

Popular Character Forms (*Súzì*) and Semantic Compound (*Huìyì*) Characters in Medieval Chinese Manuscripts

IMRE GALAMBOS

INTERNATIONAL DUNHUANG PROJECT, BRITISH LIBRARY

Traditional Chinese scholarship visualized orthographic structure using the system of *liùshū* 六書, a technical term explained as six principles of character formation. According to extant sources these categories were first put forth towards the end of the first century C.E. and have been in use ever since. Although individual categories had been sporadically criticized as early as the Sòng period, it was during the philological renaissance of the Qīng dynasty when it became clear that the entire classification system had to be reconsidered. Later on, the discovery of oracle-bone inscriptions and pre-Hàn manuscripts provided additional material for reassessing the claims of traditional historiography on the evolution of Chinese writing and the forces behind it. Among the problematic categories within the *liùshū* system was the principle of *huìyì* 會意, at times rendered into English as ‘syssemantic characters’ or simply ‘semantic compounds’.¹ This traditional category proved to be often based on folk etymologies, and most examples of it can be demonstrated to contain a phonetic component, in contrast with the traditional view that saw them as purely semantic combinations.

While modern research may be justified in doubting the impact of the *huìyì* principle at the early stages of the writing system, one cannot fail to notice the presence of numerous *huìyì*-type forms in medieval manuscripts and epigraphic sources. Some of these forms commonly feature in medieval dictionaries, while others are seen only in manuscripts and inscriptions. To be sure, for the most part these are variant forms of characters with otherwise well-attested phonetic origins, yet their occurrence in the post-Hàn period is a phenomenon that deserves our attention. In this paper, I propose to look at some of the popular or non-standard forms (*súzì* 俗字) found in medieval manuscripts and dictionaries in an attempt to reconsider the *huìyì* category from the point of view of the manuscript tradition. Rather than discussing the etymology and early development of established characters, I am specifically interested in non-standard character forms used in everyday writing, because these demonstrate that even if the *huìyì* principle did not play a major role during the early stages of the Chinese script, by medieval times it was certainly one of the key models according to which people understood orthographic structure.

I am grateful to my colleagues who have given advice and comments about earlier drafts of this paper, in particular Wolfgang Behr (Universität Zürich), Lái Guólóng (University of Florida, Gainesville), Matthias Richter (University of Colorado, Boulder), and Françoise Bottéro (CRLAO, Paris).

1. The translation of the *liùshū* principles in Western languages has its own history of over three centuries. Thus the *huìyì* principle had been explained as “*societas significatorum*” (seventeenth-century Jesuit manuscript from the Bibliothèque nationale in Paris, Fonds Latin 6277; see Lundbaek 1988: 10); “*sensum aggregantes*” (Callery 1841: 8); “combination of ideas” (Morrison 1815: xvii); “suggestive compounds” (Hopkins 1881: 18); “ideograms” (Owen 1910: 12); “logical combinations” (Wieger 1927: 17); “ideographic compounds” (Boodberg 1937: 345), etc. Modern uses include the terms “syssemantic characters” or “syssemantographs.” This latter is employed as the English equivalent of *huìyìzì* by Gilbert L. Mattos and Jerry Norman in their translation of Qíú Xīguī’s 裘錫圭 *Wénzìxué gàiyào* 文字學概要 (Qíú 1988; Qíú 2000).

THE *LIÙSHŪ* SYSTEM AND THE *HUÌYÌ* PRINCIPLE

Traditional Chinese scholarship described the principles behind the evolution of characters in terms of the *liùshū* ('six scripts' or 'six types of writing'). Although this term appears in the *Zhōulǐ* 周禮, its use with respect to character structure dates to the first century C.E. when it surfaced in three different sources. The most elaborate of these is the *Shuōwén jiězì* 說文解字 by Xǔ Shèn 許慎 (ca. 58–ca. 147), who explained the six categories in the "Post-face" 敘 in the following way:

周禮八歲入小學，保氏教國子，先以六書。一曰：指事。指事者，視而可識，察而可見，上下是也。二曰：象形。象形者，畫成其物，隨體詰詘，日月是也。三曰：形聲。形聲者，以事為名，取譬相成，江河是也。四曰：會意。會意者，比類合誼，以見指撝，武信是也。五曰：轉注。轉注者，建類一首，同意相受，考老是也。六曰：假借。假借者，本無其字，依聲託事，令長是也。

According to the *Zhōulǐ*, school begins at the age of eight. When the Protector teaches the sons of the state, he begins with the *liùshū*. The first of these is *zhǐshì* ('pointing at things'²). *Zhǐshì* characters are the ones that can be understood by looking at them, the meaning of which can be seen through observation. The characters 上 and 下 are like this. The second is *xiàngxíng* ('depicting form'). *Xiàngxíng* characters are the ones that depict objects by reproducing their physical shape. The characters 日 and 月 are like this. The third is *xíngshēng* ('form and sound'). *Xíngshēng* characters are the ones that take a thing/object to indicate the name and combine it with a [phonetic] semblance. The characters 江 and 河 are like this. The fourth is *huìyì* ('joining ideas'). *Huìyì* characters are the ones that conjoin categories to present the indicated meaning. The characters 武 and 信 are like this. The fifth is *zhuǎnzhù* ('commenting by rotation'). *Zhuǎnzhù* characters are the ones that establish categories based on a single origin and that borrow their analogous meanings from each other. The characters 考 and 老是 like this. The sixth is *jiǎjiè* ('borrowed'). *Jiǎjiè* characters are the ones that are assigned a written form, which did not exist originally, based on their pronunciation. The characters 令 and 長 are like this.

In addition to listing the names of the six categories, Xǔ Shèn provides two examples for each. He also gives a short gloss of each term, and this is the only evidence we have today of how the categories might have been understood in Hàn times.³ The other Hàn source describing the six principles is the "Yīwénzhì" 藝文志 (Record of Arts and Letters) chapter in Bān Gù's 班固 (32–92 C.E.) *Hànshū* 漢書 (History of the Hàn Dynasty), which retells the history of Chinese writing in very much the same manner as Xǔ Shèn's "Postface." Indeed, the two accounts show a number of similarities that confirm that they ultimately go back to the same source.⁴ At the same time, Bān Gù's account of the *liùshū* is more concise, only giving a list of names of the categories without examples: *xiàngxíng* 象形, *xiàngshì* 象事, *xiàngyì* 象意, *xiàngshēng* 象聲, *zhuǎnzhù* 轉注, and *jiǎjiè* 假借. In addition, we learn that these represent "the basis of character formation" 造字之本也.⁵ A third Hàn source, Zhèng Zhòng's 鄭眾 (5 B.C.E.–83 C.E.) commentary to the *Zhōulǐ* from the second half of the first

2. The translations in parentheses here are merely meant to reproduce the literal meaning of Chinese characters, rather than providing proper English terminology for the *liùshū* categories.

