Imre Galambos
Chinese Character Variants in Medieval Dictionaries and Manuscripts

Abstract: This paper compares variant characters in large-scale dictionaries from the pre-modern period with actual writing habits using a special subset of variants known as 'semantic compounds' (*huiyi* 會意) as a case study. The results show that despite their prominent presence in traditional dictionaries, only a fraction of such variants were in everyday use. Most of the forms recorded in dictionaries were preserved and handed down as part of the lexicographic tradition, to some extent irrespective of genuine writing habits. Going one step beyond recognising that only some of the documented forms were at any given time in common use, the analysis presented here measures the discrepancy between dictionaries and manuscripts as a percentage.

Large Chinese dictionaries abound in orthographic variants.[1] In fact, it is commonly remarked that the majority of characters in large dictionaries, which contain in excess of forty or fifty thousand entries, are historical variants no longer in common use. The profusion of variants in medieval dictionaries creates the impression that they were relatively common in the pre-modern period, but when we look at surviving manuscripts, such as those from Dunhuang or Turfan, we do not necessarily see that the same character was written in several different ways. We only find two or three forms at most, rather than an abundance as suggested by traditional dictionaries. The question arises whether there is indeed a discrepancy between the lexicographic tradition and actual usage. After all, one would expect dictionaries to record character forms that were in daily use, even if one understands that they often had their own agenda. This paper uses a special subset of variants known as *huiyi* 會意 ('semantic compound') characters as a case study, to show that despite their prominent presence in traditional dictionaries, only a fraction of them were in everyday use. Most of the forms recorded in dictionaries were preserved and handed down as part of the lexicographic tradition, to some extent irrespective of genuine writing habits. The main issue is how representative pre-modern dictionaries

---

1 The initial idea behind this paper was published as a brief research note in Japanese; see Galambos 2012. I am grateful to Gábor Kósa for his valuable suggestions.

were of contemporary writing habits. Going one step beyond recognising that only some of the documented forms were at any given time in common use, the analysis presented here measures the discrepancy between dictionaries and manuscripts as a percentage.

1 Semantic compounds

This paper is primarily concerned with character forms consisting of two or more components, the semantic value of which is indicative of the meaning of the word written with the character in question. In traditional Chinese usage, these are usually labelled using the term *huiyi*, one of the six principles of character formation articulated as early as the first century CE. The scope of meaning of this principle and its historical role in the development of the Chinese writing system has been the subject of heated debate in anglophone scholarship for almost a century.[2] The debate, however, has been primarily concerned with the formative stage of the script, long before the development of the 'modern' script. This paper focusses on the medieval period when the primary medium of writing was paper, and the script had been in use for at least two millennia. The focus is not on how certain characters had been initially created but what permutations they underwent at a much later stage of their existence.

David P. Branner makes the distinction between 'real' *huiyi* characters, the components of which 'contribute abstractly to the overall meaning of the word represented', and the category he calls 'portmanteau characters', the components of which can be read 'as connected words to form a phrase that defines or denotes the word'.[3] Typologically, such portmanteau characters form a distinct subset of semantic compounds, most of which came into being long after the Qin-Han 秦漢 unification of the script. In fact, the character forms examined in this paper are mostly, although not exclusively, of such portmanteau type. Chinese scholars working on Dunhuang manuscripts call them 'newly created *huiyi* characters' 新造會意字, revealing that they see them as analogous to the original *huiyi* characters.[4] In this paper, the notion of semantic compounds encompasses both Branner's portmanteau characters and those that contribute to the meaning of the word more abstractly (i.e., 'real' *huiyi* characters).

---

2 For an overview of the problem, see Lurie 2006, Handel 2006 and Galambos 2011, 398–399.
3 Branner 2011, 73. Modern examples include characters such as *wai* 歪 ('askew, crooked'), the components of which 'spell out' the phrase *buzheng* 不正 ('not straight').
4 Zhang Yongquan 2015, 144 and Huang Zheng 2005, 30.

Often there is distinction between the structure of the standard character and its variant. A point in case is the character *jia* 假 ('false, fake'), the standard form of which is a combination of the semantic component *ren* 亻 ('man') and the phonetic component *jia* 叚 ('borrow'). By contrast, its variants include the form 奡, a combination of the components *bu* 不 ('not') and *zhen* 真 ('true'), neither of which plays a phonetic role but, together, the two of them gloss the meaning of the word (i.e., *bu zhen* 不真, 'not real').[5] Therefore, even if the 'mother' character has an ordinary *xingsheng* 形聲 ('form and sound') structure, in which one of the components plays an obvious phonetic role, in the variant form both components have a semantic function.

By focussing specifically on variants that appear in manuscripts, I intend to move away from a perspective that sees characters as abstract units of the script and view them in light of the writing habits of the people who wrote them in their daily routine. It is an approach that favours concrete archaeological evidence over derivative forms of data. The basic premise of the current study is the comparison of manuscript evidence with data preserved in pre-modern dictionaries. As commonly recognised, comprehensive dictionaries often treated variants found in earlier dictionaries as part of an inherited tradition, adapting them wholescale, but also further decontextualising them. Yet there is little understanding of the extent of this phenomenon, of the proportion of variants transmitted primarily in the lexicographic tradition.

