ARTICLE

Check for updates

On the Chinese resistance to lexical borrowing: a writing-driven self-purification system

Liulin Zhang [1✉]

The Chinese language is known for its resistance to lexical borrowing. Transliterations can hardly be retained in this language that use pre-existing characters to simply transcribe the pronunciation of the source word in the donor language. This exclusion can be attributed to the ideographic nature of Chinese characters. Given the stable graphic-meaning correspondence, novel use of characters is expected to be consistent with their usage in previous literature, while the association between the graphic form and the phonetic form has always been loose, rendering it meaningless to use characters as a mere phonetic representation. Here writing is having an effect on the assimilation of loanwords, and more generally, the purist language ideology, which runs counter to the traditionally assumed derivative position of writing, thus shedding light on the implicit effect of writing on language ideology.

[1] School of Chinese Language and Literature, Soochow University, Suzhou, PR China. ✉email: liulinz@suda.edu.cn

Introduction

Among world languages, Chinese is known for its resistance to lexical borrowing. Haspelmath, Tadmor and their colleagues constructed the World Loanword Database, comprising of the words for 1460 basic meanings in 41 languages (Haspelmath and Tadmor, 2009). The finding about Mandarin Chinese presents only 25 probable and clear loanwords, accounting for 1.2% in the Chinese subdatabase (Wiebusch, 2009), far lower than any other studied language. Expanding the scope to non-basic vocabulary does not increase the ratio at all. Zhang (2017) studied the origins of the 8821 words of the HSK vocabulary list, i.e., the official guideline published by Mainland China for non-native speakers studying Chinese as a second language, and found that only 0.75% of those words fit the definition of loanwords in the strict sense (see Section "History of lexical borrowing in Chinese" for details).

As for the motivation underlying this exclusion, some researchers noticed the effect of the Chinese writing system–in the morphosyllabic writing system composed of characters, each character represents a morpheme with its own meaning–transliterations can hardly survive in the Chinese language if the original meanings of the characters interfere with the meaning of the borrowed word (Cook, 2018; T'sou and You, 2003; Wiebusch, 2009). However, in previous discussions of language ideology, orthography is typically mentioned as a battlefield of language ideologies, e.g., the etymological approach versus the phonemic approach for many languages (cf. Woolard, 1994; Brown, 1993; Hellinger, 1986; Schieffelin and Doucet, 1994), while the effect of writing on language ideology itself is barely investigated, which is conceivable under the general assumption that writing exists for the sole purpose of representing (the spoken) language (Saussure, 1916/1959: 23). For this point, the present study probes into the reason why Chinese is generally resistant to lexical borrowing, in an attempt to reveal the pervasive effect of writing on language ideology. Section "History of lexical borrowing in Chinese" briefly reviews the history of lexical borrowing in Chinese. Section "The retention of loanwords over the past 100 years" presents a quantitative study of the retention of loanwords in Chinese over the past 100 years, followed by Section "Ideographic writing and linguistic purism", which discusses the relationship between ideographic writing and linguistic purism. Section "Conclusive remarks" concludes the present study and extends the discussion to the relationship between writing and language ideology in general.

History of lexical borrowing in Chinese

Historical review. Despite the continuous documentation of Chinese over three millennia, little is mentioned regarding the contact of Chinese with other languages. The identification of loanwords depends mainly on linguists' knowledge about languages, histories, and cultures, thus speculative in nature. Zhang (1999) recognized three major periods in which a large number of words were borrowed, i.e., the spread of Buddhism that brought words from Indian languages (mainly Sanskrit), the New Culture Movement in the 1910s that brought words from Japanese and European languages, and the "Reform and Opening up" starting from 1978 that have been bringing words from English. This section will thus focus on these three periods.

The earliest written record of the Chinese language–the oracle-bone script–dates back to the thirteenth century BCE (Schuessler, 2007: 1), but the geographical area inhabited by the Han Chinese people kept expanding for over a thousand years, while neighboring peoples did not have a developed writing system, making the determination of lexical borrowing challenging. Linguists are only able to identify loanwords based on cross-linguistic comparisons and encyclopedic knowledge, e.g., some species are not indigenous to the homeland of China, so the words for them are likely to be borrowed (e.g., Norman, 1988; Norman and Mei, 1976; Schuessler, 2003; 2007; Wiebusch and Tadmor, 2009). Speculated lexical borrowings in this period include 槟榔 *bīngláng*[1] < MC *pjinlang* 'areca palm' (possibly from Malay *pinang* 'areca palm'), 象 *xiàng* 'elephant' < MC *zjangX* (possibly from Thai *chááŋ* 'elephant'), 葡萄 *pútao* 'grape' < MC *budaw* (originally written as 蒲桃 or 蒲陶, possibly from Elamite *būdawa* 'wine'), etc. Since this is the forming stage of the Chinese language, characters were created specifically for those words in writing, making it difficult for common Chinese people to tell their borrowed status (Wang, 2021).

The first period of large-scale lexical borrowing with clear records came with Buddhism. During the social turmoil after the glorious Han dynasty (202 BC–220 AD), Buddhism took off in China. Several approaches were employed to translate Buddhist concepts into Chinese. On one hand, there were still new characters created specifically to denote foreign concepts, e.g., the character 僧 *sēng* < MC *song* 'priest' was created to translate the Sanskrit word *Saṃgha* 'clergy'. On the other hand, more and more renditions resorted to pre-existing characters. When the latter approach is taken, one thing to consider is that each pre-existing character has pre-existing meanings, and thus translations either disregard those meanings by simply using the characters to transliterate the pronunciations in Sanskrit (e.g., 夜叉 *yèchā* < MC *yektsrhae* 'yakṣa: a broad class of nature-spirits in Buddhist beliefs', 舍利 *shèlì* < MC *syaeHlijH* 'sarīra: Buddhist relics', 般若 *bōrě* < MC *pannyak* 'prajñā: wisdom', etc.) or take use of the pre-existing meanings of characters, as in cases of meaning extensions, calques, or loan-based creations. For example, 业 *yè* < MC *ngjaep* originally means 'enterprise; achievement', and was used to translate 'karma' in Sanskrit. Existing morphemes also compound with each other to calque foreign words, e.g., 如 *rú* < MC *nyo* 'to follow; to resemble' compounds with 来 *lái* < MC *loj* 'to come', forming the word 如来 *rúlái* < MC *nyoloj* to calque the Sanskrit word *tathāgata*, which is an honorific title of buddha literally indicating 'the one who has thus come'. Sometimes there is a compromise between the meaning and the pronunciation. For example, the pronunciation of 禅 *chán* < MC *dzyen* is not a perfect match for *dhyāna* 'meditation; the training of mind' in Sanskrit, but this character originally means 'to worship; to abdicate', and the left part of this character is commonly seen in religion-related characters: the meaning of 禅 *chán* is more related to the intended meaning than other characters closer in pronunciation, and is ultimately chosen. Sometimes several renditions co-existed employing different approaches in translating the same word. For example, for the Sanskrit word *prajñā* 'wisdom', besides the transliteration 般若 *bōrě* < MC *pannyak*, the pre-existing compound 智慧 *zhìhuì* < MC *trjeHhwejH* 'wise + intelligent = wisdom' is also seen, which can be perceived as meaning extension and ultimately replaced 般若 *bōrě* in most contexts.

