CHINA ACROSS THE CENTURIES
Papers from a lecture series in Budapest

EDITED BY
GÁBOR KÓSA

CONTENTS

Graphic Variation in Early Chinese Writing

IMRE GALAMBOS
(University of Cambridge)

1. Introduction[1]

Reading Warring States manuscripts we are confronted with a number of graphs that are not only structurally different from modern characters and the small seal forms of the *Shuowen jiezi* 說文解字 but show variation even among themselves. While some of these graphs may be characters that have since disappeared, the majority of them are variants of known ones and represent words in much the same way as seen in transmitted sources. In early manuscripts we often see character structure varying within the same corpus or, at times, even within the same document. In most cases, the context provides enough information for deciphering the meaning of graphs, yet it is always a question whether we can link structural discrepancies with grammatical differences or shades of meaning. In other words, do graphic differences have any relevance to how the word is to be interpreted? Or are they inconsequential errors committed by scribes working in a hurry or perhaps possessing lower literacy skills?

As an example, consider the *Guodian Laozi* 郭店老子 manuscripts where the word *nan* "difficult" occurs six times (Figure 1).[2]

---

[1] Some of the discussion in this chapter relies on my book on the orthography of the early Chinese script (Galambos 2006), although most things have been re-thought and emphases have shifted. My analysis of graphs is also different and geared towards providing a more balanced and comprehensive representation of the corpus.
The current research has been carried out as part of the "Xifang hanzi xueshi" 西方漢字學史 project funded by the Ministry of Education of the PRC.

[2] Although in two of the examples one could argue that the word is used in a nominal sense, it is nevertheless the same word meaning "difficult", read in modern Chinese as *nan* and contrasted with 易 *yi* "easy," rather than the word *nan* "hardship, disaster."
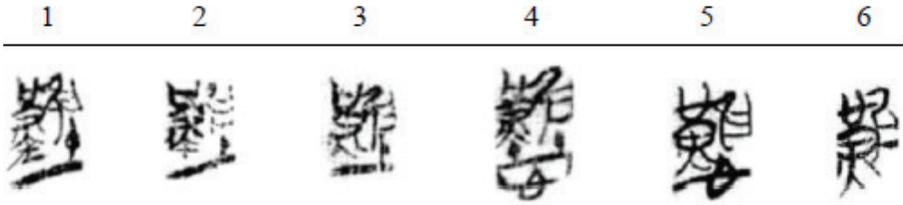
Figure 1. Six occurrences of the word *nan* "difficult"
in the *Guodian Laozi* manuscripts.[3]

The primary difference between these six graphs is in the presence or absence of the components 土 and 心. Only #1 and #2 are structurally identical, the rest of the graphs exhibit component-level (i.e. orthographic) variation. Some of them overlap with existing characters, such as 戁 (#5), which is defined in the *Shuowen* as "to revere" 敬也, and thus there it clearly represents a different word.[4] Only form #6 matches the modern structure of the character 難, which includes neither the component 土 nor 心. Now all of these five orthographically distinct configurations come from the same set of manuscripts and were written either by the same person or by several persons within the same community. Moreover, five of the six above forms, with the sole exception of #5, come from the so-called *Laozi A* manuscript, where they all appear within a few characters of each other, strongly suggesting that they were very likely written by the same person.[5]

Since these examples come from a philosophical text, it is hard to deny the possibility that orthographic differences may represent lexical variants. While it is unlikely that the scribes varied character structure independent of language to approximate the meaning of the word in context,[6] it would still be possible to argue that the words themselves were also different, even if such nuances did not survive in the received versions of the text. Or perhaps the

---

[3] Images are from *Guodian Chu mu zhujian* 郭店楚墓竹簡 (Jingmen shi bowuguan 1998).

[4] Incidentally, this structure is also attested as a variant in the Ming-dynasty dictionary *Zixue sanzheng* 字學三正 (1601) as a *guwen* 古文 form of the character 然.

[5] More precisely, they appear on slips #12, 14, 15 and 16, that is within a cluster of a few sentences.

[6] A similar idea is expressed by David Branner (2011: 94) who points out that "Chinese characters are read, not deciphered."

variant provided the opportunity to distinguish between two shades of meaning, even if both of these were normally written with the same character.[7]

The examination of graphical variation requires a context free from such complexities, where the possibility of lexical variation can be excluded. The Houma covenant texts (侯馬盟書) from the early century BC provide an ideal corpus for such a study, as they consist of hundreds of examples of the same formulaic texts. As we will see below, the covenant texts demonstrate that graphic variability was an inherent quality of early Chinese handwriting and that variant graphs in many cases did not necessarily signify lexical variance. This orthographic flexibility seems to have been a normal feature of contemporary writing habits and variants were accepted and tolerated by both readers and writers.

Part of the difficulty of getting a clear picture of scribal habits of early Chinese manuscripts is not that this body of material is separated from us by over two millennia but that we live in an age of printed — or even digital — texts and our own notions of orthographic consistency are vastly different from those in manuscript cultures. This dichotomy between printed and handwritten textual cultures is so pronounced that the main obstacles in assessing the scribal habits of the Warring States period might stem from our modern writing habits, rather than the antiquity of the manuscripts. In terms of their orthographic patterns, the medieval manuscripts from Northwest China (Dunhuang, Turfan, Khara-Khoto) discovered in the first decades of the twentieth century are closer to pre-Qin Chinese manuscripts than the books of our age.[8]

---

[7]   An example of such differentiation is the two different characters in the *Shuowen* for the word *lian* "curtain," one written as 帘 (帷也), the other as 簾 (堂簾也). Duan Yucai 段玉裁 noted that the two characters actually denoted different objects, as the first was made of fabric, the other of bamboo. This point of view, however, seems to reflect a lexicographic rationalization concerned with presenting a balanced system. If the two objects were ever differentiated in actual writing, this fine distinction did not survive and in later periods both these words were commonly written with the character 簾.

[8]   The study of variant characters in medieval China, which initially grew out of the efforts of transcribing literary texts from Dunhuang, is significantly more advanced than the same field for early Chinese manuscripts. See especially the works of Zhang Yongquan 張涌泉, e.g. Zhang 1995 and 1996.

