

Orthography-based prescriptivism: The divergence of Sino-Tibetan grammar

Liulin Zhang

liulinz@suda.edu.cn

Soochow University

Tszen Cizhen

Lasa Normal University

Article

Keywords: Sino-Tibetan, writing system, linguistic prescription, morphological evolution

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10560491/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Orthography-based prescriptivism:

The divergence of Sino-Tibetan grammar

As two major Sino-Tibetan languages, Chinese and Tibetan went on divergent evolution paths due to the distinct writing systems adopted. The ideographic Chinese characters stabilize monosyllabic morphemes for the Chinese language across time and space, but unable to represent sub-syllabic linguistic forms. Consequently, sub-syllabic morphological processes, including affixation and tonal alternation, were eliminated, giving rise to compounding. For example, the derivative **kraŋ*, composed of the root **kra* (*gra*) and the suffix **-ŋ* in Proto-Sino-Tibetan, was fused into one morpheme in Chinese when represented by a single character. In contrast, the Tibetan abugida is able to represent every single phoneme in a syllable, and accordingly, the sub-syllabic affixes in Proto-Sino-Tibetan managed to be stabilized. Evidently, writing system determines the granularity of linguistic prescription: the linguistic unit represented by a single sign is always the smallest linguistic unit that can possibly be standardized in the language.

Keywords: Sino-Tibetan; writing system; linguistic prescription; morphological evolution

1. Introduction

A fundamental assumption of modern linguistics claims that the written language exists for the sole purpose of representing the spoken language (Saussure, 1916/1959: 23), readily creating an impression that writing is always secondary to speaking. This mindset blinds people to the prescriptive effects that orthography wields on language evolution. Focusing on the divergence of Sino-Tibetan grammar, the present study brings to light those morphological features determined by orthography. Sino-Tibetan has been recognized as a language family based on systematic studies of cognates and phonological correspondences, but the dramatic morphological differences

among Sino-Tibetan languages remain inadequately explained. It is still unknown why genetically related languages exhibit such wildly divergent grammars. Taking examples of the derivatives of cognates, the present study argues that Chinese and Tibetan went on divergent paths because of the distinct writing systems that stabilize different linguistic units. In particular, the Chinese language eliminated all kinds of sub-syllabic morphemes that cannot be represented by Chinese characters, while the Tibetan segmental writing system facilitates and preservation and development of morphology. Section 2 provides a brief introduction to Sino-Tibetan languages and writing systems. Section 3 and Section 4 compare the derivational paths of the cognates **kyu* ‘hook’ and **gra* ‘large’ in Chinese and Tibetan respectively. The morphological evolution of Chinese is reconstructed in Section 5, to show how sub-syllabic morphemes were completely lost. Section 6 discusses the explicit and implicit prescriptive effects of orthography on language, while Section 7 concludes the present study and points out future directions.

2. Sino-Tibetan languages and writing systems

2.1 History of Chinese language and writing

The earliest universally recognized written records of the Chinese language are the oracle bone script dating back approximately to the 15th century BCE. Those inscriptions carved on animal bones and turtle shells were primarily used for divination. Meanwhile, there were also inscriptions on bronze vessels indicating the possessors and sometimes the relevant historical backgrounds. Ideographic nature can easily be identified from those inscriptions, as shown below in Table 1.

Table 1. Examples of ideographic signs in Chinese writing

Type	Oracle bone script	Modern Chinese character (simplified)	Meaning
Pictogram		日	sun
		月	moon
Iconic Ideogram		上	up
		下	down
Ideographic compound		明	(sun + moon) tomorrow; bright
		休	(person + wood) to rest

When a word cannot be visualized directly, a rebus might be used that employs the ideographic sign of a homonym to represent the pronunciation. For example, 𪛗 is originally the pictogram for the word **lǎg*¹ ‘wheat’, and is also used to represent the homophonic word **lǎg* ‘to come’. A general principle of such rebuses requires syllable-to-syllable correspondence: each ideographic sign is pronounced as one syllable, and as a rebus, it could only be used to represent another monosyllabic homonym. With the rebus use, many ideographic signs became multifunctional. Moreover, some rebuses compound with other ideographic signs, forming ideo-phonetic compounds. For example, the compound 𪛗 **kǐǎ* ‘wolfberry’ is composed of the pictogram 𪛗 **mók* ‘wood’ in the left as the semantic radical, and the rebus 𪛗 **kǐǎ* ‘self’ as the phonetic radical in the right. This type of compounding was particularly common in the later history of Chinese writing, especially when exploding multifunctional rebuses made reading increasingly difficult as writing was no longer limited to divination and indicating possessors. In consequence, the basic logic of ideographic writing is reinforced – virtually all Chinese characters have their own conventionalized meanings² – which crucially distinguishes Chinese writing from syllabaries, abjads, abugidas, and alphabets.

64

65 The (relatively) stable association between written signs and conventionalized meanings are key
66 to maintaining the common identity of Chinese. During the social turmoil of the Spring and
67 Autumn period (770-476 BCE) and the Warring States (475-221 BCE), Chinese writing
68 developed many regional varieties. When the first emperor of Qin (秦始皇 *Qín Shǐ Huáng*, 259-
69 210 BCE) conquered all other states and founded the first imperial dynasty in the Chinese
70 history, unifying writing is one of the first and foremost measures to centralize power, which set
71 an example for successive rulers. Rulers of later dynasties invariably attached the utmost
72 importance to the unified written language, which in turn fostered the flourishing of philology
73 concerned with the graphic forms, meanings, and pronunciations of characters. The ideographic
74 writing system was thus pushed to far-reaching areas with wildly divergent accents. In tandem
75 with the uneven phonological changes in different areas, the variations of characters'
76 pronunciations might become so dramatic that people could hardly tell those characters sharing
77 radicals were originally homonymic rebuses. Evidently, the national identity arose from the
78 common writing instead of the spoken language. The ideographic writing system has been used
79 uninterrupted until today, albeit with evolving fonts and writing materials, while the spoken
80 language had been constantly varying and changing until the 20th century when phonological
81 standardization was finally made possible by modern technology and compulsory education. The
82 pattern of “written language-national identity-spoken language” makes Chinese radically
83 different from other languages defined on the basis of mutual intelligibility (Hucker, 1975: 9;
84 Norman, 1988: 187; Zhang, 2024b, c).