The second period of large-scale lexical borrowing was in response to foreign invasion. Starting from the Opium War of 1840, China was dragged into the "century of humiliation (Kaufman, 2010)" subjugated by Western powers and Japan. The deplorable situation struck nationalist scholars that China must learn from others. Numerous books were thus translated from other languages into Chinese. This fashion culminated in the New Culture Movement (新文化运动) in the 1910s, bringing a tremendous change to the Chinese language and is thus taken as the starting point of Modern Mandarin (Wang, 1944/1984: 434; Kratochvil, 1982: 287). Translated texts naturally brought new lexical items from Europe (mainly English) and Japan. It is commonly seen that several renditions employing various

approaches co-existed for the same source word. For example, *shampoo* is translated as 香波 *xiāngbō* (literally 'fragrant ripple', meaning-pronunciation compromise) or 洗发水 *xǐfàshuǐ*/洗头膏 *xǐtóugāo* (literally 'hair-washer', loan-based creation); *vitamin* is translated as 维他命 *wéitāmìng* (literally 'to preserve his life', meaning-pronunciation compromise) or 维生素 *wéishēngsù* (literally 'the element to preserve life', loan-based creation); *bus* is translated as 巴士 *bāshì* (transliteration) or 公共汽车 *gōnggòngqìchē*/公交车 *gōngjiāochē* (literally 'public car', loan-based creation), etc. Besides, since many characters have the same pronunciation in Modern Mandarin, the same word may also have multiple ways of graphic representation, e.g., *mango* can be written as 芒果 *mángguǒ* or 杧果 *mángguǒ*; *ice-cream* can be written as 冰激凌 *bīngjīlíng* or 冰淇淋 *bīngqílín*. Apart from the approaches in translating Buddhist texts, graphic loans also appeared at this stage. Graphic loans take the form of alphabet letters, or more frequently, the Chinese-based Japanese characters known as kanji. Historically, Japanese borrowed the writing system of Chinese characters, together with a great number of words from Middle Chinese, i.e., Sino-Japanese words (Schmidt, 2009). These words were initially borrowed with Japanese adaptations to the pronunciations from Middle Chinese, later fixed as Sino-Japanese pronunciations. In the 19th century, also faced with the invasion of the West, Japan was able to modernize within a short period of time, much earlier than China. Sino-Japanese morphemes thereby compounded in new ways to translate modern terms, such as 民主 *mínzhǔ* 'people + dominate = democratic'. Some old terms from classical Chinese texts were attached to new concepts, such as 经济 *jīngjì* 'to govern + to benefit = economy'. When China was finally pushed onto the path of modernization, these modern terms consisting of Sino-Japanese morphemes were reborrowed back to China, but Chinese people read these words by their own pronunciations, which are distinct from Sino-Japanese pronunciations.

Another influx of loanwords have been affecting Modern Mandarin since the Reform and Opening-up (改革开放 *Gǎigé Kāifàng*) staring from 1978. With unprecedented intercultural communication, new lexical items are borrowed every year. The approaches of borrowing are the same as before: transliterations, morpheme-by-morpheme calques, meaning-pronunciation compromises, loan-based creations, loanblends (transliterated elements compounding with native morphemes), and graphic loans are all commonly captured. Faced with the enormous number of loanwords, there began to be explicit advocations of linguistic purism. Graphic loans taking the form of alphabet letters have sparked the most furious debate among Chinese scholars. The fact that the latest editions of *Xiàndài Hànyǔ Cídiǎn* 'Contemporary Chinese Dictionary' (《现代汉语词典》) collect English abbreviations has been harshly criticized by language purists. The government also began to be involved in the standardization of loanwords by publishing "suggested translations for foreign terms" (see the website of the Ministry of Education, http://www.moe.gov.cn/s78/A18), showing a notable avoidance of graphic loans from non-Chinese writing. In 2010, the State Administration of Press, Publication, Radio, Film and Television of China (国家新闻出版广播电影电视总局 *Guójiā xīnwén chūbǎn guǎngbō diànyǐng diànshì zǒngjú*) officially banned the use of English abbreviations, including GDP, WTO, NBA, and many other commonly-used terms.

In addition to the above-mentioned three major periods, Chinese also borrowed words from different languages sporadically, including the borrowings from Altaic languages during the Yuan dynasty (1271–1368, reigned by Mongolians) through the Qing dynasty (1636–1912, reigned by Manchurians), but these borrowings are comparatively limited in number.

It is worth mentioning that Hong Kong and Taiwan present diverse situations, distinct from Mainland China, regarding the use of loanwords. Overall, more transliterations are used in these areas than in the mainland (see Bauer, 2006; Hsieh and Hsu, 2006; Shao, 2000; among others). In Hong Kong, many words are transliterated based on Cantonese phonology, such as 多士 *duōshì* (do1si6 in Cantonese[2]) 'toast', 免治 *miǎnzhì* (min5zi6 in Cantonese) 'minced', 士多啤梨 *shìduōpílí* (si6do1be1lei4 in Cantonese) 'strawberry', 士巴拿 *shìbāná* (si6baa1naa2 in Cantonese) 'spanner' and so forth. Meanwhile, these places employ diverse standards for the transliterations of proper nouns. For example, Hillary is transliterated as 希拉里 *Xīlālǐ* in the mainland, 希拉莉 *Xīlālì* (Hei1laai1lei6 in Cantonese) in Hong Kong, and 希拉蕊 *Xīlāruǐ* in Taiwan.