3. To be exact, this only shows how Xǔ Shèn understood these terms, since there might have well been a range of competing interpretations at the time.

4. According to the introduction of the "Yīwénzhì," Bān Gù largely based this chapter on Liú Xīn's 劉歆 (ca. 46–23 C.E.) now lost *Qīlüè* 七略 (Seven Outlines). The *Qīlüè* itself, however, was based on Liú Xiàng's 劉向 (79–78 B.C.E.) *Biélù* 別錄 (Appendix); therefore the ideas recorded in the "Yīwénzhì" might go back as far as the first century B.C.E.

5. A point that is relevant to the subject matter of this paper is that Bān Gù does not actually specify whether these principles pertain to original, *ex nihilo*, character creation or to a modern (in his case, Hàn) way of understanding character structure.

century C.E., identified the six categories as *xiàngxíng* 象形, *huìyì* 會意, *zhuǎnzhù* 轉注, *chūshì* 處事, *jiǎjiè* 假借, and *xiéshēng* 諧聲, which yet again shows some discrepancy with both Xǔ Shèn's and Bān Gǔ's terminology.⁶ Thus at the source of the tradition we have three different authors from the Eastern Hàn period with three sets of names, most likely describing the same or very similar principles using slightly different nomenclature.⁷ In later centuries, however, the *liùshū* system developed into a complex conceptual framework that formed the theoretical foundation of the branch of Chinese philology dealing with the graphic shape of characters. Indeed, since the Sòng period this branch was often referred to by the name of *liùshūxué* 六書學, or the “study of the six principles of writing.”⁸

But early on, critics of individual categories raised objections regarding the etymological correctness of these principles. In the Sòng dynasty, for example, Zhèng Qiáo 鄭樵 (1104–62) noted in his *Liùshū lüè* 六書略 (Outline of the Six Scripts) that in the body of the *Shuōwén* Xǔ Shèn effectively only used the *xiàngxíng* and *xíngshēng* categories. During the Qīng dynasty, along with a renewed interest in epigraphy and textual studies, the *Shuōwén* and the *liùshū* once again became the focus of scholarly investigation and reinterpretation.⁹ Finally in modern times the rapidly growing number of newly discovered inscriptions and early manuscripts provided abundant firsthand material for creating a historically plausible narrative of the development of the Chinese script. Especially the Shāng oracle-bone inscriptions, which to this day remain the earliest surviving examples of writing in China, have come to play an important role in understanding the earlier stages of the history of the script. But equally important were the bronze inscriptions and bamboo-slip documents, not only because they provided fresh sources for research but also because these spectacular discoveries gave an impetus that drove the entire field forward.

With the adoption of Western linguistic theories for the study of the Chinese writing system, an increasing amount of attention has been directed to the phonetic nature of the script, and these new findings were also applied to explaining character etymology. As part of this new approach, the traditional categories of *liùshū*, and especially the two distinctly non-phonetic principles of *zhuǎnzhù* and *huìyì*, came under suspicion. While what *zhuǎnzhù* entailed remains a subject of scholarly debate,¹⁰ the meaning of the *huìyì* category is fairly unambiguous: it is a principle according to which two or more components are joined into a single character and their semantic values together come to represent the semantic value of the new composite character. As examples in the “Postface” to the *Shuōwén*, Xǔ Shèn gave the characters 武 and 信, of which at least the former had a literary precedent, having been explained in the *Zuǒzhuàn* 左傳 in terms of its orthography: “the meaning of ‘martial’ refers to halting weapons” 止戈爲武. Since in the *Shuōwén* we find only three more characters explicitly identified as *huìyì* compounds (i.e., 喪, 囧, and 敗), this category appears to be curiously underrepresented among the total number of over 9,000 head entries.¹¹

6. For a comparison of these three Han sources, see Galambos 2006: 56.

7. These three sources may have had the same origin. Pān Zhòngguī 潘重規 (1983: 36) points out that the “Yīwénzhì” was based on Liú Xīn's *Qīliè*, while Liú Xīn's student Zhèng Xīng 鄭興 was Zhèng Zhòng's father. Xǔ Shèn, on the other hand, was the student of Jiǎ Kuí 賈逵, whose father Jiǎ Hui 賈徽 was also the student of Liú Xīn. Thus the three sources may in fact all go back to Liú Xīn.

8. For a study of *liùshūxué* from the Sòng through the Míng periods, see Dǎng 2003.

9. The list of authors quoted in the *Shuōwén jiězì gǔlín* 說文解字詁林 has over two hundred names, whereas their works amount to over a thousand items (Qí and Zhào 1942: 48).

10. In his overview of the historical understandings of the *liùshū* categories, Qiú Xīguī (2000: 157–60) points out that at least nine different interpretations had been proposed in traditional scholarship for the *zhuǎnzhù* principle.

11. It is hard to justify why a separate category is needed for a group with only four examples. One could argue that some of the characters in the *Shuōwén* could still have been understood as *huìyì* compounds, even if not

In the past decades modern linguistic research has effectively refuted Xǔ Shèn's *hùyì* examples. Today most researchers agree, for example, that the character 信, in which the components 人 and 言 ostensibly signified a man standing by his word, the 人 component served as a phonophoric component.¹² These considerations led modern paleographers to understand the historical process of character formation as consisting of only three categories: pictographs (*xiàngxíng*), phonetic compounds (*xíngshēng*), and phonetic loans (*jiǎjiè*).¹³ This is not to say that the concept of *hùyì* compounds disappeared from common usage, as the general public remains largely unaware of these developments and continues to interpret many characters in terms of the semantic values of their constituents. This is still the case in situations where characters are being taught, either to primary school students or to foreigners. Calligraphers and seal carvers, out of respect for tradition, also place great emphasis on the *liùshū* principles, and often deliberately use *hùyì* character structures in their work.

In the West, among the most active opponents of the *hùyì* category was Peter A. Boodberg, and some of his views in this respects have been voiced in a heated academic dispute with Herrlee G. Creel, with Boodberg arguing in favor of the non-ideographic nature of Chinese writing.¹⁴ Boodberg (1937: 346) firmly denied the existence of *hùyì* characters as a class, a claim that was more recently upheld by William G. Boltz in a number of publications (e.g., 1994: 147–49, 153–54; 2006).

At the same time, while the majority of characters traditionally understood as *hùyì* compounds have been proven to include a phonophoric component, there are still compound characters that cannot be explained this way. Wolfgang Behr (2006: 84) has made a convincing point that homosomatic characters (e.g., 炎, 彘, 羸) represent *hùyì* combinations:

Even if we assumed that under a complex theory of homophonophoric series and, *a fortiori*, Old Chinese consonant clusters along with their implied morphologies yet to be fully uncovered, each and every traditional *hùyì* character could be eventually shown to include a phonophoric whose role has been eclipsed by the workings of sound change, simple graphic corruption during the clericization reform, or various other haphazard Houdini acts, homosomatic characters clearly will not yield to such a reasoning.