85

86 2.2 History of Tibetan language and writing

87 The Tibetan writing system emerged in the 7th century CE. When the empire was newly founded
88 upon the unification of the Tibetan Plateau, the minister Thönmi Sambhota (ཐོན་མི་སངས་བུ་ཁྱེད་, 618-?) was
89 sent to India to study the art of writing. Drawing inspiration from a northern variety of Indian
90 Brahmic scripts, he devised the standard writing system for the Tibetan language. Inheriting the
91 common features of abugidas, the new writing system consisted of thirty consonantal letters
92 inherently taking a default /a/, and diacritic marks are added to change the inherent vowel to a
93 different one. For example, the letter ཀ represents *ka* by default; with different vowel marks it can
94 also represent *ki* (ཀི), *ku* (ཀུ), *ke* (ཀེ), or *ko* (ཀོ). Consonantal letters can compound with each other to
95 represent consonant clusters within one syllable. In the most complex case, a syllable can be
96 broken down into seven parts in writing. For example, in the syllable བསྐྱེད་ *bsgrigs* ‘arranged’, the
97 base letter is ཀ (ga); above the base letter is the superscript ས (sa); above the superscript is vowel
98 mark for /i/; below the base letter is the subscript representing /r/, and thus སྐྱེད་ represents /sgri/;
99 before the base letter is the prescript བ (ba); after the base letter is the postscript ཀ (ga), and the post-
100 postscript ས (sa). This structure allows Tibetan writing to precisely record three-consonant clusters
101 in the beginning of syllables, and two-consonant clusters in the end of syllables. It is noteworthy
102 that not all syllables are so complex: the base letter alone can also represent a syllable. Syllables
103 are separated by dots, such as the disyllabic word བོད་ཡིག་ *bod yig* ‘Tibetan writing’.

104

105 Going with the writing system, Thönmi Sambhota also composed grammatical treatises, including
106 *Thirty Verses* (ཐུག་མཁའ་) and *The Application of Signs* (རྟགས་ཀྱི་འཇུག་པ་). Resembling the classical Sanskrit
107 grammar book *Aṣṭādhyāyī* (अष्टाध्यायी, literally means ‘Eight Chapters’) of Pāṇini (पाणिनि, about 5th
108 century BCE), *Thirty Verses* and *The Application of Signs* also use concise sutras to standardize

the grammar and orthography of the Tibetan language, including different types of morphological operations centered on verbs. With the segmental writing system, morphology is essentially closely related to orthography: derivational and inflectional affixes are all clearly represented in writing. For example, the suffix *-s*, represented by the postscript (or post-postscript) ས in writing, is normally used to mark past tense and the imperative mood, as shown in the following examples:

(1)	Meaning	Present tense	Future tense	Past tense	Imperative mood
	‘to eat’	ཟ་ <i>za</i>	བཟའ་ <i>bzav</i>	བཟས་ <i>bzas</i>	ཟོས་ <i>zos</i>
	‘to drink’	འཐུང་ <i>vt hung</i>	བཏུང་ <i>btung</i>	བཏུངས་ <i>btungs</i>	ཐུངས་ <i>thungs</i>
	‘to wear’	གོན་ <i>gon</i>	བགོན་ <i>bgon</i>	བགོས་ <i>bgos</i>	གོས་ <i>gos</i>
	‘to do’	བྱེད་ <i>byed</i>	བྱ་ <i>bya</i>	བྱས་ <i>byas</i>	བྱོས་ <i>byos</i>
	‘to speak’	ཟླ་ <i>smra</i>	ཟླ་ <i>smra</i>	ཟླས་ <i>smras</i>	ཟློས་ <i>smros</i>

The Tibetan empire collapsed in the 9th century after more than two hundred years of glory, plunging the entire plateau area into prolonged warfare. In 1247, Sakya Paṇḍita (ས་སྐུ་པ་འཛིན་, 1182-1251), the leader of a major Buddhism school in the Tibetan area, proclaimed submission to the Mongols. Subsequently, the Mongol forces swept across East Asia and established the Yuan dynasty, and thus the Tibetan region naturally became part of the Yuan realm. Thereafter, whenever the central regime changed in China, the Tibetan leaders would express their allegiance to the central authority. Meanwhile, due to the lack of a strong centralized authority to enforce linguistic norms, and the prevailing emphasis on the classical written language over spoken vernacular among monasteries, the spoken and written languages gradually diverged, giving rise to significant dialectal differences.

2.3 Studies on the Sino-Tibetan language family

The difference between Chinese writing and Tibetan writing is obvious. As a self-originated writing system, Chinese characters are derived from pictures, inherently meaningful in their own

right. Whereas the Tibetan writing system is an abugida adapted from an Indian script, in which letters and diacritic marks represent phonemes, independent of semantic value. Confronted with such obvious difference, coupled with the mutually unintelligible accents, scholars would by no means realize the affiliation between Chinese and Tibetan, until the 19th century when historical linguistics developed scientific methods to reconstruct the genetic relationships among languages.

Besides Tibetan, the Burmese writing system, first attested in the 11th century, is also derived from Indian Brahmic scripts, and thus shares the features of abugidas. In the mid 19th century, European scholars unveiled the lexical parallels between Tibetan and Burmese, and proposed the term “Tibeto-Burman” (Dettamajumdar, 2017). Later, the possible relationship between Sanskrit, Tibetan, and Chinese caught linguists’ attention, leading to the hypothesis of “Indo-Chinese” (Conrady, 1896). The term “Sino-Tibetan” was first coined in French, i.e., sino-tibétain, by Jean Przyluski (1924), and later introduced into English. From then on, linguists worldwide started systematic studies concerning the phonological correspondences and cognates among Sino-Tibetan languages (e.g., Benedict, 1972; 1976; Li, 1937/1973, 1939; Shafter, 1955, 1957, 1963). Recent phylogenetic studies employing computational approaches suggest that the Proto-Sino-Tibetan language originated around 7,000 to 8,000 years ago among millet farmers in North China (Sagart et al., 2018; Zhang et al., 2020).

Although the genetic relationship between Chinese and Tibetan has already been widely accepted, the morphological differences between Chinese and Tibetan is far from negligible: Chinese is known as an analytic language impoverished in morphology, while Tibetan is highly

agglutinative with rich fusional elements. To my knowledge, no satisfactory explanation has been provided to account for this dramatic difference. In fact, a few Chinese linguists tend to take the morphological differences as counterexamples to the proposed genetic relationship (Sun, 1998; Sun, 2007a, b; Zhang, 2013: 213-220). In an attempt to reconstruct the divergent derivational paths of cognates in Chinese and Tibetan, the present study discusses why genetically related languages could end up with dramatic typological differences, shedding light on the explicit and implicit prescriptive effects of orthography on language evolution. It is to be demonstrated that sub-syllabic morphology never got a chance to be stabilized in Chinese as ideographic characters cannot provide fine-grained representations of phonetic forms, while the Tibetan abugida allows every phoneme to be clearly represented, thus facilitating the stabilization and development of sub-syllabic morphology.

3. Derivatives of the toot **kyu* ‘hook’ in Chinese and Tibetan

3.1 Derivatives of the root **kyu* ‘hook’ in Chinese

In Chinese, the root **kyu* ‘hook’ was first represented by the character 句 **kyu* > MC *kəu/gju* > *gōu/jù*³, and later derived the characters 鉤 **kyu* > MC *kəu* > 钩 *gōu* and 勾 MC *kəu* > *gōu*.