The selective adoption of lexical items. Evidently, Chinese has been borrowing lexical items at every stage. It is hard to fathom that the total percentage of loanwords in Chinese lexicon is lower than 2% (see Wiebusch, 2009; Zhang, 2017). Chinese definitely borrows words, but the key to understanding the Chinese resistance to lexical borrowing essentially resides in the filtration of borrowed items and the definition of loanwords: many transliterations became obsolete.

As previously mentioned, there were cases that several variants employing distinct borrowing approaches co-existed for the same source word. A selective list is presented in Table 1 to demonstrate the screening of Chinese lexicon: for each source word, the obsolete variant is shaded.

From Table 4 we can conjecture a general rule of lexical selection in Chinese:

(1)   Transliterations using pre-existing characters are virtually always disfavored when there is another variant (a) using specifically-created characters; or (b) formed by native morphemes.

Accompanying this general rule is the fact that it has been less common to create new characters specifically for loanwords than before, which means that in Modern Mandarin, translations using native morphemes are virtually always preferred over transliterations.

As a result of the selective adoption, very few transliterations are retained in Modern Mandarin. Table 2 presents the results from Zhang (2017), based on the analysis of 8821 words of the HSK vocabulary list.

As previously mentioned, graphic loans from non-Chinese writing are discouraged by the government. Against this background, the HSK vocabulary list does not include any graphic loans of alphabet letters. As for graphic loans from kanji, the majority of Chinese linguists endorse that they are loanwords, but there are still researchers maintaining that at least some of those words are not loanwords in the strict sense, but should be recognized as "returning Chinese words (回归词 *huíguī qiáocí*)" (e.g., Feng, 2004: 23–28; Pan et al., 1993: 389–391).

When constructing the World Loanword Database, Haspelmath and Tadmor (2009) employed the criterion of "analyzability" to identify loanwords: if a word is analyzable (i.e., morpho-syntactically complex) within the language, it is almost certain that it was created by speakers of the language rather than borrowed from some other language. Based on this criterion, only transliterations, graphic loans of alphabet letters, and meaning-pronunciation compromises in Table 2 can be counted as loanwords, which explains why the percentage of loanwords is unusually low in Chinese in comparison with other languages.

The retention of loanwords over the past 100 years
The previous section briefly reviews the history of lexical borrowing in Chinese, and proposed a general rule that transliterations are

Table 1 The selective adoption of Chinese lexicon.

Age of Borrowing		Old Chinese–Middle Chinese		Middle Chinese–Early Mandarin		Early Mandarin–Modern Mandarin	
Source word		Elamite *būdawa* 'wine'	Xiongnu *dada* 'camel'	Sanskrit *prajñā* 'wisdom'	Sanskrit *bhikṣuṇī* 'nun'	English *shampoo*	English *telephone*
Newly-created characters	Transliteration	葡萄*pútao*	骆驼*luòtuo*				
Pre-existing characters	Transliteration	蒲桃*pútáo*	橐驼*tuótuó*	般若*bōrě*	比丘尼*bǐqiūní*	香波*xiāngbō*	德律风*délǜfēng*
	Meaning-pronunciation compromise						
	Loanblend				尼姑*nígū*		
	Calque/Loan-based creation					洗发水 *xǐfàshuǐ*	电话*diànhuà*
	Meaning Extension			智慧*zhìhuì*			

Table 2 Percentages of different types of borrowings (Zhang, 2017).

Writing	Pronunciation & Meaning	Percentage	Examples
Newly-created characters	Transliteration	0.17%	虎*hǔ* 'tiger', 江*jiāng* 'river', 葡萄*pútao* 'grape', 咖啡*kāfēi* 'coffee', etc.
Alphabet letters	Graphic loan	0%	
Pre-existing characters	Graphic loan (from kanji)	4.48%	科学*kēxué* 'sort + to study = science', 经济*jīngjì* 'to govern + to benefit = economy', 革命*gémìng* 'to remove + life = revolution', 自由*zìyóu* 'self + to allow = freedom', etc.
	Transliteration	0.26%	和尚*héshang* 'upādhyāya: priest', 英格兰*Yīnggélán* 'England', 米*mǐ* 'meter', 沙发*shāfā* 'sofa', etc.
	Meaning-pronunciation compromise	0.05%	芒果*mángguǒ* 'mango (literally 'awn fruit')', 可乐*kělè* 'cola (literally 'jolly')'
	Loanblend	0.29%	狮子*shīzi* 'lion (子*zi* is a suffix)', 车站*chēzhàn* 'station (车*chē* 'vehicle')', 啤酒*píjiǔ* 'beer (酒*jiǔ* 'alcohol')', 浪漫主义*làngmànzhǔyì* 'romanticism (主义*zhǔyì* '-ism')', etc.
	Calque/Loan-based creation	4.50%	现在*xiànzài* 'now + exist = now/present < atīta', 世界*shìjiè* 'generation + boundary = cosmos, time and space < lokadhātu', 智慧*zhìhuì* 'wise + intelligent = wisdom < prajñā', 电话*diànhuà* 'electronic + speech = telephone', 火车*huǒchē* 'fire + vehicle = train', etc.
	Meaning Extension	1.46%	将来*jiānglái* 'will + come = future > pratyutpanna', 玉米*yùmǐ* 'jade + rice = corn', 先生*xiānshēng* 'senior > sir; mister', 知识*zhīshí* 'acquaintance > knowledge', 龙头*lóngtóu* 'dragon head > faucet', etc.
Total		11.21%	

virtually always disfavored when other variants are available. To test this rule with quantitative data, this section studies the retention of different types of words for borrowed concepts over the past 100 years. We collect all the words for borrowed concepts in Lu Xun's book *Fén 'Tomb'* (《坟》), published about 100 years ago, analyze the type of borrowing for each word, and calculate the retention rate for each type. It is hypothesized that transliterations have a relatively low retention rate compared to words that are analyzable with native Chinese morphemes.

