

# Medieval Ways of Character Formation in Chinese Manuscript Culture\*

Imre Galambos

*University of Cambridge, UK*

Traditional Chinese scholarship understood the principles of character formation according to the six scripts (*liu shu* 六書) model initially set forth in Eastern Han sources towards the end of the first century CE. Although initially these categories were not intended as etymological principles, in later times they were also used to explain the origin and early development of the script. Even some modern models of the origin of Chinese writing, which generally stem from the critique of traditional views, rely on the concept of *liu shu* as they try to determine the actual number of principles at play during the formative stages of the script. All of these models, however, seem to carry the assumption that writing had been created in the distant past and then creation essentially stopped. This paper is an attempt to demonstrate that character creation, that is, the development of orthographic structure, was an ongoing process that involved a number of principles beyond the traditional *liu shu* categories. Along the same line of thought, I am trying to draw attention to the value of interpreting character forms in terms of their medieval structure, rather than disregarding what we see in an effort to find out what an archaic structure might have been at the time the character first came into being.

**Keywords:** Chinese characters, *liu shu*, medieval manuscripts, character creation, orthographic structure

---

\*I would like to thank Yum Jungsam for her valuable comments on the previous version of this paper, as well as participants of the 2012 Scripta workshop who contributed to the discussion. Likewise, I am extremely grateful to Gábor Kósa and the anonymous reviewers of *Scripta* for their numerous comments and suggestions.

## 1. Existing models of the development of Chinese writing

In the “Postface” of the *Shuowen jiezi* 說文解字, Xu Shen 許慎 (CE 58?–147?) listed the six categories and provided two examples each. To support his argument, he quoted the *Zhouli* 周禮 [Rites of Zhou] where the term *liu shu* was first used, even though there it referred not to the principles of orthographic structure but to six types or styles of writing.<sup>1</sup> Regardless whether in the *Zhouli* these meant the regional scripts of the six largest states of the Warring States period or simply six kinds of calligraphic styles, it is quite clear that they did not denote the compositional principles described by Xu Shen. Surely, this sort of quoting out of context for the sake of establishing a connection with an authoritative text was not unusual in early China where the Classics had a weight that far exceeded the literal meaning carried by the texts themselves. In any case, Eastern Han scholars like Xu Shen adopted the concept of *liu shu* and filled it with new meaning in order to explain the orthographic structure of characters. Xu Shen listed the following six categories:

- (1) *zhishi* 指事, ‘pointing at things’, e.g. 上 (*shang*; ‘above’) and 下 (*xia*; ‘below’);
- (2) *xiangxing* 象形, ‘depicting forms’, e.g. 日 (*ri*; ‘sun’) and 月 (*yue*; ‘moon’);
- (3) *xingsheng* 形聲, ‘form and sound’, e.g. 江 (*jiang*; ‘[Yangtze] River’) and 河 (*he*; ‘[Yellow] River’);
- (4) *huiyi* 會意, ‘joining ideas’, e.g. 武 (*wu*; ‘martial, war’) and 信 (*xin*; ‘trustworthy, trust’);
- (5) *zhuanzhu* 轉注, ‘commenting by rotation’, e.g. 考 (*kao*; ‘deceased, august’) and 老 (*lao*; ‘old’);
- (6) *jiajie* 假借, ‘adopting and borrowing’, e.g. 令 (*ling*; ‘order’) and 長 (*zhang*; ‘leader’).<sup>2</sup>

A parallel version of the same account also appears in the “Yiwenzhi” 藝文志 chapter of the *Hanshu* 漢書 compiled by the Eastern Han scholar Ban Gu 班固 (32–92 CE) whose list was similar but not identical to that in the “Postface” of the *Shuowen jiezi*.<sup>3</sup> In the “Yiwenzhi,” however, no

<sup>1</sup> Galambos 2006: 56–57. For a general discussion of the *liu shu* in English, see *Ibid.*: 54–61 and Bottéro 1998.

<sup>2</sup> *Shuowen jiezi*: 314.

<sup>3</sup> *Hanshu* 30.10: 1720–1721.

examples are provided and thus the categories are even more difficult to interpret than those described by Xu Shen. At the end of the list, however, the “Yiwenzhi” significantly adds the statement *zaozi zhi ben ye* 造字之本也, which can be translated either as ‘[these are] the basis of character creation’ or ‘[these are] the origins of character creation.’ In either case, the use of the verb *zao* 造 (‘to create’) may signify a shift towards interpreting the *liu shu* not as a description of the structure of existing characters in the Eastern Han period but as the principles according to which they once had been created. Although theoretically it is possible that this concluding statement in the “Yiwenzhi” should be interpreted in the present tense, that is, this is how people today (i.e. in the Eastern Han) should create characters when they learn and write them, the verb *zao* is more likely to indicate an *ab initio* creation of character structure.

With regard to the use of the phrase *zaozi* and its diachronic changes in time it is interesting to note that in the Han period the work of Cangjie 倉頡 as the inventor of the script was almost always expressed using the phrase *zuoshu* 作書 ‘made writing’. Thus the “Postface” of the *Shuowen jiezi* has ‘When Cangjie first created writing...’ 倉頡之初作書;<sup>4</sup> the *Zhonglun* 中論 [Balanced Discourses] (“Zhixue” 治學 [Ordering Learning]) writes that ‘Cangjie created writing by looking at the footprints of birds’ 倉頡視鳥跡而作書;<sup>5</sup> and the *Lunheng* 論衡 [Discussions on Balance] (“Yixu” 異虛 [Fictitious anomalies]) makes the enigmatic claim that ‘When Cangjie created writing, heaven rained grains and the ghosts wailed at night’ 蒼頡作書, 天雨穀, 鬼夜哭.<sup>6</sup> But by at least Qing times Cangjie was known as the ‘Sage who invented writing’ (*zaozi shengren* 造字聖人), and the alleged location of this legendary deed as the ‘Place where Cangjie invented writing’ (*Cangjie zaozi chu* 倉頡造字處).<sup>7</sup>

A third Eastern Han source, Zheng Zhong’s 鄭眾 (CE 5–83) commentary to the *Zhouli*, provides yet another version of the *liu shu* set, in which some of the names differ from both Xu Shen’s and Ban Gu’s. Zheng Zhong’s commentary shows how the term *liu shu* ran a full circle from having been

<sup>4</sup> *Shuowen jiezi*: 314.

<sup>5</sup> *Zhonglun* 1: 13.

<sup>6</sup> *Lunheng* 5.18: 212. In fact, in early Chinese texts, and especially the *Lunheng*, this phrase is in about half of the cases used in connection with Cangjie.

