

Linguistic Purity? Profiling Lettered Words in Chinese Dictionaries

Lan Li

The Hong Kong Polytechnic
egllan@gmail.com

Qi Gong

University Jinan University
tgongq@jnu.edu.cn

Abstract

Language use in the digital era has lead to the thriving of lettered words in non-romanized languages. Some linguistic purists have concerns for these unwanted alien elements, but using letters directly has become an irreversible process in vocabulary expansion. Lettered words have become a significant part of modern Chinese lexicon. The omnipresence of lettered words has also created a challenge to dictionary compilation. Although lettered words have entered Chinese dictionaries, but what are the selection criteria? How can they fit into macro- and micro- structures of dictionaries? This paper, from social, cultural and linguistic perspectives, probes into a list of 1000 Chinese lettered words from a recently published Chinese dictionary, analysing their origins, word class, and uses in modern Chinese. The findings show that lettered words are formed by ten word formation approaches and can be used as verbs, adjectives and adverbs though nouns taking up 75% of the total. The lettered words are widely used in different genres. The limitations of the lettered word list are also discussed. Drawing on the FUDGE factor analysis by Metcalf (2002), the authors suggest that the inclusion of lettered words in a Chinese dictionary should have practical guidelines for selection, definition, pronunciation and usage. Frequency, unobtrusiveness, diversity of use and endurance of lettered words ought to be taken into consideration.

Keywords: *lettered words, linguistic purity, etymology, classification, selection criteria*

1. Introduction

All languages show the presence of different kinds of borrowings. Rapid globalisation and English as an international lingua franca in the digital era have lead to the thriving of lettered words in non-romanized languages. Acceptance of foreign elements, however, varies from language to language. Lettered words in Chinese refer to the words consisting of English letters without Chinese translation, such as GDP, MBAUSB. Some linguistic purists have concerns for these unwanted alien elements, but using letters directly seems to have become an irreversible process in vocabulary

expansion. Lettered words are an obvious part of modern Chinese lexicon today. Lu (2013) reported that about 4000 lettered words are used in printed media, e-media and spoken language in Chinese. The official *Xiandai Hanyu Cidian* (The Contemporary Chinese Dictionary by Commercial Press) is very cautious in recording lettered words into the Chinese language league. In its 6th edition published in 2012, only 254 lettered words were documented in the appendix. This however triggered or rejuvenised a debate on linguistic purity. A number of scholars blamed the editors for recording non-standard language use in a standard Chinese dictionary, which may spoil the unity and beauty of the Chinese language (Su, 2013). They suggested substituting loan words with the creation of new words from Chinese characters and preventing lettered words from entering the language, unless they are used in some specialised areas.

Linguistic purism is a sociolinguistic issue concerning national identity therefore it is also known as linguistic protectionism. It is the practice of defining or recognizing one variety of a language as being purer or of intrinsically higher quality than other varieties. As Thomas (1991:12) observes “Purism is the manifestation of a desire on the part of a speech community (or some section of it) to preserve a language from, or rid it of, putative foreign elements or other elements held to be undesirable (including those originating in dialects, sociolects and styles of the same language). It may be directed at all linguistic levels but primarily the lexicon. Above all, purism is an aspect of the codification, cultivation and planning of standard languages”.

Linguistic purism varies in goals, forms and intensity, and is mostly carried out through language academies. Thomas (1991) noticed that some languages, such as English, Russian and Japanese, are more open to all sources of enrichment, at the same time characterized by a lack of intellectual digestion of foreign influxes. Other languages keep purism as a constant value-feature of the speech community, for example French, Arabic, Tamil and Icelandic. Most languages keep a moderate attitude to linguistic purism however with a reactive correction to a potentially dangerous trend in the development of a standard language.

Linguistic purism in Chinese writing system can be discussed at two levels: interlingual and intralingual. The interlingual is direct bowing from romanised languages mainly from English; the intralingual derives within the Chinese language, using Pinyin variants instead of characters. The following table can illustrate source of lettered words.

