LETTERED WORDS IN CHINESE: ROMAN LETTERS AS MORPHEME-SYLLABLES

Helena Riha
Oakland University

Abstract

In English individual letters are used to represent syllables, morphemes, and words in abbreviations. These uses of letters have been borrowed readily into Chinese, while the use of letters to represent phonemes in spelled words is less common. I discuss why the use of letters to represent units larger than the phoneme is more common in Chinese than their use in spelled words and what this reflects about Chinese morphology. I also argue that since abbreviations and letter-symbol words use letters as components of their structure, they show an interaction between orthography and morphology that should be recognized in morphological studies.

1.   Introduction

In English individual letters normally represent phonemes, and sequences of letters indicate the pronunciation of syllables, morphemes, and words. There are cases, however, when individual letters represent units larger than phonemes, usually syllables, morphemes, or words in examples such as *PJs* (< pajamas), *e-mail* (*e* < e(lectronic)), DJ (< disc jockey), and '*n* 'and'. Individual letters also have this function when they are used as *letter symbols* with contextual meanings, such as *A* in *grade A*, *A* in *A-line* (*dress*, *skirt*, and other women's garments), *B* in *vitamin B*, *X* in *X-ray*, and so on. These various functions of letters are marginal in English, but interestingly, they are borrowed easily into Chinese, as are words formed with letters used in these ways.

2. Describing lettered words

There are many words in contemporary Chinese that are written fully or partly with roman letters. These so-called *lettered words* (*zìmǔcí*) are words written fully or partly in roman letters rather than Chinese characters. Many are borrowings, primarily from English, and most are initialisms (e.g. *NBA*, *WTO*), acronyms (e.g. *NASA*, *NAFTA*), or *hybrid words* composed of a roman letter component and a Chinese character component (e.g. ATM机 'ATM machine'). Many hybrid words are letter-symbol words in which one or more letters are used as symbols with particular contextual meanings that are usually the same as those in English. In X光 'X-ray', for example, *X* means 'unknown identity'; in A区 'area A', *A* means 'first'. A growing number of lettered words are also natively created initialisms, such as *KTV* 'karaoke TV', and hybrid words, such as PC机 'personal computer', lit. 'PC-machine'. Acronyms and ordinary spelled words are much less common. There are no native spelling-to-sound rules for their pronunciation, making them difficult to pronounce if one does not know English.

Lettered words have gained in popularity in Chinese in the last few decades and now form an established category of new words in the language. An appendix of lettered words is included in the *Xiàndài Hànyǔ Cídiǎn* 'Dictionary of Modern Chinese', the authoritative dictionary of Standard Mandarin commonly used by the public in the People's Republic of China. Each new edition includes an ever greater number of lettered word entries in the appendix, indicating the growing influence of lettered words on the language.

3. Corpus study of lettered words in Chinese

I investigated lettered words in the *Chinese Gigaword Third Edition*, a corpus of Chinese newswires (Graff 2007), examining newswires from Xinhua, China's state news agency, during the sixteen-year period from 1991 to 2006 (Riha 2008a). My aim was to describe the characteristics of lettered words and to understand trends in their use during that time. The corpus study revealed that the use of lettered words in Xinhua newswires increased continuously from 1991 to 2006. Based on this finding, I expect that the use of lettered words will most likely continue to grow in popularity in both newswires and in other written contexts.

The most common types of lettered words in the Xinhua corpus are English initialisms and acronyms, such as *WTO* and *APEC*, and hybrid words with both roman letter and Chinese character components, such as K他命 'ketamine', pronounced [kei$^?$ta^{55}miŋ^{51}][1] in Standard Mandarin. Ordinary spelled words are much less common.

The use of initialisms, acronyms, and hybrid words is perhaps best appreciated by examining a sample of a Chinese text that contains lettered words. The text in (1) is an excerpt from an article in China's *People's Daily Newspaper*, quoted in Zhang (2005), which includes numerous lettered words.