The justification for choosing *huiyi*-type variants (as opposed to other types of variants) for the current exercise is threefold. First, *huiyi* characters are a category that has generated a considerable amount of discussion in English-speaking academia and are therefore already in the focus of attention. Secondly, they are highly visible forms which stand out among other variants with their inspired structure. In comparison with other types of variants, where the difference to the standard form may lie in the addition or omission of a stroke or a cursive simplification of parts of the character, the *huiyi*-type variants exhibit a wealth of fanciful and curious forms. Thirdly, it is certain that people at the time were very much aware of the creative possibilities in interpreting the structure of characters. The various traditional methods of dissecting and interpreting characters (e.g., *cezi* 測字, *chaizi* 拆字) show how prominent such kind of approaches were.[6]

---

5 Unsurprisingly, this gloss is a close match to the definition of the character 假 in the *Shuowen jiezi* 說文解字, which explains it as *fei zhen* 非真 ('not real').

6 On the phenomenon of 'character manipulation', see Schmiedl 2020.

2 *Huiyi*-type variants in dictionaries

Pre-modern dictionaries include a plethora of *huiyi*-type character forms, even if they are not necessarily identified as such. For the sake of the current exercise, a pool of variants is selected from four pre-modern dictionaries: 1) *Sisheng pianhai* 四聲篇海; 2) *Longkan shoujian* 龍龕手鑑; 3) *Jiyun* 集韻; and 4) *Yupian* 玉篇.[7] All are large-scale compilations in excess of 20,000 characters, which also means that they record many non-standard variants. In fact, the larger the dictionary, the higher the percentage of such variants, as the core pool of standard characters remains approximately the same. The reason for choosing more than one dictionary is to obtain a snapshot of the lexicographic tradition in general, rather than exploring the idiosyncrasies of a particular work.

The variants in the sample pool were chosen randomly, by combing through the dictionaries and identifying relevant forms. Although this is an admittedly subjective process of selection, it results in a repertoire of largely random examples. A comprehensive coverage of all relevant variants in available dictionaries would have been a monumental task and is not the aim here. Instead, the objective is to rely on a smaller body of sample data to assess the differences between dictionaries and manuscripts. A criterion when selecting the variants was that only the variants have a *huiyi* structure but not the mother character. Thus, I did not include characters such as *fei* 朏 ('light of the crescent moon'), with variants such as 胜 (月+小+生 'moon-small-born'), because the standard character 朏 itself is an unmistakable *huiyi*-type combination in which the two components (月+出) are indicative of the meaning of the word. I did select, however, several borderline cases in which the variants could be interpreted as having both a *huiyi* and *xingsheng* structure, as long as their standard character included a clearly articulated phonetic component (e.g., 仙, 妃, 念, 妊 and 聽).

Table 1 lists the sample group, which contains 37 variants of 27 characters. The rightmost column indicates their source, even if many of them appear in more than one dictionary. While the exact distribution of such variants in various dictionaries may be of interest with regards to the sources utilised by

---

7 For the *Sisheng pianhai*, I am using the *Chenghua dinghai chongkan gai bing wuyin leiju sisheng pianhai* 成化丁亥重刊改併五音類聚四聲篇海 edition (1467); for the *Longkan shoujian*, the *Xu guyi congshu* 續古逸叢書 edition with a preface from 997; for the *Jiyun*, the *Zhiyuan gengyin chongkan gai bing wuyin jiyun* 至元庚寅重刊改並五音集韻 (1290); for the *Yupian*, the Ming-dynasty edition of the *Daguang yihui Yupian* 大廣益會玉篇 in the collection of the Harvard Yenching Library.

individual lexicographers, from the point of view of the current study it is sufficient to know that they occur in at least one of the four dictionaries.

Table 1: Examples of *huiyi* variants in the four dictionaries.

Standard character	Variant	Components	Source
大 'great'		不+小 'not-small'	*Sisheng pianhai*
思 'think'		用+心 'use-mind'	*Sisheng pianhai*
老 'old'		先+人 'prior-man'	*Sisheng pianhai*
吹 'blow'		口+風 'mouth-wind'	*Sisheng pianhai*
洟 'nasal mucus'		弟(<涕)+鼻 'mucus-nose'	*Sisheng pianhai*
齋 'fast'		不+食 'not-eat'	*Sisheng pianhai*
假 'fake'		不+真 'not-real'	*Sisheng pianhai*
仙 'immortal'		人+大+山 'man-great-mountain'	*Longkan shoujian*
		人+大+止 'man-great-recess'	*Longkan shoujian*
		人+止 'man-recess'	*Longkan shoujian*
		入+山 'enter-mountain'	*Longkan shoujian*
光 'light'		日+火 'sun-fire'	*Longkan shoujian*
		火+化 'fire-transform'	*Longkan shoujian*
帆 'sail'		舟+風 'boat-wind'	*Longkan shoujian*
		風+舟 'wind-boat'	*Longkan shoujian*
計 'plan, scheme'		供+言 'provide-words' (人+共+言 'men-together-talk')	*Longkan shoujian*
蠶 'silkworm'		天+虫 'heavenly-worm'	*Longkan shoujian*
		天+天+虫 'heavenly-heavenly-worm'	*Longkan shoujian*

Table 1 (continued).

