronounced either with or without linkage. However, DNA, MRI, WTO, and CDMA are pronounced without linkage. And GSM, GPRS, GMDSS certainly can not be pronounced with linkage as there is no vowel in them.

- 3) Accent: Email, APEC, OPEC, and FORTRAN are pronounced with the accent on the first syllable (Many people incorrectly pronounce "Email" with the accent on the second syllable and it sounds unpleasant to the ear). However, UNESCO must be accented on the second syllable.

9. Homographs among Lettered Words

CD, Cd and cd in section 7 may be considered quasi-homographs, as there is similarity only among letters, but they are in different styles (in capital or in lowercase). Therefore, there will be no confusion in reading. In addition, there are many true homographs; for example, 3G is equal to 3 giga or 3rd generation, and ABC is equal to "basic knowledge" or "American Born Chinese." These homographic lettered words can be recognized only according to context, as there is similarity not only among letters but also in their styles.

10. Conversion of Lettered Words

There is a phenomenon called "conversion" in linguistics, i.e., a word that is converted from one grammatical class into another. For example, 丰富 ([en]rich) in Chinese may be converted from an adjective into a verb, so that we may say 丰富了本书的内容 ("enrich the content of this book"). Similarly, in English "email" may be converted from noun into verb, so that we may say "Can you email these materials to her?" The Chinese version of this sentence is 你能把这些材料电邮给她吗? When 电子邮件 ("email") is shortened into 电邮, it can also be used as a verb in Chinese. Even the proper noun Dell (company name) can be converted into a verb (in the meaning of "make somebody out of breath," as this

particular company provides computers according to a customer's order and the price is still much cheaper, so that no other company can compete with it).

Some lettered words may be converted too. For instance, 你的企业 IT 了吗? ("Has your enterprise been modernized by means of information technology?"); FLASH 你的生活 ("Modernize your life by means of FLASH technology"). The above advertising language is nothing new now. People often say in everyday life 吃完饭谁也别走, 咱们卡拉 OK ("Everybody stay after dinner and let us play karaoke").

11. Development of Lettered Words

Lettered words constitute a part of the Chinese lexicon. They will develop in pace with the development of society, just as other parts of the lexicon do; new words are constantly produced, and some old words will wither away simultaneously. Generally speaking, more lettered words arise and fewer die out, just as with ordinary vocabulary. All of this reflects the reality that the whole culture is more and more developed and colorful,

In addition, it is possible to borrow words from other languages besides English. If Russian words are directly absorbed, transliteration will be used, as we already have our own alphabet---Chinese Pinyin based on the Latin alphabet. The method of transliteration will be basically similar to other scripts based on the Latin alphabet (for instance English): sputnik (from Russian спутник), perestroika (from Russian перестройка). However, some discrepancies may occur: with a name like Жданов, it will be transliterated as Zhdanov in English, but it seems proper to be transliterated as Rdanov in Chinese.

*The original of this article is in Chinese (please refer to *Yuyan Wenzhi Yingyong [Application of Language and Writing]* 2002/1). The English version was translated by the author himself and polished by Prof. Mair.

- 1) LIU Yongquan: "Tantan zimuci (Talk about Lettered Words)," *Yuwen Jianshe (Language Construction)*, No. 10, 1994.
- 2) WANG Huidi: "Xinjiapo Huayu zimuci yuyong jianlun (A Brief Survey on the Use of Lettered Words in Singapore's Chinese)," *Lianhe Zaobao (United Morning Newspaper)*, August 6 and 13, 2000.
- 3) SHEN Mengying: "Qianyi zimuci rudian wenti (Preliminaries to the Entrance of Lettered Words into Dictionaries)," *Cishu Yanjiu (Study on Lexicons)*, No. 1, 2001.
- 4) JIANG Yuxing (Hong Kong): "Tan Zhongwen zimuci de jianshe (A Discussion of the Construction of Chinese Lettered Words)," *Yuwen Jianshe Tongxun (Communication on Language Construction)* (Hong Kong), No. 67, 2001.
- 5) WANG Minhua (Taiwan): "Taiwan diqu Huayu zhong suo shiyong zhi Yingwen zi (English Words Used in Chinese Language of the Taiwan Region)," *Yuwen Jianshe Tongxun (Communication on Language Construction)* (Hong Kong), No. 67, 2001.
- 6) LIU Yongquan: *Zimuci Cidian (Dictionary of Lettered Words)*, Shanghai Publishing House of Lexicons, 2001.

Since June 2006, all new issues of *Sino-Platonic Papers* have been published electronically on the Web and are accessible to readers at no charge. Back issues are also being released periodically in e-editions, also free. For a complete catalog of *Sino-Platonic Papers*, with links to free issues, visit the *SPP* Web site.

www.sino-platonic.org