As an ideo-phono compound, the character 句 **kyu* ‘hook’ is captured in the inscriptions on bronze hooks dated to the Western Zhou dynasty (1046-771 BCE), with the phonetic radical 句 **khu* lying inside the pictogram of a hook. Based on its original meaning, this character derived the verbal and adjectival use “to hook, crooked”, as shown in the following examples:

(2)	越	子	为	左	右	句	卒。
	<i>Yuè</i>	<i>zǐ</i>	<i>wéi</i>	<i>zuǒ</i>	<i>yòu</i>	<i>gōu</i>	<i>zú.</i>
	Yue	Lord	deploy	left	right	hooked	troop

‘The Lord of Yue deployed the “left and right hooked troops”.’

173 (Zuo’s Commentary on “Spring and Autumn Annals”)

(3) 句 者 毕 出, 萌 者 尽 达。

Gōu zhě bì chū, méng zhě jìn dá.

bent one all out burgeoning one all come

‘The bent sprouts all bust forth, and the burgeoning ones all come through.’

174 (Book of Rites)

175

176 No later than the 6th century BCE, the character 句**kyu* began to compound with the semantic

177 radical 金‘metal’, forming the ideo-phono compound 鉤**kyu*, also denoting ‘hook’, without the

178 derived verbal and adjectival uses, as shown in the following examples:

(4) 彼 窃 鉤 者 诛, 窃 国 者 为 诸 侯。

Bǐ qiè gōu zhě zhū, qiè guó zhě wéi zhū hóu.

those steal **hook** one die steal state one become various lord

‘Those who steal hooks are put to death, but those who steal states become feudal lords.’

179 (Zhuangzi)

(5) 以 尔 鉤 援, 与 尔 临 冲, 以 伐 崇 墉。

Yǐ ěr gōu yuán, yǔ ěr lín chōng, yǐ fá chóng yōng.

use your **hook** ladder and your siege batter to attack Chong walls

‘Take your hooked scaling ladders and your siege towers with battering rams, and assault the lofty walls of Chong.’

180 (The Classic of Poetry)

181

182 During the Han dynasty (206 BCE – 220 CE), the 𠂔 in the character 句**kyu* > *kəu* began to be

183 written as 𠂔: the character 勾 *kəu* appeared as a variant of 句 *kəu*. Meanwhile, the hook symbol

184 was sometimes used as a punctuation to indicate the ending of a sentence, and thus the word 句

185 *kəu* ‘hook’ derived the meaning ‘sentence’ through a metonymy. The meanings and pronunciations

186 of 句 *kəu* and 勾 *kəu* gradually diverged afterwards: the younger variant 勾 *kəu* > *gōu* took over

187 the verbal sense ‘to hook’, while the original character 句 *kəu* became specific for ‘sentence’ and

188 developed the pronunciation *gju* > *jù*. Besides, The character 鉤 *kəu* > *gōu* is preserved as a noun,

189 and later simplified as 钩 *gōu*.

190

3.2 Derivatives of the root *kyu ‘hook’ in Tibetan

In Tibetan, the root ལྷ ལྷ ལྷ *kyu* ‘hook’ derived words including འཇུ་བ་ *vkhyu ba* ‘crooked, not straight’, and དལྷ་བ་ *dkyu bo* ‘crooked, skewed’.

The monosyllabic morpheme *-ba* is a productive suffix commonly used to derive adjectives. Other examples include དཀར་བ་ *dkar ba* ‘white’, ཆེ་བ་ *che ba* ‘big’, ཉེ་བ་ *nye ba* ‘close, near’, ཚ་བ་ *tsha ba* ‘hot’, རྩུང་བ་ *chung ba* ‘small’, དཀའ་བ་ *dkav ba* ‘difficult’, etc. Compared to *-ba*, the monosyllabic morpheme *-bo* is more frequently used to derive nouns, e.g., ལྷ ལྷ *chu* ‘water’ - ལྷ ལྷ *chu bo* ‘river’, མི་ *mi* ‘human’ - མི་ *mi bo* ‘male’, ལྷ ལྷ *khyi* ‘dog’ - ལྷ ལྷ *khyi bo* ‘puppy’, etc. Albeit uncommon, there are indeed a few cases in which *-bo* is used to derive adjectives, such as དགའ་ *dga* ‘to rejoice’ - དགའ་བ་ *dga bo* ‘happy’.

The sub-syllabic morphemes *v-* and *d-* are indeed productive prefixes in Tibetan, but their associations with འཇུ་བ་ *vkhyu ba* ‘crooked, not straight’ and དལྷ་བ་ *dkyu bo* ‘crooked, skewed’ are not transparent. The prefixes *v-* and *d-* are primarily related to tense, as shown in the following examples:

(6) Meaning	Present tense	Future tense	Past tense
‘to roof, to cover’	འབུབས་ <i>vbubs</i>	དབུབ་ <i>dbub</i>	ཕུབ་ <i>phub</i>
‘to rob, to deprive of’	འཛོག་ <i>vphrog</i>	དཛོག་ <i>dbrog</i>	འཛོགས་ <i>vphrogs</i>
‘to release, to liberate’	འགོལ་ <i>vgrol</i>	དགོལ་ <i>dgrol</i>	བགོལ་ <i>bkrol</i>
‘to descend, to fall down’	འབོ་ <i>vbo</i>	དབོ་ <i>dbo</i>	བོ་ <i>bos</i>

Nevertheless, the meanings of འཇུ་བ་ *vkhyu* ‘to gather, to run’ and དལྷ་བ་ *dkyu* ‘to run’ do not seem related to ‘crooked, skewed’. We can only speculate that the verbal sense ‘to crook’ disappeared (or switched) after འཇུ་བ་ *vkhyu* and དལྷ་བ་ *dkyu* affixed with *ba* and *bo*.

3.3 Comparison of the derivational paths

The derivational paths of **kyu* ‘hook’ in Chinese and Tibetan can roughly be represented below in Figure 1.