Table 1 Nature of lettered words in Chinese

	Interlingual	Intralingual
Putonghua	GDP, GPA, QA, USB	HSK, GB, RMB
Cantonese	GDP, GPA, QA, USB PC (Police Constable) IP (Inspector of Police) CE (Chief executive)	

Using lettered words directly in non-romanised language can speed up communication in modern society and avoid misunderstanding caused by translations. Loan words are now used directly in most languages. However, display of foreign elements in language varies. For example, romanised languages employ the same orthographic system for loan words. Non-romanised languages, if they use alphabets, can use transliteration to avoid foreign codes, such as:

Arabic i-Pad: اى باد ; MBA: ماجستير في إدارة الأعمال

Korean i-Pad: 아이패드; MBA” 경영학석사

Hindi i-Pad: आई-पैड; MBA: एमबीए

Japanese i-Pad: i パッド; MBA: MBA

Loan words in these cases do not seem to affect orthographic purism of these languages.

Using ideographic characters, the Chinese language is in a very different writing system. Untranslated loan words or lettered words when mixed with Chinese characters look alien, undermining the orthographic purism. Some Chinese purists believe that with over five thousand years of civilization, Chinese characters can be and should be used to explain anything new, and neologisms should be introduced after translation (Lu, 2013). These scholars are concerned that the use of lettered words in Chinese would encourage non-standard use of the language and may leave the writing system in disarray. They advocate that foreign elements be banned from standard media, and Chinese dictionaries should not have non-Chinese elements.

It cannot be denied that the omnipresence of lettered words is the result of globalisation reflecting a rapid and unavoidable language change. Fig. 1 shows concordance lines of a lettered word GDP (Gross Domestic Product) used in Chinese news reports.

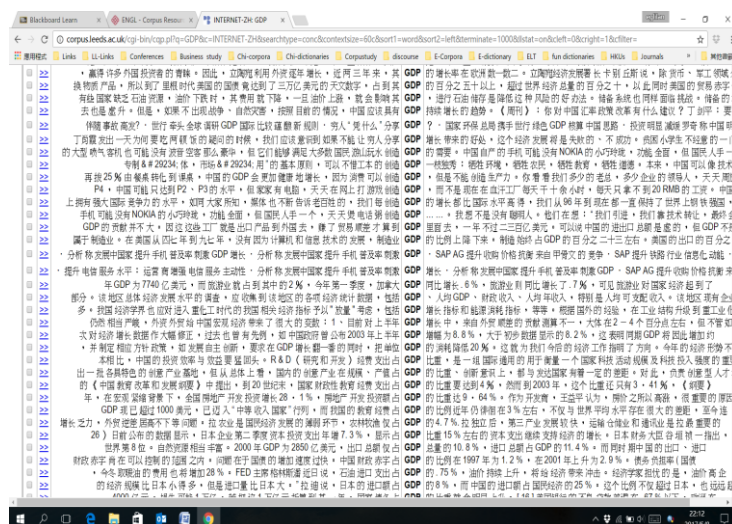


Figure 1 GDP in LeedsU online Chinese corpus

The data is from a Chinese online corpus compiled by LeedsUniversity in the UK. In this 26 million-character data set, GDP is used more often than the expression 国内生产总值. Su and Wu (2013), based on another Chinese corpus, compared the frequency of lettered words and their Chinese translations in official China Daily from 1991 to 2004.