Methodology and results. Born in Shaoxing, Zhejiang Province, Lu Xun (1881–1936) studied medical science in Japan in his early life, where he came to the realization that saving people's soul is more pressing than saving people's body. Then he dedicated himself to writing, and made active efforts to introduce Western ideas to modernize the Chinese language and culture. Mao Zedong (1940/1991) commented that Lu Xun is the chief commander of China's cultural revolution, a great writer, thinker, and revolutionist… "On the cultural front, he was the bravest and most ardent national hero without parallel in our history (Mao, 1940)." After the establishment of P. R. China, Lu Xun is the writer with the most works selected into school textbooks. *Fén 'Tomb'* (《坟》) is a self-selected collection of Lu Xun's essays published from 1907 to 1925, comprising about 180,000 characters. We found 766 words expressing borrowed concepts in this book, including 373 proper nouns, constituting our sample. The entire list of 766 words can be found as Supplementary Table S1 online. "Borrowed concepts" refer to entities and notions that clearly had foreign origins and never appeared in Chinese texts before the 19th century, but this does not necessarily mean that all these words were first translated into Chinese by Lu Xun. In fact, Lu Xun adopted many pre-existing forms, including 亚美利加*Yàměilìjiā* 'America', 英吉利*Yīngjílì* 'English', 德意志*Déyìzhì* 'Deutsch', 亚当*Yàdāng* 'Adam', etc. There are a few cases in which the same concept is expressed by several distinct words, as represented by distinct written forms. They were counted as different words instead of variants of the same word. A total of 314 words (40.99%) are transliterations, including 304 proper nouns. Besides, 405 words (52.87%) are analyzable with native morphemes represented by characters, among which 30 are proper nouns. Morpheme-by-morpheme calques (e.g., 帽架*màojià* 'hat + stand = hatstand'), loan-based creations (e.g., 教堂*jiàotáng* 'religion + hall = church'), and pre-existing words/phrases that gained new meanings (e.g., 枪*qiāng* 'spear > gun') are not distinguished for these 405 words as the exact source word is not always clear. Many of these words were borrowed from Japanese, initially coined by compounding Sino-Japanese

Table 3 The retention rate for each type of borrowing.

Word in the sample	Retention in the corpus		Examples
Transliteration (unanalyzable, 242 in total)	Retained (32.64%)		吗啡 *mǎfēi* 'morphine', 希腊 *Xīlà* 'Greece', 达尔文 *Dá'ěrwén* 'Darwin', 亚当 *Yàdāng* 'Adam', etc.
	Replaced (67.36%)	Replaced by other transliterations (63.22%)	格里累阿 *Gélǐlèi'ā* replaced by 伽利略 *Jiālìlüè* for 'Galilei', 法斯忒 *Fǎsītè* replaced by 浮士德 *Fúshìdé* for 'Faust', 华骚 *Huásāo* replaced by 华沙 *Huáshā* for 'Warsaw', etc.
		Replaced by loanblends (1.24%)	林舍亚克特美 *Línshěyàkètèměi* replaced by 意大利猞猁之眼国家科学院 *Yìdàlì shēlì zhī yǎn guójiā kēxuéyuàn* for 'Lincean Academy', 札尔 *Zhá'ěr* replaced by 沙皇 *Shāhuáng* for 'Tsar'
		Replaced by fully analyzable items (2.89%)	之不拉 *zhībúlā* replaced by 斑马 *bānmǎ* for 'zebra', 普斯多 *pǔsīduō* replaced by 平原 *píngyuán* for 'puszta', 费厄泼赖 *fèi'èpōlài* replaced by 公平竞争 *gōngpíng jìngzhēng* for 'fair play'
Loanblend (partially analyzable, 41 in total)	Retained (41.46%)		美国 *Měiguó* 'America', 德文 *Déwén* 'German (written language)', 伊凡四世 *Yīfánsìshì* 'Ivan IV of Russia', 纳尔逊传 *Nà'ěrxùnzhuàn* 'The Life of Nelson', etc.
	Replaced (58.54%)	Replaced by transliterations (0%)	—
		Replaced by other loanblends (56.10%)	罗曼派 *luómànpài* replaced by 浪漫派 *làngmànpài* for 'Romanticism', 堪勃力俱 *Kānbólìjù* replaced by 剑桥 *jiànqiáo* for 'Cambridge', 亚历山德府 *Yàlìshāndéfǔ* replaced by 亚历山大图书馆 *Yàlìshāndà Túshūguǎn* for 'Bibliotheca Alexandrina'
		Replaced by fully analyzable items (2.44%)	兴矣摩迦人 *Xìngyǐmójiārén* replaced by 民族之歌 *mínzúzhīgē* for 'Talpra Magyar'
Fully analyzable item (392 in total)	Retained (80.61%)		鱼肝油 *yúgānyóu* 'cod-liver oil', 代数 *dàishù* 'algebra', 植物学 *zhíwùxué* 'botany', 无政府主义 *wúzhèngfǔzhǔyì* 'anarchism', 中央公园 *Zhōngyāng Gōngyuán* 'Central Park', etc.
	Replaced (19.39%)	Replaced by transliterations (0%)	—
		Replaced by loanblends (0%)	—
		Replaced by other fully analyzable items (19.39%)	善种学 *shànzhǒngxué* replaced by 优生学 *yōushēngxué* for 'eugenics', 元质 *yuánzhì* replaced by 元素 *yuánsù* for 'element', 神思宗 *shénsīzōng* replaced by 唯心主义 *wéixīnzhǔyì* for 'idealism', 质学 *zhìxué* replaced by 化学 *huàxué* for 'chemistry', etc.

morphemes (see Section "Historical review"). Apart from words that are fully analyzable or completely not analyzable, there are 47 loanblends (6.14%) that are partially analyzable, including 39 proper nouns. Clearly, transliteration was mostly used for proper nouns, while native Chinese morphemes are strongly preferred by other words.

To investigate the retention of borrowed words, the words in our sample were compared to the most commonly used words expressing the same concepts in the Balanced Modern Mandarin Corpus constructed by the National Language Commission of China (国家语委 *Guójiā Yǔwěi*). This corpus contains over 100 million characters, covering five genres of texts written from 1919 to the present time. The constructors claim the corpus to be representative of the standard Modern Mandarin. The concepts expressed by 91 words, mostly uncommon proper nouns, do not appear in the corpus, and are thus excluded from analysis. For the rest 675 words, 412 (61.04%) are retained, while 263 (38.96%) have been replaced by other words. The retention rate for each type is presented below in Table 3.

A strong preference for analyzable lexical items is clearly shown in Table 3: transliterations are more likely to be replaced than calques/loan-based creations/meaning extensions. In fact, this preference is demonstrated in Lu Xun's own word choice. As previously mentioned, the book *Fén* 'Tomb' (《坟》) collects essays published from 1907 to 1925. Some foreign concepts appear in different forms across different essays. For example, 'America' appears as 亚美利加 *Yàměilìjiā* (Cantonese-based transliteration, see Section "Discussion") in *Móluó Shīlì Shuō* 'The Theories of Māratic Poetic Power' (《摩罗诗力说》, published in 1907), but as 美国 *Měiguó* (loanblend with 国 *guó* denoting 'country') in *Cóng Húxū Shuō Dào Yáchǐ* 'From

Mustache to Teeth' (《从胡须说到牙齿》, published in 1925). Meanwhile, the forms of transliterations are not stable. For example, the name 'Ibsen' is transliterated as 伊孛生 *Yībèishēng*, 伊勃生 *Yībóshēng*, and 易卜生 *Yìbǔshēng* (the variant that eventually became dominant) in different essays.