<sup>7</sup> A stele with these five characters was erected in 1784 at the Cangjie shrine near the city of Linfen 臨汾 in southwestern Shanxi.

adopted from the *Zhouli* and reinterpreted in a new context to being used in this new sense to explain the meaning of its original context.

While the six categories listed in the three Eastern Han sources are not identical, at least on the surface they appear to refer to the same set of principles.<sup>8</sup> Following the Han, scholars used these six categories to explain and discuss the orthographic structure of characters and they continue to be used even today. Nevertheless, it is not entirely clear what some of them actually meant. Especially the *zhuanzhu* category, the name of which appears with surprising consistency in the three Eastern Han sources, has been problematic—partly because so few characters are identified as belonging to this group that it is nearly impossible to prove any interpretation with conclusiveness.<sup>9</sup> It is also not clear why the system was set up in such a way that out of the total number of approximately nine thousand characters in the *Shuowen jiezi*, some of the categories (i.e. *zhishi*, *huiyi* and *zhuanzhu*) would have only a handful examples, whereas others (i.e. *xingsheng*) have thousands. There have been ongoing disputes about the nature of some of the principles and about which category particular characters belonged to.

Discussions about the structure of characters led to differences in understanding whether these principles were at work at the initial creation of characters (i.e. by Cangjie) or only later on. Thus in his *Liu shu shuo* 六書說 [Explanation of the six scripts] the Qing scholar Jiang Sheng 江聲 (1721–1799) noted that the *liu shu* ‘did not begin in Zhou times but at the initial stage of character creation’ 不始於周，而始於造字之初。<sup>10</sup>

As the result of archaeological discoveries of the first decades of the twentieth century and the introduction of Western linguistic theories, palaeographers began reconsidering the *liu shu* theory. One of the most influential models is the *san shu* 三書 (‘three scripts’) theory which sees Chinese characters as consisting of only three categories. The original version of the *san shu* theory was advanced in 1935 by Tang Lan 唐蘭 (1901–1979) whose influences, in addition to recent archaeological discoveries, included both Western linguistics and the indigenous tradition

<sup>8</sup> For a comparison of the three versions of the *liu shu*, see Bottéro 1998.

<sup>9</sup> Serruys 1957 provides an extensive overview of how the term *zhuanzhu* was understood by scholars in the past. More recent studies, in some respects superceding that of Serruys, are Sun 1991 and Zhong 2007.

<sup>10</sup> Cited in Tang 2005: 56.

of philological scholarship. He proposed that the early development of Chinese characters should be understood as having evolved from pictographs (*xiangxingzi* 象形字) to “ideographs” (*xiangyizi* 象意字) and phonetic compounds (*xingshengzi* 形聲字).<sup>11</sup> Later on, Chen Mengjia 陳夢家 (1911–1966) pointed out some of the weaknesses in Tang’s theory and advanced a modified model of *san shu* which consisted of pictographs, phonetic loans and phonetic compounds.<sup>12</sup> Finally Qiu Xigui 裘錫圭 suggested further modifications to replace the category of pictographs with that of “semantographs” (*biaoyizi* 表意字).<sup>13</sup> It is clear, however, that even the *san shu* model ultimately derives from the *liu shu* and even the term itself was coined in reference to the original *liu shu*.<sup>14</sup>

In his important monograph *The Origin and Early Development of the Chinese Writing System*, Professor William G. Boltz also identifies three stages in the early development of the script: (i) the zodiographic, (ii) the multivalent, and (iii) the determinative, demonstrating that these stages were more or less at play at the birth of other *ex nihilo* writing systems of the world.<sup>15</sup> This three-stage model to some extent overlaps with the three categories of Chinese palaeographers, although there are also some differences. For example, while one type of multivalent graph essentially matches the traditional category of phonetic loans (*jiatie*), Boltz identifies yet another derivative type which he calls ‘homosemous or parasemantic use of a graph.’ According to this derivative process, ‘a zodiograph that is already conventionally associated with one word may be used to write a second word the *meaning* of which is readily suggested by the *depictive quality* of the graph itself, regardless of any phonetic similarity or cognate relation between the two words.’<sup>16</sup>

Thus, even though (at least in popular discourse in China) the *liu*

<sup>11</sup> Tang 1935. See also, Tang 2005: 60–63.

<sup>12</sup> Chen 1956: 75–83.

<sup>13</sup> Qiu 1988: 106. For a history of the *san shu* theory, see *Ibid.*: 104–107.

<sup>14</sup> Naturally, there have been many critiques of the *liu shu* theory before the 20<sup>th</sup> century. Similarly, the *san shu* model in its different stages presented here is merely one of many advanced during the 20<sup>th</sup> century. For an overview, see Dong et al. 2007, vol. 9 “Liu shu yanjiu” 六書研究 [The Study of the *Liu shu*]; for a cultural history of *liu shu*, see, for example, Yue 2008.

<sup>15</sup> Boltz 1994: 52–72.

<sup>16</sup> *Ibid.*: 62.

*shu* remains a common way of talking about the structure of Chinese characters, modern scholarship is gradually moving away from this model towards models that better account for the growing corpus of archaeological material. As part of this shift, the six Eastern Han categories are often criticized as not reflecting the historical derivation of particular characters, which in itself is of course indicative of an approach that expects them to do so. Some of Xu Shen's explanations of character structure in the *Shuowen jiezi* are identified as mistaken, under the assumption that since he was so far removed from the formative stages of the script he could not have known as much as we do today thanks to archaeological discoveries of the past century. Yet this critique of Xu may be misguided, since we still do not fully understand Xu Shen's motivation for compiling the *Shuowen* and etymological considerations seems not to have been his major concern; it is entirely possible that he intended some of his explanations to account for contemporary character structure, rather than the earliest historical origin.<sup>17</sup>

We therefore find an ongoing tension in academic discourse between an interpretation of the *liu shu* as principles according to which Chinese characters were created at the formative stage of the script, and as descriptions of their actual orthographic structure at a particular point in time. In light of this tension, analysis of character structure in medieval manuscript texts can be a useful corrective to many of the unstated assumptions in the field, providing a fresh perspective for the analysis of Chinese character origin and structure.

## 2. Medieval character structures

Analyses concerning the structural makeup of Chinese characters often focus on the formative period of the script. The unstated assumption behind this is that once characters came into being, the process of creation essentially stopped. Consequently, any further change is simply a deviation from the "original" forms which had come into being according to an initial set of principles. This is also the reason why it is common practice

---

<sup>17</sup> For analyses of the structure and ordering principles in the *Shuowen*, see Bottéro 1998 and Bottéro and Harbsmeier 1998.

among palaeographers to trace character forms back to a presumed “original” form whenever trying to explain their structure. But the process of looking for the original form is often arbitrary and extends back in time only as far as it is convenient for the argument at hand. Sometimes the original form is found in the small seal script, sometimes on bronze inscriptions and sometimes only in the oracle-bone inscriptions.