Behr initially advanced this line of reasoning as an argument against William Boltz's "uncompromising disavowal of the existence of complex characters lacking a phonophoric element during the formative period of Chinese writing," although a footnote in the conclusions to his paper (p. 102 n. 66) acknowledges what he learned after the completion of the manuscript, namely, that Boltz "does not subsume 'homosomatic' characters under the *hùyì* category." Nevertheless, Behr made a compelling point in demonstrating that characters could also be formed by combining non-phonophoric components into a single unit. In a recent article David P. Branner (2011) describes a type of character he calls "portmanteau," the structure of which depends on reading the components "as connected words to form a phrase that defines or denotes the word." He also observes that the portmanteau characters appeared relatively late and their connection with the word they stood for was often unstable,

specifically marked as such. In most cases, however, Xǔ Shèn implied a semantic configuration without using the term *hùyì*. One such device, which appears over two hundred times in the *Shuōwén*, is the term *yì shēng* 亦聲, indicating that a semantic component also played a phonetic role in the character's structure.

12. E.g., Qiú 1988: 99; Boltz 1994: 148–49 and 2009: 106.

13. This threefold category was set up by Chén Mèngjiā 陳夢家 (1956), partly as a response to Táng Lán's 唐蘭 (1935) earlier attempt to reconsider the *liùshū* system. Qiú Xiguī (1988: 167) later suggested that the *xiàngxíng* (pictographs) category should be emended to *biǎoyìzì* 表意字 (semantographs). On the general evolution of the three-principles theory, see Qiú 1988: 163–68.

14. Boodberg 1937 and 1940 vs. Creel 1936 and 1938.

as they could represent “more than one word in records of different ages.” Although this type and its concrete examples partly overlap with the medieval *huìyì* forms I discuss in this paper, Branner makes the point that this principle is different from that of *huìyì*, where the components “contribute abstractly to the overall meaning of the word represented.” It seems to me that this distinction to some extent parallels—apart from its terminology—my own understanding of the differences between the early *huìyì* characters ostensibly created in the formative period of the Chinese script and the later ones that appeared as a subset of popular forms used in medieval manuscript culture.

POPULAR FORMS IN MEDIEVAL CHINA

The terms for variant characters in the Chinese tradition vary according to the field of study. Thus in epigraphy we have *biézi* 別字 (‘other/different characters’); in Dūnhuáng studies *súzi* 俗字 (‘popular forms’); in printed culture a variety of terms referring to “erroneous characters” (e.g., *ézi* 訛字, *wùzi* 誤字, *cuòzi* 錯字).¹⁵ A modern usage is the more comprehensive term *yìtǐzì* 異體字 (‘characters with different forms’) or, less frequently, *yìgòuzì* 異構字 (‘characters with different structure’). But what is important is that each of these terms is understood in contrast to a standard or correct form (正字 *zhèngzì*) and defines the variant character form in relation to that. One of the difficulties with this approach is that not only the degree of flexibility of the standard but even the standard itself underwent diachronic changes. As with most phenomena in history the standard form of characters evolved and was different in various points in time. Thus when it comes to the study of historical variant forms, we are confronted with a “shifting” definition that depends not only on what one looks at, but also who is looking and when (Zeng 2006: 167).

When discussing the evolution of the Chinese writing system, scholars often rely on standard character forms and treat those almost as an abstract set of characters, with little regard for the way those characters actually occur in contemporary archaeological material. It is common to think of the evolution of characters as a course moving along a timeline from ancient forms towards modern ones. While we cannot deny a temporal succession in a historical narrative, this model fails to recognize that the evolution of characters was often a complex process with countless sidesteps and backloops.

The neat line of evolution based on standard characters only makes sense from a retrospective point of view, once we know the forms that succeeded and survived in the long run. But if we step back in time and observe character forms within a particular manuscript culture, we are confronted with an incredible orthographic diversity. Certainly, in most periods there was a form that was considered standard, and at given times this standard was even officially enforced, but manuscript evidence tells us that at the same time a wide variety of non-standard forms were also in use. For example, the eighth-century dictionary *Gānlù zìshū* 干祿字書 (Character Dictionary for Seeking Official Employment) tells us that the standard form of the character 明 at the time was the form 𠄎, the left side of which was written with the component 囧. This follows the *Shuōwén*, which lists the same structure as the official way of writing the character. In manuscripts from the eighth century, however, we almost never find this form, and in most cases the left side of the character is written as 目, even in sutras commissioned by the court that are known for having been executed with exceptional care. Less frequently, in less sophisticated manuscripts we encounter the form consisting

15. For a fascinating collection of the types of mistakes in the traditional field of bibliography (版本學 *bǎnběnxué*), see Cherniack 1994: 102–25. It is worth pointing out that in printed culture character errors for the most part referred not to alternate orthography but to the use of wrong characters in a particular context.

of 日 + 月, which is of course the form that survived in the long run and became the way we write the character today. Yet according to the *Gānlù zìshū*, in the eighth century the official standard was the form 𠄎, which in reality was almost never used in contemporary manuscripts.

Characters are not independent of how they are written. By definition, they are graphic representations, and without a graphic appearance they cannot exist. Considering the array of possible orthographies for a character at a given moment in history, we should be more mindful of the variety of forms in which that character was actually written. Because generally speaking these non-standard forms are not solitary occurrences of peculiar combinations but commonly attested variants with hundreds or even thousands of examples, it would be unjustified to ignore them when studying the evolution of the script. They reflect the writing habits of literate communities and are important witnesses to the forces that governed the evolution of characters, at times revealing information that cannot be gained from the abstracted set of standard forms.¹⁶

While today, in possession of a formerly unprecedented amount of manuscript and epigraphic material, we take pride in correcting obsolete understandings and folk etymologies regarding the composition of characters (e.g., the *hùyì* category), it is important to realize that these historically “incorrect” notions of character structure were often responsible for the structure of popular character forms. We should also keep in mind that most of the medieval manuscripts we have today were written by people who were neither linguists nor paleographers. Their way of seeing the script, even if historically inaccurate, shaped the way they wrote. Therefore a crucial aspect in the development of orthography is that, beside the large historical changes of the script (i.e., the shift from oracle-bone to Warring States forms, the Qín reforms, the transition to clerical script, etc.), characters also varied on an individual level in the process of being used. Although we may have a fairly good understanding of the linguistic and paleographic forces behind the evolution of the writing system, we should not underestimate the way people saw their own script, as these notions might have been more influential in daily usage than historically accurate etymographies.¹⁷