Discussion. Language is commonly understood to be a semiotic system comprising of conventionalized (phonetic) form-meaning pairs (cf., Goldberg, 1995: 4; Langacker, 2008: 21; Saussure, 1916/ 1959: 16). Accordingly, the difference between loanwords and code-switching resides precisely in conventionality (Haspelmath, 2009: 40). Borrowing a lexical item from another language pertains to the introduction of a new form-meaning pair, and the sporadic uses by a single speaker/writer cannot guarantee the retention of the new form-meaning pair in the recipient language. Only when this form-meaning pair is frequently used throughout the entire language community can we say that it is truly integrated into the lexicon of the recipient language. Compared to alphabetic writing, the ideographic writing system of Chinese characters imposes two barriers on the language that prevent borrowed form-meaning pairs from conventional use, i.e., (i) the arbitrariness resulting from the large number of homophonic written signs, and (ii) dialectal variations regarding the pronunciations of written signs.

Unlike alphabetic writing in which each sign represents a phoneme, the signs in ideographic writing represent ideas: each sign is meaningful in its own right, and multiple signs may have the same pronunciation. There are about 400 syllables (1300 with tones) in Modern Mandarin, but *Xiàndài Hànyǔ Cídiǎn* 'Modern Chinese Dictionary' (《现代汉语词典》, the official standard in

Table 4 Dialectal variations of transliterations.

Borrowed concept	Bacon	Pushkin	France	Cromwell
Lu Xun's transliteration	培庚 Péigēng (*bEgən* in Shanghainese)	普式庚 Pǔshìgēng (*pʰusəgən* in Shanghainese)	法朗西 Fǎlǎngxī (*fɐʔlɑ̃çi* in Shanghainese)	克灵威尔 Kèlíngwēiěr (*kʰlinuEər* in Shanghainese)
The most common form in Modern Mandarin	培根 Péigēn	普希金 Pǔxījīn	法兰西 Fǎlánxī	克伦威尔 Kèlúnwēi'ěr

Table 5 Five types of purist language ideology (Thomas, 1991: 76–81).

Type of Linguistic Purism	Pure	Impure
Archaizing purism	Past literary model; linguistic material of a past golden age	Contemporary vernacular language
Ethnographic purism	Rural dialect	City speech
Elitist purism	Highly formal varieties linked to elites	Substandard and regional varieties
Reformist purism	A new style different from the past	The "pure" in the past
Xenophobic purism	Native elements	Foreign elements

Mainland China) lists over 13,000 characters. As an extreme example, over 200 characters are pronounced as *yì*, each with distinct meanings. In transliteration, translators are free to choose from an inventory of characters with same/similar pronunciations, and thus may find it challenging to stay consistent themselves. This arbitrariness explains the reason why various forms of transliterations can be captured for the same borrowed concept in Lu Xun's works (see Section "Methodology and results").

On top of the arbitrariness of character selection, the dramatic dialectal difference adds an additional layer of challenges: the same character may be pronounced differently across dialects. The Chinese language is known for involving countless mutually unintelligible varieties (Li, 2004), to the degree that many Western linguists suggest to treat Chinese as a language family, instead of a single language, as Norman (1988: 187) puts it:

(2) To the historical linguist Chinese is rather more like a language family than a single language made up of a number of regional forms. The Chinese dialectal complex is in many ways analogous to the Romance language family in Europe… it would not be surprising if we found about the same degree of diversity among the Chinese dialects as we do among the Romance languages, and in fact I believe this to be the case. To take an extreme example, there is probably as much difference between the dialects of Peking and Chaozhou as there is between Italian and French; the Hainan Min dialects are as different from the Xi'an dialect as Spanish is from Rumanian.

Lu Xun is from Shaoxing, Zhejiang Province, about 200 kilometers away from Shanghai. and his transliterations are notably influenced by the Shaoxing accent close to Shanghainese. The standard Modern Mandarin, based on the Beijing dialect, has the contrast between the back nasal coda ng[ŋ] and the front nasal coda n[n], but the Shanghai-Shaoxing dialect does not. This difference results in a series of complementary forms of transliterations, as shown below in Table 4.

In the early 20th century, communications between China and foreign cultures centered around Shanghai and Hong Kong, and thus many foreign concepts were introduced into the Chinese language through Shanghainese or Cantonese. It was not rare to see different forms used in two places to transliterate the same word. For example, the word *sofa* was transliterated as 沙发*shāfā* in Shanghai (*sofA* in Shanghainese, the variant that eventually became dominant in Mainland China), and 梳化*shūhuà* (*so1faa3* in Cantonese) in Hong Kong. For an effective communication

within the entire Chinese community, there must be some variants eliminated in the competition. When a centralized government was established in the mainland, this competition pushed the government to get involved in the standardization of loanwords based on the phonological system of Modern Mandarin (see Section "Historical review"). In this process, many Cantonese-based transliterations were driven out of the mainland, such as 亚美利加*Yàměilìjiā* (*Aa3mei5lei6gaa1* in Cantonese) 'America' in the sample (although it was once adopted by Lu Xun). Besides, it is easier for transliterations to accumulate conventionality in a small language community than in a phonologically diverse community. This explains why more transliterations are retained in Hong Kong and Taiwan than in Mainland China. Essentially, the conventionality of transliterations is built upon a conventionalized writing system corresponding to a common phonological system, otherwise there will naturally be regional variations of transliterations that lead inevitably to competitions. Accordingly, the standardization of transliterations (orthography) is possible only when there is a standard phonological system.

On the other hand, with the stable graphic-semantic association intrinsic to ideographic writing, calques, loan-based creations, and meaning extensions can easily live across different dialectal areas. Analyzable items are thus more stable than transliterations in Chinese. It can be noticed that the ideograph-based nature of Chinese writing inherently imbues the Chinese language with a self-purification system that implicitly resists direct lexical borrowing by transliteration.