---

[1] As noted in §4 on the use of lettered words in speech, the tones for roman letter names have yet to be standardized in Mandarin. I therefore use question marks in place of numerals to indicate the tones of letter names in phonetic transcriptions.

(1)

APEC记者招待会后，我约了STV的记者和一群MBA、MPA研究生朋友，讨论中国加入WTO后IT业对GDP的影响。读MBA的张小姐本来想去.COM当CEO，但觉得IT业风险大...随后大家相约关掉BP机，也不上Internet的QQ和BBS聊天，而是去了KTV唱卡拉OK。

"After the APEC press conference, I invited a reporter from STV and a group of MBA and MPA graduate student friends to discuss the impact of the IT industry on the GDP after China's entry into the WTO. Miss Zhang, an MBA student, originally wanted to go to a .COM to be a CEO, but she thinks the IT industry is too risky … Afterwards, we all decided to turn off our beepers, and we didn't chat on QQ or BBSs on the Internet. Rather, we went to a KTV to sing karaoke."[2]

The excerpt contains fifteen unique lettered words, only two of which are spelled words (*.com* and *Internet*). The thirteen other lettered words consist of one acronym (*APEC*), nine initialisms (*STV*, *MBA*, *MPA*, *WTO*, *GDP*, *CEO*, *QQ*, *BBS*, *KTV*), and three hybrid words that contain an initialism and a Chinese character component (IT业 'IT industry', lit. 'IT-industry', BP机 'beeper', lit. 'BP-machine' (BP < beeper), 卡拉OK 'karaoke' [ka^{214}la^{55}ou$^{?}$kei$^{?}$][3]). This distribution of word types is clearly different from that of most English texts, which contain primarily spelled words and relatively few initialisms, acronyms, or letter-symbol words.

The large number of initialisms, acronyms, and hybrid words in the excerpt is representative of the corpus findings as well. The list in (2) shows the ten most common roman letter words found in the Xinhua corpus (Riha & Baker 2010).

(2)   NBA, GDP, DNA, WTO, APEC, OK, H5N1, IT, IBM, CBA

Nine of the ten words are initialisms and one is an acronym (*APEC*). None of the most frequent words in the corpus are spelled words. I interpret this as further evidence of a preference in Chinese for initialisms and acronyms over spelled words.

4.   Use of lettered words in speech

In contrast to most ordinary words in Chinese, many lettered words are used primarily in writing rather than in speech. Lettered words appear frequently in news writing, technical writing, and computer-mediated communication (Zhang 2005, Gao 2007). Speakers differ, however, in how much they use lettered words in speech. The use of lettered words in the speech of Chinese speakers varies based on numerous factors, such as the speaker's level of education, English fluency, age, profession, personal interests, and

---

[2] Translation provided by the author.

[3] The word 卡拉OK 'karaoke' is a borrowing of the Japanese word カラオケ, lit. 'empty-orchestra', pronounced [karaoke] in Japanese ([oke] < orchestra). (Thanks to Kuniko Nielsen for the phonetic transcription of the Japanese word.) The Chinese characters 卡 [ka^{214}] and 拉 [la^{55}] have no meaning in this word. They are used as phonetic characters to represent the pronunciation of the first two syllables. The letters *O* and *K* are pronounced with their customary letter name pronunciation in Mandarin, [ou$^{?}$kei$^{?}$].

whether the speaker lives in an urban area. These factors can be summed up as an individual's level of participation in China's modernization (Riha 2006).

A comparison with English will help to illustrate how lettered words are primarily a product of the written language. English has a variety of words and expressions associated with a formal register that educated speakers use in writing but have few occasions to say in daily life. One category of this type is words and phrases borrowed from Latin. Speakers may vary in their pronunciation of certain Latin words and phrases because they rarely need to say them aloud even though they recognize them in writing, and because they may rarely hear them spoken. Similarly, educated Chinese speakers may recognize a variety of lettered words in print but may not necessarily need to say them aloud and may not know their pronunciation because they rarely hear the words pronounced.