Standard character	Variant	Components	Source
		神+虫 'divine-worm'	*Longkan shoujian*
聽 'listen'		耳+忍 'ear-endure'	*Longkan shoujian*
		耳+用 'ear-use'	*Longkan shoujian*
爛 'bright light'		大+明 'great-bright'	*Longkan shoujian*
念 'recite'		言+心 'speak-mind'	*Longkan shoujian*
娠 'pregnant'		女+身 'woman-body'	*Longkan shoujian*
些 'some'		少+大 'few-large'	*Jiyun*
炒 'fry'		取+火 'take-fire'	*Jiyun*
飢 'hungry'		食+乏 'food-shortage'	*Jiyun*
衄 'nosebleed'		血+鼻 'blood-nose'	*Jiyun*
飲 'drink'		冷+水 'cold-water'	*Jiyun*
妊 'pregnant'		任+女 'undertake-woman'	*Jiyun*
		女+任 'woman-undertake'	*Sisheng pianhai*
		任+身 'undertake-body'	*Sisheng pianhai*
辯 'eloquent'		巧+言 'artful-speech'	*Yupian*
妃 'concubine'		肥+女 'plump-woman'	*Jiyun*
長 'chief'		上+人 'above-person'	*Yupian*
嚏 'sneeze'		至+鼻 'reach-nose'	*Yupian*
期 'period of time; expect'		大+日 'great-day'	*Yupian*

The table shows that in many cases the components can be read together into meaningful phrases, glossing the meaning of the word written with the character. To take the first three entries from the table, the word *da* 大 ('great') is glossed by the variant's components as 'not-small' 不小; the word *si* 思 ('think') as 'use-mind' 用心; and the word *lao* 老 ('old') as 'prior-man' 先人. For this reason, the location of components in the overall composition matters, as it determines the sequence in which they are to be read.

The four dictionaries list many more similar variants in addition to the ones showcased here. Note that the components in many cases do not define the meaning of the word unambiguously and the structure only works if the reader is aware of what it intends to represent. For example, the components 不+食 ('not-eat') in the variant of the character *zhai* 齋 ('fast') could theoretically signify other possible meanings (e.g., 'starve', 'hungry', 'ill'). Therefore, the semantic value of the components is only one of the factors that contribute to the legibility of the variant. Just as important is the graphic semblance to the referent character, the preservation of its overall shape and symmetry. In addition, syntactic context and even social setting play a role in ensuring that readers recognise the variant.

As mentioned above, some of the variants can be interpreted as both *huiyi* and *xingsheng* structures. Although modern scholarship prefers to explain these as *xingsheng* combinations, the phonetic motivation of a component does not automatically disqualify it from also playing a semantic role. In many character forms, one of the components may perform a double function, even if one of these is more pronounced. Of the three variants of the character *xian* 仙 ('immortal') attested in the *Longkan shoujian* (Table 1), 㒰 (人+大+山 'man-great-mountain') and 仚 (入+山 'enter-mountain') include the component *shan* 山 ('mountain'), which is the phonetic component in the standard character. But the phonetic component is absent altogether from the form 仸 (人+大+止 'man-great-recess'), making it a pure *huiyi* composition. In contrast, the two variants that include the component 山 can be considered a *xingsheng* and *huiyi* structure at the same time. It is not hard to notice, however, the graphic link between the variants, demonstrating the role of the overall shape of the mother character in constraining the scope of possible variation.

In other cases, the variant may introduce a phonetic element that is not part of the standard character. For example, according to the *Shuowen jiezi* 說文解字, the phonetic component in the character *nian* 念 ('recite') is *jin* 今 ('now'). The *Longkan shoujian*, however, records the variant 言 (言+心 'speak-mind'), in

which the component *yan* 言 ('speak') plays a weak phonetic role, at the same time also having an evident semantic function.[8] In some cases, a component in a *huiyi*-type variant enhances the phonetic structure of the character, as it can be seen in the form 夵 (上+人 'above-person') listed in the *Yupian* as a variant for the character *zhang* 長 ('chief'). Although the two components work well together as a gloss, the component *shang* 上 ('above') also plays a phonetic role, improving the phonetically opaque structure of the standard character.

Similarly, the *Shuowen jiezi* identifies the component *chen* 辰 ('celestial bodies') as the phonetic component in the character *shen* 娠 ('pregnant'), but the *Longkan shoujian* lists the variant 娭 (女+身 'woman-body'), which replaces 辰 with the component *shen* 身 ('body'). This new component improves the standard character's phonetic transparency, compensating for the shift in pronunciation over time, but at the same time introduces a semantic dimension absent from the original component. As a result, it could be argued that in this variant the component 身 plays both a semantic and phonetic role.

There are, however, cases where the function of the component is less transparent. For example, traditional dictionaries record the variants 姙, 婂 and 雋 for the character *ren* 妊 ('pregnant'). While in the standard character the component *ren* 壬 ('ninth heavenly stem') clearly plays a phonetic role, in the variants it appears as *ren* 任 ('undertake a task'), and in this context its semantic value 'bear, undertake' is surely relevant. This is also likely to be the reason for using 任 rather than the purely phonetic 壬. At the same time, the words written with the characters 壬, 任 and 妊 have long been recognized as being related.[9] Accordingly, the component 任 in the variants merely highlights the existing semantic function of the component 壬 and thus the standard character itself could also be regarded as having both a *xingsheng* and a *huiyi* structure.