Ideographic writing and linguistic purism
The purist language ideology refers to an idea of good (pure) versus bad (impure) language varieties. Thomas (1991: 76–81) distinguished five main types of purism, i.e., archaizing purism, ethnographic purism, elitist purism, reformist purism, xenophobic purism, summarized below in Table 5.

Linguistic purism can not only be explicitly expressed by researchers and language planners, but also wield its effect implicitly (Markovic, 1984; Silverstein, 1979; Woolard, 1994), as in the case of Chinese. The resistance to lexical borrowing is a typical instantiation of xenophobic purism. However, for thousands of years, there were never explicit advocations of xenophobic purism until recent decades. In particular to the case of Lu Xun, as previously mentioned, his influence is widely recognized in the Chinese community, yet a considerable number of his transliterations have been replaced. The perennial resistance to
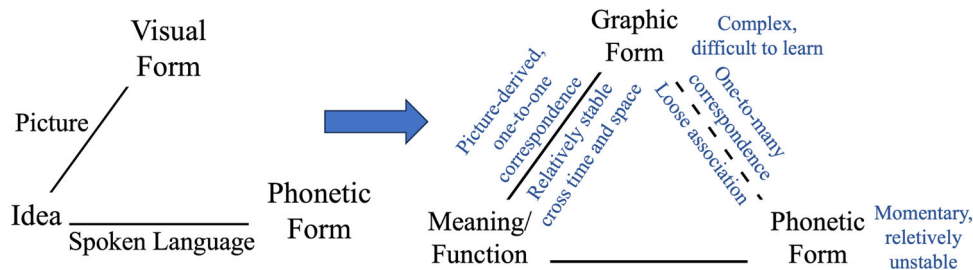
Fig. 1 The emergence of ideographic writing. The emergence of ideographic writing pertains to the development of an association between the visual forms of pictures and the phonetic forms of the spoken language.

lexical borrowing can only be attributed to an implicit xenophobic purism, incidental to the archaizing purism that is intrinsic to the Chinese language, and the archaizing purism is deeply rooted in the writing-based definition of "Chinese".

Chinese characters started from the oracle bone script. Similar to the Sumerian cuneiform script and Egyptian hieroglyphs, the oracle bone script is derived from drawing. In the beginning, pictures served as a visual expression of humans' ideas in a form to a great extent independent of speech, which expressed ideas in an auditory form (Gelb, 1952/1963: 11). When pictorial signs developed an association with the spoken language, writing emerged. However, the initial association between the earliest writing systems and the corresponding spoken languages was rather loose: pictograms and ideograms can fit divergent phonological systems, as shown below in Fig. 1.

Phonetic forms are momentary by nature, while graphic forms are stable across time and space. Before audio recording was made possible, speech was limited by time and space, thus relatively unstable. Therefore, in a picture-derived ideographic writing system, the association between the graphic form and the meaning (function) is much more stable, as recorded in texts, than the graphic-phonetic association. Owing largely to the isolated geographical environment of China, the contact between China and other civilizations was relatively limited. The ideographic writing has been evolving uninterruptedly until today, and it has always been a task of utmost importance for rulers to keep the writing unified in China. Over two millennia across the vast area of China, it is the unified written language that maintained the cultural commonality of China (cf. Hucker, 1975: 9; Norman, 1988: 2; Tu, 1994: 3-4), whereas the spoken language could not possibly achieve this goal. Meanwhile, as the ideographic writing system does not provide an accurate record of the phonetic forms, even if it was pushed to far-reaching areas to maintain the common identity, the diverse phonological systems in those areas were barely affected, and the dramatic dialectal variations mentioned in Section "Discussion" is hereby accounted for[3]. Evidently, the common identity of "Chinese" arose from the common writing, instead of the spoken language, which apparently lacked a common phonological system. The notion of the Chinese language is writing-based in nature.

The ideographic nature of Chinese characters imbues the Chinese language with a pervasive archaizing purism, which entails an implicit xenophobic purism. As the key to maintaining the common status, the use of ideographic characters is expected to conform to the stable graphic-semantic association as recorded in past literature. On the contrary, the novel use of characters that interferes with their past usage has little chance to survive. When Chinese assimilates loanwords, the meanings of pre-existing characters can be extended and compound with each other to accommodate foreign concepts, but new meanings can hardly be imposed on pre-existing characters based solely on the phonetic association as the association between ideograph and

pronunciation is never stable. Besides, archaizing purism is also inextricably intertwined with elitist purism. In general, ideographic writing systems are much more difficult to learn than phonetic writing systems (including syllabaries, abjads, abugidas, and alphabets) given the immense number of complex signs (Istrin, 1987/2002: 235). Up until the 20th century, reading and writing had been exclusive to a small group of literati elites, while the large number of laypeople were firmly bound to the land, communicating in their local accents with no access to literacy education, and were thus barely involved in the construction of the common identity. Even if people living in the border areas did use some loanwords in daily conversions, these words had little chance to reach other places, let alone to affect the lexicon of written Chinese recorded by ideographic characters.

Historically, Chinese characters also served as the major writing system in Japan, Korea, and Vietnam, forming the so-called Sinosphere (Matisoff, 1990) in which written Chinese functioned as a lingua franca. A large number of Chinese words were loaned together with the writing system of Chinese characters. In the study about loanwords in Vietnamese, Alves (2009) specifically pointed out that most Chinese words entered Vietnamese via written transmission: "a majority of Chinese vocabulary entered Vietnamese without the presence of a large bilingual community", reflecting the cross-regional stability of the ideographic writing system and the pervasive archaizing purism. During the time when Chinese characters were used as the primary writing system, these languages showed similar resistance to lexical borrowing as Chinese: transliterations from non-Sinosphere languages were disfavored. Taking Japanese as an example, Chinese characters were borrowed mainly for the graphic-semantic correspondence, but not as a representation of the phonetic form. For example, ōne is a native Japanese word meaning 'Japanese radish', consisting of two morphemes, ō 'big' and ne 'root'. The characters used to note this word is 大根, literally denoting 'big root', but the pronunciations of these two characters are dàgēn in China (dajHkon in MC), not even close to ōne (Schmidt, 2009). Admittedly, there used to be a phenomenon called ateji (当て字 'directed characters', cf. Tajima, 1998: 452–461) in Japanese, where Chinese characters were chosen based on sound instead of meaning, but just like transliterations in Chinese, many loanwords represented by ateji have fallen out of use. After the Meiji Restoration, katakana (a syllabary) began to be widely used for transliterations, driving ateji out of use. In the meantime, however, to translate modern concepts from the West, new words are formed by compounding Sino-Japanese morphemes (represented by kanji), including those reborrowed to China, e.g., 科学 kagaku 'sort + to study = science', 革命 kakumei 'to remove + life = revolution', 自由 jiyū 'self + to allow = freedom'. A functional division of writing systems can be clearly observed from the case of Japanese: katakana (a syllabary) is used for phonetic representation, while kanji (an ideographic writing system) is used for its graphic-semantic association. Evidently, across Sinosphere,