Another complicating factor regarding the pronunciation of lettered words is that their pronunciations have not been standardized and are not given in dictionaries. The appendix of lettered words in the *Xiàndài Hànyǔ Cídiǎn* provides glosses for lettered words but not their pronunciation. Presumably, speakers are expected to learn the pronunciations by listening to others rather than by consulting the dictionary.

5.   Pronunciation of lettered words

In both initialisms and letter-symbol words, roman letters are usually pronounced individually with their letter names. Roman letters have official letter name pronunciations in China that are provided in dictionaries, but these pronunciations are generally not used to pronounce roman letters in lettered words. An alternate set of letter name pronunciations modeled on English letter names has emerged for this purpose (Riha 2006, Riha 2008a).

As with the pronunciation of individual letter names and letters in initialisms, the pronunciation of many acronyms borrowed from English is based on their English pronunciation. Thus, *APEC* is pronounced approximately as [eipɛk] and *IBM* as [aibiɛm]. In cases where speakers do not know that a particular acronym is conventionally pronounced as a word in Chinese, they pronounce it as an initialism.[4] Natively created lettered words are normally also pronounced as initialisms rather than acronyms. A full discussion of the pronunciation of lettered words is provided in Riha (2008a). The main point here is that the pronunciations of lettered words have not been standardized and continue to evolve as lettered words make inroads into contemporary Chinese.

6.   Popularity of lettered words

My corpus study shows that initialisms, acronyms, and letter-symbol words are considerably more frequent in Chinese newswires than spelled words, which I interpret to mean that these types of words are generally preferred to spelled words in Chinese. These

---

[4] The same is true in English. For example, individuals who are not familiar with the *FTSE 100* (*Financial Times-Stock Exchange 100 Share Index*) would not know that it is commonly called the *Footsie* (Scott 2003) rather than being pronounced letter by letter.

types of words are preferred because they can be interpreted in terms of the most salient unit of organization in Chinese speakers' mental grammar: the Chinese *zi* (*zì*), or morpheme-syllable-character. Individual *zi* are the basic unit of metalinguistic knowledge for Chinese speakers (Riha 2008b). This view is reflected in Chao's (1968: 136-8) assertion that *zi* are the *sociolinguistic word* in Chinese, that is, "that type of unit, intermediate in size between a phoneme and a sentence, which the general public is conscious of, talks about, has an everyday term for, and is practically concerned with in various ways." Chao notes that while the sociological word in English is the familiar notion of *word*, in Chinese it is the *zi*. Larger units such as words and phrases are created from combinations of *zi*, as described by Chao (1968), Li & Thompson (1981), and Packard (2000), among others. The popular view of *zi* also appears to show that this unit is most salient for Chinese speakers. Packard (2000: 15) explains that in popular usage, the spoken morpheme-syllable and the character with which it is written are "one and the same thing" due to assumption that the spoken *zi*, or morpheme-syllable, can always be visually rendered as a written *zi*, or character.

I suggest that there is a structural correspondence between *zi* and individual letters when used in initialisms and acronyms and when used as letter symbols that makes them particularly suitable for incorporation into Chinese. Just as Chinese words and phrases are composed of sequences of morpheme-syllable-characters, these types of lettered units are composed of one or more *morpheme-syllable-letters* and possibly also one or more morpheme-syllable-characters. Chinese speakers essentially select linguistic units in English that are comparable to individual *zi* and sequences of *zi* in Chinese and integrate them directly into the spoken and written language to use alongside *zi*. Letter symbols and lettered words are borrowed, and new lettered items are subsequently created by analogy with borrowed items. For example, the letter *A* was borrowed with many of its meanings in English, one of which is 'first in order or in a series'. The letter *A* and this meaning of *A* are now used in expressions such as A餐 'set meal A' (from Hansell 1989), where *A* is used to indicate the first of several set meals on a menu. Similarly, a variety of English initialisms and acronyms have been borrowed, including *TOEFL*. The initialisms *HSK* and *PSC* [5] are most likely formed by analogy with *TOEFL* and use the same process of abbreviation, albeit for romanized Mandarin words rather than English words.