Despite these examples, it is just as common that the variant does not retain the *xingsheng* structure of the standard character. Even though the characters *chui* 吹 ('blow'), *zhai* 齋 ('fast'), *jia* 假 ('fake'), *chao* 炒 ('fry'), *ji* 飢 ('hungry') and *nü* 衄 ('nosebleed'), to name just a few, are all stereotypical *xingsheng* configurations, their variants listed in Table 1 feature a semantic-only structure with no phonetic motivation. The example of the variant 齈 of the character *nü* 衄

---

8 Interestingly, the structure 意 is also attested in the Ming-dynasty dictionary *Zhengzi tong* 正字通 as a variant of the character *yi* 意 ('intention'), although here the component 言 plays no phonetic function.

9 The *Shuowen jiezi*, for example, links 壬 with the other two characters. With regards to the character 妊, it notes that the component 壬 had a phonetic function as well, suggesting that it also had a semantic one.

('nosebleed') is particularly interesting because it replaces the phonetic component *chou* 丑 ('second earthly branch'), which had lost some of its transparency, with the semantically relevant 鼻 (*bi* 'nose'). This is a complete shift from a *xingsheng* combination to a semantic-only structure.

Naturally, the structure of medieval *huiyi*-type variants is not directly indicative of the historical development of the characters. Encoding accurate historical information from earlier stages of the script's history was clearly not an objective. Instead, the variant forms are creative interpretations exploiting the graphic possibilities inherent in the orthography and playfully reordering the characters' structure. *Huiyi*-type variants are invariably derivatives of more mainstream forms, whether those were officially recognized standard characters or merely popular forms in common use. In most cases there is a graphic connection with the mother character and the *huiyi* combination preserves a resemblance to the shape and overall symmetry of the original character.

3 *Huiyi*-type variants in manuscripts

Extracting examples of *huiyi*-type structures from medieval manuscripts is a time-consuming task. Even though manuscripts abound in non-standard forms, *huiyi*-type variants in particular are not easy to find. A pragmatic way to accelerate the search is to use dictionaries of variant characters based on pre-modern manuscripts. A straightforward choice for locating variants in Dunhuang manuscripts is the *Dunhuang suzidian* 敦煌俗字典, a dictionary of non-standard variants in the Dunhuang corpus.[10] Without question, this is an extremely useful tool, although unfortunately not large enough to paint a comprehensive image of contemporary scribal habits. As it references about 280 manuscripts, it only documents the most common variants. For this reason, even if a particular form is not attested in the dictionary, it does not necessarily mean that it does not occur in the manuscripts.

A significantly larger collection of variants is the *Koryŏ Taejanggyŏng ichʻe chajŏn* 高麗大藏經異體字典 (hereafter: *Koryŏ chajŏn*), a massive dictionary of character variants from the Korean Buddhist Canon.[11] Although the dictionary is based on texts printed in Koryŏ (Goryeo) in the first half of the thirteenth century and is thus removed from the manuscript tradition of medieval China in both

---

10 Huang Zheng 2005.
11 Yi Kyu-gap et al. 2000.

time and space, the woodblocks versions ultimately (through several incarnations) derive from Buddhist manuscripts copied in China prior to the Mongol period. Even though, as a dictionary, the *Koryŏ chajŏn* cannot be taken as directly representative of the Chinese manuscript tradition, in practical terms it is the best available resource for locating variants used in pre-modern Buddhist manuscripts. The utility of the dictionary derives from its breadth, as it includes all non-standard variants in a vast corpus. In sum, the *Dunhuang suzidian* is more relevant for our examination but is limited in its scope, whereas the *Koryŏ chajŏn* is statistically more reliable but several steps removed from medieval Chinese manuscript culture.

Both the *Dunhuang suzidian* and the *Koryŏ chajŏn* document concrete examples of variants from medieval manuscripts and prints, rather than data extracted from earlier dictionaries. They are purely descriptive in their approach, in contrast to the prescriptive agenda of many pre-modern dictionaries. This makes them especially valuable for comparing variant forms used in medieval scribal culture with those preserved in the lexicographic tradition. Using these two resources considerably speeds up the process of locating examples of *huiyi* structures. Of the original 37 variants in Table 1, only 13 can be found in the *Dunhuang suzidian* or the *Koryŏ chajŏn*; see Table 2.[12] The table records the location where the variant occurs: for the Korean Canon, it is the letter 'K' followed by the number of the text in the canon (e.g., K-1257); for the Dunhuang manuscripts, it is the pressmark.[13]

---

12 In some cases, we also find *huiyi*-type variants which are not in our original list (Table 1). For instance, under the character *ting* 聽 ('listen') the *Dunhuang suzidian* lists a variant consisting of the components *er* 耳 ('ear') and *e* 惡 ('bad'), neither of which serves a phonetic function. Although this variant does not occur in any of the four traditional dictionaries, similar forms are attested in other early dictionaries; for example, the Southern Song scholar Lou Ji's 婁機 (1133–1211) *Hanli ziyuan* 漢隸字原 records the same form occurring in Han-period stone inscriptions.
13 'DY' stands for Dunyan 敦研 (i.e., Dunhuang Research Academy 敦煌研究院); 'DB' for Dunbo 敦博 (i.e., Dunhuang Museum 敦煌博物館); 'S' for the Stein collection at the British Library; 'P' for the Pelliot chinois collection at the Bibliothèque nationale de France.