Looking at the orthography of popular forms, we immediately notice the significance of the *hùyì* principle. Throughout Chinese history, scholars have occasionally pointed out the presence of such forms, mostly as a form of criticism. A well-known example comes from Yán Zhītuī 顏之推 (531–91), who describes the appearance of popular forms during the Southern and Northern Dynasties (420–589) in the following way:

北朝喪亂之餘，書跡鄙陋，加以專輒造字，猥拙甚于江南。乃以百念為憂，言反為變，不用為罷，追來為歸，更生為蘇，先人為老，如此非一，遍滿經傳。

In the aftermath of the chaos of the Northern Dynasties, manifestations of writing became vulgar, people created characters according to their fancy, but especially the area south of the Yangzi River stood out in clumsiness and awkwardness. Thus they wrote the character 憂 (‘anxiety’) as the combination of 百 and 念 (‘hundred worries’); the character 變 (‘change, revolt’) as 言 and 反 (‘opposing one’s words’); the character 罷 (‘to dismiss’) as 不 and 用 (‘to not use’); the character 歸 (‘to return’) as 追 and 來 (‘to chase back’); the character 蘇 (‘to revive; to regain consciousness’) as 更 and 生 (‘to revitalize’); the character 老 (‘old’) as 先 and 人 (‘earlier man’). These inconsistencies widely permeated the classics and the commentaries.¹⁸

16. This situation is similar to how linguists working on phonetic reconstruction take into consideration modern spoken dialects. Similarly, it would be equally useful to include popular forms from medieval manuscripts in studying the evolution of the Chinese script.

17. The difference between these two approaches is that the first deals with characters as abstract entities with a fixed orthography, whereas the second looks at concrete examples in manuscripts and inscriptions.

18. *Yán shì jiāxùn* 顏氏家訓 (Family Instructions to the Yán Clan).

The forms brought up as examples in this description are, for the most part, forms attested in medieval manuscripts or traditional dictionaries. The character 憂, for example, is recorded in the *Lóngkān shǒujìng* 龍龕手鏡 (Hand Mirror of the Dragon Niche) as 𪛗, a form that matches Yán Zhītuī's description of 百+念; while the character 老 appears in the form 𪛗 (i.e., 先+人) in the *Wūyīn lèijù sīshēng piānhǎi* 五音類聚四聲篇海. It is also clear that in the above excerpt Yán Zhītuī strongly disapproves of these *hùyì* monstrosities and regards them as a negative trend distorting the classics and the commentaries.¹⁹

A similar complaint, with some overlapping examples, also surfaces in the *Sū shì yányì* 蘇氏演義 (Romance of Mr. Sū) by the late Táng author Sū È 蘇鶚 (fl. 890):

只如田夫民為農，百念為憂，更生為甦[sic]，兩隻為雙，神蟲為蠶，明王為聖，不見為覓，美色為艷，口王為國，文字為學，如此之字，皆後魏時流俗所撰，學者之所不用。In this manner, the character 農 ('peasant') was written as the combination of 田+夫+民 ('field, man, people'); the character 憂 ('anxiety') as 百+念 ('hundred worries'); the character 甦 (蘇, 'to revive; to regain consciousness') as 更+生 ('to revitalize'); the character 雙 ('double') as a double 隻 ('single one'); the character 蠶 ('silk worm') as 神+蟲 ('divine worm'); the character 聖 ('sage') as 明+王 ('enlightened king'); the character 覓 ('to seek') as 不+見 ('cannot see'); the character 艷 ('dazzling and colorful') as 美+色 ('beautiful appearance'); the character 國 ('country') as 口+王 ('king within his domain'); the character 學 ('learning') as 文+字 ('writing').²⁰ Characters of this type were created during the Later Wèi (386–534) by ordinary people but have been not used by scholars.²¹

All of the character formations cited by Sū È are evidenced in both traditional lexicography and medieval manuscripts. Among the material found in the Dūnhuáng cave library, manuscript Or.8210/S.388 from the Stein collection contains a number of lexicographic works, including the dictionary entitled *Zhèngmíng yàolù* 正名要錄 (Essential Record of the Rectification of Names).²² But what is more interesting for us here is that among the linguistic material in this manuscript, there is also a short list of characters that appears to be a collection of *hùyì* compounds (see Figure 1). Following the list a caption says, "To the right are all standard forms, with their corrupted vulgar forms added as a footnote" 右正行者正體，腳注訛俗。Even if the text does not explicitly mention the principle of *hùyì* in the text, most character forms appearing in the small script can be identified as folk-etymological *hùyì* compounds.²³ Although for a few characters this feature is not apparent (i.e., 齊, 弱, 俗), it seems to be the common denominator for the majority of characters on the list.

19. The Yán family continued to play a significant role in the study of popular forms, culminating in the character dictionary *Gānlù zìshū* by Yán Yuánsūn 顏元孫 (d. 714). According to the "Preface" 序 added posthumously by Yuánsūn's nephew, the celebrated calligrapher Yán Zhēnqīng 顏真卿 (708–84), the dictionary began its life with the eminent scholar Yán Shīgǔ 顏師古 (581–645), who collected character forms during his work on editing the classics. The work reached its final form when, out of respect for his late uncle, Yán Zhēnqīng used his calligraphic skills to create a final version and had it carved into stone.

20. Zhāng Yǒngquán (2001: 164) points out that the combinations 田+夫+民=農 and 文+字=學 quoted here are probably mistakes and should be read as 田+民=農 and 文+子=學, respectively. With these adjustments, we get popular forms that are widely attested in the Dūnhuáng manuscripts, whereas the original text listed unprecedented orthographies.

21. Sū È 蘇鶚, *Sū shì yányì* 蘇氏演義 (*Cóngshū jíchéng chūbiān* 叢書集成初編).

22. Nishihara (1981: 16) believes that this text is an early incarnation of the *Gānlù zìshū*.

23. It is surprising to what extent this list of characters overlaps with the character forms criticized by the two accounts of Yán Zhītuī and Sū È. This overlap suggests that the basis for this collection might have been literary, rather than purely paleographic.

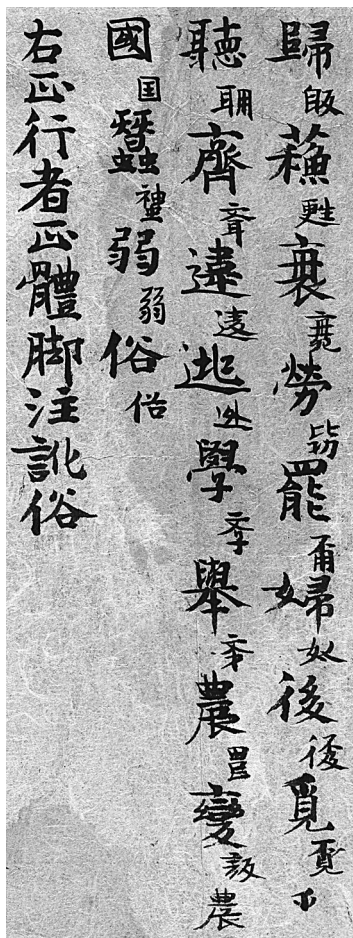


Figure 1. Part of Dūnhuáng manuscript Or.8210/S.388, listing an array of *huiyi* compounds. Conventional forms appear in large script with *huiyi* ones below in smaller.²⁴ Copyright: British Library Or.8210/S.388.