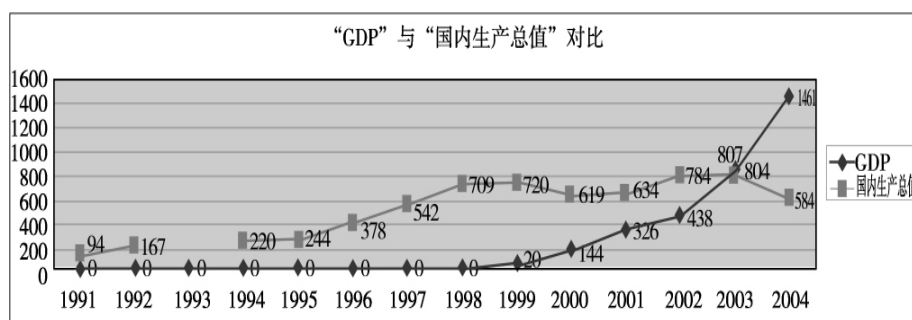


Figure 2 Comparison of GDP and 国内生产总值 in China Daily over 14 years

It is obvious that the use of lettered word GDP increased significantly after the year 2000. This is probably because of the impact of globalization and the increased level of English in China. Su and Wu (2013) also compared another word DNA. Deoxyribonucleic acid is a substance that carries unique genetic codes of living things. Although it has got a standardised Chinese translation 脱氧核糖核酸, corpus data shows that 90% of its use is in English abbreviation DNA, which is easy to say and to remember.

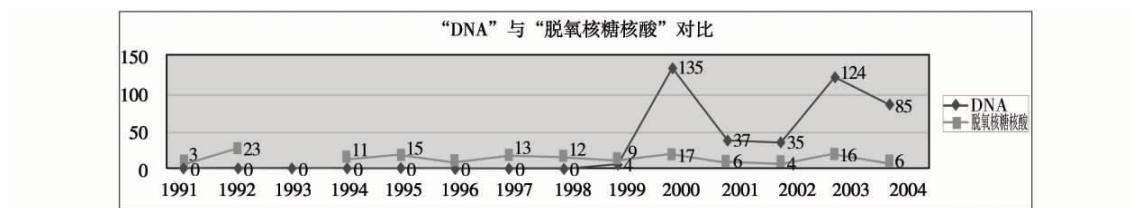


Figure 3 Comparison of DNA and 脱氧核糖核酸 in China Daily

The above examples prove that the use of lettered word is a choice by Chinese writers, a natural process of language use; it cannot be controlled by any big names or institutions. However, it should not be abused, either. This paper, based on different data, attempts to profile lettered words in Chinese, including its typology, its semantic features and lexicographic practice.

2. Data used in the study

The data used in this study is from three sources: 1) the Chinese language reports (2009-2013) by the National Language Resource Monitoring and Research Center; 2) 26 million words of online Chinese Corpus by Leeds University in the UK (<http://corpus.leeds.ac.uk/internet.html>); and 3) A list of lettered word in Xiandai Hanyu Da Cidian compiled by Gong Xuesheng (2015) published by Commerical Press. The reason to choose this list is that it has 1001 lettered words, the longest list in the published Chinese dictionary up to date.

The National Language Resource Monitoring and Research Center was set up in 2003 under the Ministry of Education's Language and Information Management Division. The aim is to provide scholars with more quantitative data of the Chinese language and its contexts, to use modern technology analyzing and managing the national language resources dynamically, and to improve the influence of the Chinese language in the world. It announced Chinese media neologisms from 2009 to 2012 but only included eleven lettered words:

Table 2 Lettered words in National language reports of neologisms

Year	Number of new word	No of loan word	%	No of lettered word	Example
2009	396	12	3%	3	70 码
2010	500	20	4%	6	Phone 时代, TA 时代, i 理财, IN 词, DNA 喷雾, PE 腐败
2011	593	15	2.5%	0	
2012	585	42	7.19%	2	VD 症候群, H 族

To have a clearer picture of lettered words in Chinese, we explored the list compiled by Gong (2015) which has 1001 lettered and numbered words.

3. Findings and discussions

3.1 Origin of lettered words

As expected, the majority of the 1001 lettered words are originated from English (84%) such as USB, CNN, BBC, API. About 10% are code-mixed with English and Chinese, for example, K 金, PH 值, IT 界. Some lettered words are shortened from Chinese Pinyin, for example, HSK, BD, GB, MM; others are English abbreviations of Chinese items such as CET (College English Test) and CNBA (Chinese National Basketball Association). There are also a number of Latin and French words directly used in Chinese now.

Word class of lettered words

Part of speech of lettered words is another observation we made. It is interesting to notice that all the lettered words are not nouns; they take up only 80%. 10% of lettered words are verbs such as CC (carbon copy), PK (player kill), PS (photoshop) and BFN (Bye for now). There are also 20 adjectives such as top, yr(your) and BT (bian tai), and 25 adverbs as WE (whenever), TID(three times a day) and TTBOMK(to the best of my knowledge). TBS (tablespoon) and TPS are quantifiers and haha, hehe are onomatopoeia.

