

***Lettered Words***  
Using Roman Letters to Create Words in Chinese

HELENA RIHA  
*The Ohio State University*

KIRK BAKER  
*The Ohio State University*

Lettered words are new borrowings and native creations in Chinese written fully or partly with Roman letters. Lettered words in which the letters fit into the Chinese visual ‘frame’ of writing, the ‘equidimensional square’, are adopted “as is,” bypassing traditional borrowing processes that employ Chinese morpheme-syllable-characters (*zì*). These include primarily initialisms, since each letter fits into one frame. Compounds containing a Roman letter as a morpheme (*X* in *X-ray*) are adapted as hybrid words that retain the letter morpheme. We hypothesize that lettered words are shorter than English words on average and use primarily upper case letters. We conducted a corpus study of lettered words in Chinese newswires to test these assumptions, finding that shorter words are most frequent, upper case is preferred, and initialisms are most common. We conclude that Roman letters are a new set of *zì* that can now be used to form words in Chinese.

***1. Introduction***

As China continues to internationalize, the Chinese language is showing the effects of this development through greater contact with English. One resulting trend is the use of a new type of borrowing called ‘lettered words’ (*zìmǔcí*). Lettered words are written fully or partly in Roman letters rather than Chinese characters. Many are borrowings, primarily from English, and most are either initialisms (*WTO*) or hybrid words composed of a roman letter constituent and a Chinese character constituent in their written form (*X* 光 ‘X-ray’, in which *X* is written as a letter and 光 ‘ray’ as a Chinese character). An increasing number are also native creations formed with the same patterns, such as *BB* ‘baby’ and *BP* 机 ‘beeper’ lit. ‘BP-machine’.

Lettered words have become increasingly popular in Chinese in the last few decades and now form an established category of new words in the language. They appear frequently in news writing, technical writing, and in computer-mediated communication (Zhang 2005, Gao 2007).

The use of lettered words in speech varies based on education, English fluency, personal interests, and age, factors that can be summed up as an individual’s level of “participation in China’s modernization” (Riha 2006).

## ***2. Lettered Words as Innovations in Chinese Morphology***

### ***2.1 Bypassing Traditional Borrowing Processes***

Borrowing in Chinese is traditionally done through the use of *zì*, Chinese morpheme-syllables and the characters used to write them

(Norman 1988). Common ‘traditional’ borrowing processes include the following (Cheung 1972, T’sou 2001):

1. Phonological adaptation: Chinese *zì* are used only for their sound value to represent the pronunciation of the foreign term but not its meaning. Their usual meanings are suppressed. This is akin to writing *Schubert* as <Shoebert>; *Shoe & bert* (*Bert* is a personal name) are used just for their sound value rather than for their meaning.

Examples of phonological adaptation include 沙发 *shāfā* ‘sofa’ and 巧克力 *qiǎokèlì* ‘chocolate’.

2. Loan translation: Foreign terms are translated using individual *zì* or combinations of *zì* that have the same literal meaning as the morphemes in the original, but the pronunciation of the original is not conveyed, as in 篮球 *lánqiú* ‘basketball’, lit. basket-ball and 绿卡 *lùkǎ* ‘green card’, lit. green-card.

3. Semantic adaptation: Aspects of the meaning of the original are expressed with Chinese *zì* but not its pronunciation, as in 天使 *tiānshǐ* ‘angel’, lit. heaven-envoy and 打字机 *dǎzìjī* ‘typewriter’, lit. hit-character-machine.

4. Combinations of phonological adaptation and semantic adaptation: Both the pronunciation and aspects of the meaning of the original are conveyed, as in 黑客 *hēikè* ‘(computer) hacker’, lit. wicked-visitor and 嬉皮士 *xīpíshì* ‘hippie’, lit. grin-cheekily-person.

The innovative development with respect to lettered words is that the either the whole word or a part of it is formed without the use of Chinese *zì* (Hansell 1989). Conversion of foreign terms and morphemes into one or more Chinese *zì* is no longer needed in certain cases. We propose that the conversion is bypassed when lettered words or morphemes have enough of the properties of Chinese *zì* that they can be used “as is” in Chinese.