My argument is that we need a mechanism for interpretation of the structure of a character in its existing state, the way it appeared to contemporary readers of manuscripts or inscriptions. In the process of being used, characters continue to change and this change is often governed by rules that may be largely independent of the principles that had been at play during the formative stages of the writing system. When we see the character 明 (*ming*; ‘bright’) clearly composed of the components 日 (*ri*; ‘sun’) and 月 (*yue*; ‘moon’) in a Warring States manuscript, an analysis of its structure need not depend on tracing it back to a bronze inscription form consisting of 囧 (*jiōng*; ‘bright’) and 月 (*yue*; ‘moon’). Its structure can be usefully interpreted based on the way it was actually written in that particular time and place. Today, under the influence of Western linguistics and its overwhelming emphasis on the phonetic aspects of writing, we all too often try to identify a phonophoric (“sound-bearing”) component of characters, and if we find none, we go back in time until we encounter an older form of the character that appears to have one. But this method is inherently arbitrary as it is governed by preconceived ideas about the nature of the Chinese script (or of writing systems in general).

While there is obviously merit in uncovering the path of the historical development of characters and thus shedding light on their earlier forms, we should also recognize the importance of the form in which they appear in archaeological material, which to some extent may reflect how people at the time understood their own script. This approach takes into consideration, and gives weight to, the orthographic diversity we typically find in archaeological material.

The traditional approach to orthographic structure (often termed ‘character etymology’) is somewhat analogous to earlier philological methodology which tried to locate a text’s Urtext, or original version. Such an attitude to textual transmission presupposed that once a text had been created by its original author, any further changes meant the corruption of the initial version and thus the modern scholar’s ultimate goal was to

reconstruct that version. In recent years, however, these linear/treelike approaches have increasingly come under criticism, as more nuanced models of textual transmission, which do not believe in a point zero at the beginning of a text's history but instead take into account the multitude of surviving witnesses and try to interpret the enormous diversity of textual variation.<sup>18</sup> Accordingly, textual or orthographic variation is not necessarily interpreted as a corruption of the original text but is recognized as potentially meaningful in the edition or manuscript where it appears. A superb example of such an analysis in the field of early Chinese texts is Matthias L. Richter's new study of the *Min zhi fumu* 民之父母 bamboo slip manuscript from the Shanghai Museum collection, in which he shows how a close attention to the textual variants in a manuscript against parallel versions in other sources may reveal valuable information about the cultural and socio-historical background of that particular version.<sup>19</sup>

While such approaches are increasingly gaining credit in textual criticism, they are still largely absent from narratives of the history of writing. Far from trying to deny the existence of a diachronic or evolutionary dimension in the history of the Chinese script, this paper merely aims to draw attention to the significance of a synchronic dimension and thereby contribute to a more balanced model.<sup>20</sup>

### 3. Character formation as seen through archaeological material

Whenever we examine the extreme variety of character forms that appear in archaeological material, we will inevitably find orthographic structures that fall outside the modern *san shu* or the traditional *liu shu* models. Such character forms are often explained as mistakes (*wu* 誤), corruptions (*e* 訛) or in some other way as derivative of more orthodox forms at the hands of people with an insufficient level of literacy or working in a careless

---

<sup>18</sup> For a critique of linear models of textual criticism and an exposition of the synchronic approach advocated in this paper, see for example, Cerquiglini 1999. For a similar attempt to interpret Chinese manuscripts from Dunhuang, see Nugent 2010.

<sup>19</sup> Richter 2013.

<sup>20</sup> Similarly, efforts to restore the "original" text have not disappeared from textual studies. For example, a convincing argument regarding the significance of reconstructing the original text of some New Testament texts is presented in Ehrman 2006: 307–324.

manner. This approach explains a manuscript character form in light of another form that may not even exist in contemporary archaeological material, essentially favoring an abstract or presumed entity over a tangibly present one.

In his *Wenzixue gaiyao* 文字學概要, after discussing the *san shu* theory, Professor Qiu makes the point that the three-category model cannot describe all of the characters because in the course of their development, the characters underwent changes that fall outside the scope of the three categories.<sup>21</sup> He writes that the model can only describe their original structure, not the way they actually look. I propose to take this idea further and apply it to the enormous diversity of character forms seen in archaeologically retrieved manuscripts and inscriptions, in contrast to the characters in edited texts transmitted down to our age through a tradition of copying. This method allows me to look at a much wider orthographic spectrum and account in a more satisfying way for variant forms and idiosyncratic usage. Likewise, I can interpret the structure of an archaeological character form without having to trace it back to another form that is more orthodox or more in line with traditional models of character formation.

Even though I believe that the principles underlying this approach are theoretically valid for any time period that produced texts written with Chinese characters, in order to maintain a controlled linguistic environment, I will refrain from citing examples that are hundreds or even thousands of years apart and will limit my observations to a specific time period. This will be the 7th–10th centuries, a period from which we have a significant corpus of material from the northwestern peripheries of the Chinese domain, namely, Dunhuang and Turfan.<sup>22</sup> Since I am only showing here a limited number of examples, I will restrict my sample corpus to the Dunhuang manuscripts.

---

<sup>21</sup> Qiu 1988: 107–109.

<sup>22</sup> Dunhuang and Turfan are cities in northwestern China. Both of them were oasis cities along the Silk Road and were important cultural centers during the Tang period. At the beginning of the 20<sup>th</sup> century, a local Daoist monk accidentally found a sealed cave with an enormous collection of manuscripts written in Chinese, Tibetan and a dozen and a half of other languages. This collection was subsequently dispersed and today parts of it are located in public and institutions around the world. Similarly, the tombs and caves in the region of Turfan yielded significant, albeit smaller, discoveries, and new items continue to be excavated.

In addition, I will only list several types of character formation, rather than providing a closed set with a specific number, be it three, six or fifteen. This way we anticipate that the list will grow as new categories are identified and we will not have to force every character form into one of the existing categories.