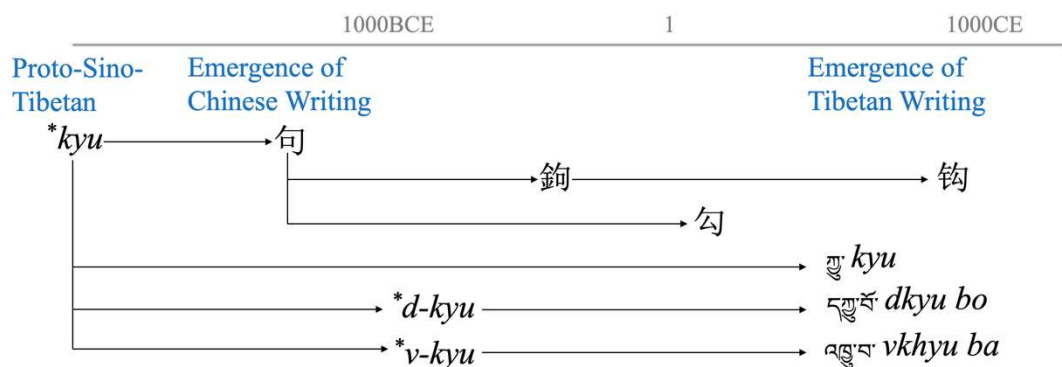


Figure 1. Derivatives of the root **kyu* ‘hook’ in Chinese and Tibetan

It is worth mentioning that 句**kyu* could be used as a noun, a verb, or an adjective in Old Chinese, as shown in examples (2) and (3). Whereas in Tibetan, different morphemes are used to derive verbs and adjectives. Furthermore, monosyllabic morphemes are stabilized by characters in Chinese, without any analyzability. The nominal sense of ‘sentence’ was originally a meaning extension from ‘hook’, but as the character 勾 was separated from 句, the pronunciations diverged accordingly. Consequently, 句**kyu* > *kəu* > *gju* > *jù* ‘sentence’ and 勾 *kəu* > *gōu* ‘to hook’ were stabilized as two individual morphemes, while the nominal sense ‘hook’ was taken over by 鉤 *kəu* > 钩 *gōu* specifically, which means that the linguistic unit recorded by a character, pronounced as one syllable, cannot be broken down into smaller linguistic units. In contrast, the segmental abugida of Tibetan is able to represent the analyzability of syllables, ensuring that sub-syllabic morphemes can also be stabilized, such as the prefixes *v-* and *d-*. In other words, even if the Tibetan prefixes *v-* and *d-* are passed down from Proto-Sino-Tibetan, they still do not have a

230 chance to be represented by Chinese characters, which stabilize monosyllabic morphemes for the
 231 Chinese language.

232

233 4. Derivatives of the root **gra* ‘large’ in Chinese and Tibetan

234 4.1 Derivatives of the root **gra* ‘large’ in Chinese

235 In Chinese, the root **gra* ‘large’ is represented by the character 夏 **gra* > MC *ya* > *xià* ‘large’,
 236 and its derivative 京 **kraŋ* > MC *kiaŋ* > *jīng* ‘big, capital’ further derived characters including
 237 鲸 MC *giaŋ* > *jīng* ‘whale’, 景 **kraŋ* > MC *kiaŋ* > *jǐng* ‘sunlight, landscape’, which further
 238 derived 影 MC *ʔiaŋ* > *yǐng* ‘shadow’, etc.

239

240 The character 夏 is written as  in the bronze inscriptions, like a person, originally denoting
 241 ‘large’, as shown in example (7). It is the name of the earliest Chinese dynasty mentioned in
 242 legends, also used as the earliest self-designation of the Han Chinese people, as shown in
 243 examples (8) and (9).

(7) 于 我 乎, 夏 屋 渠渠, 今 也 每 食 无 余。
Yú wǒ hū, xià wū qúqú, jīn yě měi shí wú yú.
 for me Particle **big** house spacious now Particle every meal no leftover
 ‘For me, in the past, the great houses were spacious; but now, after every meal, there is
 nothing left over.’

244 (The Classic of Poetry)

(8) 春 秋 内 其 国 而 外 诸 夏。
Chūn qiū nèi qí guó ér wài zhū Xià.
 spring autumn internal its state Conj. external various **Xia**
 ‘In the Spring and Autumn period, (people) regarded their own state as internal and
 considered the various Xia (states) as external.’

245 (Gongyang Commentary on “Spring and Autumn Annals”)

(9) 此 之 谓 夏 声。
Cǐ zhī wèi Xià shēng.
 this Particle call **Xia** voice
 ‘This is called the voice of Xia.’

246 (Zuo’s Commentary on “Spring and Autumn Annals”)

247
 248 The derivational relationship between 夏 *xià* ‘large’ and 京 *jīng* ‘big, capital’ is not manifested
 249 by characters or pronunciations in Modern Mandarin, but as long as we recognize the consonant
 250 clusters in Old Chinese, working in tandem with their closely related meanings, the derivational
 251 relationship is actually self-evident.

252
 253 Through rebuses, the character 京 **kraŋ > kian > jīng* serves as the phonetic radical for
 254 characters including 景 *kian > jǐng*, 鲸 *gian > jīng*, 惊 *kian > jīng*, 谅 *lian > liàng*, 掠 *lian >*
 255 *lüè*, etc., implying that they were all homonymic. However, their pronunciations in Modern
 256 Mandarin have a major difference in the onset: it is hard to fathom that the velar plosives [k] and
 257 [g] could be complementary with the lateral [r]/[l]. Linguists came up with an explanation that
 258 the initial pronunciation in Old Chinese contained a consonant cluster similar to the German
 259 word *gleich*, and the discrepancy arose as some derived characters inherited the [k]/[g] onset
 260 while others inherited the [l] onset (Gabelentz, 1881: 99; Grube, 1881: 16-17; Karlgren, 1915-
 261 1926). Therefore, the pronunciation of 京 *kian > jīng* in Old Chinese is reconstructed as **kraŋ*,
 262 suggesting a derivational relationship with 夏 **gra* ‘large’.

263
 264 In the pictographic form  , the character 京 **kraŋ* is frequently attested in the oracle bone
 265 script, indicating ‘big, capital’. Functioning as the phonetic radical, it compounded with the
 266 pictogram 日 **njit* ‘sun’ no later than the 1st century BCE, forming the ideo-phono compound
 267 景 **kraŋ* ‘sunlight, landscape’, which further compounded with the pictogram 彡 *fam* ‘hair,
 268 feathers’, forming the ideo-phono compound 影 *ʔian > yǐng* ‘shadow’. Meanwhile, 京 **kraŋ >*
 269 *kian* also compounded with the pictogram 鱼 *ŋa* ‘fish’, forming the ideo-phono compound 鲸

270 *gian* > *jīng* ‘whale’.

271

272 4.2 Derivatives of the root **gra* ‘large’ in Tibetan

273 In exact correspondence with 夏 **gra* > MC *ya* > *xià* ‘large’ in Chinese, the Tibetan word རྩ་ *rgya*

274 ‘large’ is also used to refer to the Han Chinese people. Its derivatives include རྩ་གྲྱ་ *rgyag* ‘to do, to

275 make’, རྩ་རྒྱལ་ *rkyoŋ* ‘to stretch, to extend’, and so forth. The derivation from རྩ་ *rgya* ‘large’ to རྩ་

276 *rgyag* ‘to do, to make’ and རྩ་རྒྱལ་ *rkyoŋ* ‘to stretch, to extend’ is fairly straightforward as the primary

277 function of -*g/ŋ* is to derive verbs in Tibetan, among other functions, as shown in the following

278 examples:

(10)	སྒྲ་ <i>sgra</i>	‘sound’	སྒྲ་གྲྱ་ <i>sgrag</i>	‘to sound, to make a noise’
	རྩ་གྲྱ་ <i>rgyu</i>	‘to cause’	རྩ་གྲྱལ་ <i>rgyug</i>	‘to run, to race’
	དུ་ <i>du</i>	‘smoke’	གདུག་ <i>gdug</i>	‘to smoke’
	འཁྲི་ <i>vkhri</i>	‘to tangle, to lust’	འཁྲིག་ <i>vkhrig</i>	‘to copulate’
	གན་ <i>gnya</i>	‘back of neck’	གནག་ <i>gnyag</i>	‘col, gap (in mountains)’
	བྲ་ <i>bkra</i>	‘to glitter’	བྲག་ <i>bkrag</i>	‘luster’

279

280 In addition, as orthography precisely differentiates /k/ from /g/ in the derivative རྩ་རྒྱལ་ *rkyoŋ* ‘to

281 stretch, to extend’, the -*ŋ* suffixation therein is presumably completely before the advent of

282 writing.