## *2.2 Resolving the Mismatch in the Chinese and English Writing Systems*

The Chinese and English writing systems represent different linguistic units, creating a mismatch that must be overcome to integrate lettered words into Chinese. English writing is ‘morphophonemic’, representing phonemes and words, while Chinese writing is ‘morphosyllabic’, representing syllables and morphemes (DeFrancis 1989). The visual ‘frame’ in English is orthographic *words* of different lengths separated by spaces (DeFrancis 1989). In contrast, the visual frame in Chinese is orthographic *characters* separated by spaces. No matter how visually complex characters are, each one is ‘equidimensional’ (Boodberg 1957), taking up just one imaginary square of the same size. While English writing indicates words of all types, Chinese writing indicates only morpheme-syllables and monosyllabic, monomorphemic words. Polysyllabic words containing multiple *zì* are not indicated, as illustrated in (1).

(1)

I am a painter  
我是画家

Example (1) shows that while English writing separates the words in the sentence, Chinese writing separates only the morphemes and monosyllabic words: 我 ‘I’, 是 ‘to be’, 画 ‘paint’, 家 ‘-er/-ist’. Readers of Chinese parse the strings of morphemes into words:

[我][是][画家].

Because of the mismatch in the two writing systems, integrating English words into Chinese writing is awkward. Words composed of multiple letters do not fit into Chinese frames, and the phonemic representation of the pronunciation of English words is unlike the syllabic representation of the pronunciation of Chinese words. Letters in English words are linked together to pronounce the word, while characters in Chinese are each pronounced as individual syllables, whether or not they are components of polysyllabic words. Lettered words must bridge the gap between the two writing systems to be integrated successfully into Chinese writing.

### *2.3 Congruence Between English Morphemes and Chinese Zi*

Letters used in the manner of Chinese *zì* form a new set of morpheme-syllable-characters in Chinese that we call ‘Roman letter *zì*’, or what Hansell (1989) calls the ‘Sino-alphabet’. We suggest that Roman letter *zì* have the following characteristics: (1) the letter either is or can be pronounced individually, (2) certain letters may have particular

contextual meanings, and (3) the letter fills an equidimensional square. Further, it appears that a preference exists in Chinese writing for using upper case letters more frequently than their lower case counterparts. As for pronunciation, Roman letter *zì* are pronounced with their Chinese letter names, which tend to approximate English letter names. Similar-sounding Chinese syllables are substituted for the English letter name to pronounce the letter (e.g. *U*: English [ju] > Chinese [jiəu]).

Written English words in which the Roman letters can be interpreted as having the characteristics of Chinese *zì* are assimilated most readily into Chinese. That is, cases in which *congruence* (Myers-Scotton 1993) exists between one or more English units and the Chinese frame are the preferred types of lettered words in Chinese. The criteria we propose for congruence are as follows: (1) each letter is pronounced separately from any others adjoining it, (2) letters may have independent meanings as morphemes in the word, (3) each letter fills one frame even though it is part of a larger orthographic string, and (4) words tend to be short, having only a few letters. The most common types of words that fit these criteria are initialisms such as *IBM* and *MBA* and alphanumeric combinations such as *MP3* and *SPF 15* ('Sun Protection Factor 15'). Common acronyms such as *SARS*, *AIDS*, and others are normally pronounced as words in Chinese in imitation of their English pronunciation, but if speakers do not know the customary

pronunciation of these words, they pronounce them letter by letter in the same way as other initialisms. The numerals in alphanumeric combinations are pronounced with their Chinese names.