When compared against the earlier stages of the script, the largest number of new structures seems to originate from handwriting habits related to the cursivized writing of characters. Some of these cursive forms enter the mainstream and gradually become the norm, either replacing earlier forms or being used alongside them as common variants. Even though government authority and education tend to have a stabilizing effect on the writing system, the script is also a pragmatic tool used in all aspects of daily life by people of various levels of education and because of this the actual written output we witness in archaeological material is extremely varied. While the standard is imposed from above, orthographic diversity develops as the script is used by ordinary people in situations where the emphasis is often on the content of what they write, rather than on how they do it.<sup>23</sup>

Despite the overall significance of the cursive hand in the generation of new forms, in my analysis below I will discuss less obvious categories of new structures, partly to draw attention to these and partly because of their significance in medieval manuscript culture. Even though the forms in each category are not that numerous, they involve very common characters and thus appear in manuscripts with high frequency.

### 3.1. *Archaized structures*

By the Tang dynasty, the *kai* 楷 ‘modern standard’ script form had become the absolute dominant calligraphic style (*shuti* 書體) in use, whereas more archaic styles were limited to very specific situations. Thus the “seal script” (*zhuan**shu* 篆書) was only used for carving seal inscriptions, the heading

---

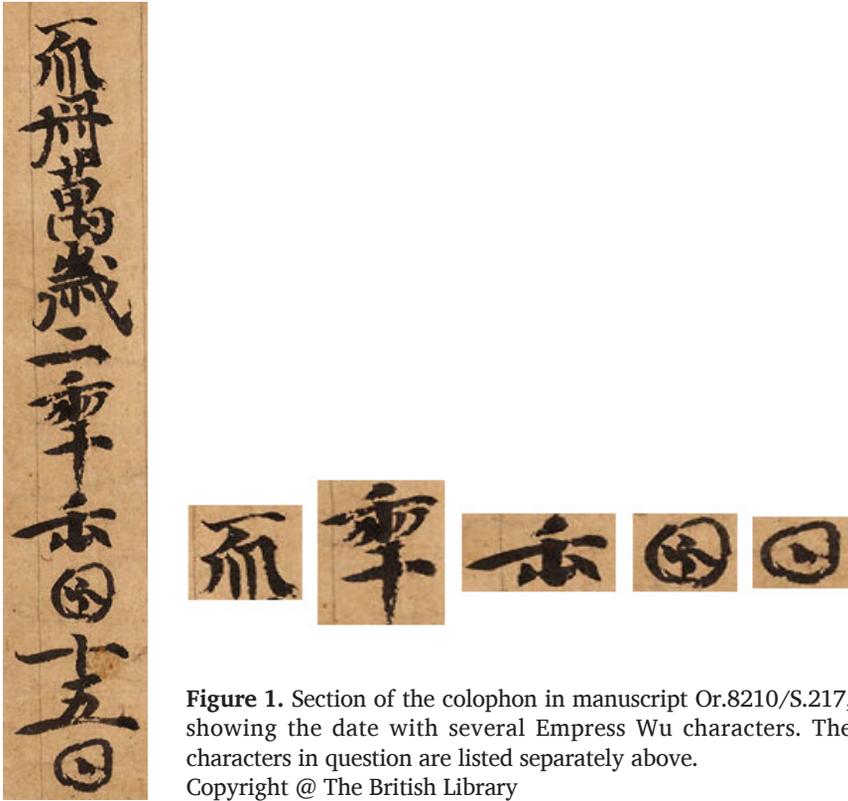
<sup>23</sup> Although aesthetic, ludic, cryptographic and “magical” considerations may all be responsible for diverse forms of the Chinese script at various points in history, we seldom see these factors at work in surviving medieval Chinese manuscripts, where for the most part orthographic diversity seems to be a “spontaneous” phenomenon connected with the everyday use of the script.

of stele inscriptions and other decorative purposes.<sup>24</sup> There are occasional cases when entire texts are written in the seal script and these are examples where the calligraphy is just as much part of the text as the words it conveys. Similarly, the “clerical script” (*lishu* 隸書) was only used as part of the art of calligraphy. Yet we occasionally find examples when earlier orthographic structures from either the seal or the clerical script appear in *kai* characters, creating variant forms otherwise not seen in the *kai* script. In some cases these examples have come down from early Chinese scripts as part of the general evolution of the script and thus they are technically not “new” forms, even if they fall outside the official standard of the time when they were written.

But there are also cases of new *kai* forms that were not the result of gradual evolution but arose out of a desire to emulate ancient character forms. These forms had been used at the earlier stages of the script but later fell out of use as they were replaced by other forms. Several such characters were introduced by Empress Wu Zetian 武則天 (r. 690–705), the only female emperor in Chinese history.<sup>25</sup> Although her actual reign lasted only 15 years and she commissioned the invention of only 18 new forms, these were extremely common ones and are thus frequently seen in medieval manuscripts. Another reason for this is that the empress was an avid supporter of Buddhism and manuscript production reached unprecedented heights during her reign. Several of the characters introduced by Empress Wu were archaic structures that had never been used in the *kai* script. For example, the character 天 (*tian*; ‘heaven’) became written as 𠄎, which resembles the way the character was written in the pre-Qin period. Similarly, the character 日 (*ri*; ‘day, sun’) became written as ☺ or ☻ which were presumably also based on archaic forms. Figure 1 one shows part of the colophon in manuscript Or.8210/S.217 where several Empress Wu characters appear together. The manuscript contains a copy of the *Guanshiyin jing* 觀世音經 [Sūtra of Avalokiteśvara], and the image shows

<sup>24</sup> For the use of the seal script on Tang stone inscriptions, see Shi 1987.

<sup>25</sup> These characters are called by a variety of different names, including *Wu Zhou xinzi* 武周新字 (‘new characters of Empress Wu’s Zhou dynasty’), *Wu Zhou wenzi* 武周文字 (‘characters of Empress Wu’s Zhou dynasty’), *Wu hou xinzi* 武后新字 (‘new characters of Empress Wu’), *Wu hou zaozi* 武后造字 (‘characters created by Empress Wu’), etc. Modern Japanese sources generally use the term *Sokuten moji* 則天文字 (‘characters of Empress Zetian’).