3.2 Formation of lettered words

The creation of lettered words in Chinese has many forms, similar to the ten word formation approaches in English, except derivation. They are shown in Table 2.

Table 3 Word formation approaches of lettered words

Word formation	Frequency	% of the total	Example
Abbreviation	751	75	GPS, GPA, VIP, HSK, CET
Acronym	82	8.2	AIDS, ASCII, SOHO, CEPA
code-mixing	133	13.3	H 股, N 股, X 光, AA 制
Compound	49	5	chatroom
Shortening	81	8	app, sync
Blending	8	0.8	G20, G8
Conversion	9	0.9	Cc, PS, VS
Eponym	2	0.2	ICQ, MSN

Lettered words are found to be used in a variety of genres. As Metcalf (2002) noticed, the more active an area is, the more neologisms will be generated in that area. The most eminent domains with lettered words are science and technology, business, education sociology and online communication.

3.3 Genre of lettered word

The new words in modern technology are fresh to the whole human race. The vast advances of computers and wireless communication have changed the life of human beings dramatically in the last twenty years. Many of them are used in everyday life but have not obtained Chinese equivalents. As a result, lettered words referring to science and technology have taken up about 20% of the 1000 list. AI (artificial intelligence), GPS (global positioning system), LAN (local area network) and CPU (central processing unit) are convenient terms for particular technology; USB (universal serial bus), LED, MP3, PSP and iPod represent ever-changing electronic gadgets; DNA, MRI (magnetic resonance imaging), CT (computerized tomography) and B超 (ultrasonic B-scan) are high-tech medical terminologies which have become part of everyday vernacular of the Chinese people.

A number of lettered words have emerged in business sector as the result of new ideas and new business transactions, such as B2B (business to business), C2C (consumer-to-consumer), A股 (RMB stocks in China), B股 (foreign currency stocks in China), H股 (Chinese stocks issued in the Hong Kong market) and N股 (Chinese stocks issued in New York). Many business job titles also borrow English abbreviations directly such as CEO (Chief Executive Officer), CFO (Chief Financial Officer), and CPA (Certified Public Accountant).

Socio-political terms have played an important role in modern languages. Politics is no doubt a very prosperous area in which language is fully utilized as an indispensable tool. The abbreviations or acronyms of many international organizations are used directly in the Chinese language, such as WTO, OPEC, WHO, BBC, CNN, even though they have Chinese translations or transliterations. Like undergoing an extremely active process of metabolism, language of politics witnesses plenty of words come and go. As everyone has noted, the post-9/11 world has exposed us to a barrage of new terms, for example, WMD (weapon of mass destruction). Many political words and expressions are employed in a creative way. As a matter of fact, new expressions of politics are often endowed with shades of subtle meanings and strong color since their purpose is mainly either to defend or to attack. C4SR (Center for Spatial Research) and C3I (Communication, Command, Control and Intelligence systems) are two new systems employed in military service.

Lettered words have shown a shift from online media to traditional media (Sue, 2013). The users of online media are mainly young people with better English. Education is one of the central topics, therefore many lettered words are used, such as

GPA, TOEFL, IELTS, GRE and SAT. The digital generation is also very open and active in language use. 181 lettered words in the list are categorised as online expressions. The majority are borrowings from English, for example, BFN (bye for now), BTW (by the way), FU (for you) and FYA (for your action/amusement).

There are also some online Chinese pinyin abbreviations such as GG, MM, DD and BD. Some are used to express intimacy such as MM, GG, others are internet thick or coarse language, for example, TMD, BD, SB. Whether these colloquial terms can have a place in dictionary is debatable; the official *Xiandai Hanyu Cidian* (The Contemporary Chinese Dictionary) excludes them all. However one thing is certain that they are a part of the modern Chinese now. Some will stay if they fit into Chinese cultural and linguistic conventions, and others will die out if they violate general rules.