2. The concept of standard in early China

Chinese definitions of character variants (*yitizi* 異體字) generally start off with the concept of a standard, claiming that a variant is meaningful only in comparison with a standard form, from which it differs.[9] Of course, the notion that variants are non-standard graphs is a relatively modern view reflecting efforts to standardize the script and eradicate variants, whereas the problem of graphical variation in pre-Qin manuscripts was not part of the picture when these terms emerged. This approach is clearly not tenable for the early script because we know next to nothing about the standards of those times. Although the use of writing presupposes a tradition of literacy acquisition, which by default enforces a normative aspect to the equation, the first references to what we could call a standard come from the end of the first century AD and make claims about a script reform that had happened more than three centuries earlier, during the early years of the short-lived Qin dynasty.[10]

Similarly, we do not know the Warring States terminology for variant graphs, if there was such. The earliest surviving term appears in the *Shuowen*, where alternate graphs are called *chongwen* 重文.[11] Yet the issue of writing characters according to a strict norm must go back earlier. In the *Shiji* 史記, we read the story of Shi Jian 石建 from the second century BC, who apparently took the issue of orthographic standard very seriously:

建為郎中令，書奏事，事下，建讀之，曰： 『誤書！「馬」者與尾當五，今乃四，不足一，上譴死矣！』甚惶恐。其為謹慎，雖他皆如是。

---

9   The *Guoyu cidian* 國語辭典, for example, defines it as "a form contrasted with a designated standard character." The foreword of the online *Yitizi zidian* 異體字字典 published by the Ministry of Education of Taiwan in 2001 claims that "a so-called 'variant character' is a character which, in comparison with a standard form, appears in textual sources with the same pronunciation and meaning as the standard character but differs from that in its graphical form." The *Hanyu dacidian* 漢語大詞典, on the other hand, offers a more balanced definition, at least from the point of view of the pre-Qin script, stating that these are "characters with identical sound and meaning but different shape."

10   This appears in the "Postface" of the *Shuowen* and the "Yiwenzhi" 藝文志 chapter of the *Hanshu* 漢書. For a comparison of the these two accounts, see Galambos 2006: 33–35.

11   Although the *Shuowen* also routinely includes *zhouwen* 籀文 and *guwen* 古文 forms, these do not appear to be variants *per se* but rather earlier forms the structure of which differed from the small seal characters he used as head entries in the dictionary.

When Jian held the post of Chamberlain for Attendants, he wrote a proposal to be submitted to the throne. Once the approval came down, Jian read it and exclaimed: "It is written incorrectly! In the character 馬 there should be five [strokes] together with the [horse's] tail. Here I only have four, missing one. I will be held responsible and punished!" He was utterly terrified. In other matters he was just as careful and alert.

According to the *Shuowen*, the small seal form of the character 馬, written as 𥃟, "depicts the shape of the horse's head, mane, tail and four legs" (象馬頭、髦尾、四足之形). Yán Shigu 顔師古 adds a comment to the *Hanshu* 漢書 version of the same incident, saying, "In the character 馬 the [stroke] that is bending downwards is the tail, which together with the four dots representing the legs makes five [strokes]" (馬字，下曲者尾，并四點爲足，凡五). The interesting thing, however, is that the character with a missing stroke is not a random mistake but an attested variants in common use at the time. The form 馬, written with three dots underneath instead of the standard four, is included in most epigraphic dictionaries.[12] Thus the mistake in this case meant that an existing variant was used in an inappropriate context. This signifies the domain-specific use of graphs in Han China, as some occasions demanded the use of highly standardized forms, while alternate forms continued to be commonly used in other social contexts.[13]

Another Western Han reference to using incorrect variants comes from the "Appendix" 附錄 to the *Zhanguoce* 戰國策 by the great Han bibliographer and editor Liu Xiang 劉向 (79–8 BC):

所校中戰國策書，中書餘卷，錯亂相糅舛 …… 本字多誤，
脫為半字，以趙為肖，以齊為立，如此字者多。

---

[12] For example, the *Beibiezi xinbian* 碑別字新編 records this form as coming from a Sui dynasty tomb inscription (Qin 1985: 144).

[13] This domain-specific use of correct vs. alternate forms during the Tang period was explicated in the preface of the *Ganlu zishu* 干祿字書, a character dictionary recording *zheng* 正 (correct), *su* 俗 (vulgar), and *tong* 通 (common) forms, each of which was considered appropriate for different occasions; see Bottéro 1996: 122–124. As an example of varying degrees of orthographic flexibility in two different versions of the same text, see Matthias Richter's study of the graphic variation in the Mawangdui *Laozi* manuscripts, which demonstrates that the two manuscripts followed different orthographic standards, *Laozi* B being more uniform (Richter 2005: 203).

> Among the Warring States historical documents stored in the imperial archives, the documents I edited had many volumes that were erroneous and disorganized. […] The original characters [in these documents] were often mistaken and omitted half of the character, writing 趙 as 肖 or 齊 as 立. There were many characters like this.

The forms cited by Liu Xiang are, once again, attested variants from the Warring States period. Seals from this period, for example, habitually write the surname Zhao 趙 as 㕚 or 㐱, leaving off the component 走. As for the character 齊, although we do not have a graph that is a perfect match for the one mentioned by Liu Xiang, there are some similar examples. For instance, the Baoshan corpus uses the graph 迮, which consists of the combination of the components 辶 and 齊. In this structure, the component 齊 appears in a highly abbreviated form, making it graphically very similar to 立. Accordingly, both variants in Liu Xiang's comment go back to common forms in use during the Warring States period. But by the end of the first century AD, these early manuscript forms were seen by some as mutilated graphs which violated the orthographic standard. This signifies an emergence of a critical attitude towards variants as we enter the Han period, which was the time of organizing and systematizing natural and human phenomena.

