

Vernacularization in Medieval Chinese: A quantitative study on classifiers, demonstratives, and copulae in the Chinese Buddhist Canon

Tak-sum Wong and John S. Y. Lee

Department of Linguistics and Translation, City University of Hong Kong, Kowloon, Hong Kong

Abstract

While studies on diachronic Chinese syntax have identified a number of linguistic changes in Medieval Chinese, they have mostly been underpinned by qualitative analyses. In the most large-scale quantitative analysis to-date, this article investigates changes in the use of classifiers, demonstratives, and copulae. Our analysis, based on the Chinese Buddhist Canon, examines over 40 million characters in texts spanning a millennium. Results suggest that from the late Eastern Han period (circa 150 CE) onwards, the vernacular style became increasingly widespread, at the expense of the literary style, as reflected by changes in the use of classifiers and demonstratives, and in the construction of nominal sentences. However, the vernacular style became less frequently used in the Northern Sung period (960–1127 CE). This reversal may shed light on the work of the Stylists, editors appointed by the Sung court to polish Buddhist texts with more literary elements.

Correspondence:

John Lee, Department of Linguistics and Translation, City University of Hong Kong, Tat Chee Avenue, Hong Kong.

E-mail:

jsylee@cityu.edu.hk

1 Introduction

The evolution of Medieval Chinese (MC), and in particular its process of vernacularization from the 2nd to 11th centuries, has been a research topic of considerable interest. In a comprehensive study, Wei (2000) identifies the features that most distinguish between the colloquial, or ‘vernacular style’, and the more ‘literary style’ of writing in MC texts. These features include significant developments in vocabulary, such as the forms of demonstratives and grammatical particles; in morphology, such as the emergence of suffixes for nouns and pronouns; and in syntax, such as the use of copulae and classifiers.

The Chinese Buddhist Canon (‘the Canon’) can shed new light on the vernacularization process, since it contains a vast quantity of text that is largely consecutive in time. It is suitable for controlled comparison between the literary and vernacular styles, because its texts are relatively homogeneous in terms of content and genre. Further, it can potentially reveal a more accurate chronology, since it exhibits more vernacular elements than contemporary secular literature. Not surprisingly, a number of studies on diachronic MC syntax have already made use of Buddhist texts. Zürcher (1977) discussed vernacular elements in Buddhist texts translated in the Eastern Han period (25–220), including changes in

the pronominal system, use of function words as well as the use of *wéi* 爲 ‘to do’ and *zuò* 作 ‘to make’. Hú (1993) traced the development of the copula *shì* 是 by comparing translations in different periods of the same text. Sun (1996) compared a 10th-century Buddhist text with materials from other periods regarding word order in prepositional phrases (PPs), disposal construction, and the grammaticalization of a number of particles. Wāng (2008) analysed the types and usage of classifiers, while Zhū (2015a, 2016) studied vocabulary and syntactic constructions, including the *wéi*-A-*suǒ*-*jiàn*-V passive construction (Zhū, 2015b) and the inessive construction (Zhu and Chen, 2008, 2013).

Most studies cited above are based on a relatively small number of texts that are deemed by the researchers as representative of the vernacular language. This methodology was adopted, since their goal was to clarify the linguistic characteristics of the vernacular. In contrast, our goal is to measure the extent to which Buddhist texts preferred the vernacular style of writing over the literary style, and how this preference evolved over time. Therefore, we focus on a number of well-known features of the vernacular language and seek to quantify their presence in all texts in the Canon. We then compare their frequency to that of their corresponding literary constructions, again from the vantage point of the entire Canon. Our main innovation is thus to exploit the statistical strength of a very large corpus to sketch a quantitative chronology of the vernacularization process. In this sense, our research methodology is closest to Hu et al. (2007), who investigated the word position of PPs in the Sheffield Corpus to evaluate the post-verbal prominence hypothesis of PP (Li and Thompson, 1974; 1975).

We focus on changes in three linguistic phenomena—the use of classifiers, demonstratives, and copulae—contrasting their vernacular and equivalent literary forms. The vernacular style requires a classifier for quantified nouns (e.g. *yī méi huā* 一枚花 *lit.* ‘one piece of flower’), while the literary style does not (e.g. *yī huā* 一花 ‘one flower’). For demonstratives, the vernacular style prefers the use of *cǐ* 此 (e.g. *cǐ yǔ* 此語 ‘these words’), whereas the literary style also

uses *shì* and *sī* 斯 (e.g. *shì yǔ* 是語 ‘these words’). In nominal sentences, the vernacular style favours the use of copulae (e.g. *Sè jí shì kōng* 色即是空 ‘Form is just emptiness.’¹), while the literary style does not (e.g. *Xiànzài jí kōng* 現在即空 ‘The present [is] just emptiness’).

We chose these three linguistic phenomena for a number of reasons. First, they are significant characteristics of the vernacular style in MC, as evidenced by the wealth of studies on their development (e.g. Liú, 1965; Zürcher, 1977; Táng, 1992; Wei, 2004; Wāng, 2008, etc.). Second, they appear frequently in the Canon in both the vernacular and the corresponding literary constructions, thus providing statistical strength in our analysis.² Third, by applying judicious restrictions on syntactic environments, they can be accurately identified based on part-of-speech (POS) annotations in our corpus (Section 3.2). This allowed us to report statistics from a very large corpus, an infeasible task if manual verification were required.

The contribution of this article is two-fold. First, in a carefully controlled comparison over 40 million characters in the Chinese Buddhist Canon, we show a rise of the vernacular over the literary style from the late Eastern Han period (2nd century CE) till the Sui-T’ang (ST) period (581–959 CE), as exemplified by classifiers, demonstratives, and copulae. This vernacularization process has hitherto been discussed mostly in qualitative terms, and has not been subject to any large-scale quantitative analysis. Second, our analysis reveals an unexpected retreat of the vernacular style in the Sung period, likely due to the editing work of the Stylists (*Rùnwénguān* 潤文官), officials charged with polishing Buddhist texts with more literary elements. To the best of our knowledge, this is the first study that points to concrete evidence for the work of the Stylists.

The rest of the article is organized as follows. After a review of the historical background in Section 2, we summarize our textual data in Section 3. We then present analyses on the use of classifier, demonstrative, and copulae in Sections 4, 5, and 6, respectively. Finally, we conclude in Section 7.

2 Background

In this section, we briefly review the literary and vernacular styles of Chinese writing (Section 2.1), and then discuss the work of the Stylists in the Sung period (Section 2.2).

2.1 Vernacular versus literary style

Literary Chinese was, up till the Written Vernacular Movement in the 1920s, the most privileged form of writing in China (Norman, 1988, p. 109). For this reason, most texts, throughout Chinese history till early 20th century, were written in Literary Chinese, even though it is modelled on the Classical prose of the late Chou and Han periods (ca 7th century BCE–220 CE), and reflects syntactic constructions of Archaic Chinese. For studies on the vernacularization process, therefore, it is more informative to focus on texts that exhibit more vernacular elements, such as religious texts and popular literature (Norman, 1988, p. 111; Zürcher, 1977).

Buddhist texts are especially valuable in this domain of study. Buddhist missionaries tried to present their teachings with straightforward, easily accessible language. Therefore, most scholars believe that the language in the Canon is more vernacular in nature and thus closer to the colloquial language at the time of the composition or translation of the text (e.g. Karashima, 1996; Norman, 1988, p. 111).

2.2 Stylists

In the history of China, the Northern Sung (NS) (960–1127) arguably provided the most support for Buddhist translation activities, including the establishment of the Institute for the Translation of the Sūtras³ (cf Sen, 2002). The translation workflow normally involved nine people (Harbsmeir, 2015). First, the master of translation read out the Sanskrit text, which was then reviewed by the philological assistant and checked by the text appraiser. After the transliterator wrote down the transliteration, the translator-scribe turned it into Chinese words. The text editor then made necessary changes to the word order. The proofreader checked the text; next, the subeditor deleted unnecessarily long expressions. Finally, the

Stylists (*Rùnwénguān* 潤文官) were responsible for polishing the texts (Huang, 1990, 1997, p. 68; Sen, 2002; Kin, 2010).

Previous research suggested that the Stylists revised the texts in three ways. First, they ensured the content conform to the ideology of the court, for example by removing any mention of the names of emperors (Wan, 2008).⁴ Second, they inserted words or clauses to smooth out the text, making them read more like fluent Chinese (Wáng, 1984, p. 190; Wan, 2008; Harbsmeir, 2015).⁵ Third, they replaced the ‘vulgar’ or vernacular elements of the texts with formal written classical language.⁶ By comparing the texts in the Sung period and those from earlier periods in terms of their literary elements, our analysis can shed light on this last aspect of the Stylists’ work.