In the next section we are going to discuss the challenges of recording lettered words in Chinese dictionaries.

Challenges to the compilation of Chinese dictionary

The value of dictionaries lies in their contribution towards eliminating obstacles in communication because the average dictionary user regards a dictionary as the most comprehensive authority on linguistic information.

The list of lettered word in Gong's Chinese dictionary (2015) has 1001 entries, providing us with a base for analysis. However the selection criteria does not seem to be clear; quite a number of words can be ruled out. Some entries are repetitive, for example, DVD and DVD 机, IT, IT 界, IT 人才, IT 时代。Other words have proper or standard Chinese translations; it seems unnecessary to enlist them in the lettered words. The examples are Ctrl 键 (控制键), Delete 键 (消除键), Enter 键 (输入键).

Hartmann (2001) regards a dictionary has three types of structures: macrostructure, microstructure and mediostructure. Macrostructure refers to the selection of words of a dictionary. With influx of lettered words, the decision on what words to include seems harder than ever before. To create a word list of lettered words, FUDGE factor analysis by Metcalf (2002) can be employed. FUDGE factors are Frequency of use, Unobtrusiveness, Diversity of users and situations, Generation of other forms and meanings and Endurance of the concept (Metcalf, 2002: 149-166). The origin of this system was from clinical medicine, in which Dr Virginia Apgar introduced a set of Apgar Score to evaluate the health condition of a newborn baby. Metcalf made great endeavor to interpret the mysterious and complex process of survival for the new words into a feasible evaluation mechanism to predict whether a word shall die or survive. In every factor, there are three levels; a word can gain zero, one and two marks respectively. The higher the mark is, the better the word should survive.

Take the word DINK for example. The acronym emerged in the late 1980s, meaning Double Incomes No Kids. As a new life-style in modern society, DINK families are surging in Western world, as well as in Hong Kong and major cities in mainland China. 'Frequency of use' in FUDGE indicates the more frequently a word is used, the more likely it will be kept in dictionary. A Google search of DINK has 1,050,000 hits. However, the figure may not be reliable because many of them are proper nouns such as names of person or institution. A search for DINK family can ensure the meaning we are looking for. With 527,000 hits, the frequency of DINK deserves a score 2.

The unobtrusiveness of DINK can be rated 2 because it is not strong-looking and is in a word family with ink, think, link, sink. The word DINK also has a high productivity to generate other forms and diversified meanings:

DINKY	Double Incomes No Kids Yet
DINKUM	Double Incomes No Kids Unbelievable Mortgage
DINKEM	Double Incomes No Kids with Excessive Mortgage
DIOK	Double Incomes One Kid
DINKWAD	Double Incomes, No Kids, With A DOG

Diversity of users and situations in the FUDGE requires the word to be used by a variety of people and not be restricted in a specific situation. DINK can be rated 2 because the corpus search reveals that it is used in a variety of genres: spoken, fiction, magazines, newspapers and academic. Endurance of the concept means if the object, event or even notion the word based on is still alive and popular. DINK, as many sociologists predicts will last for quite a number of years in modern society. The word DINK, therefore, scores 10 in the FUDGE system, and now it is the entry in 34 dictionaries.

Meaning of lettered words

Like many other words, a lettered word may also have more than one meaning. The dictionary compiler has to decide which meanings to be recorded in a Chinese dictionary and what sense order to decide.

For example, the word GG in Baidu has 11 senses ; some are from English and some originated in Chinese :

1. In computer game : GoodGame
2. Online expression: Gege (Chinese: brother)
3. Brand name: GuccioGucci
4. Pet: gougou (Chinese, dog)
5. Gan ga (Chinese, embarrassed)

6. Googlenetizen
7. Online TV: Goldenglobe
8. Girls' Generation
9. Game interactive platform
10. American soap opera: Gossip Girl
11. Hand-held gaming device: Gamegear

Senses 7 to 11 have limited frequency and are not widely used, therefore they should not be included in a dictionary entry.