283

284 4.3 Comparison of the derivational path

285 The exact correspondence in meanings suggest that the Chinese word 夏 **gra* > MC *ya* > *xià*

286 ‘large’ and the Tibetan word རྩ་ *rgya* ‘large’ are undoubtedly cognates. Besides, as previously

287 analyzed in section 4.1, the Chinese word 京 **kraŋ* > MC *kiaŋ* > *jīng* ‘big, capital’ is a derivative

288 of the root **gra*, and there happens to be a suffix -*g/ŋ* in Tibetan. Therefore, it is reasonable to

289 believe that the suffix *-*n/-ŋ* existed in Proto-Sino-Tibetan, before the advent of Chinese writing,

and along this line, the Chinese word 京 **kraŋ* ‘big, capital’ should be cognate with the Tibetan word རྒྱ་ *rkyoŋ* ‘to stretch, to extend’. Not only the root **gra*, but also the derivative **kra-ŋ*, are inherited from Proto-Sino-Tibetan. Given this awareness, the derivational paths of **gra* ‘large’ in Chinese and Tibetan can roughly be represented below in Figure 2.

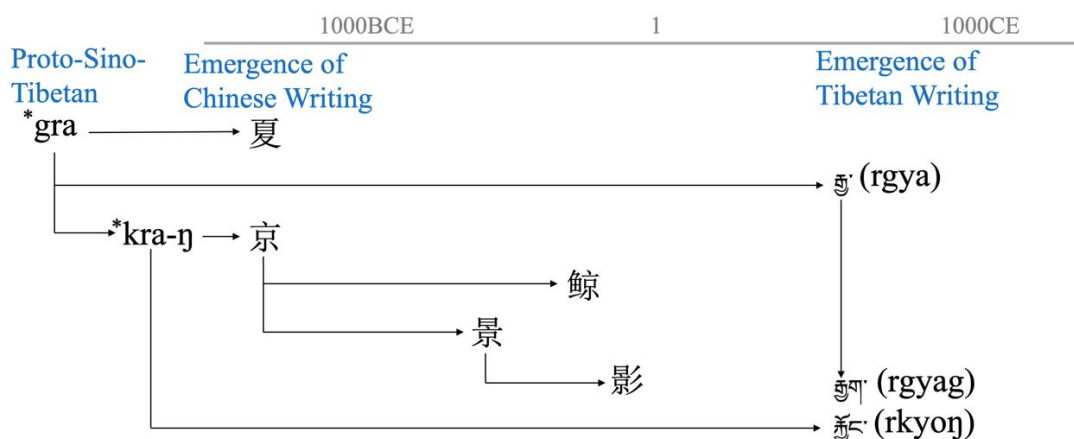


Figure 2. Derivatives of the root **gra* ‘large’ in Chinese and Tibetan

A conspicuous difference can be observed from Figure 2 between Chinese and Tibetan: the congenetic derivational suffix *-g/ŋ* is not recorded in Chinese writing, but clearly recorded in Tibetan writing. In Chinese, the monosyllabic linguistic unit represented by the Chinese character 京 **kraŋ* > MC *kiaŋ* > *jīng* ‘big, capital’ appears unanalyzable, with component morphemes fused together into one morpheme, illustrating how sub-syllabic morphemes were assimilated. Moreover, as the pronunciations of characters changed dramatically, speakers of Modern Mandarin can hardly tell 京 *jīng* is originally composed of a root and a suffix. Indeed, Chinese characters also derive, but importantly, the derivation of Chinese characters is primarily based on syllable-to-syllable homophonic relationship (see section 2.1), which cannot be taken as evidence for lexical derivation. In contrast, as the derivational suffix **-g/ŋ* is clearly represented by (post-)postscript in Tibetan writing, it is well preserved in the language. Considering the

semantic disagreement between 京 **kraŋ* > MC *kiaŋ* > *jīng* ‘big, capital’ and 𠂔 *rkyoŋ* ‘to stretch, to extend’, a conjecture can be advanced that the function of the suffix **-g/ŋ* was unstable in Proto-Sino-Tibetan, which was gradually stabilized with increased productivity by the specific written representation in Tibetan.

5. Evolution of Chinese morphology

Drawing evidence from characters sharing the same phonetic radical, it is previously analyzed that the onset of 京 **kraŋ* > MC *kiaŋ* > *jīng* ‘big, capital’ changed from **/kr/* to */k/*. Furthermore, the chronology of those characters sharing phonetic radicals allows us to infer the date when consonant clusters were lost from Chinese. Dramatic divergences in onsets are primarily observed from base characters created before the Warring States period (475 BCE), as shown below in Table 2, suggesting that cluster loss had generally completed by the 5th century BCE.

Table 2. Examples of characters sharing phonetic radicals with diverged onsets

Base character	Dating	Derived characters
京 <i>*kraŋ</i> > MC <i>kiaŋ</i> > <i>jīng</i> ‘big, capital’	No later than 10 th century BCE	惊 <i>jīng</i> , 鲸 <i>jīng</i> , 景 <i>jǐng</i> , etc.
		谅 <i>liàng</i> , 掠 <i>lüè</i> , 晾 <i>liàng</i> , etc.
各 <i>*klak</i> > MC <i>kak</i> > <i>gè</i> ‘particular’	No later than 10 th century BCE	格 <i>gé</i> , 骼 <i>gé</i> , 阁 <i>gé</i> , 膈 <i>gē</i> , etc.
		洛 <i>luò</i> , 络 <i>luò</i> , 路 <i>lù</i> , 赂 <i>lù</i> , 骆 <i>luò</i> , etc.
监 <i>*kram</i> > MC <i>kam</i> > <i>jiān</i> ‘to supervise’	No later than 10 th century BCE	檻 <i>jiàn</i> , 鉴 <i>jiàn</i> , 銓 <i>jiàn</i> , etc.
		篮 <i>lán</i> , 蓝 <i>lán</i> , 览 <i>lǎn</i> , etc.
兼 <i>*kliam</i> > MC <i>kiem</i> > <i>jiān</i> ‘to hold two or more things’	No later than 5 th century BCE	谦 <i>qiān</i> , 慊 <i>qiàn</i> , 嫌 <i>xián</i> , etc.
		廉 <i>lián</i> , 镰 <i>lián</i> , etc.