In contrast with this, we do not have sources from before the Qin about the concept of an orthographic standard against which variants or mistakes could be judged. Although it is probable that the script underwent some reforms before the Qin period, excavated manuscripts on the whole present an image of orthographic diversity. Some characters are written remarkably consistently, while others much less so. We cannot, however, by looking at this new evidence in itself, demonstrate that any of the existing forms were thought of as the "correct" ones. Similarly, we do not know when graphs were considered incorrect, as corrections evidenced in early manuscripts generally are on the level of lexical variation, that is, they involve changing the entire character and the word it stands for. In view of our lack of information on contemporary assessment of graphic variants, from a methodological point of view it makes more sense to regard variants not as forms differing from a correct or standard form but as alternate representations of the same word, none of which is more correct than the other.

At the same time, I hasten to assert that I do not claim that no standard whatsoever existed in pre-Qín scripts. This would be equivalent to saying that the script had no rules. On the contrary, convention lies at the core of any writing system and the script is by definition a system based on the associ-

ation of elements of speech with graphical signs. In early China, even those cases of variation which from our point of view appear to be random coincidences might fall within those rules and principles that Warring States scribes observed. The following story from the *Lüshi chunqiu* 呂氏春秋 is a rare example of the types of textual problems contemporary people encountered in their daily lives:

> 子夏之晉，過衛，有讀史記者曰：『晉師三豕涉河。』子夏曰：『非也，是己亥也。夫「己」與「三」相近，「豕」與「亥」相似。』至於晉而問之，則曰『晉師己亥涉河』也。

> When Zixia was going to Jin, he passed through Wei, where someone read a historical record, saying, "The Jin army and three pigs crossed the Yellow River." Zixia remarked, "That is wrong! It should say [on the day] '*jihai*' 己亥, [not 'three pigs' 三豕]. The character 己 is close to 三; and the character 豕 resembles 亥." When he arrived in Jin, he enquired about this matter, and the text indeed said: "The Jin army crossed the Yellow River on a *jihai* day."

In this story we can see how someone travelling across state boundaries in the late Spring and Autumn period is able to read manuscripts written in local scripts without particular difficulties. The ambiguity referred to here was a result of textual corruption and would have been equally problematic for local readers. At the same time, we can also see that orthographic diversity had its limits and sometimes even educated people had a hard time deciphering characters and reading manuscripts. Zixia's identification of the pair of corrupted graphs signals a normative awareness to matters of orthography and shows that concepts such as "textual mistake" or "incorrect graph" were indeed meaningful in contemporary society. His attention to normative concerns, of course, stems from his educational background and his close connection with the Ru 儒 style of learning. This tradition, with its strong prescriptive tendencies is an example how certain standards (in their case, primarily those of etiquette and classical learning) could be adhered to by followers of a restricted group, without the desire or ability to impose those on society at large.

3. Definition of graphic variation

As mentioned above, in modern China graphic variants are defined as being different from either the standard character or all other variants of the same character.[14] While this character-centric approach is feasible for the modern script, it becomes impractical when we try to apply it to the orthographic habits of the pre-Qin period. The problem is that due to the lack of an orthographically perfect form in Warring States writing it is not always possible determine, on the basis of their graphical appearance, which two forms counted as the same character. When the form 𣃟 is used in the *Guodian Laozi* to write the word *zhi* "to know," should we interpret it as the character 知 as its use suggests, or the character 智 as it would follow from its structure?[15] It would be tempting to identify it as a phonetic loan (i.e. using 智 to write 知), but that would presuppose that a "correct," or at least more common form, matching the modern character 知 was also in use, and this was clearly not the case. It is easy to identify a character when the differences are small and there is little orthographic "noise" from other similar-looking characters. But when the degree of variation increases or some of the variants overlap with other characters or their variants, the issue of identification becomes complex. Two graphs could be structurally different and yet stand for the same word, or be structurally identical and stand for different words. However, since writing is a means of recording language, each graph, regardless of its actual structure, stands for a specific word, and it is the context that permits us to identify that word. Therefore, the word, including its use in context, should be taken as the basis for determining the identity of the character. Applying this concept to variants, we can say that graphic or orthographic variants are dissimilar forms that stand for the same word.

This follows the line of reasoning of William G. Boltz who defined graphic variation as the "variation in the way the same word is written on different occurrences."[16] In addition to providing a conceptually sounder definition than the ones mentioned above, Boltz also introduced the category of "character variations," that is, cases when the same word is written with an altogether different character. Thus instead of just looking at different forms of the same *character*, we should also expand the definition to include graphic

---

[14]   See the definitions from modern Chinese dictionaries cited above.

[15]   Even structurally, however, this form is far from being identical with the modern character 智. The graphical similarity is primarily caused by the 曰 component at the bottom of the Chu form.

[16]   Boltz 1994: 159.

representations of the same *word*. Boltz used the Mawangdui 馬王堆 silk manuscripts to demonstrate that both kinds of variation were common in Han writing. Examining occurrences of the character 董 he drew attention to the fact that it "may stand theoretically for any number of words" the Han pronunciation of which approximated that of 董.[17]

This approach includes some types of variants that are usually not included in the *yitizi* category. For example, *tongjia* 通假 loan characters are generally excluded from the category of graphic variation because in such cases the character is replaced by another one. This is the reason why in modern character compendia (*wenzibian* 文字編) of Warring States manuscripts loan graphs are usually categorized, according to their graphic appearance, as different characters and are separated from other graphs that represent the same word.[18] As a result, the user may not notice these forms at all and overlook otherwise evident connections between variants.

4. Methodology of analysis

In order to document the patterns of orthographic variation, it is necessary to define the scope of the analyzed corpus and identify the boundaries of the community that used the script. As most manuscripts from early China are discovered in larger batches in tombs or sacrificial pits, each discovery by definition represents a restricted corpus connected with a single community. Sometimes this was only one family or a single person. What is important that the manuscripts were legible for all literate members of that community. The community was able to read and use all manuscripts in that group, even if some of them had come into their possession from other places.[19]

To a certain extent we would expect the scope of the corpus to limit the degree of variability. The wider we define the corpus, the higher the number of variants and their graphic diversity. If we took the people of Chu, using Olivier Venture's expression, as an "extended community," we could have all Chu manuscripts grouped together into a single corpus.[20] Although the style

---

[17] Boltz 1994: 163.