Table 6 Contrast of language ideologies based on writing.	
Ideographic writing	**Alphabetic writing**
● Common identity arises from writing, while the spoken language plays no role. ● Ideographic characters stabilize the basic semantic units of the language. ● The graphic-semantic association as recorded in past literature is always the standard for the novel use of characters.	● Common identity arises from language, while writing is secondary. ● Writing ought to serve as an accurate representation of speech. ● Letters are used merely to represent phonemes, free from the consideration of meanings.

each Chinese character has relatively stable meanings/functions, represents a smallest meaningful unit of the language that can compound with others, although it may have widely divergent pronunciations.

Conclusive remarks

In the past two millennia, the Chinese language borrowed lexical items from typologically remote languages, but a great number of the borrowed items are not retained. In particular, transliterations using pre-existing characters are virtually always disfavored when other variants are available using specifically-created characters or formed by native morphemes. In consequence, when loanwords are defined as unanalyzable, Chinese becomes notoriously resistant to lexical borrowing. It is suggested that this resistance reflects an implicit purist ideology predetermined by the employment of ideographic Chinese characters. Each character in an ideographic writing system is meaningful without recourse to the spoken language by definition—the association between the graphic form and the meaning is stable as documented in previous literature, while the association between the graphic form and the phonetic form has always been loose. Novel use of characters is expected to be consistent with the established graphic-semantic association, and conversely, the use that interferes with this graphic-semantic correspondence is likely to be filtered, when the characters serve as a mere representation of the phonetic form. This mechanism is observed not only in Chinese, but also in other languages within Sinosphere where Chinese characters were used. Thereby, ideographic writing inherently has a purification effect on the assimilation of loanwords.

While shedding light on the effect of writing on language ideology, the present finding is in sharp contrast with the traditional view that writing is secondary to the spoken language (Saussure, 1916/1959: 23). However, underlying the traditional view is precisely the implicit ideology predetermined by the nature of alphabets. Letters in alphabetic writing do not have semantic values themselves. As a mere representation of a phoneme, a letter can be used to represent all kinds of meanings with the same sound. Accordingly, alphabetic writing is expected to represent the spoken language accurately, which explains the reason why many languages had the orthographic competitions of the etymological approach versus the phonemic approach (cf. Woolard, 1994; Brown, 1993; Hellinger, 1986; Schieffelin and Doucet, 1994); such competitions cannot possibly happen in China as ideographic Chinese characters never serve as accurate representations of the pronunciations, and dialectal variations have always been dramatic. Besides, it is also related to the nature of alphabets that language is typically envisaged as a cornerstone of national identity in the European context. As previously mentioned, the common identity of Chinese exists on the basis of the common writing; but for Europeans, as writing exists for the sole purpose of representing the phonetic form of the spoken language, the credit naturally goes to the spoken language. The implicit language ideologies predetermined by ideographic versus alphabetic writing are contrasted below in Table 6.

It must be pointed out that the present discussion pertaining to the relationship between writing and language ideology is far from conclusive. In the first place, the present study focuses only on Chinese, which uses an ideographic writing system, discussions are called for on a larger scale incorporating all kinds of writing systems, e.g., syllabaries, abjads, abugidas, and alphabets. Moreover, it is mentioned in Section "Ideographic writing and linguistic purism" that the implicit linguistic purism was pervasive in Sinosphere when written Chinese functioned as a lingua franca in this area, but Chinese characters are no longer used in Korea and Vietnam now. It is of interest to see whether the implicit linguistic purism predetermined by ideographic writing still has effects in those places. As the present study hypothesizes a relationship between the type of writing and language ideology, those languages that switched writing systems are particularly worth our attention. Besides, the effect of writing does not preclude the roles of other factors in the assimilation of loanwords. As previously mentioned, within the areas where Chinese characters are used, more transliterations are used in Hong Kong and Taiwan than in the mainland. This difference is apparently related to other cultural, social, and political factors, and the interplay of all those factors awaits further investigation. In the last place, the present study looked only at the implicit linguistic purism through language use, while explicit opinions are not investigated. However, among those societies using alphabets, there are wildly divergent policies and opinions regarding linguistic purism: there are cases in which governments or national language academies explicitly advocating the purity of the national language, represented by France; there are also cases in which governments refrain from interfering with the development of language, represented by English-speaking countries (Li, 2004). The interaction between the implicit and the explicit language ideologies is undoubtedly worth exploring.

[1] In this paper, all Chinese words are immediately followed by their Pinyin annotations based on the official standard of Modern Mandarin published by the Ministry of Education of China (available at http://www.moe.gov.cn/jyb_sjzl/ziliao/). As the pronunciations of characters have been changing over time, the reconstructed pronunciations in Middle Chinese (MC) based on Baxter and Sagart (2014) are also provided when necessary.

[2] There are several systems of Romanization for Cantonese. The annotations presented in this paper are based on the de-facto standard of Jyutping (粤拼).

[3] There were a few attempts to Romanize Chinese writing in the 20th century, but none of them managed to replace the ideographic characters due mainly to the opposition of conservative intellectuals and the difficulty in promoting the phonological standard. For people who cannot speak Standard Mandarin, it is near impossible to master the orthography (spelling of words) based on Standard Mandarin. The promotion of a Chinese alphabet must be built upon the common use of a standard phonological system, which turned out to be more challenging than the promotion of the character-based literacy. According to the official report published by the Ministry of Education of the People's Republic of China (2022), the illiteracy rate had been reduced to 2.67%, while there were still 19.28% of people who could not communicate in Mandarin.