Table 2: *Huiyi*-type variants from Table 1 featuring in the *Koryŏ chajŏn* and the *Dunhuang suzidian*.

Standard character	Variant from Table 1	Koryŏ chajŏn	Dunhuang suzidian
老 'old'	㐸	㐸 K-1257	� DY.234 else DY.302 else DY.042
洟 'nasal mucus'	鼻	鼻 鼻 鼻 鼻 K-1498	
仙 'immortal'	㑪 㑪 仚 仚 仚	仚 仚 K-1257 仚 K-1498	仚 DY.029 仚 S.6825 仚 S.388
帆 'sail'	颿 颿	颿 K-1063 颿 K-1498	
蠶 'silkworm'	蚕 蚕 蝅 蝅 塩 塩 聉	蚕 蚕 K-1257 蝅 蝅 K-1257 蝅 K-0801 塩 K-0391	蝅 P.3833 蝅 S.134 蝅 S.3491 塩 S.388
聽 'listen'	聉		聉 S.388
炒 'fry'	聚	聚 K-1063 聚 K-1498	
飲 'drink'	渁	渁 K-1498	
辯 'eloquent'	誓	誓 K-0394	誓 S.530 誓 S.779
嚔 'sneeze'	鼻	鼻 鼻 K-1498 鼻 K-1063	

Not all variants aligned in the table are identical. For example, in the case of the character *xian* 仙 ('immortal'), the top component of the fourth variant in the *Longkan shoujian* is *ru* 入 ('enter'), whereas in the Korean and Dunhuang examples it is *ren* 人 ('man'). But the graphic similarity of the forms across the columns is undeniable. In general, the table demonstrates that whenever we find analogous matches, they are structurally very similar, regardless of the type of script or hand in which they appear. This attests to the stability of variant forms through time and space, as some of them emerge centuries apart in geographically distant regions (e.g., Dunhuang vs. Korea). The remarkable thing about this phenomenon is that most of the *huiyi*-type variants do not represent the usual way of writing a given character, as it is somewhat counter-intuitive that scribal cultures would sustain and transmit idiosyncratic forms that do not serve a practical function. Some of them have more strokes and thus take longer to write than the original characters (e.g., writing 羰 in place of 浼), revealing that the formation of variants was not necessarily motivated by convenience.

Taking a step further and examining the source of the variants recorded in Table 2, we can see that even within the pool of forms ostensibly collected from manuscripts, some examples derive from dictionaries. The 23 variants from the Korean Tripitaka in Table 2 are extracted from the following texts:

10 examples	K-1498	*Yiqie jing yinyi* 一切經音義, in 100 *juan*, compiled by Huilin 慧琳 during 788–807
7 examples	K-1257	*Xinji Zangjing yinyi suihan lu* 新集藏經音義隨函錄, in 30 *juan*, by Kehong 可洪 during 931–940
3 examples	K-1063	*Yiqie jing yinyi* 一切經音義, in 25 *juan*, compiled by Xuanying 玄應 after 649
1 example	K-0801	*Zhengfa nianchu jing* 正法念處經, in 70 *juan*, tr. by Gautama Prajñāruci during 538–541
1 example	K-0391	*Guoqu zhuangyan jie qianfo ming jing* 過去莊嚴劫千佛名經, in 1 *juan*, tr. by unknown translator during the Liang 梁 dynasty (502–557)
1 example	K-0394	*Wuqian wubai Foming shenzhou chuzhang miezui jing* 五千五百佛名神呪除障滅罪經

Three of these six texts are in fact dictionaries included in the Buddhist Canon and transmitted along with other texts. All three feature in their title the term *yinyi* 音義 ('sound and meaning'), referring to a type of work that provides phonetic and semantic glosses for words and terms found in Buddhist scriptures. Of

these, Huilin's *Yiqie jing yinyi* (K-1498) is a monumental work in 100 *juan* that takes up two large volumes in the modern edition of the Korean Canon. Although the thousands of variants assembled in such dictionaries supposedly derived from the medieval manuscript tradition, they also include a cumulative repertoire of forms which may not have been in common use.

Overall, we can see that out of the 23 variants from the *Koryŏ chajŏn*, only three (i.e., 蝅 from K-0801, 蟴 from K-0391 and 暜 from K-0394) come from non-lexicographic texts. Two of these are variants of the character 蠶 and consist of the components 天+虫 ('heavenly-worm') and 神+虫 ('divine-worm'). Most of the Korean variants in Table 2, however, come from various *yinyi*-type Buddhist dictionaries and thus should not be taken as representative of pre-modern writing habits. Unfortunately, removing variants attributed to *yinyi*-type dictionaries does not fully solve the problem because we can never be sure that a given variant does not *also* occur in a non-lexicographic text. As the *Koryŏ chajŏn* only records one source for each variant form, the inclusion of data from pre-modern lexicographic works potentially conceals examples of the same form occurring in other texts.