[18] For a description of this problem with the preoccupation of modern *wenzibian* with graphic shapes and character identity, see Venture 2009: 944–945.

[19] Although theoretically it is possible that someone possessed manuscripts that had been written in distant places in a script that was very different from the one used by its new owner, in practice there are no obvious cases of this in early China.

[20] On this point, see Venture 2009: 949.

of the script used in these manuscripts can indeed be confidently recognized as characteristic of the region of Chu, considering that some of the documents were written centuries apart, it would probably be difficult to talk about an actual group of people who shared these manuscripts.

Yet when we talk about the script in broad terms, it intuitively makes sense to extend it to a population larger than that of a single site or a city. While smaller communities may have had their own local traditions and idiosyncrasies, the script itself was no doubt shared, similar to a spoken language, by larger segments of the population. At the same time, the circumstances of manuscript discoveries in many ways determine what becomes available to us as a corpus. But even manuscripts from the same geographical area may differ in their social function, and there might be differences in the orthographic consistency between various types of documents (e.g. literary, philosophical, legal, administrative, divinatory).

In addition, even within the same restricted group of manuscripts we may find materials of diverse origin, coming from archaic sources or distant regions, and this eclectic composition of the corpus may complicate orthographic patterns.[21] While this is certainly true, we must also acknowledge that language in general is built in a similar manner, as it includes elements from vastly different sources, which together constitute a system and individual users are rarely aware of its complex makeup. The script is a technology, a pragmatic tool. Although it is extremely diverse and many of its elements can be traced to earlier historical strata, this does not change the fact that it can be effectively used by a single community. All archaic and foreign elements form part of the system, regardless of the particulars of their origin. In the Baoshan documents, for example, the graphs used for the word *si* "four" can be grouped into three basic types: 三, 𢀖 and 𤕝 . Of these the first one is an older form customarily seen on Western Zhou inscriptions, whereas the other two are variations of the form that matches the structure of the modern character 四.[22] The bamboo slips testify that all of these forms were used concurrently in the Baoshan region, and although they may reflect the idiosyncrasies of scribal hands, or represent remaining traces of competing traditions, they were nevertheless all valid, and therefore legible, ways of representing the word *si* "four."

I find the category of local script more useful for the analysis of graphic variation than that of a single manuscript because it is based on the concept of

---

[21] Venture 2009: 950–954.

[22] Venture (2009: 954) also notes the concurrent use of the old and new forms of the character 四 in the Shanghai Museum corpus.

a community sharing a script. The study of a single manuscript is essentially the study of one person's (or in any case a very limited number of persons') handwriting, which cannot represent "writing" at any given time and place in history, just as the utterances of one person do not represent a language or even a dialect. As speakers of the same spoken dialect understand other speakers of the same dialect, literate people using the same local script were able to read texts produced by other users of the same script. Analyzing an entire local script allows us to look at variant forms in their totality. By gathering variants together, we can discern the graphical, semantic, or phonetic constraints that were involved in representing a particular word in writing. Because one person's handwriting cannot represent the writing habits of an entire community, an analysis based on a single manuscript would always remain insufficient for documenting wider patterns. To cite a modern literary analogy, the analysis of the writings of William Faulkner or Saul Bellow will never be fully representative of twentieth-century American prose. Yet while lumping together a number of novelists from this period might obscure some of the peculiarities of individual authors, and some of these may certainly be important for authentication or other purposes, at the same time it provides the opportunity to see the characteristics of the entire genre, including common traits that are unique and distinguish the prose of this period from that of other time periods or countries.

Finally, as any typological analysis, the study of orthographic patterns should preferably be based on a statistically significant amount of data, since individual graphs by themselves are only examples of possible configurations. Solitary instances of orthographic variants are insufficient for appreciating the graphic diversity of writing a particular word because they will only show how a word *could* be written, rather than documenting the full range of how it actually was.

5. The corpus

For the sake of assessing orthographic variability and detecting its patterns, we need a corpus that has multiple instances of the same or very similar text so that words can be identified without ambiguity. Not only that, we also need enough examples to make the analysis statistically reliable. An ideal candidate for these criteria is the group of documents known as the Houma covenant texts, a group of inscribed jade and stone tablets from the fifth century BC,

possibly dating sometime between 442 and 424 BC.[23] Because the texts were written on the hard surface of the stone using a stylus and ink, they share the characteristics of bamboo slip documents. Moreover, even though they are somewhat earlier than bamboo slip manuscripts, they share a number of common features with those, which justifies studying them together. What makes them particularly interesting from the point of view of cultural history is that they are the earliest substantial body of handwritten texts in China.

The Houma covenant texts record, as part of a treaty ceremony in the state of Jin 晉, the oaths of vassals swearing allegiance to a covenant lord. They comprise hundreds of copies of the same few formulaic texts in which only the name of the oath-taker varies.[24] In essence, they are copies of formulas where each vassal inserted his name to obtain a personalized contract. This, combined with their large number, makes them ideal material for documenting the extent of graphic variation. Whenever we compare graphs that stand in parallel places in the text, we can be sure that they represent the same word, even if they structurally differ from more common forms. In contrast, manuscripts of philosophical texts offer very frew, if any, parallel passages, and even then we cannot be certain whether the graphs indeed represent the same words.

For my analysis, I select ten characters from an oath that begins with the words "Should Zhao dare not to split open his abdomen and heart in serving his lord..." (朝敢不剖其腹心以事其主). The surname Zhao here signifies the name of the oath-taker and is different in each tablet, thus I exclude it from the analysis. This leaves ten characters. While this may seem an arbitrary selection, this ensures that we are not biased in our choice and emphasize either stability or instability by picking specific characters that may support our preconception. Instead, we will also be able to document the differences between the stability of different characters, resulting in a more balanced picture of orthographic patterns.