References

Alves MJ (2009) Loanwords in Vietnamese. In: Haspelmath M, Tadmor U (eds) Loanwords in the world's languages: A comparative handbook. De Gruyter Mouton, Berlin, p 617–637

Bauer RS (2006) The stratification of English loanwords in Cantonese. J Chin Linguist 34(2):172–191

Baxter WH, Sagart L (2014) Old Chinese: A new reconstruction. Oxford University Press, New York

Brown B (1993) The social consequences of writing Louisiana French. Lang Soc 22:67–101

Cook A (2018) A typology of lexical borrowing in Modern Standard Chinese. Ling Sin 4(6):1–32

Feng T (2004) 新语探源——中西日文化互动与近代汉字术语生成 (Exloring the origin of neologism: Cultural nteractions between China, the West and Japan, and the formation of Modern Chinese terminology). Zhonghua Book Company, Beijing

Gelb IJ (1952) A study of writing (revised edition). The University of Chicago Press, Chicago & London, 1963

Goldberg AE (1995) A construction grammar approach to argument structure. University of Chicago Press, Chicago

Haspelmath M (2009) Lexical borrowing: Concepts and issues. In: Haspelmath M, Tadmor U (eds) Loanwords in the world's languages: A comparative handbook. De Gruyter Mouton, Berlin, p 35–54

Haspelmath M, Tadmor U (2009) Chapter 1: The loanword typology project and the world loanword database. In: Haspelmath M, Tadmor U (eds) Loanwords in the world's languages: A comparative handbook. De Gruyter Mouton, Berlin, p 1–34

Hellinger M (1986) On writing English-related Creoles in the Caribbean. In: Görlach M, Holm J (eds) Focus on the Caribbean. John Benjamins, Amsterdam, p 53–70

Hsieh SC, Hsu H (2006) Japan mania and Japanese loanwords in Taiwan Mandarin. J Chin Linguist 34(1):44–79

Hucker C (1975) China's imperial past: An introduction to Chinese history and culture. Stanford University Press, Stanford

Istrin VA (1987) 文字的产生和发展 (The origin and development of writing), translated into Chinese by Shaoxing Zuo. Peking University Press, Beijing, 2002

Kaufman AA (2010) The "century of humiliation," then and now: Chinese perceptions of the international order. Pac Focus 25(1):1–33

Kratochvil P (1982) Modern Chinese and linguistic change. China Q 92:687–695

Langacker RW (2008) Cognitive grammar: A basic introduction. Oxford University Press, New York

Li W-C (2004) Conflicting notions of language purity: the interplay of archaising, ethnographic, reformist, elitist and xenophobic purism in the perception of Standard Chinese. Lang Commun 24:93–133

Mao Z (1940/1991) 新民主主义论 (On New Democracy). In: Editorial Committee on Party Literature of the Central Committee (ed) 毛泽东选集 (Selected Works of Mao Zedong) Vol. 2. People's Publishing House, Beijing, pp. 680-685

Markovic M (1984) The language of ideology. Synthese 1:69–88

Matisoff JA (1990) On megalocomparison. Language 66(1):106–120

Ministry of Education of the People's Republic of China (2022) 全国普通话普及率80.72%，文盲率下降至2.67% (Mandarin speaking rate hits 80.72% nationwide, while the illiteracy rate has been reduced to 2.67%). http://www.moe.gov.cn/fbh/live/2022/54618/mtbd/202206/t20220628_641478.html. Accessed 15 Dec 2023

Norman J (1988) Chinese. Cambridge University Press, Cambridge

Norman J, Mei T-L (1976) The Austroasiatics in Ancient South China: Some Lexical Evidence. Monum Serica 32:274–301

Pan W, Yip P-C, Han Y (1993) 汉语的构词法研究 (Studies of Chinese word-formation). Student Book Co., Taipei

Saussure FD (1916) Course in general linguistics, translated into English by W. Baskin. The Philosophical Library, New York, 1959

Schieffelin BB, Doucet RC (1994) The "real" Haitian Creole: Ideology, metalinguistics, and orthographic choice. Am Ethnologist 21(1):176–200

Schmidt C (2009) Loanwords in Japanese. In: Haspelmath M, Tadmor U (eds) Loanwords in the world's languages: A comparative handbook. De Gruyter Mouton, Berlin, p 545–574

Schuessler A (2003) Multiple origins of the Old Chinese lexicon. J Chin Linguist 31(1):1–35

Schuessler A (2007) ABC Etymological Dictionary of Old Chinese. University of Hawai'i Press, Honolulu

Shao J (2000) 香港方言外来词比较研究 (A comparative study of loanwords in Hong Kong Cantonese). 语言文字应用 (Appl Linguist) 3:3–12

Silverstein M (1979) Language structure and linguistic ideology. In: Clyne PR, Hanks WF, Hofbauer CL (eds) The elements: A parasession on linguistic units and levels. Chicago Linguistic Society, Chicago, p 193–247

T'sou B, You R (2003) 当代汉语新词的多元化趋向和地区竞争 (The diversification and areal competitiveness of new words in contemporary Chinese). 语言教学与研究 (Lang Teach Res) 2:12–21

Tajima M (1998) 近代漢字表記語の研究 (Study of notation of Chinese characters in Modern Japanese). Izumi, Osaka

Thomas G (1991) Linguistic purism. Longman, London

Tu W-M (1994) The living tree: The changing meaning of Beijing Chinese today. Stanford University Press, Stanford

Wang L (1944) 中国语法理论 (Theories of Chinese grammar). Zhonghua Book Company, Beijing, 1984

Wang R (2021) 现代汉语外来词研究之研究 (On studies of loanwords in Modern Chinese). 语言科学 (Linguistic Sci) 20(4):402–410

Wiebusch T (2009) Mandarin Chinese vocabulary. In: Haspelmath M, Tadmor U (eds) World loanword database. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2130 entries (Available online at http://wold.clld.org/vocabulary/22, Accessed on 2021-11-30)

Wiebusch T, Tadmor U (2009) Loanwords in Mandarin Chinese. In: Haspelmath M, Tadmor U (eds) Loanwords in the world's languages: A comparative handbook. De Gruyter Mouton, Berlin, p 575–598

Woolard KA (1994) Language ideology. Annu Rev Anthropol 23:55–82

Zhang D (1999) 第三次浪潮——外来词引进和规范刍议 (The third wave: the assimilation and standardization of loanwords). 语言文字应用 (Appl Linguist) 3:70–76

Zhang T (2017) 基于"HSK常用词汇"的现代汉语外来词统计研究 (A statistic research of loan words in Modern Chinese based on the vocabulary of HSK words in common use). 河北大学学报 (J Hebei Univ) 42(4):14–24

Competing interests

The author declares no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-023-02556-3.

Correspondence and requests for materials should be addressed to Liulin Zhang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.