Seen in the rightmost column in Table 2 are the 13 variants occurring in the Dunhuang corpus. These come from the following manuscripts:

3 examples	S.388	*Zhengming yaolu* 正名要錄, etc.; d. u.
1 example	S.6825	*Laozi daode jing* 老子道德經 with the Xiang'er 想爾 commentary; *c.* 6th c. copy
1 example	S.134	*Shijing* 詩經, 'Qiyue' 七月; d. u.
1 example	S.3491	*Pinposaluo wanghou gong cainü gongde yi gongyang ta shengtian yinyuan bian* 頻婆娑羅王后宮綵女功德意供養塔生天因緣變; 9–10th c. copy
1 example	P.3833	Poems of Wang Fanzhi 王梵志; dated *bingshen* 丙申 year (936?)
1 example	DY.042	*Miaofa lianhua jing* 妙法蓮華經, tr. by Kumārajīva; 4–5th c. copy
1 example	DY.234	*Daban niepan jing* 大般涅槃經, tr. by Dharmakṣema 曇無讖; 4–5th c. copy
1 example	DY.302	*Weimojie suo shuo jing* 維摩詰所說經, tr. by Kumārajīva; 4–5th c. copy
1 example	DY.029	*Daban niepan jing* 大般涅槃經, tr. by Dharmakṣema; 4–5th c. copy

| 1 example | S.530 | *Da Tang Shazhou Shimen Suo falü Yibian heshang xiu gongde ji bei* 大唐沙洲釋門索法律義辯和尚修功德記碑 |
| 1 example | S.779 | *Da Fan Shazhou Shimen jiaoshou heshang Hongbian xiu gongde ji* 大蕃沙洲釋門教授和尚洪辯修功德 |

The two last items in the list record two forms of the character 辯, both of which essentially consist of the components 巧+言 (or 功+言 'merit-speak'), matching the form recorded in Table 1 as coming from the *Yupian*. These last examples are from manuscript copies of commemorative inscriptions and occur in the names of high-ranking monks Suo Yibian 索義辯 (793–869) and Hongbian 洪辯 (d. 868). Indeed, their names feature frequently in the Dunhuang corpus and are consistently written with this variant. In addition, the same variant is also commonly, but not always, used in other names. There are, however, also examples where the same variant is used to write the word *bian* 辯 ('eloquent; debate').

In the above list, three forms derive from manuscript S.388 with the *Zhengming yaolu*. In reality, this work is only one of several lexicographic lists copied onto the same scroll. One of our examples comes from the *Zhengming yaolu*, whereas the other two come from a short list of *huiyi*-type variants. Thus, once again, these three examples are based on dictionaries, rather than ordinary texts, and should be excluded from our pool of 'genuine' cases. The remaining 10 forms come from 10 different texts. Some of these forms, however, effectively have the same orthographic structure, which is why they represent only five orthographically distinct structures of four characters:[14]

老 (*lao* 'old')	𦒱 (先+人 'prior-man')
仙 (*xian* 'immortal')	仳 (人+止 'man-recess')
	仚 (人+山 'man-mountain')
蠶 (*can* 'silkworm')	蠺 (天+天+虫 'heavenly-heavenly-worm')
辯 (*bian* 'eloquent')	訬 (巧+言 'artful-speech')

Observing the dates of the manuscripts which contain the variants, we can see that variants 𦒱, 仳 and 仚 all come from manuscripts pre-dating the Tang period

---

14 The reason for this is that the *Dunhuang suzidian* often includes forms that differ from each other in terms of their calligraphic traits (e.g., balance, stroke width) but not necessarily in orthographic structure.

(i.e., DY.029, DY.042, DY.234, DY.302 and S.6825). The variant 蚕, however, appears in later manuscripts, probably dating to the ninth or tenth century. Of these, S.134 is possibly the earliest but not earlier than the beginning of the Tang. As expected, the variant 蚕 also appears in the Korean subset, which is closer in time to the majority of Dunhuang manuscripts. The overlap corroborates that this specific variant was indeed in common use in pre-modern scribal culture and is not a dictionary-only form. The only Korean variant not in the Dunhuang set is 蝅, a variant of the character 蠶. Overlaying the two sets from Korea and Dunhuang, we end up with six variants of only four characters:

老	𠁔
仙	仚, 仚
蠶	蚕, 蝅
辯	䛒

Therefore, of the 37 *huiyi*-type variants listed in Table 1 at the beginning of this paper, only these six can be linked with actual writing habits, while the rest appear to be chiefly perpetuated in the lexicographic tradition. In other words, six out of 37 (i.e., 16%) of the variants randomly chosen from traditional dictionaries are likely to have been in common use.

4 The control group

Naturally, the analysis presented above has its limitations. One of these is that the 37 variants (of 27 characters) are only a small subset of the total number of variants preserved in pre-modern dictionaries. It is likely that some variants not included in Table 1 may have been relatively common. To correct the bias of our original data set, another set of 37 variant forms was chosen as a control group, hoping to improve the accuracy of the results of the above analysis. To obtain a different set of random variants, I went through the initial lines of the *Qianziwen* 千字文 (Thousand Character Text), looked up the variants of each character in the four pre-modern dictionaries and selected the ones consisting of components which together were indicative of the meaning of the word normally written by the given character. From the point of view of the dictionaries, the sequence of characters in the *Qianziwen* is entirely random, but the characters are all commonly used ones.

As before, the chosen variants did not have to belong exclusively to the *huiyi* category, they could also be interpreted simultaneously as *xingsheng* combinations. Among the criteria for selecting variants was that their components had to contribute to the semantic value of the character unambiguously. Another criterion was to exclude variants which contained the entire mother character and thus the added other component merely disambiguated or amplified the meaning of the word (e.g., 懲 < 常, 嘆 < 悲). Table 3 lists the 37 variants of 21 standard characters chosen sequentially from the *Qianziwen*.