Before proceeding, a few words should be said about mistakes, since the tablets, as it is inevitably the case with any handwritten material, also contain some scribal errors. Although the parallel copies make it easy to detect discrepancies, in most cases these are minor differences which do not change the meaning and do not violate the grammar. For this reason, it is questionable

---

[23] This date is based on Williams 2013. We should point out, however, that the time of the covenant texts remains disputed, with dates ranging from the beginning of the fifth century BC (e.g. Zhou 1994: 126–127) to the 430s (Feng 2002).

[24] Not all of the over 5,000 excavated fragments were inscribed. In 1976, hand-tracings of over 600 tablets and 200 photographs were published in a volume called *Houma mengshu* 侯馬盟書 (Shanxi sheng wenwu gongzuo weiyuanhui 1976).

whether they should be treated as mistakes. In some cases the scribes themselves made corrections and these are sure indicators of what the scribes considered incorrect. The most common correction is the insertion of a missing character in small script between two adjacent ones. For instance, Figure 2 shows mistakes in the opening phrase of the oath. In example A, the character 其 is inserted in small script on the left side of the row; in example B, the missing character 以 is supplied.[25] In both cases the spacing makes it obvious that the small characters were inserted subsequently, after the scribe made the mistake, perhaps even after having completed the entire tablet. Seeing contemporary corrections added by the scribes, or their peers and supervisors, we can be sure that at the time these cases counted as errors, even if the omission did not result in grammatical problems (e.g. A).[26] It is noteworthy, however, there there are no examples of correcting graphic variants, as all corrections involve lexical differences.
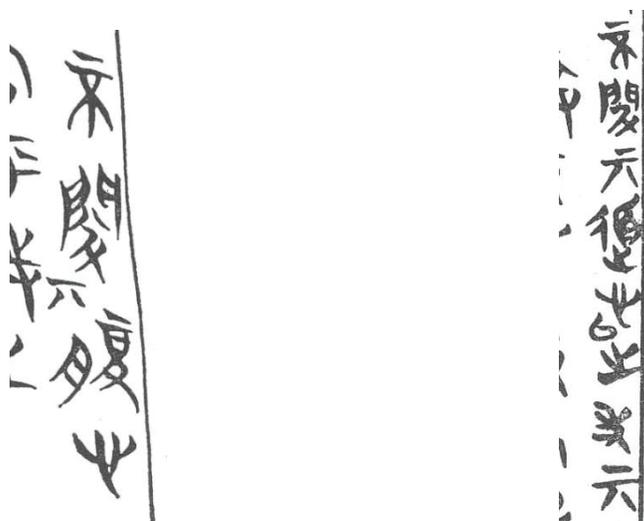

Figure 2. Examples of character insertions in the Houma covenant texts.
(A: Tablet #43; B: Tablet #64).

---

[25] Technically speaking, 以 here should replace the character 之 which was written here by mistake. Yet graphically nothing seems to indicate that this is a substitution or that the character 之 should be erased. Based on what we see on the tablet, we would have to read the two characters in succession (以之) which is grammatically problematic and, at the same time, differs from the other tablets.

[26] There are cases, however, when textual variation is left without correction. Tablet 1:54, for example, has 敢不剖其以事其主, entirely omitting the characters 腹 and 心, and thus rendering the phrase ungrammatical. Despite this, there are no signs of correction.

6. Analysis of variants

Discounting the variable surname at the beginning of the formula, the following ten words are selected for analysis:

1. *gan* "to dare" (敢)
2. *bu* "not" (不)
3. *pou* "to split open" (剖)
4. *qi* "its" (其)
5. *fu* "abdomen" (腹)
6. *xin* "heart" (心)
7. *yi* "in order to" (以)
8. *shi* "to serve" (事)
9. *qi* "its" (其)
10. *zhu* "lord" (主)

Although the character 其 occurs twice in this list with an apparently identical grammatical function, to be methodologically consistent I analyze each of these occurrences separately as if they were different words. Along the same line of thought, I only take into consideration instances of the words when they appear in this particular phrase, even though some of them (e.g. 敢, 不, 其, 以) also occur later in the text.

I limit the analysis to 100 examples of each character, even though many more tablets are available. The reason for this is practical: a comprehensive analysis would take considerably more time, without necessarily improving the reliability of our results, at least not in proportion to the amount of extra effort. I assume that 100 examples should be sufficient to document the basic patterns of graphic variation, especially if we consider that in other types of texts we would only find one or two parallel instances.

1) *gan* "to dare" (敢)

The graphs standing for the word *gan* "to dare" are fairly consistent in structure and if we were to look for a difference between them, we could single out the left bottom component which is written either as 口, 曰 or 田. An even smaller difference, which I chose to disregard in counting the variants, is that the component 𠂆 (又) is sometimes written as 𠂇 (寸), with a dot or short

stroke on the left side. Based on the variation of the left bottom component (口, 曰 or 田), we can group the variants as follows:[27]

83% ##

14% #############

3% ###

The first form occurs far more commonly than the others and, if we take into consideration that the components 口 and 曰 in early Chinese orthography are often interchangeable, we can see that this character, despite its relative complexity, is orthographically stable.

2) *bu* "not" (不)

The word *bu* "not" is also written in a fairly stable manner, and the main difference between variants is the presence or absence of a short horizontal line at the top. Using this relatively small difference as the criterion to distinguish between variants, we can have the following two groups:

59% ##

41% ####################################

Both of these forms commonly occur in the covenant texts and thus there is no reason to suppose that one of them was more "correct" than the other. Instead, both were accepted ways of writing the character and the difference seems to be ornamental, without any semantic motivation. Note that the crossing of the vertical line which is common in the Chu script is not seen in the Houma covenant texts.

---

[27] The number next to each graph is the number of its occurrences within the 100 occurrences, that is, the percentage of its frequency. To the right, the frequency is also displayed graphically in order to make the quantities more apparent and easier to compare.

3) *pou* "to split open" (剖)

This is one of the descendantless characters that did not survive in the modern script, consisting of the phonophoric 半 enclosed within the signific 門. There is no consensus on what word this character signifies and how it should be written in the modern script.[28] For our analysis here, it is not so important *which* word it writes but rather that in each instance it is the same word. There are two graphic variants in the corpus, differing in whether the signific 門 is present or not.