3 Textual Data

3.1 Textual material and dating

Our analysis was performed on the Korean Edition of the Chinese Buddhist Canon, the *Tripitaka Koreana*, since this edition is based on the most complete set of printing blocks available for the Canon. To support diachronic analysis, we obtained the time of composition or translation of each document in the Canon from the catalogue by Lancaster and Park (1979). About 9% of the documents do not have their year of publication in the catalogue, and are excluded from this study. For all the other 1,347 documents, the catalogue provides either the exact year of publication, or a period of several years, in which case we date the document at the last year.

Following Wei (2000), we divided our data into four periods: ‘Han-Chin’ (HC), covering the Eastern Han, Three Kingdoms Period, and Western Chin; ‘Eastern Chin’ (EC), covering the Eastern Chin and Northern and Southern Dynasties; ‘Sui-T’ang’ (ST), covering the Sui, T’ang, and Five Dynasties; and ‘Northern Sung’ (NS).⁷ The size of the textual material for each period is listed in Table 1.

Table 1 Size of the textual data for each historical period

Period	Abbreviation	Period	#Documents	#Characters
Han-Chin	HC	25–317 CE	273	2,852,151
Eastern Chin	EC	318–580 CE	370	16,244,877
Sui-T'ang	ST	581–959 CE	447	21,018,393
Northern Sung	NS	960–1127 CE	257	3,222,086
Total			1,347	43,337,507

3.2 Word segmentation and POS annotation

As reported in Wong and Lee (2016), we performed word segmentation and POS tagging on the entire *Tripitaka Koreana*. For word segmentation, we used the method described in Zhao et al. (2007). When trained on a corpus of Buddhist texts with 50,000 characters (Lee and Kong, 2016), supplemented with lists of Buddhist terms provided by DDBC and Chu (1996, 1998, 1999), the segmentation model achieved 96.9% precision and 98.3% recall (Wong and Lee, 2016). Most segmentation errors involve multi-character words. These errors have a relatively small effect on our study, since the classifiers, demonstratives, and copulae to be investigated are mostly single-character words.

For POS tagging, we used conditional random fields (Lafferty et al., 2001) with the CRF⁺⁺ implementation (Kudo, 2005). The POS tags are helpful for automatic retrieval of the three linguistic phenomena investigated in this study. For example, the sentence 故著論五篇 *Gù zhù lùn wǔ piān* ‘So, [he] wrote five articles’ is tagged as follows:

Word:	<i>Gù</i>	<i>zhù</i>	<i>lùn</i>	<i>wǔ</i>	<i>piān</i>
	‘so’	‘write’	‘article’	‘five’	‘piece’
POS tag:	AD	VV	NN	CD	M

The POS tagger identifies the word *gù* as an adverb (‘AD’), *zhù* as a verb (‘VV’), *wǔ* as a number (‘CD’), and *piān* as a measure word (‘M’). Usefully, it determines that *lùn* is a noun (‘NN’) rather than a verb in this context, facilitating the identification of classifier usage (*piān*) in this sentence.

On ten-fold cross-validation on a POS-tagged corpus of Buddhist texts (Lee and Kong, 2016),

the tagger performed at an accuracy of 81.86% (Wong and Lee, 2016). We further analysed tagger performance on identifying nouns (NN), verbs (VV), and prepositions (P), since they were used in constructing our data sets. The NN tags were used for identifying quantified nouns (Section 4) and nominal sentences (Section 6). The tagger recognized NN tags at 0.88 precision and 0.80 recall. These errors can be expected to appear as often in the literary instances as in the vernacular ones, since they are not correlated with the presence of classifiers and copulae. Therefore, they should have minor effects on the relative frequencies between the literary and vernacular instances.

The VV and P tags are used for identifying demonstratives preceded by verbs or prepositions (Section 5). The tagger recognized VV at 0.85 precision and 0.79 recall, and P at 0.71 precision and 0.76 recall. The relatively low performance for P can be attributed to the confusion between verbs and coverbs, which are considered prepositions in ancient Chinese. Most of the false positives of P were verbs, which still served the purpose of ensuring that the *shì* is a demonstrative. Most of the false negatives of P were also verbs; they were therefore still accounted for in the analysis in post-verbal demonstratives (cf. Section 5).

4 Classifier

When writing in the vernacular style in MC, most quantified nouns require classifiers⁸ cf. (Norman, 1988, p. 115). There are two possible word orderings. The noun may precede the number and the classifier (e.g. *huā yī méi* 花一枚 ‘flower: one piece’), which we will denote as ‘classifier-last’. The noun may also follow the number and the

Table 2 Summary of linguistic patterns of the literary style and vernacular style that are compared in this article

Pattern	Literary style	Vernacular style
Noun classifier (Section 4)	<number> + <noun> Example: <i>yī huā</i> 一花 ‘a flower’	Classifier-first: <number> + <classifier> + <noun> Classifier-last: <noun> + <number> + <classifier> Examples: <i>yī měi huā</i> 一枚花 ‘one piece of flower’; <i>huā yī měi</i> 花一枚 ‘flower: one piece’
Proximal demonstrative (Section 5)	{ <i>shì</i> 是, <i>cǐ</i> 此, <i>sī</i> 斯} [+ <noun>] Example: <i>sī yǔ</i> 斯語 ‘this speech’	<i>cǐ</i> [+ <noun>] Example: <i>cǐ yǔ</i> 此語 ‘this speech’
Nominal sentence with grammaticalized adverbs (Section 6)	<noun> + { <i>fēi</i> 非, <i>jí</i> 卽, <i>nǎi</i> 乃} + <noun> Example: <i>Xiàn zài jí kōng</i> . 現在卽空 ‘The present [is] just emptiness.’	<noun> + { <i>jí</i> 卽, <i>nǎi</i> 乃, <i>bù</i> 不, <i>fú</i> 弗, <i>fēi</i> 非, <i>wú</i> 無, <i>wù</i> 勿, <i>wèi</i> 未} + <i>shì</i> 是 + <noun> Example: <i>Sè jí shì kōng</i> . 色卽是空 ‘Form is emptiness.’

classifier (e.g. *yī měi huā* 一枚花 ‘one piece of flower’), which we will denote as ‘classifier-first’. In contrast, the literary style does not employ any classifier and uses the simple construction <number> + <noun> (e.g. *yī huā* 一花 ‘one flower’).⁹

4.1 Previous research

Although the classifier-last form appeared as early as the Shang period (16c.–11c. BCE), less than ten different classifiers were found in the oracle bone inscriptions. The classifier-first type was attested among the forty different classifiers observed in the bronze inscriptions on the vessels from the Western Chou period (1046–771 BCE), e.g. *bǎi mǔ liáng* 百畝糧 ‘(one) hundred acre of grain’ (from Huáng, 1964). However, Peyraube (1998) suggested that in Archaic Chinese, only measure words were present but not genuine classifiers. He suggested that classifiers only appeared as late as in the Han period, i. e. 2nd century BCE, while the following change of word order took place: <noun> + <number> + <measure word> → <number> + <measure word> + <noun>. The predicate <number> + <measure word> was reanalysed as a modifier and was moved to pre-nominal position in late Archaic Chinese.

Classifiers became widespread in the Han period (from 206 BCE) (Norman, 1988, p. 115). Liú (1965) identified 274 classifiers used in MC, based on about fifty texts, including some Buddhist ones, published

in the period 220–589. Wāng (2008) studied the use of classifiers in sixty-two Buddhist texts from HC to EC, while Cáo (2010) studied the use and grammatical properties of classifiers in ST. Since these studies do not compare the frequency of use and non-use of classifiers, they do not reveal changes in the extent of classifier usage.