#### 2.4 Compounds Are Integrated as Hybrid Words

English compound words are normally integrated into Chinese using one or more of the traditional borrowing processes that employs Chinese *zì* (e.g. *snowman* becomes the translation equivalent 雪人 ‘snow-person’). The newest development, however, is that compounds in which one or more letters can be interpreted as a Roman letter *zì* are integrated as hybrid words: the Roman letter *zì* are adopted directly and the rest of the compound is adapted using one or more Chinese *zì* in the traditional manner. The most common types of words adapted in this way are those that contain individual letters used as morphemes to designate categories of items, such as *X-ray* and *vitamin A*. Words in which individual letters are used iconic symbols are also adapted in this fashion (*T-shirt* and *V-shaped tube*), as are compounds containing an initialism and a spelled-out component (*ATM machine* and *IP address*). The adaptations of these examples are shown in (2).

(2)

X-ray	vitamin A	T-shirt
↓ ↓	↓ ↓	↓ ↓
X 光	维生素 A	T 恤
‘X-ray’	‘vitamin A’	‘T-shirt’
V-shaped tube	ATM machine	IP address
↓ ↓ ↓	↓ ↓	↓ ↓
V 形 管	ATM 机	IP 地址
‘V-shaped tube’	‘ATM machine’	‘IP address’

Example (2) shows that English word components that have the characteristics of Chinese *zì* are essentially treated as if they were *zì*. They are retained in their “foreign” form, while the components that are not congruent with Chinese *zì* and the Chinese visual frame are adapted in the customary fashion using Chinese *zì*.

### ***3. Corpus Study of Lettered Words***

#### *3.1 Research Questions*

We investigated lettered words in the *Chinese Gigaword Third Edition*, a corpus of Chinese newswires. We examined portions of the corpus from the People’s Republic of China (PRC) and Taiwan (Xinhua News Agency and Central News Agency, respectively) during the fifteen-year period from 1991 to 2005 to determine whether the characteristics of Roman letter *zì* and lettered words we propose hold in large sample of lettered words. The corpus study investigated the following research questions:

1. Are shorter lettered words preferred to longer ones?

English words vary in length and may contain numerous letters. Such strings of letters do not fit easily into the frames of Chinese writing, since each letter takes up one frame. We therefore predict that lettered words in Chinese will, on average, tend to contain fewer letters than words in English.

2. Are upper case letters preferred?

English writing has specific functions for upper and lower case letters, but no comparable notion of upper case and lower case exists for Chinese characters; each character has just one form. We hypothesize that Chinese writing may tend to level variation in the form of Roman letters to upper case as the default case.

### 3. Are *zì*-like words more frequent than fully-spelled words?

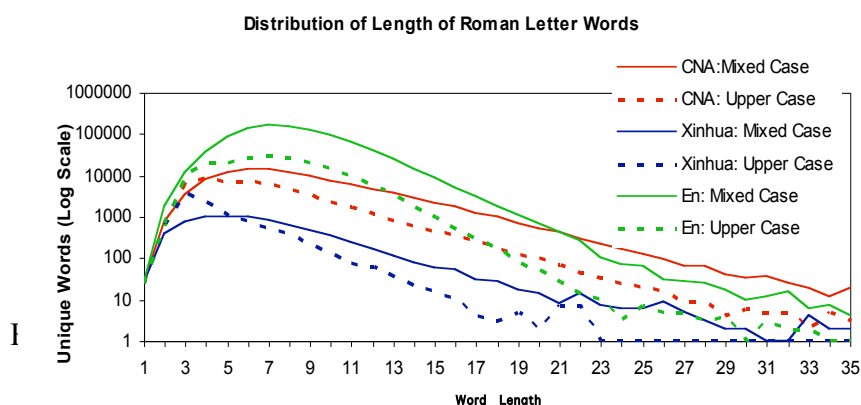
Words in English that are fully spelled out (e.g. *snowman*) tend to be integrated into Chinese using one of the traditional methods of borrowing, whereas words containing Roman letter *zì* tend to retain those elements in the Chinese adaptation of the word. Thus, we hypothesize that *zì*-like words should be more frequent in Chinese than fully-spelled words, which would tend to be adapted with Chinese *zì* using one of the traditional borrowing processes.