The loss of consonant clusters happened not only in the onset position, but also in the coda position, which entails the disappearance of consonantal affixes. Meanwhile, the simplification of the phonological system would inevitably lead to an increase in homophones, resulting in ambiguity in oral communication. Linguists suggest that the Chinese language developed tones

326 and compound words to help disambiguate (e.g., Downer, 1959; Haudricourt, 1954; Mei, 1980;
 327 Pulleyblank, 1962; Sagart, 1999: 111, 131-133; Zhang, 2024b, c).
 328
 329 Unlike sub-syllabic affixes that were never mentioned in traditional Chinese philology,
 330 derivation by tonal alternation was documented at length by philologists during the 6th-7th
 331 century. Accustomed to the logic of ideographic writing, Chinese scholars for a long time lacked
 332 effective methods for analyzing character pronunciation, until Indian phonetics was introduced to
 333 China together with Buddhism in the 1st century CE. Inspired by the Indian Abugida scripts and
 334 the sophisticated phonetic knowledge as recorded in *Aṣṭādhyāyī*, Chinese philologists began to
 335 break down the pronunciation of a character into the initial (onset) and the final (nucleus + coda)
 336 with the tone. When it came to the Sui dynasty (581-618 CE), with the thriving of tonal poetry,
 337 *Qièyùn* (《切韵》, 601 CE), the rhyme dictionary, was compiled that categorizes characters into
 338 four tones and 193 rhyme groups, and thus established a standard of Middle Chinese for reading
 339 classical texts and composing poetry. Almost at the same time, 30 volumes of *Annotations of*
 340 *Classics* (《经典释文》) put together the annotations for over 680,000 characters (over 8,000
 341 different types of characters) of classical texts. In those annotations, it is common to see the same
 342 character pronounced differently when occurring in different contexts, and the alternation
 343 between the departing tone and other tones is the most prominent, as shown below:

- (11) a. Original text from *The Classic of Documents* (《尚书·说命中》):
 树 后 王 君 公, 承 以 大夫 师 长。
Shù hòu wáng jūngong, chéng yǐ dàfū shīzhǎng.
 establish king **king** lord entrust use minister official
 ‘The sovereign and the feudal lords were established, and they were assisted by the
 great ministers and the heads of various official departments.’
 b. *Annotations of Classics* (《经典释文·尚书音义》):
 后 王, 于 方 反。
Hòu wáng, yú fāng fǎn.

king **king** yú fāng FAN

‘For the 王 wáng in 后王 hòuwáng, the initial is the same as 于 yú < MC ju, while the final is the same as 方 fāng < MC pjang (flat tone).’

- (12) a. Original text from Zuo’s *Commentary on “Spring and Autumn Annals”* (《左传·成公二年》):

四 王 之 王 也,
Sì wáng zhī **wàng** yě,
four king Particle **to be king** Particle
树 德 而 济 同 欲 焉。
shù dé ér jì tóng yù yān.
establish virtue Conj. fulfill common desires Particle

‘As for the kingship of the Four Kings, they established virtue and fulfilled the common desires of the people.’

- b. *Annotations of Classics* (《经典释文·春秋左氏音义》):

之 王, 于 况 反。
Zhī **wàng**, yú kuàng fǎn.
Particle **to be king** yú kuàng FAN

‘For the 王 wàng in 之王 zhī wàng, the initial is the same as 于 yú < MC ju, while the final is the same as 况 kuàng < MC xjwang (departing tone).’

344

345 The character 王 wáng/wàng is pronounced with a flat tone when used as a noun in example

346 (11), but pronounced with a departing tone when used as a verb in example (12), illustrating a

347 derivation by tonal alternation. This type of derivation made Chinese characters multi-functional:

348 one character could be used to represent multiple words, including the original word and the

349 derivatives, contingent upon the context. Noteworthy, the annotations in *Annotations of*

350 *Classics* do not always match the phonological system in *Qièyùn*, and many derived words

351 vanished amid the fusion of the Han Chinese people with the nomadic peoples from the north

352 (mainly Mongolians and Manchurians). It is estimated that about 50% characters used to have

353 more than one pronunciation according to *Annotations of Classics*, but this ratio is only about

354 10% in Modern Mandarin (Zhang & Qin, 2016; Zhang, 2024b). Apparently, the language contact

355 brought by migration enhanced the one-to-one correspondence between single characters and

356 monosyllabic morphemes, despite the fact that more and more characters became homophones as

357 the phonological system was further simplified in the meantime.

358

359 The overflowing homophones necessitated other tools to help disambiguate, which became
360 increasingly pressing since social development kept bringing new concepts. The ultimate
361 solution that the Chinese language adopted was compounding: monosyllabic morphemes
362 represented by single characters began to compound with each other, as shown in the following
363 examples:

(13) 其 敢 以 许 自 为 功 乎?

Qí gǎn yǐ Xǔ zì wéi gōng hū?

Particle are use Xu **self** is achievement Particle

‘How would I dare to claim the annexation of the state of Xu as my own achievement?’

364 (Zuo’s Commentary on “Spring and Autumn Annals”, 4th century BCE)

(14) 修 己 而 不 责 人。

Xiū jǐ ér bù zé rén.

cultivate **self** Conj. not blame others

‘Cultivate yourself and do not blame others.’

365 (Zuo’s Commentary on “Spring and Autumn Annals”, 4th century BCE)

(15) 与 人 方便, 自己 方便。

Yǔ rén fāngbiàn, zìjǐ fāngbiàn.

give others convenience **self** convenience

‘To give convenience to others is to give convenience to yourself.’

366 (The Plum in the Golden Vase, 16th century CE)

367

368 Derived from pictures, Chinese characters were initially created to represent individual words (see
369 section 2.1). The monosyllabic morphemes represented by most characters could stand alone as
370 words in Old Chinese, such as 自 *dzih > MC dzi > zì ‘self’ in example (13) and 己 *kǐə > MC ki >
371 jǐ ‘self’ in example (14). However, an increasing number of morphemes, represented by different
372 characters, became homophonic to them in the simplification of phonology. In consequence, these
373 two synonymous morphemes compound with each other, forming the compound 自己 zìjǐ ‘self’
374 in example (15). Admittedly, this type of compounding appeared before the unification of Qin in
375 221 BCE, but its frequency increased significantly during wartime after the 8th century. An
376 increase in compounds’ frequency necessarily entails a decrease in the syntactic freedom of their

constituents: the monosyllabic morphemes represented by single characters are no longer free morphemes all the time, and in extreme cases, could eventually turn to bound morphemes. It needs to be noted that compounding happened mainly in the spoken language, which was not the concern of traditional philology, while a perennial archaizing purism pervaded formal and literary writing (Zhang, 2024c).

In general, the morphological processes occurring in the history of the Chinese language can be summarized below in Table 3.

Table 3. Morphological processes in the history of the Chinese language

Morphological process	Source	Historical period	Philological record	Example
Sub-syllabic affixation	Proto-Sino-Tibetan	Lost before the 5 th century BCE	No	suffix <i>*-ŋ</i>
Tonal alternation	Self-developed	5 th century BCE – 8 th century CE	Rich	王 <i>wáng</i> ‘king’ – 王 <i>wàng</i> ‘to be king’
Compounding	Self-developed	Accelerated after the 8 th century	No	自 <i>zì</i> + 己 <i>jǐ</i> = 自己 <i>zìjǐ</i> ‘self’

6. Discussion

The divergent evolution paths of Chinese and Tibetan clearly instantiate the explicit and implicit prescriptive effects of orthography on morphology.