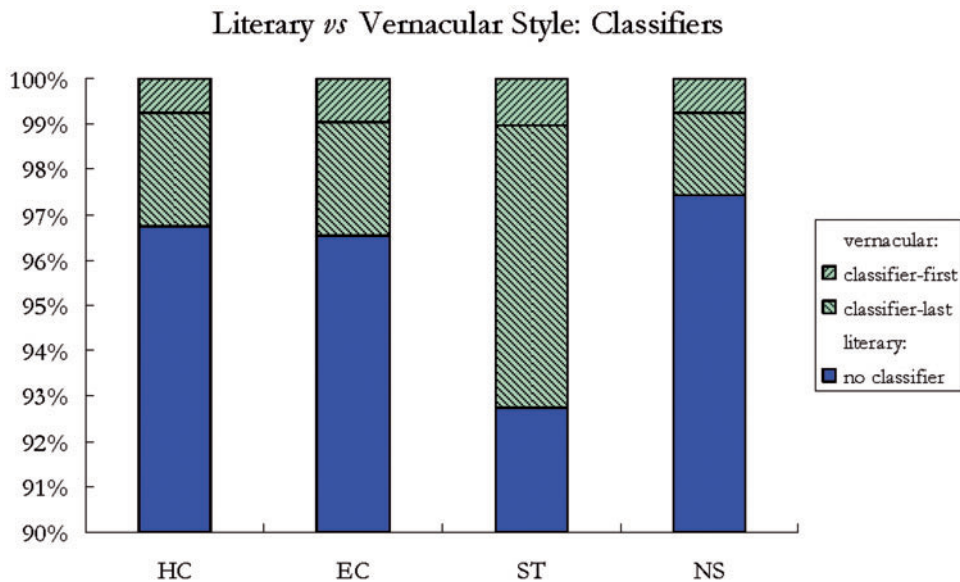
4.2 Data

Raw frequency counts alone do not accurately quantify the extent of classifier usage; they must be normalized by the number of nouns. In addition, the set of nouns and classifiers must be carefully chosen. We excluded abstract nouns, which do not take classifiers, to avoid inflating the frequency of the literary style. We also avoided classifiers with significant semantic content, such as those specifying units of mensuration¹⁰ or currency,¹¹ because their use may reflect semantic considerations rather than a preference for the vernacular style.

To derive our data set, we examined the 237 classifiers in the list by Liú (1965) and the 67 classifiers in that of Wāng (2008). We included those that express a collective concept¹² or serve to ‘actualize shape boundary’¹³ (from Bisang, 1999; Gerner, 2013, p. 66), but excluded those that indicate a currency or unit of mensuration and those that mean ‘type, kind’.¹⁴ For the final set of 110 classifiers left, we searched for nouns that participate in the

Table 3 Most frequent nouns and their classifiers

Type	Noun	Most frequent classifiers
Classifier-first	義 <i>yì</i> 'doctrine'	句 <i>jù</i> 'sentence' 偈 <i>jì</i> 'gatha'
	經 <i>jīng</i> 'sūtra'	行 <i>háng</i> 'column' 段 <i>duàn</i> 'section'
	欄橋 <i>lánshuǐ</i> 'barrier'	重 <i>chóng</i> 'layer'
	車 <i>jū</i> 'chariot'	乘 <i>shèng</i> 'piece' 兩 <i>liàng</i> 'piece'
Classifier-last	水 <i>shuǐ</i> 'water'	滴/滴 <i>dī</i> 'drop' 分 <i>fēn</i> 'share'
	衆 <i>zhòng</i> 'crowd'	人 <i>rén</i> 'people'
	邑 <i>yì</i> 'city'	戶 <i>hù</i> 'family'
	兄弟 <i>xiōngdì</i> 'brothers'	人 <i>rén</i> 'people'
	菩薩 <i>púsà</i> 'bodhisattva'	人 <i>rén</i> 'people' 名 <i>míng</i> 'name'
	女 <i>nǚ</i> 'lady'	人 <i>rén</i> 'people'

**Fig. 1** Use of classifier over different historical periods

vernacular patterns in Table 2, where <noun> refers to a word tagged as 'NN', <number> refers to a word tagged as 'CD', and <classifier> refers to a word from the lists of Liú (1965) and Wāng (2008) as described above. There was a total of 3,010 classifier-first instances and 2,852 classifier-last instances, involving 831 different nouns. The most frequent combinations are shown in Table 3.

To retrieve corresponding samples for the literary style, i.e. non-use of classifier, we then searched for the literary pattern in Table 2 on these 831 nouns. There were 201,018 such instances.

4.3 Analysis

Figure 1 shows the average proportion of vernacular style (i.e. use of classifier) versus literary style (i.e. no classifier) in the texts of the Chinese Buddhist Canon. Throughout all four historical periods, the literary style dominated; on average, over 90% of the quantified nouns in a text do not use any classifier.

4.3.1 From HC to ST

The proportion of nouns in the vernacular style remained stable between the HC and EC periods, at

3.2%. The classifier was more frequently placed after the noun than before. The extent of classifier usage increased substantially in the ST period, to about 7.5% (the difference is significant at $p=0.009$ by χ^2 -test). Much of this change was due to the increased frequency of the classifier-last pattern in the ST period.

To verify this trend for individual nouns, we computed similar statistics for each of the 831 nouns. A majority (56%) of the nouns saw the proportion of the literary style decreased from the HC and EC period to the ST period. The words *huā* 花/華 ‘flower’ and *shǒu* 手 ‘hand’ can serve to illustrate. For *huā*, the proportion of vernacular style was 15% during EC; for *shǒu*, the proportion was

- (1) 持此一花。 (K981, from HC)
Chí cǐ yī huā.
 hold this one flower
 ‘Holding this flower.’

- (2) 有爲相如二頭三手不可得。 (K631, from EC)
Yǒuwéi xiàng rú èr tóu sān shǒu bùkědé.
 productive form like two head **three hand** unobtainable
 ‘The productive form is like two heads and three hands
 (that both of them are) unobtainable.’

only 3% in HC. Some example sentences are:

By the time of the ST period, the vernacular style was more widely adopted. For *huā*, the share of the vernacular style increased from 15 to 22%; meanwhile, for *shǒu*, the figure increased to 5%. Examples of the vernacular style included both classifier-last and classifier-first constructions, such as:

- (3) 應取蓮花一百八枚 (K289, from ST)
Yīng qǔ liánhuā yībǎibā méi.
 should take lotus 108 **piece**
 ‘Should take 108 lotuses.’

- (4) 持七莖花 (K802, from ST)
Chí qī jīng huā.
 hold seven **stem** Flower
 ‘Holding seven flowers.’

- (5) 長一搥手半 (K310, from ST)
cháng yī zhā shǒu bàn
 long one cl. hand half
 ‘one and a half hands long’

4.3.2 From ST to NS

As shown in Fig. 1, the vernacular style retreated in the NS period; the proportion of nouns with classifier returned to 3%, similar to the HC and EC periods (the difference is significant at $p=0.0005$ by χ^2 -test). This phenomenon is unexpected, since no previous research in MC suggested any halt or reversal of the vernacularization process during Sung. The retreat of the vernacular style was thus likely a special phenomenon in the translation of the Buddhist Chinese Canon. More specifically, it may be due to the Stylists (see Section 2.2), who was given greater prominence and support from the court in the Sung period. If that is so, then the removal of classifiers appears to be one aspect of their editing work.

We again verified this trend for individual nouns and, indeed, for 64% among the fifty most frequent nouns, the literary style increased from the ST to the NS period. The words *huā* ‘flower’ and *shǒu* ‘hand’ were no exception. In the NS period, the proportion of vernacular style decreased to 8% for *huā* and to 0% *shǒu*. Some

- (6) 持以一花 (K1488, from NS)
Chí yī huā.
 hold by **one** **flower**
 ‘Hold one flower.’

- (7) 具三十四手 (K1214, from NS)
jù sānshísi shǒu.
 possess **thirty-four** **hand**
 ‘Possess thirty-four hands.’

example sentences are:

We now turn to another linguistic feature—the use of proximal demonstratives—to see if it exhibits the same trend regarding the literary and vernacular styles.

5 Proximal Demonstrative

Following general practice in Classical Chinese research (cf. Pulleyblank, 1994, pp. 85–90), we use the term ‘demonstrative’ to cover both determiner and pronoun. Frequently used demonstratives in the Chinese Buddhist Canon include the proximal demonstratives *shì* 是, *cǐ* 此 and its dialectal variation, *sī* 斯 (Wei, 2004, pp. 78–82); the distant demonstrative *bǐ* 彼; as well as *ěr* 爾, which can be either proximal or distant¹⁵ (Wáng, 1980, p. 366).

The proximal demonstratives offer a sharp contrast between the literary and vernacular styles: the vernacular style prefers *cǐ* (e.g. *cǐ yǔ* 此語 ‘this speech’), while *cǐ*, *sī*, and *shì* were all used in the literary style (e.g. *sī yǔ* 斯語 ‘this speech’). Since *cǐ* is common in both literary and vernacular language, there is no clear-cut dichotomy between the two styles. As will be discussed below, the overall development was that *cǐ* gradually took on the meaning of *shì* and replaced it. Hence, the proportion of *sī* and *shì* among proximal demonstratives can serve as an indicator of the vernacularization process.

5.1 Previous research

In Archaic Chinese, *shì* and *cǐ* carried distinct meaning: *shì*, which developed from *shí* 時, was a temporal deictic expression, while *cǐ* was a spatial deictic expression (Huáng 1983); further, *cǐ* was observable, while *shì* was not (Wei 2004, p. 78). In late AC, *shì* already started its grammaticalization to a copula.