 61% ##

 39% ####################################

Both forms are common, although the first one is used almost twice as often as the other. Apparently, even in the case of the second form that omitted the signific, the intended word remained unambiguous, as the phonetic component preserved the link with the pronunciation of the word.

4) *qi* "its" (其)

We see two main forms of the character 其: the archaic form originally created to represent the word *ji* "basket" (written in the modern script as 箕), and the abbreviated form 六 which comprises only the lower part of the archaic form. Beside these two main types, there is also a number of smaller variations:

---

[28] The editors of the *Houma mengshu* volume write the character using its clericized version as a combination of 門+半, noting that its meaning approximates that of the words written today with the characters 判, 剖 or 布, and see it as referring to opening up and displaying something in public, see Shanxi sheng wenwu gongzuo weiyuanhui 1976: 36.

81% ##

12% ###########

3% ###

2% ##

2% ##

By far the most common of these five forms is the first one, also frequently encountered in bamboo-slip manuscripts from the Warring States period. Contrasting the high percentage of this single form with the possible configurations (four other types in this case), we can say that the word *qi* "its" appears in the corpus in an orthographically consistent way.

5) *fu* "abdomen" (腹)

This word has the largest number of graphic variants in the corpus. There are twenty-three forms listed below, differing from each other in at least one component. This breakdown does not actually reflect the full scope of its variability because for practical reasons I had to limit myself to component-level variation; differentiation between forms that include or omit individual strokes would have produced a long and unmanageable list. But we should nevertheless keep in mind that for some other characters (e.g. 敢) we set the definition of what constitutes variation stricter and made finer distinctions. In this case, on the other hand, a number of nuances are disregarded to avoid clutter. For example, the component 复 is taken as a single unit, regardless of its internal variation.

28% ##########################

21% ####################

13% #############

8% ########

6% ######

4% ####

2% ##

2% ##

2% ##

2% ##

2% ##

1% #

1% #

1% #

1% #

1% #

1% #

1% #

1% #

1% #

1% #

This is the example where the notion that the word identifies the character and that graphic variation represents different ways of writing the same word becomes important. While the most common form is structurally identical to the modern form (肉+复=腹), the second one includes an additional 彳 component, associating it with the character 復. Although the exact match of the modern character 復 appears only once (the last one on the list), most of the forms here represent a version of the characters 腹 or 復, or at times a combination of the two. The reason for this, beside the obvious phonetic and structural connection between 腹 and 復, may be that both of these characters appear in the oath in each other's vicinity, representing the words *fu* "abdomen" and *fu* "to restore," respectively. Thus they may influence each other's structure.

As for the degree of graphic variability of the word *fu* "abdomen" in the corpus, we should point out that the most common form is used in less than a third of the cases (28%). In other words, even if that was the preferred form, the word was more often written in other ways.

6) *xin* "heart" (心)

The word *xin* "heart" appears in the covenant texts in complete orthographic uniformity. One of the reasons for this is probably that 心 is a single-component character with a pictographic origin. Having said that, it is still possible to identify smaller differences, such as whether the two side strokes are written out in full (e.g. 屮) or are joined into a single one that runs across the middle section (e.g. 屮). Such minute discrepancies, however, shall not be treated here.

屮   100%   ###

7) *yi* "in order to" (以)

The character 以 occurs in two distinct forms in the covenant texts, one of which omits the signific 口, and the other that includes it.

ᐣ   78%   ##

ᐣ   22%   ####################

The first form was apparently much more common than the second, although the second was not rare either—it is also attested in other pre-Qin manuscripts. The second form matches the structure of the later character 台, glossed in the *Shuowen* as "happy, joyful" 說也, evidently referring to the word written today as 怡 (*yi* "pleasant"). Yet here even this form unquestionably stands for the word written in later orthography with the character as 以.

8) *shi* "to serve" (事)

The forms below are differentiated according to their top component, whether it includes a horizontal line or not, whether it is written with a 口 or 曰, etc. Nevertheless, these differences are minor and despite the seeming variability of the character, on a component-level we should consider it relatively consistent.

31% ############################

31% ############################

27% #########################

8% #######

1% #

1% #

1% #
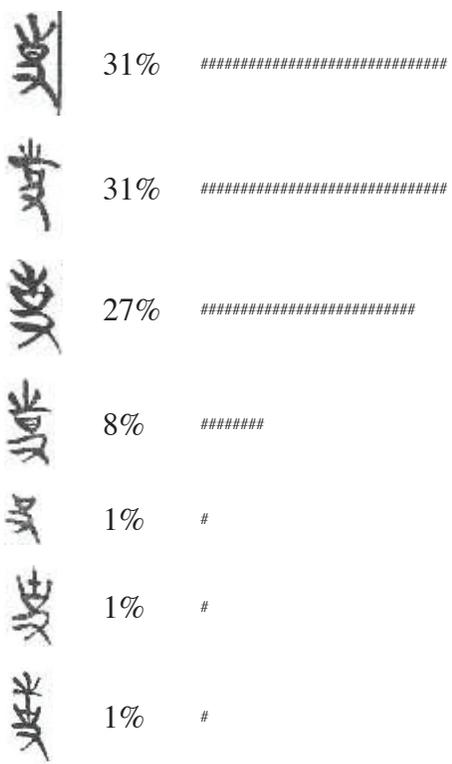
The first two forms both appear 31% of the time ,which means that none of them was more preferred than the other. The form 夎 (1%) coincides with how the character 吏 was written in the Warring States period but this is to be expected as the characters 事, 吏, 史 and 使 are related and this connection was still evident in the pre-Qin scripts.

9) *qi* "its" (其)

This is the second occurrence of the word *qi* "its" in our segment of ten words and although it serves an identical grammatical function (i.e. third person possessive pronoun), it is interesting to contrast the two sets of variants with each other.