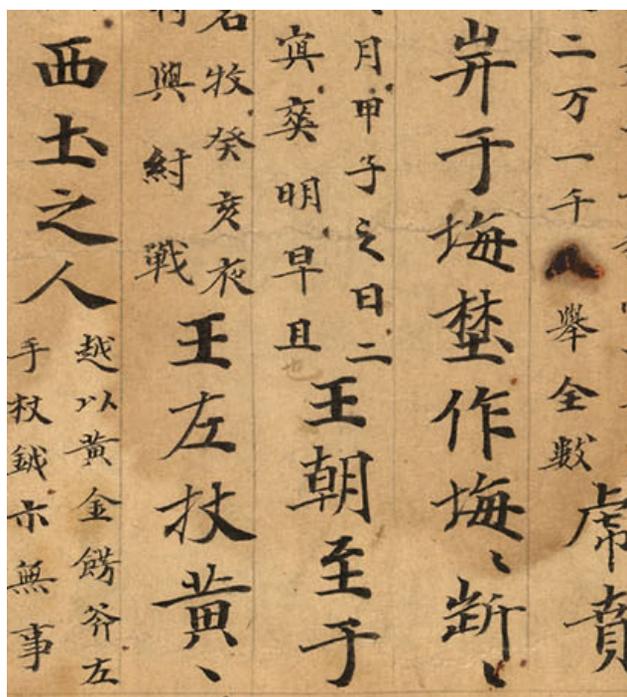


**Figure 1.** Section of the colophon in manuscript Or.8210/S.217, showing the date with several Empress Wu characters. The characters in question are listed separately above.  
Copyright @ The British Library

the date from the colophon which says: ‘the 15<sup>th</sup> day of the 1<sup>st</sup> month of the 2<sup>nd</sup> year of the Tiance wansui reign’ 天冊萬歲二年正月十五日, which was in the spring of 696.

On the image, the characters 天, 年, 正, 月 and 日 are written as Empress Wu characters (as shown separately on the right). Of these five characters 天, 月 and 日 are archaized forms, and possibly 年 as well.

Another group of archaized characters that appear in Tang dynasty manuscripts are the *liguding* 隸古定 (‘clericized’) forms in the manuscripts of the *Guwen Shangshu* 古文尚書 [Exalted documents in the ancient script]. Clericization is a process that originally happened during the transition from the scripts of the Warring States to the modern script of the Qin-Han period, which at the time meant the clerical script. This was a major step in the history of the Chinese script which produced orthographic structures



**Figure 2.** Section of the *Guwen Shangshu* in manuscript Or.8210/S.799, showing several clericized characters.

Copyright @ The British Library

which were used throughout the dynastic period and are in use today. In terms of their structural components the majority of characters followed their pre-Qin forms, even though many of them acquired a new structure. There were, however, Warring States manuscripts that were discovered during the Han, well after the transition and when these were transcribed into the contemporary clerical script, in some cases the transcribers made an effort to preserve the original orthography of the characters. In the philological tradition such copies of the *Shangshu* 尚書 came to be known under the name *Guwen Shangshu* or *Liguding Shangshu* 隸古定尚書.

Among the Dunhuang manuscripts, there are several copies of the *Guwen Shangshu*. One of these is manuscript Or.8210/S.799, written in an even and beautiful calligraphy, a testimony to the great amount of care and attention that went into producing the manuscript. On the image shown in Figure 2, there are several clericized characters that emulate the structure

of Warring States forms. Thus the full-size characters in the second line read 戰于牧野作牧誓 (‘fought at Muye and composed the “Oath of Mu”’), in which the last two characters comprise the title of the *Shangshu* chapter and are accordingly marked with diacritic marks. In this short string of text the characters 戰, 牧, 野 and 誓 are clericized forms of pre-Qin characters written in the *kai* script.

The study of *guwen* character forms reached new heights during the Song dynasty when scholars also became interested in collecting antique inscriptions and vessels. There were several important works on pre-Qin forms which relied on excavated material. These works, however, usually did not clericize the ancient forms into the *kai* script but kept their original form.<sup>26</sup>

### 3.2. Folk-etymological structures

*Huiyi* is one of the original categories in the *liu shu* system which Xu Shen explained in the Postface of the *Shuowen jiezi* the following way:

會意者，比類合誼，以見指撝，武信是也。

*Huiyi* characters are the ones that conjoin categories to present the indicated meaning. The characters 武 and 信 are like this.

Accordingly, this category comprises characters that consist of two or more semantic components the meaning of which is relevant to the meaning of the entire character.<sup>27</sup> Some scholars believe that an additional criterion is that for real *huiyi* characters none of the components can function as a phonophoric component but others disagree with this.<sup>28</sup> In either case, modern linguists have successfully shown that most of the traditional examples of *huiyi* compositions are in fact *xingsheng* characters (composed of a semantic component and a phonophoric component) in which the phonophoric component had lost its transparency. There is, however, still

<sup>26</sup> A classic example of such a work from the early Song is Guo Zhongshu’s 郭忠恕 (?–977) *Hanjian* 汗簡 which reproduces a multitude of archaic forms in their original form.

<sup>27</sup> Translating the word *yi* 意 in *huiyi* 會意 as ‘meaning’ inevitably discards other possible translations (e.g. ‘idea’, ‘intention’) which may have been part of the original sense of the term *huiyi*.

<sup>28</sup> Examples of the former opinion are Boltz 1994: 147–149, Boltz 2006; examples of the latter, Handel 1998 and Behr 2006.

disagreement whether the category existed at all during the formative stage of the script or is a later invention of folk etymology.<sup>29</sup>

Nevertheless, surviving Tang-dynasty manuscripts from Dunhuang and Turfan show a considerable number of character forms composed according to the *huiyi* principle. Some of the examples are so-called *suzi* 俗字 ('vulgar forms'), which were in daily use throughout the medieval period.<sup>30</sup> Even though these *suzi* were non-standard forms that in principle did not satisfy the requirements of official writing, they were extremely common and in many cases represented the dominant way of writing particular characters at the time. In other words, they had their own norm outside the official standard. For example, the character 蘇 (*su* 'to revive, regain consciousness') was often written as 甦, which consisted of the components 更 (*geng* 'again') and 生 (*sheng* 'live'), which if read together as two words would mean 'to regenerate, revive,' thereby matching the meaning of the character 蘇.<sup>31</sup>

David P. Branner calls such forms 'portmanteau characters' and points out that they are different from traditional *huiyi* configurations, as in these the components are read in order as a phrase 'to give the meaning of the word represented by the whole character.'<sup>32</sup> Thus the form 思 appears in the dictionary *Sisheng pianhai* 四聲篇海 [Sea of Writings Arranged according to the Four Tones] as a variant of the character 思 (*si* 'to think'). In this orthographic formation, the components 用 (*yong* 'to use') and 心 (*xin* 'mind, heart') are read as a phrase, and their composite meaning 'to use the mind' approximates the meaning of the whole character. To cite another example, also from the *Sisheng pianhai*, the form 齋 is listed as a variant of the character 齋 (*zhai* 'to fast'), and the components 不 (*bu* 'not') and 食 (*shi* 'to eat') together signify the meaning of the whole character. In both cases, however, it is apparent that the portmanteau variant approximates the overall appearance of the original character, creating a form that still falls within its general symmetry. (While there are quite a few *huiyi* and portmanteau forms in medieval dictionaries, research suggests that

<sup>29</sup> See for example, Behr 2006, Boltz 2006, Handel 1998 and Handel forthcoming.