There is no consensus on the usage of these three demonstratives in the Eastern Han period. On the one hand, Dobson (1964), Chan (1973), and Ushijima (1971) noted very frequent use of *cǐ* in all syntactic positions. In contrast, according to Zürcher (1977) and Wei (2004, p. 84), *shì* appeared frequently, but *cǐ* and *sī* much more rarely. During the Six Dynasties (222–589), *shì* was still used but was not as common as in Eastern Han. After the end of 4th century, *cǐ* was dominant, but some texts still use *shì* frequently.

5.2 Data

To compute the frequency of the use of *shì*, *cǐ*, and *sī* as demonstratives, the simple string search would not suffice, since *shì* can also be used as copula.¹⁶

Even manual verification can be difficult, since its usage as copula or as demonstrative is not always clear, such as in the following bridge example:

- (8) 聖道支是不生法 (K1, from ST)
Shèng dào zhī shì bùshēng fǎ.
 holy way branch SHI nirvāna dharma

Since the Canon was not punctuated, *shì* can be interpreted above as either a demonstrative, i.e. in the sense of ‘(For) the holy way branch, this (is) a dharma of nirvāna.’; or as a copula, i.e. in the sense of ‘The holy way branch is a dharma of nirvāna.’ Without the language intuition of MC, it is often difficult for contemporary experts to objectively discern the function of *shì* in these sentences. To avoid the need for manual verification, we focus on linguistic environments that unambiguously rule out the use of *shì* as a copula. In general, a copula rarely follows a verb (e.g. in Example (9) below), and never forms a verb–verb compound. The same can be said about any *shì* that follows a preposition. Hence, we restricted our analysis to post-verbal and post-prepositional *shì*, *cǐ*, and *sī*.

5.3 Analysis

We identified post-verbal and post-prepositional *shì*, *cǐ*, and *sī* by searching for occurrences of these words following a word tagged as ‘VV’ or as ‘P’. There are a total of 297,266 instances. Figures 2 and 3 show the proportion of the three proximal demonstratives following a verb or a preposition, respectively. In all periods and in both the post-verbal and post-prepositional positions, *sī* was the least used of the three. Notably, *cǐ* was present in substantial numbers in HC, lending evidence to the claims of Dobson (1964), Chan (1973), and Ushijima (1971) discussed in Section 5.1.

5.3.1 From HC to ST

The share of *shì*/*sī* progressively fell from the HC to ST periods, apparently being gradually replaced by *cǐ*. In the post-verbal position, they retreated from 60% in HC to 32% in ST (Fig. 3) (the difference is significant at $p = 9.7 \times 10^{-12}$ by χ^2 -test); in the post-prepositional position, from over 76% in HC

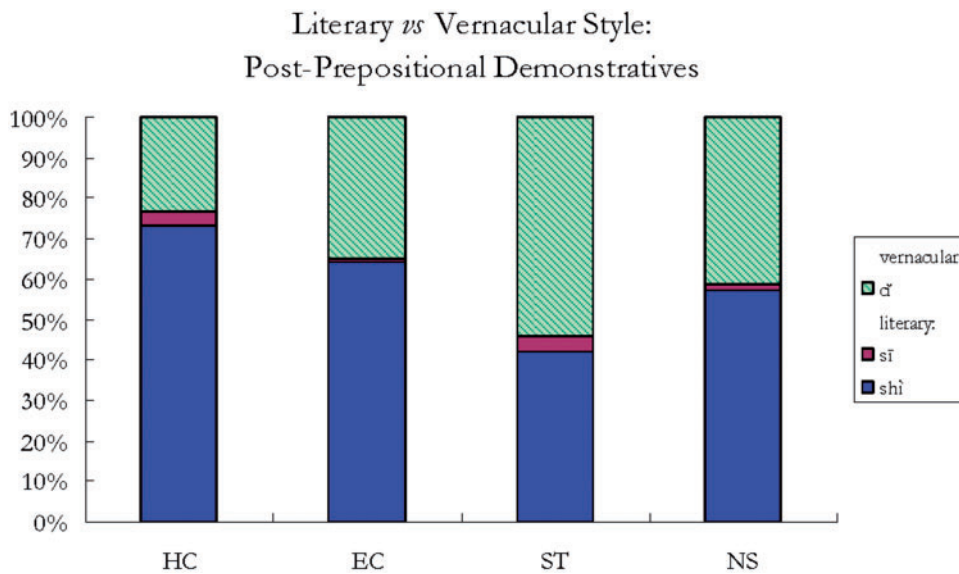


Fig. 2 Use of post-prepositional proximal demonstratives over different historical periods

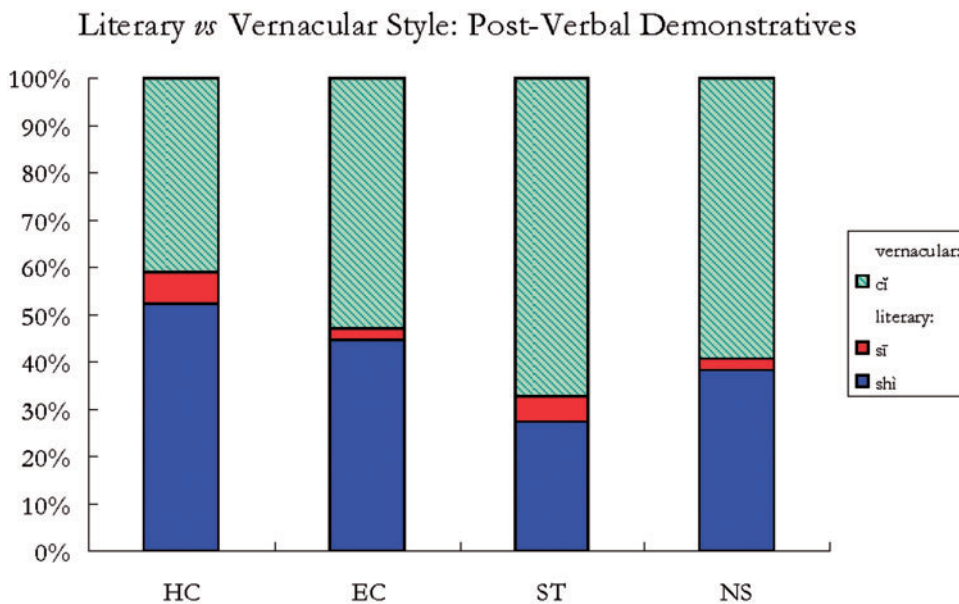


Fig. 3 Use of post-verbal proximal demonstratives over different historical periods

to 46% in ST (Fig. 2) (the difference is significant at $p = 4.0 \times 10^{-14}$ by χ^2 -test). By the time of the ST period, *cǐ* was already more frequently used than *shì*. Although the use of *cǐ* cannot be exclusively

attributed to the vernacular style, the overall trend of *shì* and *sī*, which are clearly the literary style, suggests again that Buddhist texts became more and more vernacular until the T'ang period.

We verified this general trend for each of the fifty most frequent nouns in the post-verbal position. Among these fifty nouns, 92% saw a decrease in the proportion of literary demonstratives from HC and EC to the ST period. Likewise, among the fifty most frequent nouns in the post-prepositional period, 80% experienced a similar decrease. Take the nouns *bǐqiū* 比丘 ‘bhikṣu’ and *xīn* 心 ‘heart’ as examples. The use of post-verbal *shì* for *xīn* and *bǐqiū* was 49 and 35%, respectively, in HC. Examples include:

- (9) 已得是心法義 (K363, from HC)
 Yǐ dé shì xīn fǎ yì.
 already obtain this heart dharma doctrine
 ‘Has obtained the dharma and doctrine of this heart already.’

- (10) 在是比丘兩邊而坐 (K1017, from HC)
 Zài shì bǐqiū liǎng biān ér zuò.
 at this bhikṣu two side CONN sit
 ‘Sit at two sides of this bhikṣu.’

The proportion of the literary style decreased for both *xīn* and *bǐqiū* in ST, the former to 38% and the latter to 19%. The following are examples from the

- (11) 卽說此心爲自性清淨 (K586, from ST)
 Jí shuō cǐ xīn wéi zìxìng qīngjìng.
 that.is speak this heart COP own.nature pure.and.clean
 ‘That is to say this mind is pure and clean in nature.’

ST period that adopted the vernacular style:

- (12) 我等今應問此比丘 (K388, from ST)
 Wǒděng jīn yīng wèn cǐ bǐqiū.
 1PL now should ask this bhikṣu
 ‘We should ask this bhikṣu now.’