79%   ##

9%   #########

6%   ######

5%   #####

1%   #

Not surprisingly, we have the same five variants as in the first set and the distribution is very similar. The frequency of the most common form is 79%, whereas in the first set it was 81%; the second form here occurs 9% of the time, as opposed to 12% earlier. This also corroborates the statistical reliability of our results with respect to the entire corpus, and that these numbers may be taken as a fairly accurate prediction for the distribution of variants in a similar group of material.[29]

---

[29]   A potential candidate would be the Wenxian covenant texts, which are still pending publication. For English language studies of the Wenxian texts, see Weld 1997, Williams 2005 and 2011.

10) *zhu* "lord" (主)

This is the last word of the ten selected for the analysis. Unlike in the modern standard form, all forms have an additional ⌐⌐ component on the top, and there is also some difference in the bottom component. Once again, this is one of the more stable characters that are separated here into two larger groups only on account of the presence of absence of a small horizontal line in above the component 主. In reality, this is a minor difference and cannot compare to the component-level variation seen in some of the other characters above.

55%   ##

43%   ##

1%   #

1%   #
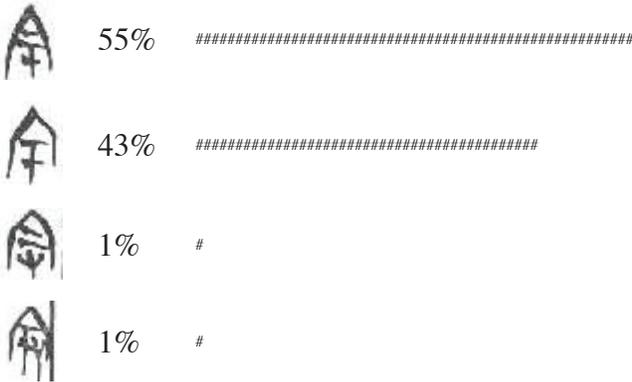
More significant are the changes in the other two forms which include the component 丫 (又), as this would certainly count as component-level variation. However, these variants are so rare (1% each) that they can be safely disregarded. All in all, we can say that the word *zhu* "lord" is used in this place of the oath texts consistently and the only variation results from the presence or absence of a horizontal stroke.

7. The degree of graphic variability

Having completed an orthographic analysis of 1,000 graphs from the Houma corpus (i.e. hundred occurrences of ten characters), one of the most apparent impressions we are left with is that different words were written with characters of varying degree of graphic consistency. Some of them were very stable (e.g. the uniform orthography of the character 心 used to write the word *xin* "heart"), while others exhibited a high degree of variability (e.g. the diverse variants used for the word *fu* "abdomen"). Most other characters were located

between these two extreme points. On the whole, this condition in itself testi-fies to the flexibility of orthographic standards in the state of Jin at the end of the Spring and Annals period.

Statistically speaking, there is a certain degree of unpredictability in why some characters vary more and others less. While structural complexity may play an important role in variability, this is not always the case. Even relative-ly simple characters can be written with an alternate structure. The character 以, for example, which in later manuscripts is usually written with the form 㠯 as a single-component character, in 22% of the cases appeared in the cov-enant texts as 㠯, with an additional 口 signific at the bottom. Although this variant is also attested in pre-Qin inscriptions, it is usually less expected to be seen in manuscripts. As a counter-example, the relatively complex character 敢 was written in a fairly consistent manner, and the variants had only minor discrepancies.

In several cases we can witness the concurrent use of archaic and modern variants of the same character. One such example is the two basic forms of the character 其 used for the possessive pronoun *qi* "its." The old form, also attested on oracle bones and early bronze inscriptions was written as 箕, or in a slightly abbreviated way as 丌, whereas the form used with highest frequency was 丌, which is also the form typical for Warring States manuscripts. The presence of forms associated with different time periods together within the same corpus demonstrates that the script is a complex system with a long tra-dition and that at any given point in history it included traces of earlier stages of its history.

The comparison of the two examples of the word *qi* "its" used in the same grammatical function but in two different places confirms that context does not affect orthography. To some extent, this also validates the notion that there is no immediate link between character structure and the meaning of the word it is used to represent, but that the connection between word and its written form is a matter of convention or habit. In the process of reading, a character is processed as a symbol pointing to a word in the language, and at this stage structure is not relevant anymore, provided that it does not impede legibility and the character is recognizable.[30] As to the distribution of graphic variants of the word *qi* "its," we have seen that this was very similar in the two examples, which is also an indication that we can expect the same regularity each time the word is used in the corpus.

Another pattern that emerges from the analysis of the 1,000 graphs in the covenant texts is that for most words there is a dominant written form,

---

[30] On this, see also Branner 2011: 93.

representing the most common way of writing that word. While this at first seem like a trivial point to make, it corroborates the notion that Jin scribes adhered to their own writing conventions, even if these were less strict than what we are accustomed to today. As we enter into the dynastic period and the script undergoes recurring rounds of normalization, many of these dominant forms become accepted as the standard forms, securing their place in the modern script, while less common forms are discontinued.

With regard to the internal rules of graphic variation, we can establish the absolute stability of the phonetic component. While in highly variable characters (e.g. 腹) we can see a great degree of fluidity with regards to the inclusion or omission of various semantic components, the phonetic component remains intact and is present in all configurations. Naturally, the phonetic component itself may exhibit some variability on the level of its subcomponents (if there are any) or individual strokes but as a unit it is always present and remains recognizable. Although I expect to see some deviation from this neat pattern if we include more material in our analysis, in the case of the above 1,000 graphs this rule emerged with 100% consistency.[31] It is interesting to observe, however, that among the 1,000 graphs we cannot identify with confidence a single case of *jiajie* 假借 borrowing, even though the profusion of such graphs in pre-Qín manuscripts has become one of the basic tenets of modern palaeography. While when taken separately, variant forms such as 復 and 腹 may appear to be phonetic loans, within the context of their aggregate forms it becomes clear that a match with the graphic appearance of another word (in this case, *fu* "again," written today as 復) is coincidental.