<sup>30</sup> Galambos 2011 examines the *huiyi* forms among the *suzi* from Dunhuang.

<sup>31</sup> The form 甦 is also used today as a standard character both in China and Japan.

<sup>32</sup> Branner 2011: 73. Wolfgang Behr calls these 'intrasyntactic characters' but does not classify them as a subcategory of *huiyi* (Behr 2009: 301).

the majority of these was only transmitted as part of the lexicographic tradition rather than being used in daily practice.<sup>33</sup>)

Furthermore, several of the 18 forms introduced by Empress Wu are also semantic compounds. Thus the character 照 (*zhao* ‘to illuminate’), which was the personal name of the empress, was changed to be written as 𡩺, that is, a combination of 明 (*ming* ‘to brighten’) and 空 (*kong* ‘sky’); whereas the new form of the character 地 (*di* ‘the Earth’) was written as 𡩻, a tripartite combination of 山 (*shan* ‘mountain’) + 水 (*shui* ‘water’) + 土 (*tu* ‘soil’).

But there are also cases in which the semantic compounds are not unorthodox forms but represent the ordinary way of writing a given character. Thus the typical way of writing the character 明 (*ming* ‘bright’) during the Tang was either as 明 (i.e. the way we write it today) or 𡩼 (目 [*mu* ‘eye’] + 月 [*yue* ‘moon’]), or somewhat less commonly as 𡩽 (目+目) or 𡩾 (月+月).<sup>34</sup> All of these forms are classic examples of *huiyi* compounds. Even if we were to argue that the component on the left side is a corrupted or distorted form of what was once a phonophoric element, that would only prove that an earlier form of this character used to be a *xingsheng* character. This argument would have no bearing on the actual forms in front of us. Thus regardless of their origin, in terms of their then-current orthographic structure, the Tang dynasty forms are semantic compounds in which both components play a semantic role and none of them has a phonophoric function. Similarly, for this particular time period, most of the traditional examples of *huiyi* characters (as well as many traditional *xingsheng* characters), in which the phonetic function of one of the components had become obscure, would have to be categorized as semantic compounds because of their contemporary structure.

<sup>33</sup> On the comparison of the abundance of *huiyi* variants in medieval dictionaries with actual manuscript data, see Galambos 2012.

<sup>34</sup> At the same time, some medieval dictionaries such as the *Ganlu zishu* 干祿字書 followed the *Shuowen jiezi* as the benchmark of orthographic norm and asserted that it was the form 𡩼 that was the standard, even though this form is completely absent from contemporary manuscripts. The form 𡩾 (明) would still be different from the character 朋 used to write the word *peng* (‘friend’), as the latter usually slanted with its top to the right side (e.g. 𡩾), making it quite distinct from the upright 朋 (明).

### 3.3. Taboo characters

Imperial name taboos were yet another source for generating hitherto unattested orthographic forms. The custom of name taboo goes back to pre-Qin times yet there are still many questions related to how strictly these rules were observed in actual usage.<sup>35</sup> In essence, the practice meant the avoidance of writing the personal names of members of the imperial family of the reigning dynasty, adding ever new names to the list until the dynasty ended and the whole system began anew with a new ruling house. The corpus of medieval manuscripts shows that there were many inconsistencies in the system and because of this taboo characters are not a fully reliable method for dating manuscripts. Still, there are patterns which allow us to make some general observations.

There were two main methods to avoid writing a tabooed name. One of them involved the replacement of the particular character with another one of similar meaning. For example, the characters *shi* 世 ‘world’ and *min* 民 ‘people’ in the name of Li Shimin 李世民 (598–649; r. 626–649), also known as Emperor Taizong 太宗, were habitually substituted with the characters 代 and 人, respectively. There was, however, also another way to avoid writing the tabooed characters, namely, to omit a stroke and thereby create a new orthographic form which was recognizable yet different from the original character. Thus the character 民 would be written as 𠃉, and the character 世 as 𠃊.

In principle only the second method produced new orthographic forms, which were used in a very restricted time frame. In reality, however, because of the length of the Tang dynasty, some of the forms were used for hundreds of years. In addition, in many cases the taboo characters were copied along with the texts even after the end of the Tang dynasty, thus producing a relatively large body of manuscripts which contained such incomplete forms. Moreover, since many of the characters were extremely common (e.g. 世 and 民), the actual number of such forms is quite large, even though the taboos were not consistently observed even during the Tang.

Besides, sometimes the tabooed character was replaced even when it

---

<sup>35</sup> On the origins of the name taboos in China, see Peng 1999, Lin 2008 and van Ess 2008.

occurred as a component in another character. A complex case involved the vulgar form of the character 構, which on top of the right hand component consisted of 世+舟, where the component 世 was replaced to observe the taboo of the personal name of Taizong with the component 云, creating an entirely new structure.<sup>36</sup>

### 3.4. Ligatures

In the medieval manuscripts from Dunhuang there are also some composite graphs used for writing disyllabic words.<sup>37</sup> These are ligatures that represent two-character compositions as a single graph. Although such formations—known as *hewen* 合文 ('joined graphs')—have been common in early China, there are a number of new structures that only occur from the medieval period onward, especially in Buddhist manuscripts. The most common of such ligatures is the graph 𠄎, standing for the word *pusa* 菩薩 ('Bodhisattva'), which abbreviates the two original characters by only writing the two top components, i.e. the two 'grass radicals'. Another common example is the word *puti* 菩提 ('Bodhi, enlightenment'), which could be written as 𠄎 (Fig. 3) and 𠄎. While the first graph is clearly the combination of the original characters (with 菩 abbreviated to the grass radical), the other two are less easily recognizable. The third graph remarkably matches the ligature for the word *pusa* seen in the first example, which also means that without the context it was impossible to tell them apart. Perhaps the function of the dot in the lower part of the second graph is to disambiguate the graph and distinguish it from the ligature used for the word for *pusa*.<sup>38</sup> Ligatures of the word *puti* also commonly occur as part of the Chinese name of Subhūti, one of the Buddha's foremost disciples, which was transcribed into Chinese as Xuputi 須菩提. In the Chinese transcription the last two syllables coincide with the Chinese word for 'enlightenment' (*puti* 菩提) and thus manuscripts at times wrote the name with a ligature for the last two syllables. But there is no attested ligature incorporating all three syllables of the name.

<sup>36</sup> See Dou 2007: 242–243.

<sup>37</sup> As far as we can tell from the available material, ligatures in medieval manuscripts were only used for disyllabic words, never for tri- or tetrasyllabic ones.