This is not to say that only letter symbols, initialisms, and acronyms can be borrowed into Chinese and that "regular" spelled words cannot because they lack a correspondence between the individual letters and *zi*. As pointed out by Thomason (2001: 63), a language can borrow "anything" from another language, and my study certainly does show instances of spelled words in the Xinhua corpus. Chinese speakers simply appear to have a greater affinity for the use of letters as morpheme-syllable-letters than the ordinary use of letters as symbols for phonemes. Letters used in this alternate manner can be interpreted by analogy with similar units in Chinese speakers' native language system, whereas letters used to represent phonemes cannot.

---

[5] *HSK* stands for *Hànyŭ Shuǐpíng Kǎoshì*, the romanized name of 汉语水平考试, the proficiency test of Standard Mandarin for foreign learners. *PSC* stands for *Pŭtōnghuà Shuǐpíng Cèshì*, the romanized name of 普通话水平测试, the proficiency test of Standard Mandarin for native speakers of Chinese varieties and other languages in China.

7.   Roman letters as a new type of *zi*

Roman letters have the following functions in Chinese, among others: 1. to write foreign spelled words (*Microsoft*, *Olympics*), 2. to write romanized Chinese morphemes and words (*Běijīng*, *Shànghǎi*), 3. to write foreign initialisms and acronyms (*IBM*, *NAFTA*), 4. to create initialisms based on romanized Chinese words (*HSK*, *PSC*), 5. to write Chinese syllables that do not have a corresponding written character (*K* in K他命 'ketamine'), and 6. to have independent meanings as morphemes that in many cases are the same as those in English (*A* means 'first', 'top', 'best', etc.; *X* means 'unknown identity'). The manner in which letters are used in all of these functions except the first two, writing spelled words, gives them the characteristics of *zi* and allows them to be interpreted in the same way as *zi*, as described below.

Chinese *zi* have the following features: 1. one or more pronunciations as a syllable, 2. one or more associated meanings or grammatical functions, and 3. a self-contained written form that fits into the imaginary *equidimensional square* of Chinese writing (Boodberg 1957). This imaginary square requires each *zi*, no matter how complex its internal structure, to fit into a square of the same size. In addition to ordinary *zi* with these three features, there is also a small number of *zi* used as phonetic characters with no associated meaning or grammatical function. These *zi* are normally used to write loanwords (T'sou 2001). Examples include 咖 [ga^{55}] and 喱 [li^{35}], used in writing 咖喱 'curry'.

Example (3) shows the two *zi* in the compound meaning 'gingko'. Each is written in an equidimensional square. The second *zi* is more complex internally than the first, but both fit into a square of the same size.

(3) | 白 | 果 | 'ginko'; 白 [pai^{35}] 'white', 果 [kuo^{214}] 'fruit'

When individual letters are used to represent morphemes or words (e.g. PC机 'PC', lit. 'PC-machine'), syllables (BB 'ba_by_'), letter symbols (维生素A 'vitamin A'), or phonetic symbols (K他命 'ketamine'), they also have the three features of *zi*. They have a letter name pronunciation that in many cases is just one syllable in length, an independent written form that can fit into an equidimensional square, and an independent meaning, grammatical function, or phonetic function, as shown in (4).

(4) | e | 书 | 'e-book'; e [ji$^?$] 'e(lectronic)', 书 [şu^{55}] 'book'

Example (4) shows that just like the Chinese *zi* 书 'book', which is pronounced as a syllable and has an associated meaning and an independent written form, the letter-morpheme-character *e* also has these same three features: it is pronounced as a syllable, has an associated meaning, and has an independent written form. Letters that can be interpreted as morpheme-syllable-characters in this manner have become a new set of *zi* in Chinese. These *roman letter zi* can stand alone as morphemes and words; they are pronounced as individual units, usually one syllable in length; and they can be combined in ways that are similar to those of *zi* to create new words and phrases (Riha 2008a).