Table 3: Control group of 37 additional *huiyi*-type variants from the four dictionaries.

Standard character	Variant	Components	Source
天 'heaven'		一+先 'one-prior'	*Yupian*
		佛+國 'Buddha-kingdom'	*Sisheng pianhai*
地 'earth'		山+水+土 'mountain-water-soil'	*Yupian*
		土+水+山 'soil-water-mountain'	*Sisheng pianhai*
		水+土 'water-soil'	*Jiyun*
洪 'flood'		水+山+竝(=並) 'water-mountain-merge'	*Longkan shoujian*
		水+生+生+竝(=並) 'water-grow-grow-merge'	*Longkan shoujian*
歲 'year; harvest; Jupiter'		山+火+火 'mountain-fire-fire'	*Sisheng pianhai*
		山+木+山 'mountain-tree-mountain'	*Longkan shoujian*
		山+收 'mountain-harvest'	*Longkan shoujian*
陽 'Yang, sun, southern slope'		阜+火 'abundance-fire'	*Sisheng pianhai*
騰 'gallop'		炎+頁+馬 'flame-head-horse'	*Longkan shoujian*
雨 'rain'		宀+水 'cover-water'	*Longkan shoujian*
水 'water'		井+泉 'well-spring'	*Sisheng pianhai*

Table 3 (continued).

Standard character	Variant	Components	Source
昆 'mass, many; brothers'		四+弟 'four-younger brothers'	*Sisheng pianhai*
		豕+豕+虫+虫 'pig-pig-worm-worm'	*Sisheng pianhai*
岡 'mountain ridge'		山+四 'mountain-four'	*Longkan shoujian*
		山+四+止 'mountain-four-stop'	*Sisheng pianhai*
珍 'precious'		西+玉 'western-jade'	*Sisheng pianhai*
國 'kingdom'		囗+王 'enclosed-king'	*Longkan shoujian*
		囗+八+方 'enclosed-eight-directions'	*Jiyun*
		囗+八+土 'enclosed-eight-lands'	*Yupian*
罪 'crime'		自+辛 'self-painful'	*Jiyun*
		自+卒 'self-die'	*Longkan shoujian*
愛 'love'		心+及 'heart-reach'	*Sisheng pianhai*
體 'body'		身+本 'body-foundation'	*Yupian*
		身+豊(<豐) 'body-plentiful'	*Jiyun*
		身+骨 'body-bones'	*Sisheng pianhai*
歸 'return'		去+來 'leave-come'	*Longkan shoujian*
四 'four'		二+二 'two-two'	*Yupian*
五 'five'		又+一 'another-one'	*Longkan shoujian*
惟 'think'		佳+心 'good-mind'	*Longkan shoujian*
得 'gain'		見+寸 'see-inch'	*Jiyun*
信 'trust'		人+口 'person-mouth'	*Yupian*
		言+心 'speak-heart'	*Jiyun*
		言+小 'speak-small'	*Jiyun*
		子+心 'child-heart'	*Sisheng pianhai*

As before, comparing the variants of the control group with data in the *Koryŏ chajŏn* and *Dunhuang suzidian* results in a significantly smaller set of variants. If, in addition, we exclude lexicographic data, then we are left with six variants of only four characters occurring in similar form in either, or both, of these two sources.[15] Table 4 shows these variants along with their sources.

Table 4: *Huiyi*-type variants from Table 3 matching non-lexicographic variants in the *Koryŏ chajŏn* and the *Dunhuang suzidian*.

Standard character	Variant from Table 3	*Koryŏ chajŏn*	*Dunhuang suzidian*
地 'earth'	坔	坔 K-0934 坔 K-1058	坔 P.2151
岡 'mountain ridge'	崗	崗 K-0648	
國 'kingdom'	囯 圀	囯 K-0056 圀 K-0934	囯 F-96 圀 P.2151
體 'body'	躰 軆	躰 K-0549	躰 F-96 軆 Dunyan-119 體 Zhedun-027

The variants 囯 (<國) and 躰 (<體) are similar, although not identical, to the modern simplified forms 国 and 体, currently in use in Mainland China and Japan. The reason for this is that such forms, and other similar ones, remained popular for many centuries over a vast geographical area and were eventually chosen as substitutes for their more complex standard characters. That both

---

15 I also excluded as lexicographic data the so-called *guwen* 古文 ('ancient script') variants handed down as part of the *Liguding Shangshu* 隸古定尚書 (S.799). The reason for this is that this text transcribes pre-Qin (i.e., pre-221 BCE) character forms into the 'modern' script and effectively preserves them as a specialised set of palaeographic data. The *guwen* variants certainly did not represent the way people normally wrote at the time when the manuscripts were written.

Japan and then China – independently – chose the same forms attests to them having been in widespread use throughout East Asia.