Many English abbreviations are polysemic. The online Acronymfinder shows that PS can have 245 meanings, NA has 101, APP120, TMD 40, TSP 131 and TBS 142. Which meanings can enter into Chinese can also be assessed by FUDGE factors.

Pronunciation

Pronunciation is another challenge when lettered words enter into Chinese. Is the borrowing orthographic or phonetic? The dictionaries with lists of lettered word seem to have avoided this issue by providing meanings only. Xinshiji Xinciyu Dacidian by Kang and Liu (2016, Shanghai Lexicographic Press) collected 158 lettered and numbered words occurred during the years 2000-2015. Phonetic information is provided by adding pinyin to Chinese characters in code-mixed words. For example, E 时代[E shidai], DNA 指纹 [DNA zhiwen], S 股[S gu] and N 婚 (N hun). The lettered words without phonetic information seem to have the assumption that all dictionary users understand English or at least know English alphabet, but this is not likely the case. Is it necessary to provide phonetic symbols for the letters? English or Chinese? Or both? The following words are some examples which may cause confusion:

MP3 em pi: san / em pi: θri:
 APP ai pu / ei pi: pi:
 VS vi es / vɜ: s θ s
 GGdʒi: dʒi: / gege
 MM emem / meimei / mumu / mama ?

Mediostructure

Mediostructure is cross-reference in a dictionary. It can be used to establish relations among different components of a dictionary and create textual cohesion. For example, ATM can have a cross-reference to the Chinese entry 自动柜员机 and IELTS can be redirected to 雅思。 Various structural components can interact by mediostructure. There is no doubt that lettered words in printed Chinese dictionaries will face problems of indexing. How to count the number of strokes ? How to index them in the Radical

List? Where to arrange them in the Phonetic Guide? These questions could be answered in electronic media.

4. Conclusions

There is no absolute pure language in the globalized world today, therefore it is not surprising to see more and more lettered words enter into the Chinese language. We should have an open mind to this language change. Using lettered words in Chinese is a natural process of modern communication and cannot be stopped by anyone. However, given the fact that Chinese has a different orthographic system, lettered words should be carefully observed, rather than abused.

It is lexicographers' job to provide dictionary information of lettered words to Chinese users. The challenges are at macro-, micro and medio-levels. FUDGE factor analysis is a good guide for word selection. Each lettered word can also follow lexicographic conventions with a proper microstructure. 'Although no dictionary would encompass all the information a dictionary user would need or like to have, the lexicographer should try to satisfy the average dictionary user' (Zgusta, 1989).

References

- Béjoint, H. (2010). *The lexicography of English*. Oxford: Oxford University Press.
- Hartmann, R.R.K. (2001). *Teaching and Researching Lexicography*. Harlow: Longman Pearson Education.
- Li, Y. and Li, W. (eds.) (2014). *Language situation in China*. Beijing: The Commercial Press.
- Liu, J. and H. Tao (2012). *Chinese under globalization: Emerging trends in language use in China*. Singapore; Hackensack, N.J.: World Scientific.
- Metcalf, A. (2002). *Predicting New Words: the secret of their success*. New York: Houghton Mifflin Company.
- Thomas, G. (1991). *Linguistic purism*. London & New York: Longman.
- Zgusta, L. (1989). The role of dictionaries in the genesis and development of the standard. In F.J. Hausmann et al. (eds.) *Dictionaries. An international encyclopedia of lexicography* (pp. 70–79). Berlin: De Gruyter.
- 社科院(2012)现代汉语词典。北京：商务印书馆。
- 亢世勇, 刘海润 (2016)新世纪新词语大词典。上海：上海辞书出版社。
- 龚学胜(2015) 现代汉语大词典。北京：商务印书馆。
- 宋子然(2016) 汉语新词新语大辞典。上海：上海辞书出版社。
- 陆俭明 (2013) “汉语危机”论可以休也。中国语文， 3/1, 80 — 82。

国家语言资源监测语料库年度新词语报告

http://www.cssn.cn/zt/zt_xkzt/xkzt_yyxzt/hyddxcyyj/

苏新春,吴晓芳 (2013) 字母词的生命力与局限性。北华大学学报,14/2,11-15。