5.3.2 From ST to NS

Similar to the case of the classifier, the NS period presented an unexpected development. During this period, the proportion of *shì/sī* increased again, to a

level closer to the period of EC. Specifically, it rose to 55% in the post-prepositional position (Fig. 2) (the difference is significant at $p=0.002$ by χ^2 -test), and to 40% in the post-verbal position (Fig. 3) (the difference is significant at $p=0.034$ by χ^2 -test). Again, since no previous research suggested a reversal of the vernacularization process in NS, this phenomenon is likely unique to the Canon, and may reflect the editing of the Stylists for the demonstratives.

As before, we verified the general trend for each of the fifty most frequent nouns in the post-verbal position. Totally 78% of these nouns saw an increase in the use of literary demonstratives from ST to NS. The corresponding figure for the fifty most frequent nouns in the post-prepositional position is 66%. Returning to our running example, the proportion of the literary style increased in NS for both *xīn* and *bǐqiū*, to 45 and 42%, respectively. Some example sentences include:

- (13) 但知是心。 (K1499, from NS)
 Dān zhī shì xīn.
 only know this heart
 ‘Only know this mind.’

- (14) 知是比丘爲護淨戒不傷草葉。 (K1495, from NS)
 Zhī shì bǐqiū wèi hù jìngjiè
 know this bhikṣu for.sake.of guard pure.commandments
 bù shāng cǎo yè.
 NEG harm grass leaf
 ‘(The king) knew that this bhikṣu did not hurt the grass and leaves for the sake of guarding the pure commandments.’

6 Nominal Sentence with Grammaticalized Adverbs

The emergence of the copula is a well-known phenomenon of the vernacularization process. To express ‘<noun1> is <noun2>’, the literary style does not use any copula and prefers the formulation ‘<noun1> zhě 者 <noun2> yě 也’. However, *zhě* or *yě* can be omitted (Peyraube, 1994); for example,

- (15) 那先者諸沙門師。 (K1002, from EC)
Nàxiān zhě zhū shāmén shī.
 Nāgasena PRT all śramaṇa teacher
 ‘Nāgasena was the teacher of a group of monks.’

In contrast, the vernacular style prefers using the copula *shì* 是, yielding the pattern ‘<noun1> *shì* <noun2>’, for example:

- (16) 爾時此母是我之親。 (K218, from EC)
Ēr shí cǐ mǔ shì wǒ zhī qīn.
 this time this mother COP 1SG GEN close.relative
 ‘At this time, this mother belongs to my family.’

For negative nominal sentences, the literary style uses the pattern ‘<noun1> *fěi* <noun2>’, while the vernacular style uses ‘<noun1> negator + *shì* <noun2>’, such as:

- (17) 現世罪人非佛弟子 (K761, from HC)
Xiànshì zuìrén fēi Fó dìzǐ.
 present.world offender NEG Buddha disciple
 ‘The offenders in the present world are not followers of Buddha.’

- (18) 藁非是蟲 (K22, from ST)
Gǎo fēi shì chóng.
 date NEG COP insect
 ‘A date is not an insect.’

- (19) 種姓不是婆羅門 (K651, from EC)
Zhǒngxìng bù shì Póluómén.
 caste NEG COP Brāhmaṇa
 ‘The class is not Brāhmaṇa.’

The copula *shì* can also be placed at the end of a sentence due to syntactic influences of the original Indic languages (Cao and Yu, 2000; Wu, 2014). Since this construction reflects the word order of the original language rather than a choice between the vernacular and the literary style, these nominal sentences are excluded from our study.

6.1 Previous research

It has been arguing the existence of copula in Archaic Chinese. While some scholars asserted that there was no copula (e.g. Wáng, 1980), others attempt to show that copulae have been frequently used as early as in pre-Classical Chinese (Peyraube, 1994). In any case, although the copula used in modern Mandarin, *shì*, was attested as early as in late Warring States period (3rd century BCE), how and when it emerged remains an open question in Chinese linguistics (Féng, 2003; Wang, 2003). In a quantitative study, Xiè (2006) counted the number of vernacular (i.e. copular) and literary (i.e. non-copular) constructions in two classics, *Zuǒzhuàn* 左傳 and *Shǐjì* 史記. The ratio of the literary construction decreased from 99.3% in *Zuǒzhuàn*, composed in late 5th century BCE, to 87.4% in *Shǐjì*, finished in 122 BCE. Lee and Kong (2016) found the ratio of the literary construction decreased from 22.6% in selected texts from Vimalakīrti Sūtra (circa 200 CE), to 4.8% in selected texts from Surangama Sūtra (circa 705 CE). Wang (2003) also compared different versions of translations of the Vimalakīrtinirdeśa-sūtra to study the emergence of the copula.

In a comparison between translations of the same Buddhist texts in different eras, Hú (1993) concluded that *shì* has become extensively used in nominal sentence in translations from the HC period, especially after an adverb. Some studies have focused specifically on negative nominal sentences. According to Táng (1992), *fēi shì* appeared as early as in Eastern Han period (25–220), and *bù shì* in the Three Kingdoms period (220–280). Two further trends were noted in a quantitative study on nine documents from the EC and ST period. First, *fēi shì*, which was two times more frequent than *bù shì* in the EC period, was supplanted by *bù shì* by the ST period. Second, from the Three Kingdoms period on, replacing *fēi* with *fēi shì* is fairly common. However, in Xiè (2012, p. 200)’s study, *fēi*-construction constitutes 80–90% among the negative nominal sentences, and its dominance did not weaken over time.

6.2 Data

A majority of negative nominal sentences in the literary style use the *fěi*-construction, i.e. with the adverb *fěi* 非 between two nouns, but without the copula. Most instances in the vernacular style use the *shì*-nominal construction with pre-modification by the negative adverbs *bù* 不, *fú* 弗, *fěi* 非, *wú* 無, *wù* 勿, and *wèi* 未 (Norman, 1988, p. 98–9; Xiè, 2012, p. 200).¹⁷ Hence, most negative nominal sentences can be reliably retrieved.

Positive nominal sentences, however, are more challenging to identify. Among the non-copular (literary) instances, only a small minority, such as (15), uses the markers *zhě* and *yě*; the others exhibit few linguistic cues. Most copular (vernacular) instances contain the word *shì* but, as explained in Section 5.2, due to its process of grammaticalization, one often cannot objectively determine whether it is serving as a copula or as a demonstrative, especially in the large number of bridge examples. The most common way to test whether a Chinese word is a verb is to ask whether it can be modified by an adverb; indeed, the distinguishing feature between the syntactic distribution of verbs and that of demonstratives is that a verb can be modified by an adverb but a demonstrative cannot (cf Chao, 1968, p. 767). The fact that *shì* is preceded by an adverb indicates the vanishing of the resumptive function of the demonstrative *shì* (Hóng, 1957); hence, pre-modification by an adverb implies that *shì* must be a copula rather than a demonstrative. We adopted this test in creating our data set of positive nominal sentences in the vernacular style. Specifically, we included all sentences with the *shì*-nominal construction with pre-modification by the adverbs *jí* 卽 and *nǎi* 乃 ‘namely, just’, for example:

- (20) 色卽是空。 (K20, from ST)
 Sè *jí* *shì* *kōng*.
 form **namely** cop empty
 ‘Form is just empty.’ (translation by
 Venerable et al., 2006)

These two adverbs were chosen because they are among the most frequent grammaticalized adverbs, and they impart relatively little change in meaning to

the sentence. Importantly, throughout all historical periods in our corpus, both the literary style (i.e. use these two adverbs alone) and the vernacular style (i.e. using them with copulae) are widely attested. The corresponding set of literary-style positive nominal sentences thus includes all sentences where the copula *shì* is absent, and the adverbs *jí* or *nǎi* stand alone (Norman, 1988, p. 95)¹⁸, such as:

- (21) 因緣卽空。 (K1499, from NS)
 Yīnyuán *jí* *kōng*.
 causes **Namely** emptiness
 ‘Causes is just emptiness.’

6.3 Analysis

Figure 4 shows the average proportion of literary (i.e. non-copular) versus vernacular (i.e. copular) style for negative nominal sentences in a text, based on 39,308 instances found in the Canon that fit the criteria explained above. Figure 5 shows similar statistics for nominal sentences with the grammaticalized adverbs *jí* and *nǎi*, based on 6,007 instances.

6.3.1 From HC to ST

As shown in Fig. 5, among nominal sentences with *jí* and *nǎi*, the literary style became less common from the HC (46%) to the EC (11%) period, mostly on account of the substantial drop in the use of *nǎi* alone and increase in *jíshì*. Between the EC and ST period, the proportion of the various vernacular and literary constructions remained largely the same (the difference between HC and ST is significant at $p = 0.009$ by χ^2 -test). Taking the noun *púsà* 菩薩 ‘bodhisattva’ as an example, the following sentence from the HC period employed the literary style:

- (22) 爾乃菩薩 (K77, from HC)
 Ěr *nǎi* *púsà*.
 2sg **EMPH** bodhisattva
 ‘You are a bodhisattva.’