<sup>38</sup> For more examples of ligatures in the medieval Chinese manuscripts, see Galambos 2010.



**Figure 3.** Ligature writing the word *puti* 菩提 ('Bodhi, enlightenment') in the phrase *zheng da puti* 證大菩提 ('verify the great enlightenment') in manuscript P.2173.

All such examples were technical terms associated with the Buddhist tradition and in this sense resembled the ligatures used in religious contexts in some Western traditions.<sup>39</sup>

### 3.5. Assimilated forms

An important category of character formation is assimilation, which means that a character is modified from its original form under the influence of another character. These are often called *leihuazi* 類化字 ('analogical character formations') in modern scholarship, although the term has not gained universal popularity.<sup>40</sup> The most obvious cases of assimilation are those where one of the characters in a binome or a common phrase changes under the influence of the other, typically by acquiring the same semantic component. For example, the word *fenghuang* 鳳凰 ('phoenix') in medieval manuscripts and inscriptions was at times written as 鳳皇 but the second character gradually became written with the 几 radical around it, undoubtedly under the influence of the character preceding

<sup>39</sup> In the Russian orthodox tradition, for example, many of the names (e.g. Christ, Mother of God) are commonly written as ligatures.

<sup>40</sup> See, for example, Zhang 1995.



**Figure 4.** The phrase *xiangma chesheng* 象馬車乘 ('chariots drawn by elephants and horses') in manuscript Or.8210/S.12, showing the similarity of the characters 象 and 馬.

it. At one point, the rationalized form 鷓 was also in use but it never became widespread.<sup>41</sup> This phenomenon of matching the radicals in a binome is probably why many of the “true” binomes ended up with the same radical, even though in the manuscript tradition they perhaps even more commonly appear with dissimilar radicals or without radicals at all. Thus the words *hudie* 蝴蝶 ('butterfly') and *zhizhu* 蜘蛛 ('spider') commonly appear in an array of combinations in which either one or both characters are written without the 虫 radical.

Another example is the medieval form of the character 象 (*xiang* 'elephant') which is customarily written in Sui-Tang Buddhist manuscripts as 𧈧, the lower part of which is very similar to the character 馬 (*ma* 'horse'). Figure 4 shows a case when the two characters are adjacent and form the phrase *xiangma chesheng* 象馬車乘 ('chariots drawn by elephants and horses'). Although the four dots underneath are written in a cursive manner as a single line, the similarity between the two forms is obvious, especially when compared against the standard form of the character 象.<sup>42</sup>

The same principle is also at work in specialized forms that result from

<sup>41</sup> This form, composed of the phonophoric *huang* 皇 and 鳥 'bird' as semantic element, is seen on the Northern Qi 北齊 (550–577) stele “Gao Rui xiusi bei” 高叡修寺碑 [Stele commemorating the renovation of a monastery by Gao Rui]; see Qin 1985: 147.

<sup>42</sup> For a detailed study of *leihuazi* in the Dunhuang manuscripts, see Zhang 1995.



Figure 5. The phrase *zhen chi xi* 縝絺絺 (‘wore an unlined robe made of either fine or coarse material’) in manuscript Dx.2144, showing how the first character acquires the 糸 radical as a result of being influenced by the following two characters.

the orthographic influence of the immediate context in a manuscript. For example, in manuscript Dx.2144, a small fragment of the *Lunyu* 論語 [Analects of Confucius], the phrase *zhen chi xi* 縝絺絺 (‘wore an unlined robe made of either fine or coarse material’)<sup>43</sup> is written as 縝絺絺, with the first character which means ‘to wear an unlined robe’ is replaced by the homophonous *zhen* 縝 (‘fine, dense’), most likely to match the structure of the following two characters, both of which are written with the 糸 radical (Fig. 5).<sup>44</sup>

Because the replacement character is itself an existing standard character, its use could be explained by the traditional model as a phonetic loan (*jiajie* 假借) graph, yet its graphic connection with the following two characters reveals that in this particular context the form 縝 should rather be understood as an orthographic variant of the character 縝, still representing the same verb. While we may dismiss this example as a mistake, it nevertheless illustrates the workings of the principle of assimilation in actual practice, and it is not difficult to see how in some cases such “mistakes” became widely used as variants. When the same phenomenon occurs as part of an artistic device in poetry, it is an intentional graphic pun which naturally cannot be regarded a mistake.

<sup>43</sup> Translation from D. C. Lau (1992: 89).

<sup>44</sup> We can see, however, that the case may be even more complex as there is not only a graphical but also a semantic influence from the following character (i.e. 絺) which has the meaning ‘fine’ and is thus synonymous with the semantic value of the character 縝.

## 4. Summary

The five categories introduced in this paper all originate from the medieval manuscript tradition (although it is also possible that the same principles may have been at play during earlier stages of the script) and illustrate some of the orthographic patterns observable in the Dunhuang corpus. All of these five examples differ from the influential *san shu* model advanced by modern scholars to explain the formative process of the history of Chinese characters. The aim of this paper was to demonstrate that if we look at concrete examples of character forms as they appear in manuscripts, we can identify several other principles that are not accounted for in modern theories of the origin of Chinese characters. Yet these were existing forms for which we have hundreds or rather thousands of examples in the surviving corpus of medieval manuscripts. Consequently, it should be possible to analyze them in their existing form, rather than discounting them as derivative (i.e. corrupt) forms of characters that otherwise conform to the *san shu* theory.

Many of these medieval forms were short-lived and when we look at the history of Chinese writing from a broad perspective stretching across many centuries, they tend to be left out because they were discontinued. In effect, they are evolutionary cul-de-sacs. To be sure, a historical development itself is a conceptual notion that only makes sense in retrospect, while the actual changes and developments that happen in real life, whether in the past or the present, are significantly more complex and often appear to be random. The orthographic diversity evidenced in medieval manuscripts reminds us of the complex paths of the development of the Chinese script.

## References

### Primary sources

- Hanshu* 漢書. Compiled by Ban Gu 班固, commentary by Yan Shigu 顏師古. Beijing: Zhonghua shuju, 1962.
- Lunheng* 論衡. Composed by Wang Chong 王充. Huang Hui 黃暉, *Lunheng jiaoshi* 論衡校釋, Changsha: Shangwu yinshuguan, 1938.
- Shuowen jiezi* 說文解字. Compiled by Xu Shen 許慎. *Shuowen jiezi fu jianzi*

說文解字附檢字, Beijing: Zhonghua shuju, 1963.