As previously mentioned, grammatical treatises emerged at the same time with the Tibetan writing system that detailed the morphology of the Tibetan language including sub-syllabic affixes, while congenetic affixes were never mentioned in Chinese philology, and thus never explicitly taught in literacy education. This difference can largely be attributed to the writing systems. Derived from pictures, the phonetic representations provided by Chinese characters are

relatively coarse-grained, simply unable to record the inner structure of syllables. When fine-grained phonetic tools were finally imported to China, sub-syllabic morphemes had already long disappeared. In stark contrast, with a segmental abugida, Tibetan orthography inherently necessitates an awareness of both phonemes and morphemes: writers must understand the morphophonemic principles of the script in order to write correctly. Self-evidently, the metalinguistic awareness of linguistic units is closely related to orthography, which determines the salience of those units in prescriptive linguistic studies, thus wielding an explicit influence on language evolution through literacy education.

Compared to the explicit prescriptive effect, the implicit influence appeared to be more fundamental, pervading the history of the Chinese language. The decline of tonal alternation offers a prominent example: the phenomenon thoroughly described in philology did not survive the impact of language contact, while the one-to-one correspondence between characters and monosyllabic morphemes was enhanced. It is crucial to recognize that the rise of tonal alternation illustrated the implicit effect of ideographic writing itself – the development of tones correlated with the loss of consonant clusters, and the boundaries between syllables (the pronunciations of single characters) are thus clearly delineated – the implicit effect came into operation even before the advent of philology. Taking 京 **kraŋ* > MC *kiaŋ* > *jīng* ‘big, capital’ as an example (see section 4.2), the simplification of the initial cluster from **kraŋ* to **kiaŋ* would have passed largely unnoticed in speech, yet it produced a syllable with more distinct boundaries, making it less susceptible to misperception alongside neighboring syllables. On top of clear-cut morpheme boundaries, the Chinese language ultimately resorted to compounding, in which monosyllabic morphemes stabilized by single characters compound with each other to help

disambiguate. Clearly, those morphological processes not represented in writing – whether or not they were recorded in philological sources – were gradually lost over the course of Chinese linguistic history, including sub-syllabic affixation and tonal alternation, leaving compounding as the most productive morphological process.

As a system of auditory symbols, speech is essentially momentary, unable to be delivered across time and space with the same permanence as scripts. In addition, the phenomenon of visual dominance is widely observed in human cognition: when visual and auditory stimuli conflict, the response to the visual stimulus typically overrides the auditory one (Colavita, 1974; Hecht & Reiner, 2009). Therefore, those linguistic units represented in writing are naturally more likely to be noticed, studied, and thus stabilized than those not represented. In this sense, writing system determines the granularity of linguistic prescription: the linguistic unit represented by a single sign is always the smallest linguistic unit that can possibly be standardized in the language. It is not language that picks the matching script, but orthography that determines the basic unit of linguistic norms, thus shaping the form of the language. Genetically related languages might diverge due to the distinct writing systems adopted, while the adoption of writing system is ultimately shaped by geopolitical factors, as exemplified by Chinese and Tibetan.

7. Conclusive remarks

Based on a comparison of the derivational paths of cognate roots in Chinese and Tibetan, the present study argues that the loss of sub-syllabic morphology in Chinese can be attributed to an influence of the ideographic writing system that is relatively coarse-grained in phonetic representation, shedding light on the explicit and implicit prescriptive effects of writing on

language evolution.

It must be admitted that the geopolitical environment of China provided a special context allowing the written language to fully wield its influence. The self-originated ideographic writing system was the only writing system in a vast monolithic area without any competition for at least two millennia, and thereby became the cornerstone of the common identity across regional accents (see section 2.1). In other cradles of ancient writing, like the Nile Valley and Mesopotamia, the picture-derived scripts simply did not manage to persist as Chinese characters did, so naturally, we are unable to observe similar prescriptive effects from them. In a general sense, orthography relies on stable centralized power and literacy education to wield long-lasting influence, so that the norms of the written language can ultimately be extended to the spoken language of all areas within the sphere of influence. Moreover, as abugidas and alphabets offer more fine-grained phonetic representations than ideographic writing, thus more sensitive to phonetic variations and change, the orthography based on them is inherently more difficult to retain than ideographic writing, let alone to affect language evolution. Turning back to Chinese and Tibetan, the dialectal differences of Chinese have always been dramatic as phonological standardization is beyond the ability of ideographic characters, while the collapse of the Tibetan empire in the 9th century gave rise to significant dialectal differences.

Since the present study is based primarily on the standard phonological system – the phonological system of *Qièyùn*, the standard pronunciation of Modern Mandarin as represented by Pinyin (拼音), the phonology of classical Tibetan as recorded in the orthography – data from other varieties are necessary to make further generalizations. Preliminary observations suggest

that the loss of consonant cluster is thoroughgoing in Chinese, but the preservation of tonal alternation does have remarkable regional differences. As for Tibetan, those varieties spoken in the colder northeastern areas close to the Qilian mountains largely preserve the consonants of classical Tibetan, while those varieties spoken in the warmer southwestern areas, including the city of Lhasa, tend to reduce consonant clusters and develop tones. Moreover, the evolution paths of other languages, especially Burmese and Japanese, are called for to attest the hypothesized prescriptive effects of orthography on language evolution. Among Sino-Tibetan languages, Burmese also has a long history of documentation, and the Burmese writing system is also a segmental abugida adapted from Indian scripts, yet the derivatives of cognate roots have not been analyzed together with Chinese and Tibetan. Besides ideographic Chinese characters, syllabaries cannot show the analyzability of syllables either, and Japanese is the only language that has been using syllabaries for over 1,000 years, so an in-depth investigation is of interest regarding its evolution: Japanese surely does not have sub-syllabic morphology, but it is unknown whether this feature is related to the writing system. Last but not least, psycholinguistic studies may be able to provide helpful insights as to how writing managed to affect the salience of linguistic units in mental representation, and thus to shape the evolution of languages.

Abbreviations:

Conj.	Conjunctive
FAN	Marker indicating the spelling of the initial and the final (with tone) in traditional Chinese philology
MC	Middle Chinese

Endnotes:

¹ The asterisk * is conventionally indicative of reconstructed forms in Proto-Sino-Tibetan and Old Chinese.

² Exceptions are observed from the transliterations of foreign words and inseparable disyllabic terms (联绵词), such as 蝴蝶 *húdié* ‘butterfly’, 蜘蛛 *zhīzhū* ‘spider’, 参差 *cēncī* ‘irregular’, etc. which are rather limited and unstable in use (see Author, 2024a).

³ The pronunciation in Middle Chinese (MC) is reconstructed based on the phonological system of *Qīèyùn* (《切韵》, 601 CE), while the pronunciation in standard Modern Mandarin is transcribed in Pīnyīn (拼音).