A form that appears here and is also commonly seen in Dūnhuáng manuscripts is the variant form of the character 覓 ('to seek'), written as 覓 (不+見: 'cannot see'). The variant for the character 後 ('behind') is also evidenced in epigraphic material as 後 (彳 + 不 + 及), the left side of which signifies moving, and the right side the inability to catch up or to reach something.²⁵ An even more widespread combination that, with a slight modification, came to be part of our modern orthography is the popular form of the character 國 ('country'), written as 国 (口+王), which indicates a monarch residing inside his domain. The character 歸 ('to return') at the beginning of the list appears as 販, which is a combination of the characters 自 and 反, signifying 'to go back from somewhere'. We have seen above that Yán Zhītuī listed a different form for the same character, that written as 追+來 ('to chase back'). This

24. The images of Dūnhuáng manuscripts are used with permission. The copyright holder for the Stein manuscripts (with pressmarks beginning with 'Or.8210/S.') is the British Library, and for the Pelliot manuscripts (beginning with "Pelliot chinois") the Bibliothèque nationale de France. The images of most manuscripts mentioned in this paper are available online on the website of the International Dunhuang Project (<http://idp.bl.uk>).

25. This form occurs on a Táng dynasty tombstone (see Qín 1985: 88).

demonstrates that there could be more than one folk-etymological orthography for a single character.²⁶

In contrast, the character 蘇 ('to revive; to regain consciousness') matches the description of Yán Zhītuī and Sū È, written as 甦 (更 + 生 = 'to revitalize'). This form occurs, for example, in a "transformation text" (*biànwén* 變文) from Dūnhuáng called *Lúshān Yuán gōng huà* 廬山遠公話 (The Story of Sir Yuán from Lúshān) in manuscript Or.8210/S.2073 (Figure 2). The character occurs in the phrase "came to senses after a long while" 良久乃蘇, perfectly fitting the meaning of the word in context. Yet the character 蘇 appears once more in the same manuscript, only this time in the word *sūlào* 蘇酪, which was a kind of cheese-like dairy product, more commonly written as 酥酪. It is interesting that in this place the character is written with the same *hùyì* orthography (更 + 生), even though the word it represents has nothing to do with the concept of "revival." This demonstrates that although this formation owes its etymology to the combination of the semantic values of characters 更 and 生, it has become equivalent to the character 蘇, regardless of its usage in context. In other words, the compound form is linked with the character, not the word it represents.

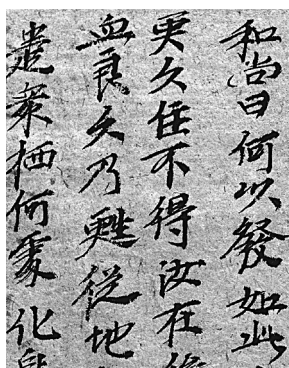


Figure 2. Section of manuscript Or.8210/S.2073 where the character 蘇 (fifth character in third line) is used for the verb 'to revive', and is written as the semantic combination 更 + 生 (>甦). Copyright: British Library Or.8210/S.2073.

A similar case is when the character 切 is written as the combination of 下 + 刀 ('to start cutting'), obviously referring to the semantic field of the character associated with the verb 'to cut'. This is, of course, the basic meaning of 切, and the *Shuōwén* glosses it with the word *cǎn* 刊 ('to cut'), explaining that it is made up of the semantic component 刀 ('knife') and the phonetic component 七 ('seven'). Yet we see that in manuscript Gānbó 078 the character appears written as 下刀 (下 + 刀 = 'to start cutting') in the word *yíqiè* (一切 'all, every'), even though in this place its semantic value is unrelated to cutting.

Medieval dictionaries contain a surprisingly large number of *hùyì* combinations, even if many of these never occur in written material outside of the domain of lexicography. To be sure, the number of such lexicographic ghosts shrinks as newly discovered texts become available; yet one can easily see that some of the fancier dictionary forms are unlikely ever to occur in manuscripts or epigraphic sources. Having said this, we now have a sufficient amount of material to demonstrate that *hùyì* forms were indeed commonly used in everyday

26. The form consisting of the combination of 自 and 反 is not merely a lexicographic oddity but also appears in ordinary manuscripts. Examples of this are Or.8210/S.2832 and Or.8210/S.4624, both of which use the same form in continuous text.

life and they are a regular feature of medieval manuscripts. Figure 3 shows several additional characters from the Dūnhuáng corpus.²⁷

學	孛	Pelliot chinois 2721	文+子 (‘educated child’)
塵	𡗗	Pelliot chinois 2133	少+土 (‘a small amount of soil’)
聖	聖	Dūnyán 194	明+王 (‘enlightened king’)
蠶	蠶	Or.8210/S.134	天+天+虫 (‘heavenly worm’)
掘	掘	Pelliot chinois 3873	扌+入+土 (‘penetrating the soil with hands’)
逃	逃	Or.8210/S.328	外+辵 (‘go beyond the outer [edge]’)

Figure 3. Examples of *hùyì* forms from the Dūnhuáng corpus

Most of these examples are also documented in lexicographic sources, and a popular form of the character 塵 closely resembling that in Figure 3 (少+土) has come down to our days as the simplified form of the same character: 尘 (小+土).

An interesting case of the use of *hùyì* characters during the medieval period is the characters used in Zhuàng writing, occasionally referred to in Chinese sources.²⁸ One of these is the late-twelfth-century gazetteer called *Lǐngwài dàidá* 嶺外代答 (In Lieu of Answers from Beyond the Passes), written by Zhōu Qūfēi 周去非 (ca. 1135–89) during his tenure in Guǐlín. In a section called “Popular Forms” 俗字, the author recorded some of these orthographic peculiarities he collected in the Guǎngxī area (Zhōu 1999: 161–62):

廣西俗字甚多:如囊音矮,言矮則不長也。壘音穩,言大坐則穩也。癸音動,言瘦弱也。歪音終,言死也。髻音臘,言不能舉足也。佻音嫻,言小兒也。媯徒架切,言姊也。門音欄,言門橫闕也,音礪,言岩崖也。余音涸,言人在水上也。炎音魅,言没人在水下也。鬚音鬚,言多髭。研東敢切,言以石擊水之聲也。大理國間,有文書至南邊,猶用此囿字,囿武后所作國字也。