### 3.2 Results

We measured the length of lettered words in the Xinhua (PRC) and Central News Agency (CNA, Taiwan) newswires, as well as the length of English words in a corpus of English newswires. Lettered words were categorized as all ‘upper case’, which we hypothesize to be the preferred case, and ‘mixed case’, which represents what we hypothesize to be a dispreferred case, since lower case letters are used.

Figure 1 shows the results of the study.

@@ Insert Figure 1 here



We find that Chinese has a greater number of short upper case word types relative to mixed case. Compared to English, we find a greater concentration of short lettered words in Xinhua and CNA. Xinhua shows a preference for more short word types relative to CNA. The upper case peaks are eight letters for English, five for CNA, and four for Xinhua.

In both Xinhua and CNA, the most frequent instances of lettered words are upper case, borrowed abbreviations with four or fewer letters, as shown in Table 1.

@@ Insert Table 1 here

Top 10 Most Frequent Lettered Word Instances
Xinhua: NBA, GDP, DNA, WTO, APEC, OK, H5N1, IT, IBM, CBA
CNA: WTO, APEC, SARS, IC, NBA, KTV, WHO, DRAM, IBM, GDP

Table 1. Most frequent lettered words in *Chinese Gigaword 3<sup>rd</sup> Edition*

### 3.3 Discussion

Our corpus study shows that initialisms and acronyms are better integrated into Chinese than fully-spelled words. Initialisms and

acronyms resolve the mismatch between the Chinese and English writing systems most successfully. We also find that upper case letters are most frequent in writing lettered words. They may be preferred in part because they may appear more *zì*-like than lower case letters. Finally, short lettered words are preferred to longer words, which may take up too many Chinese frames. The most frequent lettered words are shorter than the average length of words in English.

#### ***4. Conclusions***

We conclude that lettered words present a unique paradox in modern Chinese. On the one hand, they represent a striking innovation in Chinese borrowing and word formation since they use Roman letter elements as morphemes, contradicting the longstanding tradition that foreign words “must be made Chinese in form and character” (Pasierbsky 1989:102), or *sinicized*, before they can enter Chinese. On the other hand, lettered words that are *zì*-like ultimately *are* sinicized since their structure is the same as that of traditional Chinese words formed with Chinese *zì*. Lettered words challenge established notions of what is native in Chinese, what types of units occupy the Chinese lexicon, and what types of units can combine to form Chinese words. Lettered words reveal the creative use of a new set of morpheme-syllable-characters in Chinese, Roman letter *zì*, and also the flexibility of word formation processes in their ability to incorporate new elements into the language.

## 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.
- Cheung, Samuel. 1972. *Cantonese as Spoken in Hong Kong*. Hong Kong: The Chinese University of Hong Kong.
- DeFrancis, John. 1989. *Visible Speech: The Diverse Oneness of Writing Systems*. Honolulu: University of Hawai'i 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 Donald. 1989. *Lexical Borrowing in Taiwan*. Ph.D. dissertation, University of California at Berkeley.
- Myers-Scotton, Carol. 1993. *Dueling Languages: Grammatical Structure in Code-Switching*. Oxford: Clarendon Press.
- Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.
- Pasierbsky, F. 1989. "Adaptation Processes in Chinese". *Language Adaptation* ed. by Florian Coulmas, 90-103. Cambridge: Cambridge University Press.
- Riha, Helena. 2006. "The Pronunciation of Lettered Words in Mandarin Chinese". Paper presented at the conference New Ways of Analyzing Variation (NWAV 35), Columbus, Ohio, November 2006.
- 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, & 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". *Words in Asian Cultural Contexts: Proceedings of the 4<sup>th</sup> Asialex Conference (Asialex 2005), Singapore, 1-3 June 2005* ed. by Vincent B Y Ooi, Anne Pakir, Ismail Talib, Lynn Tan, Peter K W Tan, & Ying Ying Tan, 380-92. Singapore: National University of Singapore.