8.   Conclusion

In this article I have shown that Chinese speakers have a preference for using letters to represent morpheme-syllables, the basic linguistic unit in Chinese, rather than phonemes, as is normally done in English. English words in which letters represent morpheme-syllables are uniquely suited to the *zi*-based (morpheme-syllable-based) grammar of Chinese and are commonly used in the contemporary language. This use of letters is exploited in Chinese by interpreting letters in a Chinese manner, that is, as individual *zi*; sequences of letters are similarly interpreted as sequences of *zi*.

In the use of letters as roman letter *zi*, there is an interaction between orthography and morphology that has yet to be recognized in morphological studies. Roman letters used as morpheme-syllable-characters fit well into Chinese in part because each letter fits into the equidimensional square of Chinese writing in the same way as individual Chinese characters. The written form of roman letter *zi* is an integral part of their functionality in Chinese.

As familiarity with English continues to increase in Chinese society, there may be less reliance on the use of roman letters by analogy with Chinese *zi*, and more ordinary spelled words may be used. Nonetheless, lettered words will most likely continue to be primarily initialisms, acronyms, and letter-symbol words since they have the characteristics of Chinese *zi*.

References

Boodberg, Peter A. 1957. The Chinese script: An essay in nomenclature (the first hecaton). *Bulletin of the Institute of History and Philology Academia Sinica* (Taipei) 39:113-120.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese.* Berkeley: University of California Press.

Chinese Academy of Social Sciences, Dictionary Editing Group. 2005. 现代汉语词典 *Xiàndài Hànyǔ Cídiǎn* [Dictionary of Modern Chinese]. Beijing: Commercial Press.

Gao, Liwei. 2007. *Chinese Internet language: A study of identity constructions.* Munich: LINCOM Europa.

Graff, Dave. 2007. *Chinese gigaword third edition.* Linguistic Data Consortium, Philadelphia.

Hansell, Mark D. 1989. *Lexical borrowing in Taiwan.* University of California, Berkeley dissertation.

Li, Charles N., and Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar.* Berkeley: University of California Press.

Packard, Jerome. 2000. *The morphology of Chinese.* Cambridge: Cambridge University Press.

Riha, Helena. 2006. The pronunciation of lettered words in Mandarin Chinese. Paper presented at NWAV 35 (New Ways of Analyzing Variation 35), Columbus, Ohio.

Riha, Helena. 2008a. *Lettered words and roman letter characters in Chinese writing: A study of alphabetic writing in Chinese newswires.* The Ohio State University dissertation.

Riha, Helena. 2008b. Aronoff and sociological words. (Letter to *Language*.) *Language* 84, 1-2.

Riha, Helena, and Kirk Baker. 2010. Lettered words: Using Roman letters to create words in Chinese. In *Variation and Change in Morphology: Selected papers from the 13th International Morphology Meeting, Vienna, February 2008* (IMM 13), ed. by Wolfgang U. Dressler, Dieter Kastovsky, Hans Christian Luschützky, and Franz Rainer. Amsterdam: John Benjamins.

Scott, David L. 2003. *Wall Street words: An A to Z guide to investment terms*. Boston: Houghton Mifflin.

Thomason, Sarah G. 2001. *Language contact: An introduction*. Washington, DC: Georgetown University Press.

T'sou, Benjamin K. 2001. Language contact and lexical innovation. *New terms for new ideas: Western knowledge and lexical change in late imperial China* (= *Sinica Leidensia* 52), ed. by Michael Lackner, Iwo Amelung, and Joachim Kurtz, 35-56. Leiden: Brill.

Zhang, Yihua. 2005. Lexical concerns about neologisms in Chinese lexicography: A cognitive approach to the motivated structure of new words and expressions. In *Words in Asian Cultural Contexts: Proceedings of the 4th Asialex Conference (Asialex 2005)*, ed. by Vincent BY Ooi, Anne Pakir, Ismail Talib, Lynn Tan, Peter KW Tan, and Ying Ying Tan, 380-92. Singapore: National University of Singapore.