There are a great many popular character forms in the Guangxi area. For example, 囊 is pronounced as the character 矮, meaning ‘short’, i.e., ‘not tall’; 壘 is pronounced as the character 穩, meaning ‘sitting cross-legged’, i.e., ‘steady’; 癸 is pronounced as the character 勑, meaning ‘feeble’; 歪 is pronounced as the character 終, meaning ‘death’; 髻 is pronounced as the character 臘, meaning ‘cannot lift one’s feet’; 佻 is pronounced as the character 嫻, meaning ‘little son’; 媯 is pronounced *daɪH* (徒架切),²⁹ meaning ‘elder sister’; 門 is pronounced as the character 欄, meaning a ‘horizontal bolt on the door’; 岳 is pronounced as the character 礪, meaning a ‘cliff’; 余 is pronounced as the character 涸, meaning ‘a man on the surface of the water’; 炎 is pronounced as the character 魅, meaning ‘a man sunken beneath the water’; 鬚 is pronounced as the character 鬚, meaning ‘abundant facial hair’; 研 is pronounced *tamX* (東敢切), meaning ‘the sound of water hitting the water’. In the Kingdom of Dali there are documents in the southern part of region still using the character 囿, which is an [alternate] form of the character 國, created by Empress Wǔ.

While these characters were used to write a non-Chinese language, the presence of the *hùyì* principle in the characters is evident. As for the final claim that the Empress Wǔ character form 囿 had been used centuries after the end of her reign (705), this is also corroborated by epigraphic material from Yúnnán.³⁰

27. Examples taken from the *Dūnhuáng súzìdiǎn* 敦煌俗字典 (Huáng 2005).

28. On the use of Chinese characters in Zhuàng manuscripts, see Holm 2009.

29. Transcription is based on Baxter 2000.

30. Apparently this character form proved to be especially lasting in Yúnnán, having been used for centuries. On this, and on the spread and survival of the form 囿 in Yúnnán after the end of the reign of Empress Wǔ, see

FOLK ETYMOLOGY VS. HISTORICAL EVOLUTION

What a particular element in a character form represents is one of the basic questions in a paleographic analysis; yet at times the answer is not straightforward. When the element is visible and corresponds to something already known, it generally poses no difficulty to establish its identity. But in many cases the element does not resemble anything we might recognize as a meaningful unit within the writing system, or appears to be something that does not make sense. The usual solution to this problem is to trace the etymology of the element historically and see from what it evolved and what it used to represent in the past. Such analyses often result in claims that elements that are easily recognizable graphically in fact represent something else; that is, they have evolved from other elements and acquired a different physical appearance. A simple example for such a mental “redirection” is the case of the radical 肉 (‘meat’), which in the *kǎi* script is commonly written as 月 (‘moon’). Although we can easily recognize the component 月 graphically, based on etymological considerations we identify it with the radical 肉, even if this is clearly a departure from what we perceive visually. The case of the character 明 is a less obvious example for most users who do not realize that the component 日 here might in reality be standing for another component, as etymological considerations suggest.³¹ At the same time historical etymologies can also be part of the general knowledge of society, as is the case with the radical 肉 being written as 月 in a series of characters. Accordingly, many modern dictionaries still list characters with this component under the radical 肉. The character 服, however, etymologically derives from the character 舟 (‘boat’); yet most dictionaries no longer retain this information and list it under the component 月.

My contention is that the orthography of a particular character form as it appears in medieval manuscripts carries in itself not only the genetic information of its historical evolution but also the contemporary (i.e., Táng or Sòng) interpretation of what character structure it represented. When a Táng scribe wrote the character 明 as the now obsolete combination of 目 (‘eye’) and 月 (‘moon’), he was no doubt more aware of the connotations involved in this orthography than of the actual evolution of this character during and before the Warring States period.

The call to understand characters according to their existing graphic structure instead of the character’s etymology (attested or imagined) is by no means a modern invention. In medieval lexicography the *Lóngkān shǒujǐng* is one of the earliest dictionaries that makes an effort to analyze character structure not in terms of historical etymology but according to the way they appear in manuscripts, and it records an incredible variety of popular forms in common use.³² The same type of approach is noted by Bottéro (1996) to have been followed by Hán Xiàoyàn 韓孝彦 and Hán Dào zhāo 韓道昭, the father and son who compiled the

Zhāng Nán 1992. On the use of Empress Wǔ characters for dating, see also Drège 1984 and Galambos 2004. We should also mention that, although far from being popular forms, among the eighteen characters introduced by Empress Wǔ, several were undoubtedly constructed on the basis of the *hùyì* principle. For example, the character 照 (‘radiance’) written as 𠄎, a combination of 明 + 空 (‘bright sky’); the character 人 written as 𠄎, the combination of 一 + 生 (‘one life’); the character 地 written as 𠄎, a combination of 山 + 水 + 土 (‘mountains, waters, and earth’).

31. For an etymological discussion of the character 明, see Boodberg 1940: 270–73.

32. When the identity of a popular form is not immediately apparent from its form, the *Lóngkān shǒujǐng* uses the formula A音B to indicate that character A should be read as character B. This is, therefore, not simply a phonetic gloss on character A but its complete identification with character B. For example, the Empress Wǔ form for the character 人 is 𠄎, a combination of the characters 一 (‘one’) and 生 (‘life’). The dictionary lists the form 𠄎 and says: “古文, 音人” (this is the *gǔwén* form and is pronounced as the character 人). But although it merely claims to give the pronunciation in such cases, in reality it is consistent in providing the character that is the standard form used for the same word; it never gives a homophonous character that is semantically unrelated.

mammoth dictionary *Wǔyīn lèijù sishēng piānhǎi*. Boltz (2000: 474) summarizes this point in his review of Bottéro's book the following way:

[Daozhao] has classified a number of characters according to the graphic structure apparent to him instead of according to a character's known historical source or traditional understanding. For example, he has classified 畏 (*wei*) under 田 'field' instead of the "ghost-head" classifier 田 under which it has a genuine lexical affinity, thus severing it from its historical and etymological link to *gui* 鬼. The general tendency, not surprisingly, is to see the development of character classification moving away from its historical grounding and toward an observation and empirical pragmatism, not unlike the modern tendency in some quarters to spell English words impressionistically "as they are pronounced," for example, "night," "phantasy," and "through" as "nite," "fantasy," and "thru."

It is clear that a move towards such "empirical pragmatism" is one of the features of folk-etymological explanations of character structure.³³ While in most cases the historical identity of character components remains evident to contemporary users, in cases where the component loses its functional transparency, a new graphically based orthography may de-obscure the structure and provide a more meaningful solution.³⁴

The fundamental difference between these medieval *hùyì* characters and the "original" ones ostensibly created at the dawn of character formation is that, while the latter are often seen as created at the "formative stage" of Chinese writing, the medieval ones were generally the result of graphic assimilation of obscure structures into a composition that "made sense." In other words, these popular orthographies started off with compositions where either a component was too complex or the phonetic component was no longer immediately detectable. Thus under the character 愛 ('love') the *Lóngkān shǒujìng* records the variant 爰, made up from the combination 及 + 心 ('reaching with the heart'). This orthography is a gloss of the word *ài* ('love') while retaining a graphical similarity to the original character. Although the character 愛 is a *xíngshēng* formation, the original phonophoric component 无 of its small seal form lost its transparency during the process of clericization, and the popular form 爰 in the *Lóngkān shǒujìng* represents a purely semantic combination with no phonophoric component.