By the EC period, most were written in the vernacular style, such as:

(23) 如是衆生乃是菩薩佛法因緣 (K76, from EC)

Rúshì zhòngshēng nǎi shì púsà fófǎ
 like.this living.beings EMPH COP bodhisattva buddhadharma
yīnyuán.
 causes
 'Thus, such living beings are (due to) bodhisattva, buddha-
 dharma and causes.'

The negative nominal sentence followed a rather different trend. As shown in Fig. 4, the share of the literary style is consistently much higher (94.5–96%) across all three historical periods than nominal sentences with *jí* and *nǎi*, corroborating the dominant position of the *fěi* construction observed by Xiè (2012). Second, from HC to ST, unlike all other constructions examined in this study, there was no drop in the literary style; in fact, there was a slight increase in usage of the *fěi*-construction. Among the vernacular style, *fěi shì* is the most common throughout all three historical periods; we observed only a slight gain of the *bù shì* construction that was discussed by Táng (1992).

6.3.2 From ST to NS

As shown in Fig. 5, there was an increase in the literary style for all adverbs from ST to the NS

period (the difference is significant at $p=0.006$ by χ^2 -test). The most substantial increase was observed for the adverb *jí*, from 15% in ST to 32% in NS, apparently at the expense of *jí shì*. An example for *nǎi* is shown below for our running example with *púsà*:

(24) 出此了知乃菩薩。(K991, from NS)

Chū cǐ liǎozhī nǎi púsà.
 beyond this thorough.knowledge EMPH bodhisattva
 'One who goes beyond this thorough knowledge is
 bodhisattva.'

The *fěi* construction, already dominant at 96% in the ST period, further increased to close to 99% in the NS period (Fig. 4) (the significance is at $p=0.218$ by χ^2 -test). Again, this phenomenon, against the current of the vernacularization process, may be best explained by editing work by the Stylists. It should be pointed out, however, that the increase in literary style appeared to have already begun earlier, in the EC period in the case of nominal sentences with grammaticalized adverbs, and in HC or earlier in the case of negative nominal sentences.

Literary vs Vernacular Style: Negative Nominal Sentences

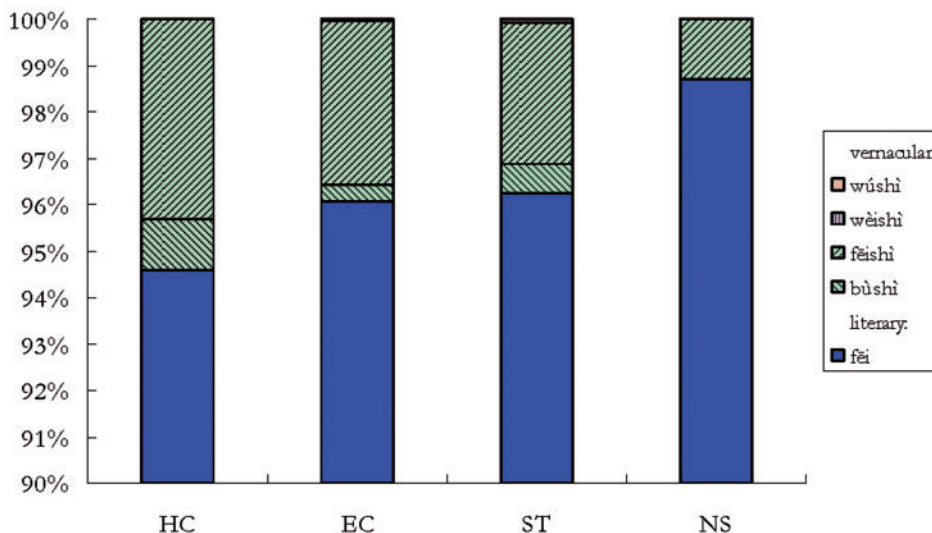


Fig. 4 Use of the copula *shì* in negative nominal sentences over different historical periods

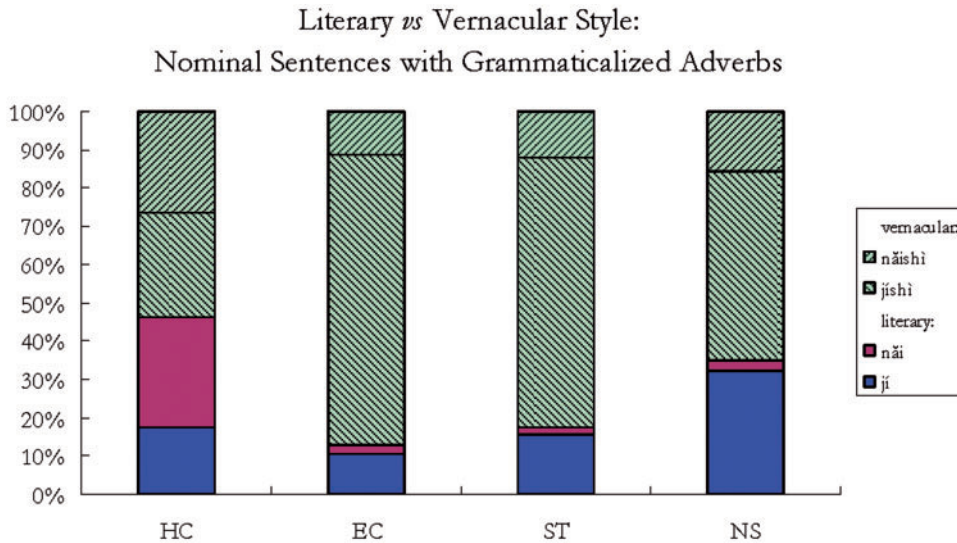


Fig. 5 Use of the copula *shì* in positive nominal sentences with grammaticalized adverbs over different historical periods

7 Conclusion

We have presented a quantitative analysis on linguistic changes in MC, focusing on the use of classifiers, demonstratives, and copulae. Most previous studies on this topic tend to emphasize qualitative analyses, based on manual examination of a relatively small number of example texts. In contrast, we leveraged automatic analysis to cover the breadth of the Chinese Buddhist Canon. This methodology, however, is not without its own challenges. In cases of linguistic ambiguity, such as the possible roles of *shì* as demonstrative and as copula, accurate and complete retrieval is difficult. We show how, by making judicious restrictions on its linguistic environment to disambiguate its usage, one can still reliably retrieve a large amount of samples.

Our results confirmed the tilt from the literary towards the vernacular style, as exemplified by the use of classifier, demonstrative, and copula, from the HC to the ST period. The proportion of quantified nouns with classifier increased from 3% to 8%; the use of the literary demonstratives, *shì/sī* fell 30%; among nominal sentences with grammaticalized adverbs, the proportion of the non-copular construction also fell more than 30%. The only exception was

negative nominal sentences, in which the literary style increased throughout this period. Based on over 40 million characters in the Chinese Buddhist Canon, this analysis is the largest corpus-driven study to-date on these three linguistic phenomena.

During the NS period, there was an unexpected retreat in the vernacular style. The proportion of the vernacular constructions for classifiers, demonstratives, and nominal sentences with grammaticalized adverbs all returned to their level in the HC and EC periods. This phenomenon likely reflected the impact of the Stylists, who were appointed by the imperial court to polish the Buddhist texts with more literary elements. Our analysis is the first to present evidence of their work, in particular the lexical replacement of proximal demonstratives, and the removal of classifiers and copulae.

For the study of language change, this article has illustrated the feasibility, even for a relatively low-resource language, of complementing qualitative analyses with quantitative evidence from large corpora. We have laid a foundation for future work on comparative analysis on other linguistic phenomena in MC to paint a fuller picture of its vernacularization process. It is hoped that similar approaches will be applied on the development of ancient Chinese in

other historical periods, as well as on other historical languages.

Funding

The work described in this article was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 155412).