*Zhonglun* 中論. Compiled by Xu Gan 徐幹. Xu Xianglin 徐湘霖, *Zhonglun jiaozhu* 中論校注. Chengdu: Bashu shushe, 2000.

### Secondary sources

Behr, Wolfgang (2006). “‘Homosomatic Juxtaposition’ and the Problem of ‘Syssemantic’ (*Huiyi*) Characters.” In Françoise Bottéro and R. Djamouri (eds.), *Écriture chinoise: Données, usages et représentations*, Paris: École des Hautes Études en Sciences Sociales, Centre de Recherches Linguistiques sur l’Asie Orientale. 75–114.

Behr, Wolfgang (2009). “In the interstices of representation: Ludic writing and the locus of polysemy in the Chinese sign.” In Alex de Voogt and Irving Finkel (eds.), *The Idea of Writing: Play and Complexity*. Leiden: E.J. Brill. 281–314.

Boltz, William G. (1994). *The Origin and Early Development of the Chinese Writing System*. American Oriental Series, vol. 78. New Haven: American Oriental Society.

Boltz, William G. (2006). “Phonographic motivation in the formation of compound Chinese characters: The case of *wu* 武.” In Françoise Bottéro and R. Djamouri (eds.), *Écriture chinoise: Données, usages et représentations*, Paris: École des Hautes Études en Sciences Sociales, Centre de Recherches Linguistiques sur l’Asie Orientale. 56–73.

Bottéro, Françoise (1998). “La vision de l’écriture de Xu Shen à partir de sa présentation des *liushu*.” *Cahiers de Linguistique - Asie Orientale* 27:2. 161–191.

Bottéro, Françoise & Christoph Harbsmeier (1998). “The *Shuowen jiezi* dictionary and the human sciences in China.” *Asia Major* 21:1. 249–271.

Branner, David Prager (2011). “Portmanteau characters in Chinese.” *Journal of the American Oriental Society* 131:3. 73–82.

Cerquiglini, Bernard (1999). *In Praise of the Variant: A Critical History of Philology*. Baltimore: John Hopkins University Press.

Chen Mengjia 陳夢家 (1956). *Yinxu buci zongshu* 殷墟卜辭綜述. Shanghai: Shanghai renmin chubanshe.

Dong Lianchi 董連池 et al. (2007). *Shuowen jiezi yanjiu wenxian jicheng (Gudai juan)* 說文解字研究文獻集成(古代卷). Beijing: Zuoja chubanshe.

Dou Huaiyong 竇懷永 (2007). *Dunhuang wenxian bihui yanjiu* 敦煌文獻避諱研究.

- Ph.D. dissertation, Zhejiang daxue.
- Ehrman, Bart D. (2006). *Studies in the Textual Criticism of the New Testament* (New Testament Tools and Studies). Leiden, London: Brill.
- van Ess, Hans (2008). "The origin of posthumous names in *Shih-chi* 14." *Chinese Literature: Essays, Articles, Reviews* 30. 133–144.
- Galambos, Imre (2006). *Orthography of Early Chinese Writing: Evidence from Newly Excavated Manuscripts (490–221 BC)*. Budapest Monographs in East Asian Studies. Budapest: Eötvös Loránd University, Department of East Asian Studies.
- Galambos, Imre (2010). "Scribal notation in medieval Chinese manuscripts: The *hewen* (ligature) and *chongwen* (duplication) marks." *Manuscript Cultures in Asia and Africa* 2. 5–9.
- Galambos, Imre (2011). "Popular character forms (*súzi*) and semantic compound (*huiyi*) 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—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.
- Handel, Zev (1998). "In defence of huiyi: The case for the phonetic-less character." Unpublished book review of Boltz 1994.
- Handel, Zev (Forthcoming). "Does Xǔ Shèn's huiyi category reflect historical reality? An argument for the existence of compound graphs lacking phonophorics."
- Lau, D. C. (1992). *Confucius: The Analects*. Hong Kong: The Chinese University Press. (First published in 1997 by Penguin Books Ltd.)
- Lin Shengjie 林聖傑 (2008). "Xi Zhou yifa yanjiu: Cong jinwen yu wenxian guan cha" 西周諡號研究—從金文與文獻觀察. *Ciji daxue renwen shehui kexue xuekan* 慈濟大學人文社會科學學刊 7. 89–121.
- Nugent, Christopher (2010). *Manifest in Words, Written on Paper: Producing and Circulating Poetry in Tang Dynasty China*. Cambridge, MA: Harvard University Asia Center Press.
- Peng Yushang 彭裕商 (1999). "Yifa tanyuan" 諡法探源. *Zhongguo shi yanjiu* 中國史研究 2. 3–11.
- Qin Gong 秦公 (1985). *Xinbian bei biezi* 新編碑別字. Beijing: Wenwu chubanshe.
- Qiu Xigui 裘錫圭 (1988). *Wenzixue gaiyao* 文字學概要. Beijing: Shangwu

yinshuguan.

- Richter, Matthias L. (2013). *The Embodied Text: Establishing Textual Identity in Early Chinese Manuscripts* (Studies in the History of Chinese Texts.) Leiden, Boston: Brill.
- Serruys, Paul L.-M. (1957). "The Study of the chuan-chu in Shuo wen." *Bulletin of the Institute of History and Philology* 29. 131–195.
- Shi Anchang 施安昌 (1987). *Tangdai shike zhuanwen* 唐代石刻篆文. Beijing: Zijincheng chubanshe.
- Sun Yongchang 孫雍長 (1991). *Zhuanzhu lun* 轉注論. Changsha: Yuelu shushe.
- Tang Lan 唐蘭 (1935). *Guwenzixue daolun* 古文字學導論. Beijing: Beijing daxue.
- Tang Lan 唐蘭 (2005). *Zhongguo wenzixue* 中國文字學. Shanghai: Shanghai guji chubanshe. (First edition 1949)
- Yue Shanyue 岳山岳 (2008). *Liu shu yu Zhongguo chuantong wenhua* 六書與中國傳統文化. Shanghai: Shanghai sanlian shudian.
- Zhang Yongquan 張涌泉 (1995). "Dunhuang wenshu leihuazi yanjiu" 敦煌文書類化字研究. *Dunhuang yanjiu* 敦煌研究 4. 71–79.
- Zhong Ruxiong 鍾如雄 (2007). *Hanzi zhuanzhuxue yuanli* 漢字轉注學原理. Chengdu: Dianzi keji daxue chubanshe.

Imre Galambos  
Asian and Middle Eastern Studies  
University of Cambridge, England  
<iig21@cam.ac.uk>

[Received 30 March 2014;  
revision received 19 August 2014;  
accepted 25 August 2014]