An interesting example of an intentional use of the *hùyì* principle is the name of Hóngbiàn 洪辯, the "bishop" of Dūnhuáng during the first half of the ninth century (manuscript Pelliot chinois 3720). The character 辯 ('to debate') in his name invariably appears in manuscripts written as 巧 + 言 ('to talk skillfully'). This particular form is relatively common: it appears not only in manuscripts but also in virtually every medieval dictionary.³⁵ But what makes it interesting is that it is used consistently in a name.³⁶ Since this is a monastic name, we can be certain that Hóngbiàn himself was not only well aware of the implications of this *hùyì* combination but also cherished them.

33. Technically speaking, the term folk etymology is not entirely correct because we are not dealing with etymology, or even etymography, here. Manuscript evidence suggests that, for literate people in medieval China, *hùyì* combinations meant not the evolution of a character but its current structure and semantic implications as they saw in it.

34. This function of "de-obscuring" is described in Rundblad and Kronenfeld (2003: 121–22) as one of the key elements of folk etymology.

35. Another interesting *hùyì* variant for this character that appears in the *Wǔyīn lèijù sishēng piānhǎi* is the combination of 金 + 言 ('golden speeches'), with 金 appearing atop 言.

36. There are also numerous modern examples of non-standard characters preserved and used consistently in personal and place names, especially in Japan. The character 嶋, for example, a variant of 島, commonly occurs in Japanese surnames. The Buddhist temple Honkokuji 本國寺 in Kyōto is to this day written with the Empress Wǔ form of the character 國 (圀).

At the same time these *hùyì* examples are not arbitrary structures made up of two or more characters that convey the meaning of the original character. Generally speaking, a precondition is the graphic similarity to the standard form.³⁷ In this sense this process of *hùyì* character creation differs from the early stages of the Chinese script, where this was probably not the case. But in medieval manuscript culture, popular characters of the *hùyì* type employed folk-etymological glosses whenever the overall balance of the original character permitted this. This principle also limited the potential number of such *hùyì* forms, as they were not ad hoc creations but part of a tendency to turn running hand forms into semantically meaningful combinations, while maintaining their general balance. The degree of tolerated divergence was determined by the legibility of the new form, as readers still had to be able to recognize it as a variant of the conventional form. A structure too dissimilar would have impeded immediate recognition and made the variant form impractical.

We should also point out that *hùyì* characters were only one of the possible configurations among popular character forms. There were cases where the phonophoric component of the standard form was replaced by another one, with a similar pronunciation but usually fewer strokes. The popular form of the character 髀 ('thigh'), for example, is sometimes written in the Dūnhuáng manuscripts as 𦍋, where the phonophoric 卑 is replaced with the nearly homophonous component 𦍋. The *Lóngkān shǒujìng* in addition records the form 𦍋, where the phonophoric is the component 𦍋. In other cases the popular orthography of the character lost its phonetic information without becoming a semantic compound. For example, the character 礙 ('obstacle'), with the phonophoric 疑, commonly appears in manuscripts written as 𦍋 (matching the right side of the character 得), a form that already appears in Hàn dynasty inscriptions. At the same time, this form (𦍋) seems to carry no phonetic information that would be relevant to the pronunciation of the character (礙).³⁸ In sum, the components in many popular forms used in medieval manuscripts did not form meaningful compositions, and at times are not even identifiable.³⁹

Therefore these *hùyì* characters represent a special case of popular forms, which initially owe their existence to a graphical semblance to possible semantic configurations. If this orthographic potential becomes apparent for the users of the script, they may exploit it through emphasizing the individual components in an attempt to create a more logical and transparent character structure. The exception to this rule are the characters created during the reign of Empress Wǔ, as most of these are ideologically motivated formations that have no graphical connection with their standard forms. They are artificial creations prescribed and enforced by a political authority, rather than developed gradually through everyday usage, as it was the case with popular character forms.

CONCLUSIONS

The medieval *hùyì* formations described in this paper are non-standard or popular forms that appear in the Dūnhuáng manuscripts. The key point in my methodology was to include these popular forms in the study of the evolution of the script, as, from the perspective of

37. Naturally the new form sometimes emulates the graphical shape not of the standard *kǎi* form but of the different types of running hand versions.

38. This form, or rather its variant which in addition includes the 石 radical, came to be accepted as the modern simplified form of the character 礙, now written in mainland China as 碍.

39. We should also note that one of the commonly seen cases is when the semantic component in a character was replaced by another, often synonymous one, while maintaining the *xíngshēng* structure of the whole character. The character 體, for example, was also commonly written with the component 身 on its left side, instead of the component 骨 that appears in its standard form.

medieval manuscript culture, they represented one of the several possible ways of writing a given character. While their semanticization of character structure has often been described as a folk etymological approach to orthography, they nevertheless reflected how people at the time understood the composition of characters, even if that was historically often inaccurate. We commonly regard folk-etymological changes as a form of corruption, in which the “original” structure of a character becomes compromised. In reality, however, they are adjustments that serve to rationalize a structure that no longer seems valid or has lost its transparency. From this point of view these are not corruptions but rather improvements.

My proposition is that in our study of the history of writing we should be able to look at the state of affairs synchronically, at specific moments in time, and analyze what we see as a comprehensive and self-contained system, instead of tracing back character structure in time to different stages of evolution. The Dūnhuáng manuscripts provide us with ample source material to do this for the medieval period. What we see in this corpus is that a large number of non-standard forms were in use, and many of these had a structure that could be classified as semantic compounds. In other words, *hùyì* forms were an inherent feature of the writing system during this period, even if they were occasionally criticized by contemporary scholars as being incorrect.