It is conspicuous that two of the six variants are character forms introduced during the reign of Wu Zetian 武則天 (r. 690–705). These are 埊 (山+水+土), used in place of the character *di* 地 ('earth'), and 圀 (囗+八+方), a variant of the character *guo* 國 ('kingdom').[16] As these forms had a political and religious motivation and, at least initially, were primarily used within her reign, it is questionable whether they can be taken as representative of everyday writing habits in the medieval period. They did not, however, disappear altogether and were transmitted to other parts of East Asia, including Korea, Japan and the Nanzhao 南詔 kingdom, where some of them remained part of written culture for centuries. If we include the Wu Zetian forms in the results, then the portion of variants in actual use comprises 16% of the control group, matching the results obtained from our initial group. If we exclude them, we are left with only four variants, representing 11% of the group. In either case, the control group's results are entirely consistent with those of the initial group.

5 Conclusions

This paper assessed to what extent character variants documented in traditional dictionaries reflected actual writing habits in the medieval period. As part of the exercise, 37 *huiyi*-type variants were chosen from four pre-modern dictionaries. These were then matched against the *Dunhuang suzidian* and the *Koryŏ chajŏn* to evaluate the degree of overlap between data preserved in traditional dictionaries and roughly contemporary scribal habits. Although a portion of the variants had corresponding forms in the *Koryŏ chajŏn* and the *Dunhuang suzidian*, many of the matches occurred in dictionary-type works that had been incorporated into the Buddhist Canon or transmitted along with other types of manuscripts. Naturally, these variants were not directly representative of the way people wrote in real life. After eliminating such 'lexicographic noise', only six variants of four characters (i.e., 老, 仙, 蠶 and 辯) remained. Consequently, of the original 37 variants, only six (16%) seemed to have been in actual use.

In the second step, to validate the results of the above analysis, I assembled a control group of another 37 variants. This second analysis similarly yielded six variants (16%) of four characters (i.e., 地, 岡, 國 and 體). As two of these vari-

---

16 On the origin of each character, see Bottéro 2013.

ants (i.e., 埊<地 and 圀<國) were character forms introduced during the reign of Wu Zetian, they primarily occurred in manuscripts and inscriptions dating to the end of the seventh and beginning of the eighth centuries, when she was on the throne. They did not fully disappear, however, but remained in use in various parts of East Asia, which is why they should not be dismissed outright. If we exclude these two forms from the results, we are left with only four variants, comprising 11% of the initial pool of variants in the control group. Whether we go with 11% or 16%, the results of the control group are fully in accord with those obtained from the first group of variants, attesting to the statistical reliability of the data.

In sum, the analysis presented in this paper shows that only a fraction (11%–16%) of the *huiyi* variants preserved in traditional dictionaries were in daily use. Other types of variants are likely to exhibit a similar pattern, even if specific percentages may be slightly different. Clearly, lexicography had its own tradition, and large dictionaries, aspiring to gather a comprehensive set of character forms ever in existence, relied just as much on earlier dictionaries as on epigraphic and manuscript sources. A variant featuring in an influential dictionary became part of the lexicographic tradition and from there on was likely to be included in large dictionaries. Lexicographic knowledge was cumulative and precedence-based, gradually incorporating even relatively rare variants.

References

Bottéro, Françoise (2013), 'Les graphies énigmatiques de l'impératrice Wu Zetian', *Études chinoises*, 32/2: 67–99.

Branner, David Prager (2011), 'Portmanteau Characters in Chinese', *Journal of the American Oriental Society*, 131/3: 73-82.

Galambos, Imre (2011), 'Popular Character Forms (*súzì*) and Semantic Compound (*huìyì*) Characters in Medieval Chinese Manuscripts', *Journal of the American Oriental Society*, 131/3: 395–409.

Galambos, Imre (2012), 'Shahon no jittai kara mita jisho kijutsu: Kaii moji wo rei to shite' 写本の実態から見た字書記述一会意文字を例として, in Ishizuka Harumichi 石塚晴通 (ed.), *Kanji jitaishi kenkyū* 漢字字体史研究, Tokyo: Bensei shuppan, 303–315.

Galambos, Imre (2020), *Dunhuang Manuscript Culture: End of the First Millennium* (Studies in Manuscript Cultures, 22), Berlin: De Gruyter.

Handel, Zev (2016), 'Does Xǔ Shèn's *huìyì* Category Reflect Historical Reality? An Argument for the Existence of Compound Graphs Lacking Phonophorics', in Ding Bangxin et al. (eds), *New Horizons in the Study of Chinese: Dialectology, Grammar and Philology*, Hong Kong: Chinese University of Hong Kong, 581–598.

Huang Zheng 黃征 (2005), *Dunhuang suzidian* 敦煌俗字典, Shanghai: Shanghai jiaoyu chubanshe.

Lurie, David B. (2006), 'Language, Writing, and Disciplinarity in the Critique of the "Ideographic Myth": Some Proleptical Remarks', *Language and Communication*, 26/3–4: 250–269.

Schmiedl, Anne (2020), *Chinese Character Manipulation in Literature and Divination: The Zichu by Zhou Lianggong (1612–1672)* (Prognostication in History, 3), Leiden: Brill.

Yi Kyu-gap 李圭甲 et al. (2000), *Koryŏ Taejanggyŏng ich'e chajŏn* 高麗大藏經異體字典, Seoul: Koryŏ Taejanggyŏng Yŏn'guso.

Zhang Yongquan 張涌泉 (2015), *Dunhuang suzi yanjiu* 敦煌俗字研究, Shanghai: Shanghai jiaoyu chubanshe.