References:

Author (2024a)

Author (2024b)

Author (2024c)

Benedict PK (1972) *Sino-Tibetan: A conspectus*. Cambridge: Cambridge University Press.

Benedict PK (1976) Sino-Tibetan: Another look. *Journal of the American Oriental Society* 96(2): 167-197.

Colavita FB (1974) Human sensory dominance. *Perception & Psychophysics* 16(2): 409-412.

<https://doi.org/10.3758/BF03203962>

Conrady A (1896) *Eine Indo-Chinesische causative-Denominativ-Bildung und ihr*

Zusammenhang mit den Tonaccenten. Leipzig: Otto Harrassowitz.

Dettamajumdar S (2017) History of the scientific study of the Tibeto-Burman languages of north-east India. *Indian Journal of History of Science* 54(2), 420-445.

<https://doi.org/10.16943/ijhs/2017/v52i4/49265>

Downer GB (1959) Derivation by tone-change in Classical Chinese. *Bulletin of the School of*

513 *Oriental and African Studies* 22, 258-290. <https://doi.org/10.1017/S0041977X00068701>
 514 Gabelentz G (1881) *Chinesische Grammatik: Mit Ausschluss des niederen Stiles und der*
 515 *heutigen Umgangssprache*. Leipzig: T.O. Weigel.
 516 Grube W (1881) *Die sprachgeschichtliche Stellung des Chinesischen*. Leipzig: T.O. Weigel.
 517 Haudricourt AG (1954) Comment reconstruire le chinois archaïque. *Word* 10(2-3), 351-64.
 518 <https://doi.org/10.1080/00437956.1954.11659532>
 519 Hecht D, Reiner M (2009) Sensory dominance in combinations of audio, visual and haptic
 520 stimuli. *Experimental Brain Research* 193, 307-314. <https://doi.org/10.1007/s00221-008->
 521 1626-z
 522 Hucker C (1975) *China's imperial past: An introduction to Chinese history and culture*. Stanford,
 523 CA: Stanford University Press.
 524 Karlgren B (1915-1926) *Études sur la phonologie chinoise*. Uppsala: K. W. Appelberg.
 525 Li F (1939/2013) Zàng-Hàn xì yǔyán yánjiū fǎ [Research methods of Sino-Tibetan languages].
 526 *Hàn-Zàng Yǔyán Xuébào [Journal of Sino-Tibetan linguistics]* 7(1), 1-10.
 527 Li F (1937/1973) Languages and dialects of China. *Journal of Chinese Linguistics* 1(1): 1-13.
 528 <https://doi.org/10.2307/1790133>
 529 Mei TL (1980) Sishēngbiéyì zhōng de shíjiān céngcì [Strata in the use of tones to differentiate
 530 meanings]. *Zhōngguó Yǔwén [Studies of the Chinese Language]* 29(6), 427-433.
 531 Norman J (1988) *Chinese*. Cambridge: Cambridge University Press.
 532 Przyluski J (1924) Langues sino-tibétaines. In: Meillet A, Cohen M (eds) *Les langues du monde*
 533 (pp. 361–384). Paris: Librairie Ancienne Édouard Champion.
 534 Pulleyblank EG (1962) The consonantal system of Old Chinese. *Asia Major* 9, 206-265.
 535 Sagart L, Jacques G, Lai Y, Ryder RJ, Thouzeau V, Greenhill SJ, List JM (2018) Dated language

536 phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National*
 537 *Academy of Sciences of the United States of America (PNAS)* 116(21), 10317-10322.
 538 <https://doi.org/10.1073/pnas.1817972116>
 539 Sagart L 1999. *The roots of Old Chinese*. Amsterdam: John Benjamins.
 540 Saussure F (1916/1959) *Course in General Linguistics*. Translated by Wade Baskin. New York:
 541 Philosophical Library.
 542 Shafter R (1955) Classification of the Sino-Tibetan languages. *Word* 11(1), 94-111.
 543 <https://doi.org/10.1080/00437956.1955.11659552>
 544 Shafter R (1957) *Bibliography of Sino-Tibetan Languages*. Wiesbaden: Otto Harrassowitz.
 545 Shafter R (1963) *Bibliography of Sino-Tibetan Languages: Volume 2*. Wiesbaden: Otto
 546 Harrassowitz.
 547 Sun H (1998) Guānyú Hànzàng yǔ fēnlèi yánjiū de huígù yǔ cúnzài wèntí [Review of the
 548 typological studies on Sino-Tibetan languages]. *Mínzú yǔwén [Minority Languages of*
 549 *China]* (3), 42-46.
 550 Sun Y (2007a) Shàngǔ hànǔ cíhuì gòunǐ xīpíng (shàng) [Comments on the reconstruction of
 551 Archaic Chinese (I)]. *Jiānghàn Dàxué Xuébào (Rénwén Kēxué Bǎn) [Journal of Jiangnan*
 552 *University (Humanities and Social Sciences)]* (3), 39-46.
 553 Sun Y (2007b) Shàngǔ hànǔ cíhuì gòunǐ xīpíng (xià) [Comments on the reconstruction of
 554 Archaic Chinese (II)]. *Jiānghàn Dàxué Xuébào (Rénwén Kēxué Bǎn) [Journal of Jiangnan*
 555 *University (Humanities and Social Sciences)]* (4), 68-75.
 556 Zhang H, Ji T, Pagel M, Mace R (2020) Dated phylogeny suggests early Neolithic origin of
 557 Sino-Tibetan languages. *Scientific Reports* 10, 20792. [https://doi.org/10.1038/s41598-020-](https://doi.org/10.1038/s41598-020-77404-4)
 558 77404-4

Zhang Z, Qin M (2016) Lùn hànyǔshǐ shàng ‘xué’ ‘xiào’ de yīnbiàn gòucí” [Phonological derivations of xue and xiao in the history of Chinese language]. *Dōngnán Dàxué Xuébào (Zhéxué Shèhuì Kēxué Bǎn)* [Journal of Southeast University (Philosophy and Social Science)] 18(1), 137-145.

Zhang Z (2013) *Hànyǔ Biànshēng Gòucí Yánjiū* [A study on derivation by initial-change in Classical Chinese]. Beijing: China Book Press.

Figures legends

Figure 1. Derivatives of the root **kyu* ‘hook’ in Chinese and Tibetan

Figure 2. Derivatives of the root **gra* ‘large’ in Chinese and Tibetan

Tables

Table 1. Examples of ideographic signs in Chinese writing

Table 2. Examples of characters sharing phonetic radicals with diverged onsets

Table 3. Morphological processes in the history of the Chinese language

Data Availability

The supplementary material is openly available at <https://zenodo.org/records/22013583>, including the philological records of the Chinese characters analyzed in the paper. Other qualitative data are drawn from dictionaries publicly available online, such as Hàndiǎn (汉典, <https://zdic.net>) and the Tibetan-Chinese dictionary (<http://www.tibetans.cn>).

Ethical Statements

Ethical Approval:

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

582 **Informed Consent:**

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