REFERENCES

- Baxter, William H. 2000. *An Etymological Dictionary of Common Chinese Characters*. [Preliminary draft of 28 October 2000.] <http://www-personal.umich.edu/~wbaxter/etymdict.html> (last accessed 16 November 2010).
- Behr, Wolfgang. 2006. “Homosomatic Juxtaposition” and the Problem of “Syssemantic” (*Hùyì*) Characters. In *Écriture chinoise: Données, usages et représentations*, ed. Françoise Bottéro and R. Djamouri. Pp. 75–114. Paris: École des Hautes Études en Sciences Sociales, Centre de Recherches Linguistiques sur l’Asie Orientale.
- Boltz, William. 1994. *The Origin and Early Development of the Chinese Writing System*. American Oriental Series, vol. 78. New Haven: American Oriental Society.
- . 2000. Review of Bottéro 1996. *JAOS* 120: 471–74.
- . 2006. Phonographic Motivation in the Formation of Compound Chinese Characters: The Case of Wu 武. In *Écriture chinoise: Données, usages et représentations*, ed. Françoise Bottéro and R. Djamouri. Pp. 56–73. Paris: École des Hautes Études en Sciences Sociales, Centre de Recherches Linguistiques sur l’Asie Orientale.
- . 2009. Orthographic Variation in Early Chinese Manuscripts. *Acta Orientalia Scientiarum Hungaricae* 62: 89–113.
- Boodberg, Peter A. 1937. Some Proleptical Remarks on the Evolution of Archaic Chinese. *Harvard Journal of Asiatic Studies* 2: 329–72.
- . 1940. “Ideography” or Iconolatry? *T’oung Pao* 35: 266–88.
- Bottéro, Françoise. 1996. *Sémantisme et classification dans l’écriture chinoise: Les systèmes de classement des caractères par clés du Shuōwén jiězì au Kāngxī zìdiǎn. Mémoires de l’Institut des Hautes Études Chinoises*, vol. 37. Paris: Collège de France, Institut des Hautes Études Chinoises.
- Branner, David P. 2011. Portmanteau Characters in Chinese. *JAOS* 131: 73–82.
- Callery, Joseph-Marie. 1841. *Systema phoneticum scripturae Sinicae*. Macao.
- Chén Mèngjiā 陳夢家. 1956. *Yīnxū bǔcí zōngshù* 殷墟卜辭綜述. Shànghǎi: Shànghǎi Rénmín Chūbǎnshè.
- Cherniack, Susan. 1994. Book Culture and Textual Transmission in Sung China. *Harvard Journal of Asiatic Studies* 54: 5–125.

- Creel, Herrlee Glessner. 1936. On the Nature of Chinese Ideography. *T'oung Pao* 32: 85–161.
- . 1938. On the Ideographic Element in Ancient Chinese. *T'oung Pao* 34: 265–94.
- Dǎng Huáixīng 黨懷興. 2003. *Sòng Yuán Míng liùshūxué yánjiū* 宋元明六書學研究. Běijīng: Zhōngguó Shèhuì Kēxué Chūbǎnshè.
- Drège, Jean-Pierre. 1984. Les caractères de l'impératrice Wu Zetian dans les manuscrits de Dunhuang et de Turfan. *Bulletin de l'École française d'Extrême-Orient* 73: 339–54.
- Galambos, Imre. 2004. Dunhuang Characters and the Dating of Manuscripts. In *The Silk Road: Trade, Travel, War & Faith*, ed. Susan Whitfield. Pp. 72–79. Chicago: Serindia.
- . 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.
- Holm, David. 2009. A Typology of Readings of Chinese Characters in Traditional Zhuang Manuscripts. *Centre de Recherches Linguistiques sur l'Asie Orientale* 38: 245–92.
- Hopkins, Lionel Charles. 1881. *The Six Scripts*. Amoy.
- Huáng Zhēng 黃征. 2005. *Dūnhuáng sùzìdiǎn* 敦煌俗字典. Shànghǎi: Shànghǎi Jiàoyù Chūbǎnshè.
- Lundbaek, Knud. 1988. *The Traditional History of the Chinese Script: From a Seventeenth Century Jesuit Manuscript*. Aarhus: Aarhus Univ. Press.
- Morrison, Robert. 1815. *A Dictionary of the Chinese Language*. Macao: East India Company Press.
- Nishihara Kazuyuki 西原 一幸. 1981. Tōdai kaisho jisho no seiritsu: *Ganshi jiyō kara Kanroku jisho, Gokei monji e* 唐代楷書字書の成立:『顏氏字樣』から『干祿字書』『五經文字』へ. *Kinjō gakuin daigaku ronshū (Kokubungaku hen)* 金城学院大学論集(国文学編) 23: 1–27.
- Owen, George. 1910. *The Evolution of Chinese Writing*. Oxford: Horace Hart.
- Pān Zhòngguī 潘重規. 1983. *Zhōngguó wénzìxué* 中國文字學. Táiběi: Dōngdà.
- Qí Pèiróng 齊佩瑢 and Zhào Yīntáng 趙蔭棠. 1942. *Zhōngguó wénzìxué gàiyào* 中國文字學概要. Běijīng: Guólì Huáběi Biānyìguǎn.
- Qín Gōng 秦公. 1985. *Bēi biézi xīnbiān* 碑別字新編. Běijīng: Wénwù Chūbǎnshè.
- Qiú Xīguī 裘錫圭. 1988. *Wénzìxué gàiyào* 文字學概要, tr. Gilbert L. Mattos and Jerry Norman. Běijīng: Shāngwù Yīnshūguǎn.
- . 2000. *Chinese Writing*. Berkeley: Society for the Study of Early China and the Institute of East Asian Studies, Univ. of California, Berkeley.
- Rundblad, Gabriella, and David B. Kronenfeld. 2003. The Inevitability of Folk Etymology: A Case of Collective Reality and Invisible Hands. *Journal of Pragmatics* 35: 119–38.
- Táng Lán 唐蘭. 1935. *Gǔwénzìxué dǎolùn* 古文字學導論. Běijīng: Běijīng Dàxué. Republished in Jīnán (1981) by Qílǚ Chūbǎnshè.
- Wieger, Léon. 1927. *Chinese Characters: Their Origins, Etymology, History, Classification and Signification*. Xianxian: Catholic Mission Press. (First pub. 1915.)
- Zēng Liáng 曾良. 2006. *Sùzì jí gǔjí wénzì tōnglì yánjiū* 俗字及古籍文字通例研究. Nánchāng: Bǎihuāzhōu Wényì Chūbǎnshè.
- Zhāng Nán 張楠. 1992. Wǔ Zhōu xīnzi guó zài Yúnnán de liúchuán kǎoshì 武周新字“囿”在云南的流傳考釋. *Gùgōng Bówùyùàn Yuànkān* 故宮博物院院刊, no. 3: 60–61.
- Zhāng Yǒngquán 張涌泉. 2001. Cóng yǔyán wénzì de jiǎodù kàn Dūnhuáng wénxiàn de jiàzhí 從語言文字的角度看敦煌文獻的價值. *Zhōngguó Shèhuì Kēxué* 中國社會科學, no. 2: 155–65.
- Zhōu Qùfēi 周去非. 1999. *Língwài dàidá jiàozhù* 嶺外代答校注. Yáng Wǔquán 楊武泉, ed. Zhōngguó Wàijiāotōng Shǐjí Cóngkān 中外交通史籍叢刊. Běijīng: Zhōnghuá Shūjú.