References

- Bisang, W.** (1999). Classifiers in East and Southeast Asian languages: counting and beyond. In Gvozdanoviæ, J. (ed.), *Numeral Types and Changes Worldwide*. Berlin: Mouton de Gruyter, 113–85.
- Cáo, F.** (2010). *A Study of Measure Words in Táng and Wú dài Dynasties*. Doctoral dissertation, Nánkai University.
- Cao, G. and Yu, H.** (2000). The influence of translated later Han Buddhist Sutras on the development of the Chinese disposal construction. *Cahiers de Linguistique Asie Orientale*, 29(2): 151–78.
- Chan, S.** (1973). *Shishuōxīnyǔ yǔfǎ yánjiū* [A Study of the Syntax in Shishuōxīnyǔ]. Taipei: Student Book.
- Chao, Y. R.** (1968). *A Grammar of Spoken Chinese*. Berkeley; Los Angeles: University of California Press.
- Chu, C.** (1996). Vocabulary of Buddhist Sutras of the West Chin Dynasty. Technical Report. Taipei: National Chung Cheng University.
- Chu, C.** (1998). The Lexicology of Buddhist Sutra in Ancient China (II) — Three Kingdoms. Technical Report. Taipei: National Chung Cheng University.
- Chu, C.** (1999). The Lexicology of Buddhist Sutra in Ancient China (III) — Eastern Han Dynasty. Technical Report. Taipei: National Chung Cheng University.
- Dobson, W. A. C. H.** (1964). *Late Han Chinese: A study of the Archaic-Han shift*. Toronto: University of Toronto Press.
- Féng, S.** (2003). The copula in declarative sentences in Archaic Chinese. *Gǔhànyǔ Yánjiū* [Research in Ancient Chinese Language], 2003(1): 30–6.
- Gerner, M.** (2013). *A Grammar of Nuosu*. Berlin: Mouton de Gruyter.
- Harbsmeir, C.** (2015). Early Chinese buddhist translators on translation: a brief introduction with textual data. In Aussant, É. (ed.), *La Traduction Dans L’histoire Des Idées Linguistiques: Représentations Et Pratiques*. Paris: Librairie orientaliste Paul Geuthner, pp. 259–73.
- Hóng, C.** (1957). Lùn Nánběicháo yǐqián Hànyǔ zhōng de xìcí [On the copula in Chinese before the Northern and Southern Dynasties]. *Yǔyán Yánjiū* [Studies in Language and Linguistics], 1957(2): 1–22.
- Hú, X.** (1993). Cóng Jiūmóluóshí de fōjīng chóngyìběn yǔ yuányìběn de duìbǐ kàn xìcí “shì” de fāzhǎn [The development of the copula “shì”: contrasting the original translation with the re-translation by Kumārajīva]. *Journal of Social Science of Hunan Normal University*, 1993(3): 118–21.
- Hu, X., McLaughlin, J., and Williamson, N.** (2007). Syntactic positions of prepositional phrases in the history of Chinese: using the developing sheffield corpus of Chinese for diachronic linguistic studies. *Literary and Linguistic Computing*, 22(4): 419–34.
- Huang, C.** (1990). Sòngdài de Yījīng Rùngwénguān yǔ Fójiào [The Stylists translating Sutras in Sung Dynasty and Buddhism]. *The National Place Museum Research Quarterly*, 7(4): 13–31. <http://ccbs.ntu.edu.tw/FULLTEXT/JR-MAG/hs04.htm> (accessed 18 October 2017)
- Huang, C.** (1997). *Běisòng Fójiàoshǐ lùn’gǎo* [A Manuscript of the History of Buddhism in Northern Sung]. Taipei: Commercial Press.
- Huáng, S.** (1983). Xiānqín gǔhànyǔ zhǐshìcí yánjiū [A study in the demonstratives in Pre-Ch’in Chinese]. *Yǔyán Yánjiū* [Studies in Language and Linguistics], 1983(2): 136–57.
- Huáng, Z.** (1964). Cóng Jiǎwén, Jīnwén liàngcí de yìngyòng, kǎochá Hànyǔ liàngcí de qǐyuán yǔ fāzhǎn [A study in the origin and development of classifiers in Chinese by the use of classifiers in oracle bone inscription and bronze script]. *Zhōngguó Yǔwén* [Studies of the Chinese Language], 1964(6): 432–41.
- Kieschnick, J.** (2015). *A Primer in Chinese Buddhist Writings: Volume One: Foundations, Translation Key*. Stanford: Stanford University. Retrieved April 4, 2018, from <https://religiousstudies.stanford.edu/people/john-kieschnick/primer-chinese-buddhist-writings>.
- Karashima, S.** (1996). On vernacularisms and transcription in early Chinese Buddhist scriptures. In Zürcher, E. and Karashima, S. (eds), *Vernacularisms in Medieval Chinese texts*. Philadelphia: University of Pennsylvania, pp. 32–45.

- Kawabe, H.** (2011). *‘Shin’-yaku Hannyashingyou: Kokoro no kura wo hiraku kagi* [Translating Prajñā-pāramitā-hṛdaya-sūtra by Mind: The Key to Open the Chest of the Mind]. Tōkyō: Magazine House Company Limited.
- Kin, B.** (2010). *Kanbun tou Higashiajia* [Classical Chinese Literature and East Asia]. Tōkyō: Iwanami Shoten.
- Kudo, T.** (2005). CRF⁺⁺: yet another CRF toolkit. User’s manual and implementation available at <http://crfpp.googlecode.com/svn/trunk/doc/index.html>
- Lafferty, J., McCallum, A., and Pereira, F. C. N.** (2001). Conditional random fields: probabilistic models for segmenting and labeling sequence data. In Brodley, C. E. and Danyluk, A. P. (eds), *Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001)*. San Francisco: Morgan Kaufmann Publishers Inc, pp. 282–9.
- Lancaster, L. and Park, S.** (1979). *The Korean Buddhist Canon: A Descriptive Catalogue*. Berkeley: University of California Press. http://www.acmuller.net/descriptive_catalogue/ (accessed 18 October 2017).
- Lee, J. and Kong, Y. H.** (2016). A dependency treebank of Chinese Buddhist texts. *Digital Scholarship in the Humanities*, 31(1): 140–51.
- Li, C. N. and Thompson, S.** (1974). An explanation of word order change SVO < SOV. *Foundations of Language*, 12(2): 201–14.
- Li, C. N. and Thompson, S.** (1975). The semantic function of word order: a case study in Mandarin. In Li, C. N. (ed.), *Word Order and Word Order Change*. Austin: University of Texas Press, pp. 163–95.
- Liú, S.** (1965). *Wèijìnnánběicháo Liàngcí Yánjiū* [Studies of classifiers in the Wei, Chin, Northern and Southern Dynasties]. Peking: Zhonghua Book Company.
- Norman, J.** (1988). *Chinese*. Cambridge; New York, NY: Cambridge University Press.
- Peyraube, A.** (1994). Problems relating to the history of different copulas in ancient Chinese. In Chen, M. Y.-C. and Tzeng, O. J. L. (eds), *In Honor of William S.-Y. Wang: Interdisciplinary Studies on Language and Language Change*. Taipei: Pyramid Press, pp. 383–404.
- Peyraube, A.** (1998). Shànggǔ, Zhōnggǔ Hànyǔ liàngcí de lìshǐ fāzhǎn [The historical development of classifiers in Archaic Chinese and Medieval Chinese]. *Yǔyánxué Lùncóng* [Essays on Linguistics], 21: 99–122.
- Pulleyblank, E. G.** (1994). *Outline of Classical Chinese grammar*. Vancouver: University of British Columbia Press.
- Sen, T.** (2002). The revival and failure of Buddhist translations during the Song dynasty. *T’oung Pao*, 88: 27–80.
- Sun, C.** (1996). *Word-Order Change and Grammaticalization in the History of Chinese*. Stanford: Stanford University Press.
- Táng, Y.** (1992). Zhōnggǔ ‘shì’ zì pànduàn jù shù yào [A précis in shì-judgement sentences in Mediaeval Chinese]. *Zhōngguó Yǔwén* [Studies of the Chinese Language], 1992(5): 394–9.
- Ushijima, T.** (1971). *Kango bunpou ron: Chuuko hen* [Dissert on Chinese Syntax (Medieval Chinese)]. Tōkyō: Taishukan Publishing Company Limited.
- Venerable Y., Owens, M. C., and Romaskiewicz, P. M.** 2006. *Prajñāpāramitā Heart Sūtra*. California: Buddha’s Light Publishing.
- Wan, C.** (2008). Fódǎn hànyì liúchéng lǐ ‘guòdùxìng wénběn’ de yǔwén jǐngguān (Dì yī bù): Yījīng wéntí, yìcháng zǔzhī yǔ yījīng liúchéng [The linguistic landscape of transient text in the flow of Chinese Buddhist translations (part 1): genre of the translated text, organization of the translation forum and flow of translation]. *Satyābhisamaya*, 44: 103–42.
- Wáng, L.** (1980). *Hànyǔshǐ Gǎo* [A Manuscript on the History of the Chinese Language]. Peking: Zhonghua Book Company.
- Wang, S.** (2003). The historical evolution of ‘shì’ in ‘Wéi Mó Jié Suǒ Shuō Jīng’. *Kaohsiung Normal University Journal*, 15: 369–85.
- Wáng, W.** (1984). *Fódǎn Hànyì zhī yánjiū* [A Study of the Translation of Buddhist Texts]. Taipei: Heavenly Lotus Publishing Company Limited.
- Wáng, Y.** (2008). *Zhōnggǔ Fódǎn liàngcí yánjiū* [A Study of Classifiers in Buddhist Texts from the Medieval Period]. Doctoral dissertation, Nankang Normal University.
- Wei, P.** (2000). The position of the Eastern Han and Six Dynasties in the history of Chinese grammar. *Chinese Studies*, 18: 199–230.
- Wei, P.** (2004). *Hàn-Wèi Liùcháo chēngdàicí yánjiū* [A study of pronouns in Han-Wei Six Dynasties]. Taipei: Institute of Linguistics, Academia Sinica.
- Wong, T. and Lee, J.** (2016). A dependency treebank of the Chinese Buddhist canon. In Calzolari, N. (ed.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation*. Paris:

- European Language Resources Association, pp. 1679–83.
- Wu, C.** (2014). *Passive Sentences in Diamond Sūtra—Research on the Comparison of Sanskrit and Chinese Versions*. Master thesis, Graduate Institute of Religious Studies, Nanhua University.
- Xiè, Z.** (2006). A comparative research on declarative sentences in *Zuǒzhuàn* and *Shǐjì*. *Journal of Chongqing University of Arts and Sciences (Social Sciences Edition)*, 5(3): 45–9.
- Xiè, Z.** (2012). *Zhōngguóhànyǔ Pànduànjù Yánjiū* [Studies in Affirmative Sentence in Middle Chinese]. Chengdu: Bashu Publishing House.
- Zhao, H., Huang, C., and Li, M.** (2007). An improved Chinese word segmentation system with conditional random field. In Ng, H. T. and Kwong, O. O. Y. (eds), *Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing*. Stroudsburg, PA: Association for Computational Linguistics, pp. 162–5.
- Zhū, Q.** (2015a). ‘R wéi A suǒ jiàn V’-shì bèidòngjù de zuìzǎo shǐyòng niándài [The earliest attestation of the passive structure ‘R wéi A suǒ jiàn V’]. In Hóng, B. and Wú, F. (eds), *Linguistic Essays in Honor of Professor Tsu-Lin Mei on his 80th birthday*. Peking: Capital Normal University Press, pp. 339–91.
- Zhū, Q.** (2015b). Cóng Fójīng yīnyì zīliào kàn ‘wèi’ zì de zàoziyì [A look at the character-meaning of ‘wèi’ from the phonology and semantics as observed in Buddhist text]. In Xú, S., Liáng, X., and Sōng, J. (eds), *Fójīng yīnyì yánjiū: Dì Sān Jiè Fójīng Yīnyì Yánjiū Guójì Xuéshù Yántǎohuì lùnwénjí* [Studies in phonology and semantics of Buddhist texts: Proceedings of the Third International Conferences of Phonology and Semantics of Buddhist Texts]. Shanghai: Shanghai Lexicographical Publishing House, pp. 98–103.
- Zhū, Q.** (2016). Shìlùn Fójīng fānyì duì Hànyǔ hūcǐ ‘wèi’ zhī chǎnshēng de kěnéng yíngxiǎng [A preliminary study on the possible influence of the emergence of Chinese vocative “wèi”]. In Zhū, Q., Wāng, W., and Dōng, Z. (eds.), *Hànyǔ lìshǐ yǔyánxué de chuánchéng yǔ fāzhǎn* [The Inheritance and Development of Chinese Historical Linguistics]. Shanghai: Fudan University Press, pp. 443–505.
- Zhu, Q. and Chen, W.** (2008). The inessive structure in Archaic and Medieval Chinese: an evolutionary study of inessive demonstrative uses from Archaic to Early Modern Chinese. In Xu, D. (ed.), *Space in Language of China*. New York, NY: Springer International Publishing, pp. 249–66.
- Zhu, Q. and Chen, W.** (2013). Lùn Shàng, Zhōnggǔ Hànyǔ de “zhī-chù jiégòu”: Hànyǔ chùsuǒ zhīdài fāngshì cóng shànggǔ dào jìndài biànhuà de qiáliáng [On the inessive structure in Archaic and Mediaeval Chinese: the bridge of evolution of the inessive demonstrative uses from Archaic to Early Modern Chinese]. In Xú, D. (ed.), *Zhōngguó jìngnèi yǔyán de kōngjiān biǎodá* [Expressing Space by Languages in China]. Peking: World Book Inc, pp. 281–322.
- Zürcher, E.** (1977). Late Han vernacular elements in the earliest buddhist translations. *Journal of the Chinese Language Teachers Association*, 12(3): 177–203.

Notes

- 1 Translation by Kawabe (2011, p. 30).
- 2 The vernacular style is characterized by many other phenomena, e.g. the A-not-A form of polar questions, or pre-posed pronoun-object (e.g. *bù zhī shā* 不之殺 *lit.* ‘Not it kill’ versus *bù shā zhī* 不殺之 ‘Not kill it’). However, they are less frequent and therefore less suitable for this study.
- 3 In Chinese, *Yìjīngyuàn* 譯經院. Later renamed as *the Institute for the Transmission for the Dharma* (in Chinese, *Chuánfǎyuàn* 傳法院).
- 4 ‘On his majesty: ... order ... (a brunch of officials) ... always inspect. If there are taboo words, just polish at any time.’ (In original Chinese, 敕云：... 宜令太子太傅... 等，時爲看閱，有不隱便處，即隨時潤色。 , from T2502.)
- 5 Zhì Pán 志磐, dated 1258 – 69, gave two examples in the *Heart Sūtra*, where the clause ‘度一切苦厄’ was inserted, and the expression ‘是故’ was inserted into the clause as ‘是故空中...’ (In original Chinese, 如《心經》：「度一切苦厄」一句，元無梵本。又「是故空中.....」一句，「是故」兩字元無梵本。 , from T2035).
- 6 Their work was described as such: 良以筆受在其油素，文言豈無俚俗。儻不失於佛意，何妨刊而正之？ ‘It is very much that the translated texts are just the spoken words from the translator-scribe. How could it be possible that the texts are not vulgar? If one can remain faithful to Buddha’s message, why not delete and correct [the vulgarity]?’ (K1074).
- 7 Since no work in the 12th century is attested in the *Tripitaka Koreana*, likely due to warfare from the end of the Northern Sung Dynasty, and all three texts in the 13th century were compiled by Koreans, we excluded these two centuries from our study and labelled the last historical period as NS instead.

- 8 In contrast to modern Sinitic languages, classifiers in MC rarely appear with nouns that are modified by a demonstrative rather than a number. The use of classifiers without a noun is common in MC only for mensuration (e.g. *wǔshí lǐ* 五十里 ‘fifty miles’), and is excluded from our study. We also exclude analysis of classifiers for verbs, e.g. in the construction <verb> + <number> + <classifier>.
- 9 The pattern <noun>+<number>, while frequent in Pre-Classical Chinese, hardly occurred in MC.
- 10 E.g., *chàng liù yǒu* 鬯六卣 ‘sacrificial spirits: six jars’ (Huáng, 1964).
- 11 E.g., *huà bèi shí péng* 畫貝十朋 ‘colourful shells: ten pieces’ (Huáng, 1964).
- 12 E.g., *Yì shì èr tuán* 邑示二屯 ‘Yì examined two pairs (of scapulas)’ (Huáng, 1964).
- 13 *Jū èr bǐng* 車二丙 ‘chariot: two pieces’. This type is a marked construction. The main purpose of using classifier in the context of this type is only to actualize the boundary of nouns. The meaning is clear when the classifier is absent (*èr jū* ‘二車’ or *jū èr* ‘車二’). In the pre-Han period, this type was the exception rather than the rule—numerals normally occurred immediately in front of the noun (Norman, 1988, p. 115), for example, *sān bái shǐ* 三白豕 ‘three white pigs’ (from Huáng, 1964).
- 14 These ten classifiers are *zhǒng* 種, *zhòng* 衆, *lèi* 類, *pǐn* 品, *kē* 科, *bèi* 輩, *cáo* 曹, *děng* 等, *liú* 流, *mén* 門.
- 15 Originally proximal, but usages of distant demonstrative were attested in the HC period.
- 16 For example, in the sentence *Cǐ shì cháng fǎ*. 此是常法 ‘This is the constant law.’ (translation from Kieschnick, 2015) (K647, from EC). In addition, *shì* can also serve as adverb or sentence final particle.
- 17 Another frequently used copula, *wéi*, has only twenty-two instances of negator+*wéi*. In addition, in the word sequence negator+*wéi*, it is also possible that *wéi* functions as a preposition, which may lead to an overestimation of the presence of the copula, *wéi*. For these reasons, *wéi* was excluded from our study.
- 18 In this context, others view *jí* as a copula (Dobson, 1964, p. 72).