8. Conclusions

This chapter looked at the phenomenon of graphic variation in early Chinese writing. I have tried to show that the concept of a standard at this time had not arisen yet, at least not in its explicitly prescriptive form documented in Han sources. At the same time, as is the case with any writing system, the pre-Qin script was a complex system which was based on regular associations between words and written graphs, and this network of conventions could indeed be interpreted as a kind of standard. We may assume that scribal traditions were maintained by local groups and handed down from teacher to student in the process of literacy acquisition. Character structure in general, however, was more

---

[31]   Cf. Martin Kern's (2002: 162) observation on the nature of textual variants that "only a fraction of all these variants show no immediately apparent phonological relation."

flexible at this time than how it is portrayed in Han and later sources, and a word could be written in more than one way.

Orthographic flexibility also has practical implications for reading early Chinese manuscripts. One of these is that when analyzing graphs in manuscripts where we do not have access to parallel versions of the same text, assumptions solely based on orthographic considerations should be treated with caution. Thus we should be careful not to use structural differences as our main criterion for making judgments about the grammatical function of the word and its meaning in context, the provenance and dating of the manuscript, or even the identification of scribal hands. While orthography has much to add to each of these analyses, individual configurations do not represent the whole range of orthographic possibilities and thus should be used only alongside additional supporting evidence. Sufficient attention should also be given to other considerations, such as calligraphic attributes (style of strokes, spacing of characters, etc.) or the archaeological context of the discovery.

The risk of attributing too much significance to a particular variant or its structural configuration is especially obvious in the case of philosophical texts where the modern reader is already in a difficult position trying to interpret the meaning of the text. A contrast with a transmitted counterpart may further emphasize the uniqueness of a manuscript variant and invite the reader to try to rationalize its orthographic structure. The analysis of variant graphs presented in this chapter, however, tells us that a certain degree of flexibility is to be expected and that this was an inherent part of the script in early China. The analysis of graphic variants may be useful for reading a text as long as it helps to identify the words written with those graphs. Beyond that point specific orthographic details do not contribute to the meaning of words in context.

Primary Sources

Ganlu zishu 干祿字書. By Yan Yuansun 顏元孫. Shanghai: Shangwu yinshuguan, 1936.

Hanshu 漢書. By Ban Gu 班固; commentary by Yan Shigu 顏師古. Beijing: Zhonghua shuju, 1962.

Lüshi chunqiu 呂氏春秋. By Lü Buwei 呂不韋. Shanghai: Zhonghua shuju, 1936.

Shiji 史記. By Sima Qian 司馬遷. Beijing: Zhonghua shuju, 1959.

Shuowen jiezi 說文解字. By Xu Shen 許慎. Beijing: Zhongguo shudian, 2011.

Zhanguoce 戰國策. By Liu Xiang 劉向. Shanghai: Shanghai guji chubanshe, 1978.

Zixue sanzheng 字學三正 (1601). By Guo Yijing 郭一經. Beijing: Beijing chubanshe, 1997.

Secondary Sources

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

Bottéro, Françoise 1996. *Sémantisme et classification dans l'écriture chinoise: Les systèmes de classement des caractères par clés du* Shuowen jiezi *au* Kangxi zidian. Paris: Collège de France, IHEC.

Branner, David Prager 2011. "Phonology in the Chinese Script and its Relationship to Early Chinese Literacy." In: Li Feng and David Prager Branner (eds.) *Writing and Literacy in Early China*. Seattle: University of Washington Press, 85–137.

Feng Shi 馮時 2002. "Houma, Wenxian mengshu niandai kao 侯馬、溫縣盟書年代考." *Kaogu* 考古 8: 69–73.

Galambos, Imre 2006. *Orthography of Early Chinese Writing: Evidence from Newly Excavated Manuscripts*. [Budapest Monographs in East Asian Studies 1.] Budapest: Department of East Asian Studies, Eötvös Loránd University.

Jingmen shi bowuguan 荊門市博物館 (ed.) 1998. *Guodian Chu mu zhujian* 郭店楚墓竹簡. Beijing: Wenwu chubanshe.

Kern, Martin 2002. "Methodological reflections on the analysis of textual variants and the modes of manuscript production in Early China." *Journal of East Asian Archaeology* 4.1–4: 143–181.

Qin Gong 秦公 1985. *Beibiezi xinbian* 碑別字新編. Beijing: Wenwu chubanshe.

Richter, Matthias 2005. "Towards a profile of graphic variation." *Asiatische Studien / Études Asiatiques* LIX. 1: 169–207.

Shanxi sheng wenwu gongzuo weiyuanhui 山西省文物工作委員會 (ed.) 1976. *Houma mengshu* 侯馬盟書. Shanghai: Wenwu chubanshe.

Venture, Olivier 2009. "Looking for Chu people's writing habits." *Asiatische Studien / Études Asiatiques* LXIII. 4: 943–957.

Weld, Susan R. 1997. "The covenant texts from Houma and Wenxian." In: Edward L. Shaughnessy (ed.) *New Sources of Early Chinese History: An Introduction to the Reading of Inscriptions and Manuscripts*. [Early China Special Mongraph Series 3.] Berkeley: Institute of East Asian Studies and Society for the Study of Early China, 125–160.

Williams, Crispin 2005. "A Methodological Procedure for the Analysis of the Wenxian Covenant Texts." *Asiatische Studien / Études Asiatiques* LIX.1: 61–114.

Williams, Crispin 2011. "Early references to collective punishment in an excavated Chinese text: analysis and discussion of an imprecation from the Wenxian covenant texts." *Bulletin of the School of Oriental and African Studies* 74: 437–462.

Williams, Crispin 2013. "Dating the Houma Covenant Texts: The significance of recent findings from the Wenxian Covenant Texts." *Early China* 35–36: 247–275.

Zhang Yongquan 張涌泉 1995. *Hanyu suzi yanjiu* 漢語俗字研究. Changsha: Yuelu shushe.

Zhang Yongquan 張涌泉 1996. *Dunhuang suzi yanjiu* 敦煌俗字研究. Shanghai: Shanghai jiaoyu chubanshe.

Zhou Fengwu 周鳳五 1994. "Houma mengshu niandai wenti chongtan 侯馬盟書年代問題重探." *Zhongguo wenzi* 中國文字 19: 111–135.